Reinsurance and Portfolio Optimisation under Model Uncertainty

Yi Xia

A thesis submitted for the degree of Doctor of Philosophy

School of Mathematics, Statistics and Actuarial Science

University of Essex

9^{th} April, 2026

Acknowledgements

I would like to express my deepest gratitude to Dr Junlei Hu and Dr Peng Liu for your invaluable guidance, patience and care throughout my doctoral journey. Your encouragement and scholarly insight have shaped this work and helped me grow both professionally and personally. Without your support, I would not have been able to achieve what I have.

I am immensely grateful to my parents, Youyun and Li, for your constant love and for believing in me throughout my life and studies. Your unwavering support has been the foundation of my journey, and having them by my side at the conclusion of this learning path means more than words can express. To my lovely sister, Lin, thank you for filling my life with joy and always thinking I am the best. Your positivity has been a constant source of strength.

Finally, I would like to extend my heartfelt thanks to a dearest friend, Lyon, whose spiritual support helped me navigate stress and doubt. Your kindness, understanding and patience in weathering my emotions have been invaluable, and your respect for my choices has meant a great deal.

The research was conducted under the supervision of Dr Junlei Hu and Dr Peng Liu, whose guidance and contributions have been invaluable throughout this process.

Chapter 2 has been published in the European Journal of Operational Research and is a joint work with Dr Tolulope Fadina, Dr Junlei Hu, and Dr Peng Liu. Chapter 3 is a joint work with Dr Tao Gao and Dr Junlei Hu and has been submitted for publication. Chapter 4 is a joint work with Dr Peng Liu and Professor Steven Vanduffel.

Abstract

Reinsurance and portfolio decisions are challenging under conditions of model uncertainty and non-convex risk measures. This thesis develops robust and tractable frameworks by explicitly accounting for dependence ambiguity, distributional uncertainty, and non-convexity in objective functions.

First, for worst-case reinsurance for multiple lines of business with fixed marginal distributions and unknown dependence structure, Value-at-Risk (VaR) and Range Value-at-Risk (RVaR) are used to evaluate the risk in finding forms of optimal ceded loss function. For the VaR-based model with only two risks, the limited stop-loss reinsurance treaty is optimal for each line of business. Second, this thesis adapts Homotopy Optimisation with Perturbations and Ensembles (HOPE) to multi-line, non-convex reinsurance, achieving high-quality solutions far faster than conventional grid search on VaR-based problems, thereby closing the theory-practice gap. Third, the analysis presented in this thesis derives distributionally robust bounds for broad distortion risk metrics under some uncertainty sets, which are characterised by moment constraint, probability constraint via Wasserstein ball, or unimodality constraint, thereby identifying worst-case distributions. The numerical results for the application in portfolio optimisation are also derived to show the power of the theory.

Keywords: Optimal reinsurance; Portfolio optimisation; Multivariate risk; Dependence uncertainty; Global optimisation; Non-convex optimisation; Numerical optimisation; Homotopy method; Robust distortion risk metrics; Mean variance; Wasserstein metrics.

Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

Reinsurance is a risk transfer tool, that allows insurers to cede portions of their liabilities to third parties, thereby improving capital efficiency and limiting exposure to large losses. The early theoretical foundations, especially in relation to the variance minimisation approach of Borch (1960) and the expected utility framework of Arrow (1963), clarified the trade-offs between risk reduction and retention. In practice, however, the direct applicability of these classical models is limited by several realistic complications, such as the uncertain or unknown dependence between risks and the non-convexity of objective functions. These issues make the design of optimal reinsurance problems substantially more challenging than classical theory suggests. Many of the same difficulties arise in portfolio optimisation, wherein decision-makers must allocate capital across correlated assets while controlling tail risk and accounting for model mis-specification. The mean–Gini (MG) approach proposed by Shalit and Yitzhaki (1984) exemplifies a constructive alternative to variance-based methods. The Global Financial Crisis of 2008 demonstrated the practical consequences of underestimating dependence and model uncertainty, as fragile models and excessive reliance on specific distributional assumptions can produce catastrophic mispricing and unexpected losses. Consequently, contemporary research, and this thesis, move beyond point-estimate models and toward robust, distributionally aware methods that incorporate investor preferences, uncertain dependence structures, and the stochastic nature of losses and returns. Doing so requires both principled modelling, to capture realistic dependence and ambiguity, and computationally tractable algorithms, to solve high-dimensional, often non-convex optimisation problems.

This thesis investigates optimal reinsurance with an unknown dependence structure, its related computational problem for high-dimensional optimisation, and portfolio optimisation under conditions of model uncertainty.

Chapter 2 focuses on an optimal reinsurance design in the presence of dependence uncertainty across multiple insurance risks. Classical approaches often assume independence or impose specific copula models. These approaches may lead to misalignment that can result in false results. Recent developments in robust risk aggregation, such as Cheung et al. (2014), Bernard et al. (2020), and Cai and Chi (2020), spur the evolution of strategies that solve problems with an unknown dependence structure in relation to risk. Within this framework, we derive optimal reinsurance policies that minimise worst-case Value-at-Risk (VaR) and Range-Value-at-Risk (RVaR), subject to budgetary constraints and general premium principles. A key step in solving the optimisation problem is the extension of classical aggregation bounds, such as Makarov (1981), Rüschendorf (1982), and Blanchet et al. (2023), to settings where marginal tail distributions exhibit convexity or concavity. These structural assumptions enable closed-form solutions for classes of optimal contracts, including capped quota-share and piecewise-linear policies, across arbitrary dimensions. The resulting designs provide tractable and theoretically justified reinsurance strategies given the prevailing uncertainties of the model.

The optimisation problem arising in the numerical illustration of Chapter 2 is inherently non-convex. In that particular case, the problem structure permits a discretisation over a finite grid, allowing the use of grid search techniques to locate global optima. However, such an approach is only feasible for low-dimensional problems with an exploitable structure. In more general, non-convex settings, especially those with continuous or high-dimensional control variables, grid search optimisation methods become computationally intractable or unreliable. This limitation motivates the second project, which addresses the broader class of non-convex reinsurance optimisation problems.

In Chapter 3, we develop scalable numerical methods to overcome the computational challenges posed by non-convexity. Specifically, we adapt the Homotopy Optimisation with Perturbations and Ensembles (HOPE) algorithm. Its global optimisation approach was originally developed for non-linear PDEs by Liao (1992) and later extended to non-convex problems by Lin et al. (2023). HOPE combines homotopy continuation with stochastic perturbations and ensemble learning to navigate complex solution landscapes more efficiently. We apply this method to reinsurance models involving multiple lines of business and VaR-based objectives, demonstrating its ability to produce optimal solutions with significantly reduced computational effort when compared to grid search methods. The results establish HOPE as a potentially viable and scalable approach for actuarial optimisation under conditions of non-convexity.

Chapter 4 turns to robust risk metrics and portfolio optimisation under conditions of distributional ambiguity. Recent advances in distributionally robust optimisation under distorted expectations, Cai et al. (2025) have shown how applying distortion functions to probabilities can capture a decision maker's risk attitude more flexibly than traditional expected utility approaches. In this context, distributionally robust optimisation (DRO) has become a powerful tool (see Ben-Tal et al. (2009) and Esfahani and Kuhn (2018)), where the ambiguity sets are defined through moment constraints or Wasserstein distances. We develop new bounds for a wide class of distortion risk metrics, including non-monotonic and discontinuous distortion functions, under ambiguity sets constrained by the mean, variance, Wasserstein distance, and unimodality. Our methods generalise the prior findings of Li et al. (2018) and Bernard et al. (2024)), offering sharper worst-case estimates for practical metrics such as the Gini deviation, inter-quantile differences, and GlueVaR. These results are applied to portfolio optimisation under conditions of model uncertainty to derive the best portfolios in the worst-case scenario for different distortion risk metrics. The numerical results are also given to illustrate the power of our theory.

In summary, this thesis proposes and tests a novel theoretical and computational toolkit for reinsurance and risk management under conditions of uncertainty. To this end, it contributes tractable solutions, generalised bounds, and scalable algorithms that bridge actuarial theory and practical implementation, offering robust strategies for managing risk and portfolio within complex and ambiguous environments.

CHAPTER 2

Optimal reinsurance with multivariate risks and dependence uncertainty

2.1 Introduction

Reinsurance serves as a fundamental risk management tool for insurance companies and has attracted considerable interest from both practitioners and academics. The study of optimal reinsurance design originates from the pioneering work of Borch (1960), which has inspired extensive research focused on minimising insurer risk exposure or maximising the expected utility of terminal wealth for risk-averse insurers, see for instance, Arrow (1963), Van Heerwaarden et al. (1989) and Chi and Lin (2014). In recent years, Value-at-Risk (VaR) and Expected Shortfall (ES) have become widely used risk measures to assess insurance risk and determine regulatory capital. Consequently, a growing body of literature has addressed optimal reinsurance under VaR-based or ES-based criteria. For an overview of the theory and methods of reinsurance, refer to Albrecher et al. (2017) and the review by Cai and Chi (2020). Alternative approaches to optimal reinsurance, such as those based on probabilities of disaster, have also received attention. For example, Tan et al. (2020) showed that under a mean–TVaR premium principle, the optimal reinsurance contract assumes a dual excess-of-loss structure.

There are very few works that study optimal reinsurance design under multivariate risks, whereas problems in the univariate setting have been widely studied. In practice, insurers manage multiple lines of business, each potentially reinsured via separate contracts with one or more reinsurers. This motivates a holistic study of reinsurance design in a multivariate context, taking into account aggregated risk across lines of business. In this paper, we aim to minimise certain risk measures of the aggregate retained risk in each business line by reinsurance. The challenge lies not only in treating individual marginal distributions but also in understanding the unknown dependence between these risks.

Only a limited number of papers consider multivariate reinsurance problems. Under the assumption of identically and independently distributed risks, Denuit and Vermandele (1998) demonstrated that the stop-loss treaty is optimal for each line of business with respect to the stop-loss order and under expected premium principles. Their results were extended to certain positively dependent risks by Cai and Wei (2012). Zhu et al. (2014) examined optimal reinsurance that minimises capital requirements and showed that one- or two-layer reinsurance contracts are optimal under general premium principles, regardless of the dependence structure. Cheung et al. (2014) addressed the problem under complete dependence uncertainty by considering the worst-case scenario for general law-invariant convex risk measures and found that the optimal policy remains the stop-loss type. A key simplification arises from the comonotonicity of the worst-case dependence structure. Bernard et al. (2020) further explored the problem in the context of optimal coinsurance across multiple interdependent policyholders from an expected utility perspective, revealing that insurance supply and demand may not necessarily align.

Following the framework of Cheung et al. (2014), we consider a min–max reinsurance design problem for multivariate risks without assuming a known dependence structure. This assumption is motivated by the practical observation that, while marginal distributions can often be estimated with reasonable accuracy, estimating the dependence structure is significantly more difficult and, in some cases, infeasible. Mis-specification of the dependence structure can lead to severe consequences in risk management; see McNeil et al. (2015). In practice, data on correlated products are often collected independently, leaving no clear information about their joint distribution; see Embrecht et al. (2013). To address this, we adopt a robust optimisation approach and analyse the worst-case scenario. Robust optimisation is comprehensively reviewed in Ben-Tal et al. (2009), with applications in various fields, see in Ben-Tal and Nemirovski (2008), Polak et al. (2010), Bertsimas et al. (2011), Gabrel et al. (2014) and Asimit et al. (2017)). Benati and Conde (2022) introduced a robust framework considering expectations, risk, and regret. Most recently, model uncertainty in optimal reinsurance has been studied by Chi et al. (2022), Boonen and Jiang (2024), and Cai et al. (2024). Among robust formulations, the min–max criterion is particularly well suited to settings where dependence information is scarce, as it ensures the chosen reinsurance strategy remains effective under the least favourable dependence scenario. While this may lead to strategies that are more conservative than otherwise necessary when the realised dependence is benign, it nonetheless eliminates the risk of severe under-reserving that can arise from committing to a mis-specified copula model.

Compared to Cheung et al. (2014), we employ VaR and RVaR to quantify retained risk and use them as the basis for our optimisation criteria. Specifically, we minimise the worst-case VaR or RVaR of total risk exposure after reinsurance across business lines, subject to a budget constraint on total reinsurance spending. The non-convexity of these risk measures introduces analytical challenges. Our approach builds on existing expressions for worst-case VaR and RVaR in the context of aggregation under known marginals and unknown dependence. Foundational results can be found in Makarov (1981) and Rüschendorf (1982) for bivariate risks, and in Wang et al. (2013), Embrecht et al. (2013), Bernard et al. (2014), and Jakobsons et al. (2016) for higher dimensions. Our analysis extends recent results by Blanchet et al. (2023), which apply to distributions with monotonic densities. We generalise these expressions to allow for convex or concave tails, a key technical contribution of this work. Importantly, the worst-case dependence structure under VaR or RVaR is not comonotonic, but rather exhibits a combination of joint mixability, as shown by Wang and Wang (2016), and mutual exclusivity, as studied by Dhaene and Denuit (1999), corresponding to a strongly negative dependence regime.

Due to the limitations of existing results in the VaR and RVaR literature, we restrict our feasible set of ceded loss functions to those that are convex or concave when the number of risks exceeds two. Similar constraints appear in Cai et al. (2008) and Cheung (2010). We also consider general premium principles satisfying mild assumptions, encompassing the expected value principle as a special case. Under these settings, we derive optimal reinsurance policies for both VaR- and RVaR-based criteria. For instance, we identify piecewise, linear ceded loss functions, generalising stop-loss and bounded linear contracts that extend quota-share arrangements. In the special case of two risks, the convexity/concavity requirements can be relaxed, and we establish that limited stop-loss contracts are optimal for the VaR-based problem. For RVaR, we omit cases with two risks or convex-tailed marginals due to a lack of applicable robust aggregation results.

We also study quota-share reinsurance under conditions of dependence uncertainty. When the budget constraint is non-binding, the optimal contract for each business line is either 'full' or 'no coverage', indicating that the insurer may opt not to reinsure certain business lines, which is similar in spirit to Bernard et al. (2020), where the insurer selectively withholds insurance supply given dependent risks.

The remainder of this chapter is organised as follows: Section 2.2 presents the model setup and notation; Sections 2.3 and 2.4 explore the problems of reinsurance based on VaR and RVaR, respectively; while Section 2.5 discusses the special case of quota-share reinsurance with dependence uncertainty; and Section 2.6 presents analytical and numerical results for the case of two risks, using the findings of Section 2.3. Proofs of all results are provided in Section 2.7.

2.2 Model description and notation

Consider an insurer operating n distinct business lines over a given time period. For each line of business i, the aggregate loss is represented by a random variable X_i, defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ (Ω is the sample space, that the set of all elementary outcomes ω describing the uncertain state of the world over the period, for example all claims; $\mathscr{F}$ is a σ-algebra of subsets of Ω called events; $\mathbb{P} : \mathscr{F} \to [0,1]$ is a probability measure). These losses have cumulative distribution functions $F_i(x) = \mathbb{P}(X_i \leqslant x)$ and satisfy $0 < \mathbb{E}(X_i) < \infty$. The total loss of the insurer without reinsurance is $S_n = \sum_{i=1}^n X_i$. To mitigate substantial losses, the insurer implements reinsurance strategies for each line. Let $f_i(X_i)$ denote the portion ceded to the reinsurers and $T_{f_i}(X_i) := X_i - f_i(X_i)$ represent the retained loss, where f_i and T_{f_i} are termed the ceded and retained loss functions, respectively.

To preclude ex-post moral hazard, reinsurance arrangements must satisfy incentive compatibility. Consequently, both f_i and T_{f_i} are non-decreasing and non-negative on $[0,\infty)$ for $i = 1,\ldots,n$, equivalent to Lipschitz continuity:

$$0 \leqslant f_i(y) - f_i(x) \leqslant y - x, \ \ 0 \leqslant x \leqslant y, \ \text{ and } \ 0 \leqslant f_i(x) \leqslant x, \ \ x \geqslant 0.$$

This restricts admissible ceded loss functions to the domain:

$$\begin{aligned}\mathscr{D}^n = \Big\{ \mathbf{f} = (f_1,\ldots,f_n) : \, & 0 \leqslant f_i(y) - f_i(x) \leqslant y - x,\\ & 0 \leqslant x \leqslant y,\\ & 0 \leqslant f_i(x) \leqslant x,\\ & x \geqslant 0, \quad \text{for } i = 1,\ldots,n \Big\}.\end{aligned}$$

The insurer compensates the reinsurers via premiums determined by the principles π_i for each line. Let $\mathscr{X}_1$ be the space of non-negative random variables with finite mean in $(\Omega,\mathscr{F},\mathbb{P})$ and $\mathscr{X} \subset \mathscr{X}_1$. Then $\pi_i : \mathscr{X} \to \mathbb{R}^+$. Following Chi and Tan (2013), we analyse premium principles fulfilling:

(i) *Distribution invariance*: $\pi_i(Y) = \pi_i(Z)$ for $Y, Z \in \mathscr{X}$ when $F_Y = F_Z$;

(ii) *Risk loading*: $\pi_i(Y) \geqslant \mathbb{E}(Y)$ for $Y \in \mathscr{X}$;

(iii) *Preserving stop-loss order*: $\pi_i(Y) \leqslant \pi_i(Z)$ if $\mathbb{E}\left((Y-d)_+\right) \leqslant \mathbb{E}\left((Z-d)_+\right)\ \forall d \in \mathbb{R}$;

(iv) *Continuity*: $\lim_{n\to\infty} \pi_i(Y_n) = \pi_i(Y)$ when $\lim_{n\to\infty}$ ess-sup$|Y_n - Y| = 0$ for $Y_n, Y \in \mathscr{X}$, $n \geqslant 1$, and $\lim_{d\to\infty} \pi_i(Y \wedge d) = \pi_i(Y)$.

Although we assume π_i satisfies (i)-(iii), continuity (iv) ensures the existence of optimal ceded functions. Common principles satisfy constraints (i)-(iii), including expected value, Wang's, Dutch, and others (see in Wang et al. (1997) for Wang's principle and Young (2004) for a survey).

Post-reinsurance total risk exposure is given by: $S_n^{\mathbf{f}} = \sum_{i=1}^n \left[T_{f_i}(X_i) + \pi_i(f_i(X_i))\right]$. For a risk measure $\rho : \mathscr{X} \to \mathbb{R}$, the insurer seeks optimal $\mathbf{f} \in \mathscr{D}^n$ minimising $\rho(S_n^{\mathbf{f}}(X_1,\ldots,X_n))$ under budget constraint $\sum_{i=1}^n \pi_i(f_i(X_i)) \leqslant P$ ($P > 0$). This defines:

$$\mathscr{D}^n(P) = \left\{\mathbf{f} \in \mathscr{D}^n : \sum_{i=1}^n \pi_i(f_i(X_i)) \leqslant P\right\}.$$

This multivariate problem faces practical challenges in estimating the dependence structures of $X_1,\ldots,X_n$. Mis-specification can severely affect risk management (McNeil et al. (2015)), motivating the analysis under dependence uncertainty where the marginals are fixed but the dependence is unknown. The uncertainty set is given by:

$$\mathscr{E}_n(\mathbf{F}) = \{(X_1,\ldots,X_n) : X_i \sim F_i,\ i = 1,\ldots,n\},$$

where $\mathbf{F} = (F_1,\ldots,F_n)$. We solve for worst-case optimal reinsurance and find $\mathbf{f} \in \mathscr{D}^n(P)$ minimising

$$\sup_{(X_1,\ldots,X_n)\in\mathscr{E}_n(\mathbf{F})} \rho(S_n^{\mathbf{f}}(X_1,\ldots,X_n)). \tag{2.1}$$

Given the importance of the regulatory authorities of VaR and ES, we focus on these risk measures and their robust counterpart, RVaR. Defining for $\alpha \in (0,1]$:

$$\mathrm{VaR}_\alpha(X) = F^{-1}(\alpha) = \inf\{x : F(x) \geqslant \alpha\},$$

and for $\alpha \in [0,1)$:

$$\mathrm{VaR}_\alpha^+(X) = F_+^{-1}(\alpha) = \inf\{x : F(x) > \alpha\},$$

with $\inf\emptyset = 0$. For $\alpha \in [0,1)$, ES is defined as below in McNeil et al. (2015) and Föllmer and Schied (2016) :

$$\mathrm{ES}_\alpha(X) = \frac{1}{1-\alpha}\int_\alpha^1 F^{-1}(t)\mathrm{d}t.$$

RVaR Cont et al. (2010) bridges VaR and ES. For $0 \leqslant \beta < \beta + \alpha \leqslant 1$,

$$R_{\beta,\alpha}(X) = \frac{1}{\alpha} \int_{\beta}^{\beta+\alpha} F_{+}^{-1}(1-t)\mathrm{d}t.$$

Key relations between VaR, ES, and RVaR include:

$$\mathrm{ES}_\alpha(X) = R_{0,\alpha}(X), \;\; \mathrm{VaR}_\alpha(X) = \lim_{\beta \downarrow 0} R_{1-\alpha,\beta}(X), \;\; \mathrm{VaR}_\alpha^{+}(X) = \lim_{\beta \downarrow 0} R_{1-\alpha-\beta,\beta}(X).$$

We examine VaR- and RVaR-based reinsurance. Due to limitations in robust aggregation, we sometimes restrict to convex/concave-ceded functions:

$$\mathscr{D}_1^n(P) = \{\mathbf{f} = (f_1, \ldots, f_n) : \mathbf{f} \in \mathscr{D}^n(P),\; f_i \text{ convex},\; i = 1, \ldots, n\},$$

$$\mathscr{D}_2^n(P) = \{\mathbf{f} = (f_1, \ldots, f_n) : \mathbf{f} \in \mathscr{D}^n(P),\; f_i \text{ concave},\; i = 1, \ldots, n\}.$$

Special ceded functions, that are central to our results, are defined as below:
Let $\mathbf{a} = (a_1, \ldots, a_n)$, $\mathbf{b} = (b_1, \ldots, b_n)$, $\mathbf{c} = (c_1, \ldots, c_n)$, $\mathbf{d} = (d_1, \ldots, d_n)$.

- *Limited stop-loss*:

  $l_{a,b}(x) := (x-a)_+ - (x-b)_+$ $(0 \leqslant a \leqslant b \leqslant \infty)$.

  Set $\mathbf{l}_{\mathbf{a},\mathbf{b}} = (l_{a_1,b_1}, \ldots, l_{a_n,b_n})$ with domain:

$$\mathscr{A}(P) = \{(\mathbf{a}, \mathbf{b}) : \mathbf{l}_{\mathbf{a},\mathbf{b}} \in \mathscr{D}^n(P),\; 0 \leqslant a_i \leqslant b_i \leqslant \infty,\; i = 1, \ldots, n\}.$$

- *Generalized stop-loss*:

  $h_{a,b,c,d}(x) := c(x-a)_+ + d(x-b)_+$ $(0 \leqslant a \leqslant b \leqslant \infty, 0 \leqslant c, d \leqslant c+d \leqslant 1)$.

  Set $\mathbf{h}_{\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d}} = (h_{a_1,b_1,c_1,d_1}, \ldots, h_{a_n,b_n,c_n,d_n})$ with domain:

$$\mathscr{A}_1(P) = \{(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}) : \mathbf{h}_{\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d}} \in \mathscr{D}^n(P), 0 \leqslant a_i \leqslant b_i \leqslant \infty,\; 0 \leqslant c_i,$$
$$d_i \leqslant c_i + d_i \leqslant 1,\; i = 1, \ldots, n\}.$$

- *Proportional-minimum*:

$g_{a,b}(x) := a\min(x,b)$ $(0 \leqslant a \leqslant 1,\, 0 \leqslant b \leqslant \infty)$.

Set $\mathbf{g}_{\mathbf{a},\mathbf{b}} = (g_{a_1,b_1}, \ldots, g_{a_n,b_n})$ with domain:

$$\mathscr{A}_2(P) = \{(\mathbf{a},\mathbf{b}) : \mathbf{g}_{\mathbf{a},\mathbf{b}} \in \mathscr{D}^n(P),\ 0 \leqslant a_i \leqslant 1,\ 0 \leqslant b_i \leqslant \infty,\ i = 1, \ldots, n\}.$$

Subsequent sections demonstrate that optimal reinsurance over $\mathscr{D}^n(P)$, $\mathscr{D}_1^n(P)$, and $\mathscr{D}_2^n(P)$ simplifies to optimisation over:

$$\mathscr{L}^n(P) = \{\mathbf{l}_{\mathbf{a},\mathbf{b}} : (\mathbf{a},\mathbf{b}) \in \mathscr{A}(P)\},\ \ \mathscr{H}^n(P) = \{\mathbf{h}_{\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d}} : (\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d}) \in \mathscr{A}_1(P)\},$$

$$\mathscr{G}^n(P) = \{\mathbf{g}_{\mathbf{a},\mathbf{b}} : (\mathbf{a},\mathbf{b}) \in \mathscr{A}_2(P)\}.$$

Throughout, we assume $X_i \sim F_i$ for $i = 1, \ldots, n$.

2.3 Optimal reinsurance under Value-at-Risk

To evaluate the quantity $\sup_{(X_1,\ldots,X_n)\in\mathscr{E}_n(\mathbf{F})} \mathrm{VaR}_\alpha(S_n^{\mathbf{f}}(X_1, \ldots, X_n))$, we built on a recent advance in robust risk aggregation presented in Blanchet et al. (2023), which is formalized in Proposition 2.1. We generalise the findings on marginals with monotone densities from Theorem 2 in Blanchet et al. (2023) to include marginals whose tail distributions are either convex or concave. This extension is key in determining optimal reinsurance strategies. Importantly, convex/concave tail behaviour allows for probability mass at the distribution's extrema, which is instrumental in constructing optimal ceded loss functions. For $\alpha \in (0,1)$, a distribution F is said to be concave beyond its α-quantile if the function $\frac{(F(x)-\alpha)_+}{1-\alpha}$ is concave on its support; analogously, F is convex beyond its α-quantile if $\frac{(F(x)-\alpha)_+}{1-\alpha}$ is convex on $(-\infty, F^{-1}(1))$ (equivalently, if this function is convex on its support and $F(F_+^{-1}(\alpha)) = \alpha$ or 1). We denote the sets of such distributions by $\mathscr{M}_{cx}^\alpha$ and $\mathscr{M}_{ce}^\alpha$, corresponding to convex and concave tails, respectively. Defining

$$\Delta_n = \left\{ \boldsymbol{\gamma} \in (0,1) \times [0,1)^n : \sum_{i=0}^{n} \gamma_i = 1 \right\},\ \text{where } \boldsymbol{\gamma} = (\gamma_0, \gamma_1, \ldots, \gamma_n).$$

Proposition 2.1. *Let $\alpha \in (0,1)$ and $\mathbf{F} \in (\mathscr{M}_{cx}^{\alpha})^n \cup (\mathscr{M}_{ce}^{\alpha})^n$. Then*

$$\sup_{(X_1,\ldots,X_n)\in\mathscr{E}_n(\mathbf{F})} \mathrm{VaR}_{\alpha}^{+}\left(\sum_{i=1}^{n} X_i\right) = \inf_{\boldsymbol{\gamma}\in(1-\alpha)\Delta_n} \sum_{i=1}^{n} R_{\gamma_i,\gamma_0}(X_i).$$

Using the above result, we now demonstrate that the optimal reinsurance problems defined in the sets $\mathscr{D}^n(P)$, $\mathscr{D}_1^n(P)$, and $\mathscr{D}_2^n(P)$ can be reduced to optimisation problems in $\mathscr{L}^n(P), \mathscr{H}^n(P)$, and $\mathscr{G}^n(P)$, respectively.

Theorem 2.2. *Assuming that each $F_i^{-1}(\cdot)$ is continuous on $(0,1)$ and let $\alpha \in (0,1)$, then the following holds:*

(i) If $n = 2$ or $n - \sum_{i=1}^{n} F_i(F_i^{-1}(\alpha)) \leqslant 1-\alpha$, or $\sum_{i=1}^{n}(F_i(F_i^{-1}(1)-)-\alpha) \leqslant 1-\alpha$, then

$$\inf_{\mathbf{f}\in\mathscr{D}^n(P)} \sup_{(X_1,\ldots,X_n)\in\mathscr{E}_n(\mathbf{F})} \mathrm{VaR}_{\alpha}(S_n^{\mathbf{f}}(X_1,\ldots,X_n)) = \inf_{(\mathbf{a},\mathbf{b})\in\mathscr{A}(\mathbf{P})} \inf_{\boldsymbol{\gamma}\in(1-\alpha)\Delta_n} L(\mathbf{a},\mathbf{b},\boldsymbol{\gamma}),$$

where $L(\mathbf{a},\mathbf{b},\boldsymbol{\gamma}) = \sum_{\mathbf{i=1}}^{\mathbf{n}} \left\{\mathbf{R}_{\gamma_i,\gamma_0}(\mathbf{X_i} - \mathbf{l}_{\mathbf{a_i},\mathbf{b_i}}(\mathbf{X_i})) + \pi_{\mathbf{i}}(\mathbf{l}_{\mathbf{a_i},\mathbf{b_i}}(\mathbf{X_i}))\right\}$.

(ii) If $\mathbf{F} \in (\mathscr{M}_{cx}^{\alpha})^n$, then

$$\inf_{\mathbf{f}\in\mathscr{D}_{\mathbf{1}}^{\mathbf{n}}(\mathbf{P})} \sup_{(X_1,\ldots,X_n)\in\mathscr{E}_n(\mathbf{F})} \mathrm{VaR}_{\alpha}(S_n^{\mathbf{f}}(X_1,\ldots,X_n)) = \inf_{(\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d})\in\mathscr{A}_{\mathbf{1}}(\mathbf{P})} \inf_{\boldsymbol{\gamma}\in(1-\alpha)\Delta_n} H(\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d},\boldsymbol{\gamma}),$$

where $H(\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d},\boldsymbol{\gamma}) = \sum_{\mathbf{i=1}}^{\mathbf{n}} \left\{\mathbf{R}_{\gamma_i,\gamma_0}(\mathbf{X_i}) - \mathbf{R}_{\gamma_i,\gamma_0}(\mathbf{h}_{\mathbf{a_i},\mathbf{b_i},\mathbf{c_i},\mathbf{d_i}}(\mathbf{X_i})) + \pi_{\mathbf{i}}(\mathbf{h}_{\mathbf{a_i},\mathbf{b_i},\mathbf{c_i},\mathbf{d_i}}(\mathbf{X_i}))\right\}$.

(iii) If $\mathbf{F} \in (\mathscr{M}_{ce}^{\alpha})^n$, then

$$\inf_{\mathbf{f}\in\mathscr{D}_{\mathbf{2}}^{\mathbf{n}}(\mathbf{P})} \sup_{(X_1,\ldots,X_n)\in\mathscr{E}_n(\mathbf{F})} \mathrm{VaR}_{\alpha}(S_n^{\mathbf{f}}(X_1,\ldots,X_n)) = \inf_{(\mathbf{a},\mathbf{b})\in\mathscr{A}_{\mathbf{2}}(\mathbf{P})} \inf_{\boldsymbol{\gamma}\in(1-\alpha)\Delta_n} G(\mathbf{a},\mathbf{b},\boldsymbol{\gamma}),$$

where $G(\mathbf{a},\mathbf{b},\boldsymbol{\gamma}) = \sum_{\mathbf{i=1}}^{\mathbf{n}} \left\{\mathbf{R}_{\gamma_i,\gamma_0}(\mathbf{X_i}) - \mathbf{R}_{\gamma_i,\gamma_0}(\mathbf{g}_{\mathbf{a_i},\mathbf{b_i}}(\mathbf{X_i})) + \pi_{\mathbf{i}}(\mathbf{g}_{\mathbf{a_i},\mathbf{b_i}}(\mathbf{X_i}))\right\}$.

Observe that Theorem 2.2(i) reveals the limited stop-loss policy is optimal when $n = 2$ or the tails of the marginals are not overlapping. This contrasts with Cai and Wei (2012) and Cheung et al. (2014), where the stop-loss policy is optimal. This discrepancy arises because, in these cases, the worst-case dependence is counter-comonotonic on the tails (i.e., negatively dependent). In Theorem 2.2(ii), under convexity constraints on the ceded loss functions, the optimal solution takes the form $h_{a,b,c,d}(x) = c(x-a)_+ + d(x-b)_+$ for each line of business, which generalises stop-loss as a special case. This piecewise linear form involves three segments: no coverage below a, partial coverage (at rate c) between a and b, and increased coverage (rate $c+d$) beyond b, suggesting that stop-loss may no longer be optimal. In Theorem 2.2(iii), the optimal solution is given by $g_{a,b}(x) = a\min(x,b)$, which encompasses the quota share as a special case, here referred to as *limited quota-share*. Interestingly, this may outperform regular quota share from the insurer's viewpoint, as it allows higher reimbursement for losses below a cap, while still maintaining affordability.

We now derive the optimal ceded loss functions using Theorem 2.2 and provide a result on the existence of optimal parameters.

Proposition 2.3. *The parameters for the optimal ceded loss functions $\mathbf{l}_{\mathbf{a},\mathbf{b}}, \mathbf{h}_{\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d}}, \mathbf{g}_{\mathbf{a},\mathbf{b}}$ in Theorem 2.2(i)-(iii) are determined by:*

(i) For $\mathbf{l}_{\mathbf{a},\mathbf{b}}$,

$$(\mathbf{a}^*, \mathbf{b}^*) \in \arg \inf_{(\mathbf{a},\mathbf{b}) \in \mathscr{A}(\mathbf{P})} \left\{ \inf_{\boldsymbol{\gamma} \in (1-\alpha)\Delta_n} L(\mathbf{a}, \mathbf{b}, \boldsymbol{\gamma}) \right\};$$

(ii) For $\mathbf{h}_{\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d}}$,

$$(\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*, \mathbf{d}^*) = \arg \inf_{(\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d}) \in \mathscr{A}_1(P)} \left\{ \inf_{\boldsymbol{\gamma} \in (1-\alpha)\Delta_n} H(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}, \boldsymbol{\gamma}) \right\};$$

(iii) For $\mathbf{g}_{\mathbf{a},\mathbf{b}}$,

$$(\mathbf{a}^*, \mathbf{b}^*) = \arg \inf_{(\mathbf{a},\mathbf{b}) \in \mathscr{A}_2(P)} \left\{ \inf_{\boldsymbol{\gamma} \in (1-\alpha)\Delta_n} G(\mathbf{a}, \mathbf{b}, \boldsymbol{\gamma}) \right\}.$$

Furthermore, the existence of these optimal parameters is ensured by the continuity of each π_i, for $i = 1, \ldots, n$.

In the case of $n = 2$, we adopt an alternative approach to obtain a simpler form for the optimal ceded loss functions. This is valuable in numerical analyses, as it allows the original non-convex problem to be reformulated into a series of convex ones, as shown in Section 2.6. Based on Makarov (1981) and Rüschendorf (1982), we establish:

Proposition 2.4. *Assume $n = 2$ and that F_1^{-1} and F_2^{-1} are continuous on $(0,1)$. Then,*

$$\inf_{(f_1,f_2)\in D^2(P)} \sup_{(X_1,X_2)\in\mathscr{E}_2} \mathrm{VaR}_\alpha(S_2^{\mathbf{f}}(X_1,X_2)) = \inf_{(a_1,a_2,b_1,b_2)\in\mathscr{A}(P)} \inf_{t\in[0,1-\alpha]} L_1(a_1,a_2,b_1,b_2,t),$$

where

$$\begin{aligned} L_1(a_1,a_2,b_1,b_2,t) = {} & \mathrm{VaR}_{\alpha+t}(X_1 - l_{a_1,b_1}(X_1)) + \mathrm{VaR}_{1-t}(X_2 - l_{a_2,b_2}(X_2)) \\ & + \pi_1(l_{a_1,b_1}(X_1)) + \pi_2(l_{a_2,b_2}(X_2)). \end{aligned}$$

Moreover, the pair $(l_{a_1,b_1}, l_{a_2,b_2})$ constitutes the optimal ceded loss function under the worst-case scenario if:

$$(a_1,a_2,b_1,b_2) \in \arg \inf_{(a_1,a_2,b_1,b_2)\in\mathscr{A}(P)} \left\{ \inf_{t\in[0,1-\alpha]} L_1(a_1,a_2,b_1,b_2,t) \right\}.$$

It is important to note that this section focuses solely on the budget constraint. However, additional constraints can be incorporated. As an example, a profitability condition can be introduced, such as: $\mathbb{E}(S_n^{\mathbf{f}}(X_1,\ldots,X_n)) \leqslant (1-\theta)P_1$, where $P_1 > 0$ denotes the total premium collected by the insurer over the period and $0 < \theta < 1$ signifies the insurer's target profit margin. The results derived in this section remain applicable if we redefine the admissible set $\mathscr{D}^n(P)$ as follows:

$$\left\{ \mathbf{f} \in \mathscr{D}^n : \sum_{i=1}^n \pi_i(f_i(X_i)) \leqslant P, \quad \mathbb{E}(S_n^{\mathbf{f}}(X_1,\ldots,X_n)) \leqslant (1-\theta)P_1 \right\},$$

provided that condition (iii) on π_i is strengthened to preserve the stop-loss order of the differences. Specifically, for any $Y, Z \in \mathscr{X}$, the inequality $\pi_i(Y) - \mathbb{E}(Y) \leqslant \pi_i(Z) - \mathbb{E}(Z)$ must hold whenever $\mathbb{E}\left((Y-d)_+\right) \leqslant \mathbb{E}\left((Z-d)_+\right) \quad$ for all $d \in \mathbb{R}$.

2.4 Optimal reinsurance under Range Value-at-Risk

In this section, we investigate the optimal reinsurance problem based on Range Value-at-Risk (RVaR). RVaR serves as a link between VaR and ES, as both can be viewed as special (i.e., limiting) cases of RVaR. Furthermore, RVaR belongs to a class of robust risk measures in the sense described by Cont et al. (2010). Following the approach used for VaR, we begin by generalising Theorem 1 from Blanchet et al. (2023), extending the setting from marginal distributions with decreasing tail densities to those whose tails are concave.

Proposition 2.5. *Let α, β be such that $0 \leqslant \beta < \beta + \alpha \leqslant 1$, and suppose that $\mathbf{F} \in (\mathscr{M}_{ce}^{1-\beta-\alpha})^n$. Then:*

$$\sup_{(X_1,\ldots,X_n)\in\mathscr{E}_n(\mathbf{F})} R_{\beta,\alpha}\left(\sum_{i=1}^n X_i\right) = \inf_{\boldsymbol{\gamma}\in(\beta+\alpha)\Delta_n, \gamma_0\geqslant\alpha} \sum_{i=1}^n R_{\gamma_i,\gamma_0}(F_i). \tag{2.2}$$

The following theorem shows that RVaR-based optimal reinsurance problems in domains $\mathscr{D}^n(P)$ and $\mathscr{D}_2^n(P)$ can be equivalently transformed into problems over the domains $\mathscr{L}^n(P)$ and $\mathscr{G}^n(P)$, respectively.

Theorem 2.6. *Let α, β satisfy $0 \leqslant \beta < \beta + \alpha \leqslant 1$. Then the following two statements hold:*

(i) If $\beta = 0$ or $n - \sum_{i=1}^n F_i((F_i)_+^{-1}(1-\beta-\alpha)) \leqslant \beta + \alpha$, then

$$\inf_{\mathbf{f}\in\mathscr{D}^n(P)} \sup_{(X_1,\ldots,X_n)\in\mathscr{E}_n(\mathbf{F})} R_{\beta,\alpha}\left(S_n^{\mathbf{f}}(X_1,\ldots,X_n)\right) = \inf_{(\mathbf{a},\mathbf{b})\in\mathscr{A}(\mathbf{P})} \inf_{\boldsymbol{\gamma}\in(\beta+\alpha)\Delta_n,\gamma_0\geqslant\alpha} L(\mathbf{a},\mathbf{b},\boldsymbol{\gamma}),$$

where $L(\mathbf{a},\mathbf{b},\boldsymbol{\gamma})$ is defined in Theorem 2.2.

(ii) If $\mathbf{F} \in (\mathscr{M}_{ce}^{1-\beta-\alpha})^n$, then

$$\inf_{\mathbf{f}\in\mathscr{D}_2^{\mathbf{n}}(\mathbf{P})} \sup_{(X_1,\ldots,X_n)\in\mathscr{E}_n(\mathbf{F})} R_{\beta,\alpha}\left(S_n^{\mathbf{f}}(X_1,\ldots,X_n)\right) = \inf_{(\mathbf{a},\mathbf{b})\in\mathscr{A}_2(P)} \inf_{\boldsymbol{\gamma}\in(\beta+\alpha)\Delta_n,\gamma_0\geqslant\alpha} G(\mathbf{a},\mathbf{b},\boldsymbol{\gamma}),$$

where $G(\mathbf{a},\mathbf{b},\boldsymbol{\gamma})$ is also specified in Theorem 2.2.

Theorem 2.6 indicates that the structure of the optimal reinsurance strategy under the RVaR framework mirrors that under the VaR model, except that the results are not available for the case $n = 2$ or when the marginal distributions have convex tails.

The result in part (i) of Theorem 2.6 partially extends the findings of Cheung et al. (2014), which deal with convex risk measures and expectation-based premium principles. Furthermore, it is notable that, under convex risk measures, the worst-case dependence between multiple risks results in comonotonicity. In contrast, RVaR, being non-convex except when it reduces to ES, yields a worst-case dependence structure characterised by a mixture of mutual exclusivity and joint mixability, implying a form of negative dependence. For further details, see Blanchet et al. (2023) and Jakobsons et al. (2016).

We now present a result that specifies the optimal reinsurance strategies and the conditions for their existence.

Proposition 2.7. *The parameters defining the optimal ceded loss functions* $\mathbf{l}_{\mathbf{a},\mathbf{b}}$ *and* $\mathbf{g}_{\mathbf{a},\mathbf{b}}$ *in parts (i) and (ii) of Theorem 2.6 are determined as follows:*

(i) For $\mathbf{l}_{\mathbf{a},\mathbf{b}}$*, the optimal parameters satisfy*

$$(\mathbf{a}^*,\mathbf{b}^*) \in \arg \inf_{(\mathbf{a},\mathbf{b})\in\mathscr{A}(P)} \left\{ \inf_{\boldsymbol{\gamma}\in(\beta+\alpha)\Delta_n,\gamma_0\geqslant\alpha} L(\mathbf{a},\mathbf{b},\boldsymbol{\gamma}) \right\};$$

(ii) For $\mathbf{g}_{\mathbf{a},\mathbf{b}}$*, we have*

$$(\mathbf{a}^*,\mathbf{b}^*) = \arg \inf_{(\mathbf{a},\mathbf{b})\in\mathscr{A}_2(P)} \left\{ \inf_{\boldsymbol{\gamma}\in(\beta+\alpha)\Delta_n,\gamma_0\geqslant\alpha} G(\mathbf{a},\mathbf{b},\boldsymbol{\gamma}) \right\}.$$

Additionally, the existence of such optimal parameters is ensured by the continuity of the pricing functions π_i *for* $i = 1,\ldots,n$.

2.5 Quota share reinsurance under dependence uncertainty

In this section, we examine the most basic form of reinsurance, namely quota share reinsurance, where each ceded loss is a fixed proportion of the original loss. Specifically, the ceded loss takes the form $f_i(X_i) = w_i X_i$, for $i = 1,\ldots,n$, where the retention levels w_i are in $[0,1]$. For further details on quota share reinsurance, see, e.g., Albrecher et al. (2017). We say that a premium principle π is homogeneous if $\pi(\lambda X) = \lambda\pi(X)$ holds for all $\lambda > 0$ and $X \in \mathscr{X}$.

Throughout this section, we assume that π satisfies the following three properties: distribution invariance, safety loading, and homogeneity. These properties are satisfied by many commonly used premium principles such as the expected principle, standard deviation principle, Wang's principle, and the Dutch principle; see Young (2004) for more examples.

Under quota share reinsurance, the insurer's total retained risk is given by:

$$S^{\mathbf{w}}(X_1,\ldots,X_n) = \sum_{i=1}^{n}\left((1-w_i)X_i + w_i\pi_i(X_i)\right),$$

where $\mathbf{w} = (w_1,\ldots,w_n)$ denotes the vector of quota share levels. We define the admissible domain for $\mathbf{w}$ under the budget constraint as:

$$\Lambda_n(P) = \left\{\mathbf{w}\in[0,1]^n : \sum_{i=1}^{n} w_i\pi_i(X_i) \leqslant P\right\}.$$

Observe that $\mathbf{0}\in\Lambda_n(P)$ and that $\Lambda_n(P)$ is a convex and compact set, with its boundary denoted by $\partial\Lambda_n(P)$. By applying Propositions 2.1 and 2.5, we obtain the following results:

Proposition 2.8. *Let $\alpha\in(0,1)$. If either $n=2$, or $\mathbf{F}\in(\mathscr{M}_{cx}^{\alpha})^n\cup(\mathscr{M}_{ce}^{\alpha})^n$ and each $F_i^{-1}(\cdot)$ is continuous on $(0,1)$, then*

$$\begin{aligned}
&\inf_{\mathbf{w}\in\Lambda_n(P)}\sup_{(X_1,\ldots,X_n)\in\mathscr{E}_n(\mathbf{F})}\mathrm{VaR}_\alpha\left(S^{\mathbf{w}}(X_1,\ldots,X_n)\right)\\
&=\inf_{\mathbf{w}\in\partial\Lambda_n(P)}\inf_{\boldsymbol{\gamma}\in(1-\alpha)\Delta_n}\sum_{i=1}^{n}\left((1-w_i)R_{\gamma_i,\gamma_0}(X_i)+w_i\pi_i(X_i)\right).
\end{aligned}$$

Furthermore, there exists an optimal quota-share allocation $\mathbf{w}^*\in\partial\Lambda_n(P)$.

Proposition 2.9. *Given any α,β with $0\leqslant\beta<\beta+\alpha\leqslant 1$, if $\beta=0$ or $\mathbf{F}\in(\mathscr{M}_{ce}^{1-\beta-\alpha})^n$, then*

$$\begin{aligned}
&\inf_{\mathbf{w}\in\Lambda_n(P)}\sup_{(X_1,\ldots,X_n)\in\mathscr{E}_n(\mathbf{F})}R_{\beta,\alpha}\left(S^{\mathbf{w}}(X_1,\ldots,X_n)\right)\\
&=\inf_{\mathbf{w}\in\partial\Lambda_n(P)}\inf_{\boldsymbol{\gamma}\in(\beta+\alpha)\Delta_n,\gamma_0\geqslant\alpha}\sum_{i=1}^{n}\left((1-w_i)R_{\gamma_i,\gamma_0}(X_i)+w_i\pi_i(X_i)\right).
\end{aligned}$$

Again, there is an optimal solution $\mathbf{w}^*\in\partial\Lambda_n(P)$.

Now suppose that the budget is sufficiently large, i.e., $P\geqslant\sum_{i=1}^{n}\pi_i(X_i)$. In this case, we have $\Lambda_n(P)=[0,1]^n$. Define $\Lambda_n^0=\{\mathbf{w}\in[0,1]^n : w_i\in\{0,1\},\ i=1,\ldots,n\}$. Then, for each line of business, the optimal quota share level is either full reinsurance ($w_i=1$) or no reinsurance ($w_i = 0$). This suggests that, under such a budget, the insurer adopts only extreme positions, potentially declining to reinsure certain business lines. Interestingly, Bernard et al. (2020) provides a parallel result where, based on preference and risk profile, the insurer chooses not to offer coverage to some policyholders.

Corollary 2.10. *If $P \geqslant \sum_{i=1}^{n} \pi_i(X_i)$, then the optimal quota-share solutions $\mathbf{w}^*$ in both propositions 2.8 -2.9 satisfy $\mathbf{w}^* \in \Lambda_n^0$.*

2.6 Examples

Our principal findings demonstrate that determining the optimal ceded-loss functions in the worst-case reinsurance model with dependence uncertainty reduces to minimising a fixed deterministic function. To illustrate the utility of these results, this section presents both analytical and numerical resolutions of selected multi-risk optimal reinsurance problems where dependence uncertainty is present. Given that both VaR and ES are standard risk measures for insurance risk quantification and regulatory capital determination, and noting that ES has been studied in Cheung et al. (2014), we focus on the VaR setting.

2.6.1 Analytical solutions

We first address the analytical solution of the core optimisation problem (2.1) for the case $\rho = \text{VaR}$ and $n = 2$, i.e.,

$$\min_{\mathbf{f} \in \mathscr{D}^2(\mathbf{P})} \max_{(X_1, X_2) \in \mathscr{E}_2(\mathbf{F})} \text{VaR}_\alpha\left(S_2^{\mathbf{f}}(X_1, X_2)\right). \tag{2.3}$$

Proposition 2.4 allows us to reformulate (2.3) equivalently as:

$$\min_{(a_1, a_2, b_1, b_2) \in \mathscr{A}(P)} \min_{t \in [0, 1-\alpha]} L_1(a_1, a_2, b_1, b_2, t), \tag{2.4}$$

where

$$L_1(a_1,a_2,b_1,b_2,t) = \mathrm{VaR}_{\alpha+t}\big(X_1 - l_{a_1,b_1}(X_1)\big) + \mathrm{VaR}_{1-t}\big(X_2 - l_{a_2,b_2}(X_2)\big) \tag{2.5}$$

$$+ \pi_1\big(l_{a_1,b_1}(X_1)\big) + \pi_2\big(l_{a_2,b_2}(X_2)\big). \tag{2.6}$$

Proposition 2.11. *Assume for $i = 1,2$ that π_i satisfies distribution invariance, risk loading, and preserves stop-loss ordering. If $\mathscr{A}(P) \neq \emptyset$, then for any $t \in [0, 1-\alpha]$:*

(i) $\pi_1\big(l_{u_1,v_1}(X_1)\big) + \pi_2\big(l_{u_2,v_2}(X_2)\big) \leqslant \pi_1\big(l_{a_1,b_1}(X_1)\big) + \pi_2\big(l_{a_2,b_2}(X_2)\big)$*;*

(ii) $(a_1,a_2,b_1,b_2) \in \mathscr{A}(P)$*;*

(iii) $L_1(u_1,u_2,v_1,v_2,t) \leqslant L_1(a_1,a_2,b_1,b_2,t)$ *for all* $(a_1,a_2,b_1,b_2) \in \mathscr{A}(P)$*.*

Here, $v_1 = \mathrm{VaR}_{\alpha+t}(X_1)$, $v_2 = \mathrm{VaR}_{1-t}(X_2)$, $u_1 = v_1 - l_{a_1,b_1}(v_1)$, and $u_2 = v_2 - l_{a_2,b_2}(v_2)$. Moreover,

$$\min_{(a_1,a_2,b_1,b_2)\in\mathscr{A}(P)} \min_{t\in[0,1-\alpha]} L_1(a_1,a_2,b_1,b_2,t) = \min_{(u_1,u_2)\in\mathscr{A}_v(P,v_1,v_2)} \min_{t\in[0,1-\alpha]} L_2(u_1,u_2,t),$$

with $L_2(u_1,u_2,t) = \sum_{i=1}^{2}\Big(u_i + \pi_i\big(l_{u_i,v_i}(X_i)\big)\Big)$, and

$$\mathscr{A}_v(P,v_1,v_2) := \{(u_1,u_2)\colon \ell_{u_i,v_i} \in \mathscr{D}^2(P),\ 0 \leqslant u_i \leqslant v_i,\ i = 1,2\}.$$

Example 2.12. Suppose each X_i is exponentially distributed with mean $\lambda_i > 0$, with a density function $f_i(x) = \frac{1}{\lambda_i}e^{-\frac{x}{\lambda_i}}$, $\quad F_i(x) = 1 - e^{-\frac{x}{\lambda_i}}$, for $x > 0$ and $i = 1,2$. Assume that each premium principle is in the form $\pi_i(X_i) = w_i + (1+\delta_i)\mathbb{E}[X_i]$, with $w_i \geqslant 0$, $\delta_i > 0$. Note that, if $w_1 + w_2 > P$, then $\mathscr{A}_v(P,v_1,v_2) = \emptyset$. We thus restrict to $w_1 + w_2 \leqslant P$ to consider non-trivial cases.

Proposition 2.13. *Under $w_1 + w_2 \leqslant P$, Table 2.1 provides the optimal solution*

$$(a_1^*,a_2^*,b_1^*,b_2^*,t^*) = \arg\min_{(a_1,a_2,b_1,b_2)\in\mathscr{A}(P)} \min_{t\in[0,1-\alpha]} L_1(a_1,a_2,b_1,b_2,t)$$

in the following cases:

(i) $(1+\delta_1)\lambda_1 = (1+\delta_2)\lambda_2$, *and* $\sum_i (w_i+\lambda_i) - (1+\delta_1)(1-\alpha)\lambda_1 \leqslant P$.

(iia) $(1+\delta_1)\lambda_1 > (1+\delta_2)\lambda_2$, $(1-\alpha)(1+\delta_1) < 1$, *and* $\sum_i (w_i+\lambda_i) - (1+\delta_1)(1-\alpha)\lambda_1 \leqslant P$.

(iib) *Same inequality on* $(1+\delta_1)\lambda_1 > (1+\delta_2)\lambda_2$, *but* $(1-\alpha)(1+\delta_1) \geqslant 1$, *and* $w_2+\lambda_2 \leqslant P$.

(iiia) $(1+\delta_1)\lambda_1 < (1+\delta_2)\lambda_2$, $(1-\alpha)(1+\delta_2) < 1$, *and* $\sum_i (w_i+\lambda_i) - (1+\delta_2)(1-\alpha)\lambda_2 \leqslant P$.

(iiib) *Same as above except* $(1-\alpha)(1+\delta_2) \geqslant 1$ *and* $w_1+\lambda_1 \leqslant P$.

Case	Sub-case	$(a_1^*, a_2^*, b_1^*, b_2^*, t^*)$
(i)		$t^* \in \mathscr{T}$; $a_i^* = -\lambda_i \ln\big(1/(1+\delta_i)\big)$, $i=1,2$; $b_1^* = -\lambda_1 \ln(1-\alpha-t^*)$, $b_2^* = -\lambda_2 \ln t^*$.
(ii)	(iia)	$t^*=0$, $a_i^* = -\lambda_i \ln\big(1/(1+\delta_i)\big)$, $b_1^* = -\lambda_1 \ln(1-\alpha)$, $b_2^* = \infty$;
	(iib)	$t^*=0$, $a_1^* = b_1^* = -\lambda_1 \ln(1-\alpha)$, $a_2^* = -\lambda_2 \ln\big(1/(1+\delta_2)\big)$, $b_2^* = \infty$.
(iii)	(iiia)	$t^* = 1-\alpha$, $a_i^* = -\lambda_i \ln\big(1/(1+\delta_i)\big)$, $b_1^* = \infty$, $b_2^* = -\lambda_2 \ln(1-\alpha)$;
	(iiib)	$t^* = 1-\alpha$, $a_1^* = -\lambda_1 \ln\big(1/(1+\delta_1)\big)$, $b_1^* = \infty$, $a_2^* = b_2^* = -\lambda_2 \ln(1-\alpha)$.

Table 2.1: Optimal solutions in Example 2.12.

For other configurations, a numerical procedure is necessary to determine the optimal.

The table yields several insights. Cases (i)–(iii) are distinguished by comparing $(1+\delta_1)\lambda_1$ to $(1+\delta_2)\lambda_2$. Since $\lambda_i = \mathbb{E}[X_i]$ and δ_i is the risk-loading factor, the product $(1+\delta_i)\lambda_i$ reflects the relative cost of reinsuring X_i. If reinsuring X_1 is more expensive ($(1+\delta_1)\lambda_1 > (1+\delta_2)\lambda_2$), then case (ii) applies: $b_1^* < \infty$ and $b_2^* = \infty$, meaning that X_1's reinsurance is capped but X_2's is not. Conversely, if $(1+\delta_1)\lambda_1 < (1+\delta_2)\lambda_2$, then in case (iii) X_2 is capped while X_1 is uncapped. When $(1+\delta_1)\lambda_1 = (1+\delta_2)\lambda_2$ (case i), reinsuring both risks is equally costly and the solution is non-unique, given that it depends on the insurer's risk tolerance α and the loading factors.

Sub-case (iib) occurs when $(1+\delta_1)\lambda_1 > (1+\delta_2)\lambda_2$ and $(1-\alpha)(1+\delta_1) \geqslant 1$ (i.e. α small and/or δ_1 large). This means cherishing high risk tolerance or high reinsurance premium for X_1, making it optimal not to reinsure X_1 (i.e. $a_1^* = b_1^*$). Similarly, sub-case (iiib) implies $a_2^* = b_2^*$ when α is small and/or δ_2 is large.

2.6.2 Numerical examples

In this section, we address the optimisation problem (2.4) using a numerical method, specifically the grid search approach. To begin, we discretise each random variable X_i into m observations, denoted by the vector $\mathbf{x}_i := (x_{i1}, x_{i2}, \ldots, x_{im})^T$, for $i = 1,2$. These observations x_{ij}, where $i = 1,2$ and $j = 1,2,\ldots,m$, can be generated either from historical data or via Monte Carlo simulation.

The corresponding reinsurance indemnity function $l_{a_i,b_i}(X_i)$ is then represented by the vector $\mathbf{y}_i := (y_{i1}, y_{i2}, \ldots, y_{im})^T$, where each y_{ij} denotes the reinsurance payment associated with x_{ij}, for all $i = 1,2$ and $j = 1,2,\ldots,m$. Without loss of generality, we assume that the observations are sorted in ascending order, i.e., $x_{i1} \leqslant x_{i2} \leqslant \ldots \leqslant x_{im}$ for each $i = 1,2$. Due to the monotonicity of the function $l_{a_i,b_i}(X_i)$, it follows that the reinsurance payments also satisfy $y_{i1} \leqslant y_{i2} \leqslant \ldots \leqslant y_{im}$ for $i = 1,2$.

Example 2.14. We adopt the setup of Example 2.12 with parameters $\lambda_1 = 8000$, $\lambda_2 = 3000$, $\delta_1 = 0.8$, $\delta_2 = 0.3$, illustrating case (ii). A Monte Carlo sample of size $m = 200$ is generated for each X_i.

For each fixed $t \in [0, 1-\alpha]$, the local optimum of (2.4) is computed by solving the linear programme:

$$\begin{aligned} &\min_{(\mathbf{y}_1,\mathbf{y}_2)\in\mathbb{R}^m\times\mathbb{R}^m} x_{1q_1} - y_{1q_1} + x_{2q_2} - y_{2q_2} + \sum_{i=1}^{2}\left(w_i + \tfrac{1+\delta_i}{m}\mathbf{1}^T\mathbf{y}_i\right) \\ \text{s.t.}\quad & \mathbf{0} \leqslant \mathbf{A}\mathbf{y}_i \leqslant \mathbf{A}\mathbf{x}_i, \ \text{for } i = 1,2, \\ & \sum_{i=1}^{2}\left(w_i + \tfrac{1+\delta_i}{m}\mathbf{1}^T\mathbf{y}_i\right) \leqslant P, \end{aligned} \tag{2.7}$$

where $\mathbf{1}$ is the m-vector of ones, $q_1 = \lceil m(\alpha + t)\rceil$, $q_2 = \lceil m(1-t)\rceil$, and $\mathbf{A}$ is the difference matrix enforcing monotonicity.

A global optimum is obtained by evaluating (2.7) over a grid of $t \in [0, 1-\alpha]$. Since q_1 and q_2 take only finitely many values, this is finite.

Figure 2.1 depicts the numerical results. The left side of the figure displays scatter plots of the optimal indemnity functions at $\alpha = 0.95$ (sub-case iia), revealing that $l_{a_1^*,b_1^*}(X_1)$ is

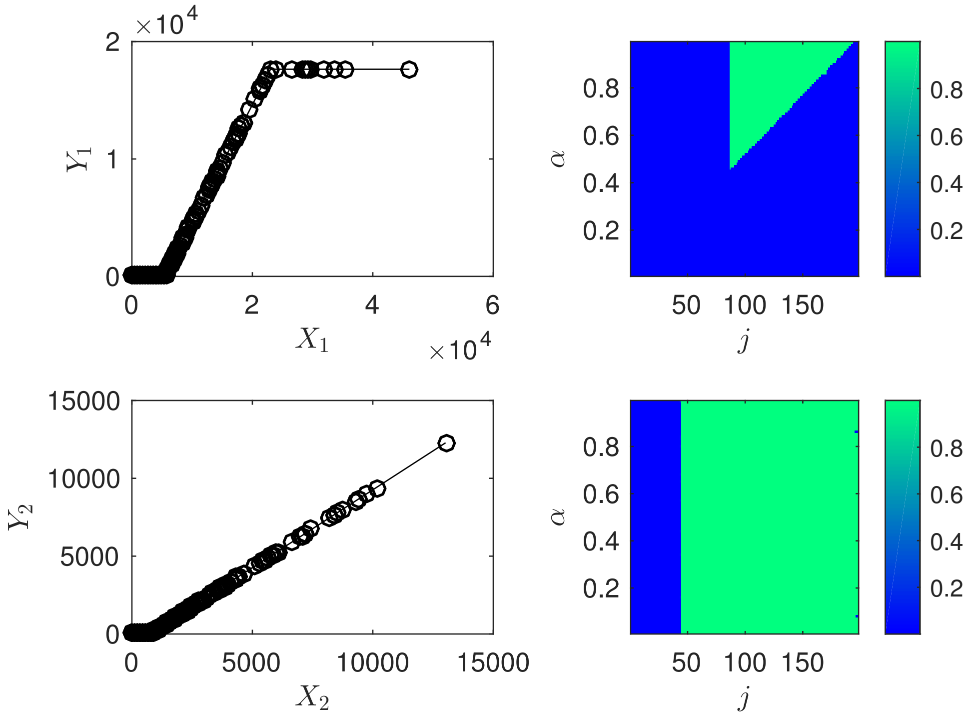

Figure 2.1: Illustration of the numerical optimal reinsurance solution found in Example 2 using scatter plots and heat maps. (Scatter plots: optimal solution for $\alpha = 0.95$; Heat maps: the gradient (colour scale) of the scatter plots for optimal solutions under different α values.)

a limited stop-loss capped at its $\mathrm{VaR}_{0.95}$, whereas $l_{a_2^*,b_2^*}(X_2)$ acts as an uncapped standard stop-loss ($b_2^* = \infty$).

It is also worthwhile to examine how the shape of $l_{a_1^*,b_1^*}(X_1)$ and $l_{a_2^*,b_2^*}(X_2)$ varies with respect to α, as this illustrates how the insurer's reinsurance decision changes in response to different levels of risk aversion. To this end, we compute $l_{a_1^*,b_1^*}(X_1)$ and $l_{a_2^*,b_2^*}(X_2)$ for 99 distinct values of α (ranging from $\alpha = 0.01$ to $\alpha = 0.99$ with a step size of 0.01). To provide a more compact visualisation of these 99 pairs of solutions, heat maps are used, as shown in the right-hand panels of Figure 2.1. Each heat map displays the gradient of $l_{a_i^*,b_i^*}(X_i)$ evaluated at the points X_{ij}, where $j = 1, 2, \ldots, 199$. Due to the definition of the function l, the gradient in different segments of $l_{a_i^*,b_i^*}(X_i)$ is either 0 or 1. As observed from the heat maps, $l_{a_2^*,b_2^*}(X_2)$ consistently retains the same shape, transitioning from a horizontal line to an upward-sloping line with a turning point located at $a_2^* = -\lambda_2 \ln\left(\frac{1}{1+\delta_2}\right)$. In contrast, the shape of $l_{a_1^*,b_1^*}(X_1)$ evolves as α changes. When α is large (i.e., $\alpha > 1 - \frac{1}{1-\delta_1}$), there are two turning points, as the gradient changes from 0 to 1 and then back to 0 once X_{1j} reaches $\mathrm{VaR}_\alpha(X_1)$. Conversely, when α is small, the gradient remains 0 for all X_{1j}, implying that $l_{a_1^*,b_1^*}(X_1) = 0$, which aligns with the analytical result derived in Example 1.

2.7 Proofs of results

All the proofs of the results in Sections 2.3-2.6 are presented in this section.

2.7.1 Proofs of results in Section 2.3.

To prove Proposition 2.1, the following lemma is applied.

Lemma 2.15. *For $\mathbf{F}^{(k)} = (F_1^{(k)}, \dots, F_n^{(k)}), k \geqslant 1$ and $\mathbf{F} = (F_1, \dots, F_n)$, suppose that, for each i, $(F_i^{(k)})_+^{-1}(t) \to (F_i)_+^{-1}(t)$ uniformly over $t \in (\alpha, 1)$, then:*

$$\lim_{k\to\infty} \sup_{(X_1,\dots,X_n)\in\mathscr{E}_n(\mathbf{F}^{(k)})} \mathrm{VaR}_\alpha^+\left(\sum_{i=1}^n X_i\right) = \sup_{(X_1,\dots,X_n)\in\mathscr{E}_n(\mathbf{F})} \mathrm{VaR}_\alpha^+\left(\sum_{i=1}^n X_i\right).$$

Proof. By Theorem 4.6 of Bernard et al. (2014), we have:

$$\sup_{(X_1,\dots,X_n)\in\mathscr{E}_n(\mathbf{F}^{(k)})} \mathrm{VaR}_\alpha^+\left(\sum_{i=1}^n X_i\right) = \sup_{U_1,\dots,U_n\sim U[0,1]} \text{ess-inf} \sum_{i=1}^n (F_i^{(k)})_+^{-1}(\alpha + (1-\alpha)U_i),$$

and

$$\sup_{(X_1,\dots,X_n)\in\mathscr{E}_n(\mathbf{F})} \mathrm{VaR}_\alpha^+\left(\sum_{i=1}^n X_i\right) = \sup_{U_1,\dots,U_n\sim U[0,1]} \text{ess-inf} \sum_{i=1}^n (F_i)_+^{-1}(\alpha + (1-\alpha)U_i).$$

It follows from the uniform convergence that:

$$\begin{aligned}
&\sup_{t_1,\dots,t_n\in(0,1)} \left| \sum_{i=1}^n (F_i^{(k)})_+^{-1}(\alpha + (1-\alpha)t_i) - \sum_{i=1}^n (F_i)_+^{-1}(\alpha + (1-\alpha)t_i) \right| \\
&\leqslant \sum_{i=1}^n \sup_{t\in(0,1)} \left| (F_i^{(k)})_+^{-1}(\alpha + (1-\alpha)t) - (F_i)_+^{-1}(\alpha + (1-\alpha)t) \right| \to 0
\end{aligned}$$

as $k \to \infty$. Hence,

$$\begin{aligned}
&\lim_{k\to\infty} \sup_{U_1,\ldots,U_n\sim U[0,1]} \text{ess-inf} \sum_{i=1}^{n} (F_i^{(k)})^{-1}(\alpha+(1-\alpha)U_i) \\
&= \sup_{U_1,\ldots,U_n\sim U[0,1]} \text{ess-inf} \sum_{i=1}^{n} (F_i^{-1})_+(\alpha+(1-\alpha)U_i).
\end{aligned}$$

This completes the proof. $\square$

Proof of Proposition 2.1. We first focus on $\mathbf{F} \in (\mathscr{M}_{cx}^{\alpha})^n$. For $i = 1,\ldots,n$, and $k \geqslant 1$, if $F_i(F_i^{-1}(1)-) < 1$, let

$$F_i^{(k)}(x) = \begin{cases} F_i(x), & x < F_i^{-1}(1) - 1/k \\ F_i(x) + k(1 - F_i(F_i^{-1}(1)-))(x - F_i^{-1}(1) + 1/k), & F_i^{-1}(1) - 1/k \leqslant x < F_i^{-1}(1) \\ 1, & x \geqslant F_i^{-1}(1) \end{cases},$$

and if $F_i(F_i^{-1}(1)-) = 1$, let $F_i^{(k)} = F_i$. Clearly, each $F_i^{(k)}$ admits an increasing density beyond its α-quantile. Hence, in light of Theorem 2 of Blanchet et al. (2023),

$$\sup_{(X_1,\ldots,X_n)\in\mathscr{E}_n(\mathbf{F}^{(k)})} \text{VaR}_{\alpha}^{+}\left(\sum_{i=1}^{n} X_i\right) = \inf_{\boldsymbol{\gamma}\in(1-\alpha)\Delta_n} \sum_{i=1}^{n} R_{\gamma_i,\gamma_0}(F_i^{(k)}).$$

Note that $\sup_{t\in(0,1)} |(F_i^{(k)})_+^{-1}(t) - (F_i)_+^{-1}(t)| \leqslant 1/k$, which implies that $(F_i^{(k)})_+^{-1}(t) \to (F_i)_+^{-1}(t)$ uniformly on $(0,1)$. Consequently, by Lemma 2.15, we have:

$$\begin{aligned}
\sup_{(X_1,\ldots,X_n)\in\mathscr{E}_n(\mathbf{F})} \text{VaR}_{\alpha}\left(\sum_{i=1}^{n} X_i\right) &= \lim_{k\to\infty} \sup_{(X_1,\ldots,X_n)\in\mathscr{E}_n(\mathbf{F}^{(k)})} \text{VaR}_{\alpha}\left(\sum_{i=1}^{n} X_i\right) \\
&= \lim_{k\to\infty} \inf_{\boldsymbol{\gamma}\in(1-\alpha)\Delta_n} \sum_{i=1}^{n} R_{\gamma_i,\gamma_0}(F_i^{(k)}) \\
&= \inf_{\boldsymbol{\gamma}\in(1-\alpha)\Delta_n} \sum_{i=1}^{n} R_{\gamma_i,\gamma_0}(F_i),
\end{aligned}$$

where the last inequality is implied by the fact that $\sup_{\boldsymbol{\gamma}\in(1-\alpha)\Delta_n} |R_{\gamma_i,\gamma_0}(F_i^{(k)}) - R_{\gamma_i,\gamma_0}(F_i)| \leqslant 1/k$. Therefore, the claim holds for $\mathbf{F} \in (\mathscr{M}_{cx}^{\alpha})^n$.

We next focus on the case $\mathbf{F} \in (\mathscr{M}_{ce}^{\alpha})^n$. For $i = 1,\ldots,n$ and $k \geqslant 1$, if $q_i := F_i((F_i)_+^{-1}(\alpha)) -$

$\alpha > 0$, let

$$F_i^{(k)}(x) = \begin{cases} F_i(x), & x < (F_i)_+^{-1}(\alpha) \\ F_i(x) - q_i + kq_i(x - (F_i)_+^{-1}(\alpha)), & (F_i)_+^{-1}(\alpha) \leqslant x < (F_i)_+^{-1}(\alpha) + 1/k \\ F_i(x), & x \geqslant (F_i)_+^{-1}(\alpha) + 1/k \end{cases},$$

and if $q_i = 0$, let $F_i^{(k)} = F_i$. Clearly, each $F_i^{(k)}$ admits a decreasing density beyond its α-quantile. By Theorem 2 of Blanchet et al. (2023), we have:

$$\sup_{(X_1,\ldots,X_n)\in\mathscr{E}_n(\mathbf{F}^{(k)})} \mathrm{VaR}_\alpha^+\left(\sum_{i=1}^n X_i\right) = \inf_{\boldsymbol{\gamma}\in(1-\alpha)\Delta_n} \sum_{i=1}^n R_{\gamma_i,\gamma_0}(F_i^{(k)}).$$

The rest of the proof is the same as the proof of $\mathbf{F} \in (\mathscr{M}_{cx}^{\alpha})^n$. Hence, we omit it. □

Proof of Theorem 2.2. By Lemma 4.5 of Bernard et al. (2014), the continuity of $F_1^{-1},\ldots,F_n^{-1}$ over $(0,1)$ implies that for $\alpha \in (0,1)$,

$$\sup_{(X_1,\ldots,X_n)\in\mathscr{E}_n(\mathbf{F})} \mathrm{VaR}_\alpha^+(S_n^{\mathbf{f}}(X_1,\ldots,X_n)) = \sup_{(X_1,\ldots,X_n)\in\mathscr{E}_n(\mathbf{F})} \mathrm{VaR}_\alpha(S_n^{\mathbf{f}}(X_1,\ldots,X_n)).$$

We first focus on case (i). In light of Theorem 2 of Blanchet et al. (2023), we have:

$$\begin{aligned} &\inf_{\mathbf{f}\in\mathscr{D}^{\mathbf{n}}(\mathbf{P})} \sup_{(X_1,\ldots,X_n)\in\mathscr{E}_n(\mathbf{F})} \mathrm{VaR}_\alpha(S_n^{\mathbf{f}}(X_1,\ldots,X_n)) \\ &= \inf_{\mathbf{f}\in\mathscr{D}^{\mathbf{n}}(\mathbf{P})} \inf_{\boldsymbol{\gamma}\in(1-\alpha)\Delta_n} \sum_{i=1}^n \left\{R_{\gamma_i,\gamma_0}(X_i - f_i(X_i)) + \pi_i(f_i(X_i))\right\} \\ &= \inf_{\boldsymbol{\gamma}\in(1-\alpha)\Delta_n} \inf_{\mathbf{f}\in\mathscr{D}^{\mathbf{n}}(\mathbf{P})} \sum_{i=1}^n \left\{R_{\gamma_i,\gamma_0}(X_i - f_i(X_i)) + \pi_i(f_i(X_i))\right\}. \end{aligned}$$

Next we show that $\mathscr{D}^n(P)$ can be replaced by $\mathscr{L}^n(P)$. Evidently,

$$\inf_{\mathbf{f}\in\mathscr{D}^{\mathbf{n}}(\mathbf{P})} \sum_{i=1}^n \left\{R_{\gamma_i,\gamma_0}(X_i - f_i(X_i)) + \pi_i(f_i(X_i))\right\} \leqslant \inf_{(\mathbf{a},\mathbf{b})\in\mathscr{A}(\mathbf{P})} L(\mathbf{a},\mathbf{b},\boldsymbol{\gamma}). \tag{2.8}$$

Next, we show the converse inequality. Let $a_i = F_i^{-1}(1-\gamma_0-\gamma_i) - f_i(F_i^{-1}(1-\gamma_0-\gamma_i))$ and

$b_i \in [\mathrm{VaR}_{1-\gamma_0-\gamma_i}(X_i), \infty]$ be the solution of $g_i(s) = 0$, where

$$g_i(s) := \int_{1-\gamma_0-\gamma_i}^{1} l_{a_i,s}(F_i^{-1}(t))\mathrm{d}t - \int_{1-\gamma_0-\gamma_i}^{1} f_i(F_i^{-1}(t))\mathrm{d}t.$$

Note that b_i may take the value of ∞. The existence of b_i is guaranteed by the continuity of g_i over $[a_i, \infty]$ and the fact that $g_i(F_i^{-1}(1-\gamma_0-\gamma_i)) \leqslant 0$ and $g_i(\infty) \geqslant 0$. One can easily check that, for the above, having chosen a_i and b_i,

$$\int_s^1 l_{a_i,b_i}(F_i^{-1}(t))\mathrm{d}t \leqslant \int_s^1 f_i(F_i^{-1}(t))\mathrm{d}t, \tag{2.9}$$

for $s \in (0,1)$. By Theorem 2.57 in Föllmer and Schied (2016), we have that for any $u \in \mathbb{R}$,

$$\mathbb{E}((l_{a_i,b_i}(X_i) - u)_+) \leqslant \mathbb{E}((f_i(X_i) - u)_+).$$

Hence, it follows from the property (iii) of π_i that: $\pi_i(l_{a_i,b_i}(X_i)) \leqslant \pi_i(f_i(X_i))$. Moreover, combination of (2.9) and the fact that $g_i(b_i) = 0$ yields

$$R_{\gamma_i,\gamma_0}(l_{a_i,b_i}(X_i)) \geqslant R_{\gamma_i,\gamma_0}(f_i(X_i)).$$

Consequently,

$$\sum_{i=1}^{n} \left\{ R_{\gamma_i,\gamma_0}(X_i - f_i(X_i)) + \pi_i(f_i(X_i)) \right\} \geqslant L(\mathbf{a}, \mathbf{b}, \boldsymbol{\gamma}).$$

We next verify $(\mathbf{a}, \mathbf{b}) \in \mathscr{A}(P)$ for our constructed $(\mathbf{a}, \mathbf{b})$. Using the fact that $\mathbf{f} \in \mathscr{D}^{\mathbf{n}}(\mathbf{P})$, we have:

$$\sum_{i=1}^{n} \pi_i(l_{a_i,b_i}(X_i)) \leqslant \sum_{i=1}^{n} \pi_i(f_i(X_i)) \leqslant P.$$

Hence, for any $\mathbf{f} \in \mathscr{D}^{\mathbf{n}}(\mathbf{P})$, there exists $(\mathbf{a}, \mathbf{b}) \in \mathscr{A}(\mathbf{P})$ such that:

$$\sum_{i=1}^{n} \left\{ R_{\gamma_i,\gamma_0}(X_i - f_i(X_i)) + \pi_i(f_i(X_i)) \right\} \geqslant L(\mathbf{a}, \mathbf{b}, \boldsymbol{\gamma}),$$

which combined with (2.8) implies that (2.8) is an equality. This completes the proof of (i).

We next consider case (ii). Note that for $\mathbf{f} \in \mathscr{D}_{\mathbf{1}}^{\mathbf{n}}(\mathbf{P})$, we have $T_{f_i}(0) = 0$, and $T_{f_i}(x)$ is a concave, continuous, and increasing function on $[0, \infty)$. If $T_{f_i} = 0$, the distribution of $T_{f_i}(X_i)$ is convex beyond its α-quantile. Next we suppose $T_{f_i} \neq 0$. Direct computation gives:

$$\mathbb{P}(T_{f_i}(X_i) \leqslant x) = \mathbb{P}\left(X_i \leqslant T_{f_i}^{-1}(x)\right) = F_i(T_{f_i}^{-1}(x)),\ x \geqslant 0,$$

where $T_{f_i}^{-1}(x) = \sup\{t \geqslant 0 : T_{f_i}(t) \leqslant x\}$ for $x \geqslant 0$. Note that $T_{f_i}^{-1}$ is convex on $[0, T_{f_i}(\infty)]$. Hence $\frac{(F_i(T_{f_i}^{-1}(x))-\alpha)_+}{1-\alpha}$ is convex on $(-\infty, T_{f_i}(F_i^{-1}(1)))$. Hence, in light of Proposition 2.1,

$$\begin{aligned}
&\inf_{\mathbf{f}\in\mathscr{D}_{\mathbf{1}}^{\mathbf{n}}(\mathbf{P})} \sup_{(X_1,\ldots,X_n)\in\mathscr{E}_n(\mathbf{F})} \mathrm{VaR}_\alpha(S_n^{\mathbf{f}}(X_1,\ldots,X_n)) \\
&= \inf_{\mathbf{f}\in\mathscr{D}_{\mathbf{1}}^{\mathbf{n}}(\mathbf{P})} \inf_{\boldsymbol{\gamma}\in(1-\alpha)\Delta_n} \sum_{i=1}^n \left\{R_{\gamma_i,\gamma_0}(X_i - f_i(X_i)) + \pi_i(f_i(X_i))\right\} \\
&= \inf_{\boldsymbol{\gamma}\in(1-\alpha)\Delta_n} \inf_{\mathbf{f}\in\mathscr{D}_{\mathbf{1}}^{\mathbf{n}}(\mathbf{P})} \sum_{i=1}^n \left\{R_{\gamma_i,\gamma_0}(X_i - f_i(X_i)) + \pi_i(f_i(X_i))\right\}.
\end{aligned}$$

Next we show that $\mathscr{D}_1^n(P)$ can be replaced by $\mathscr{H}^n(P)$. We will only show

$$\inf_{\mathbf{f}\in\mathscr{D}_{\mathbf{1}}^{\mathbf{n}}(\mathbf{P})} \sum_{i=1}^n \left\{R_{\gamma_i,\gamma_0}(X_i - f_i(X_i)) + \pi_i(f_i(X_i))\right\} \geqslant \inf_{(\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d})\in\mathscr{A}_{\mathbf{1}}(\mathbf{P})} H(\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d},\boldsymbol{\gamma}). \tag{2.10}$$

as the inverse inequality is trivial. Let

$$a_i = \begin{cases} F_i^{-1}(1-\gamma_0-\gamma_i) - \frac{f_i(F_i^{-1}(1-\gamma_0-\gamma_i))}{(f_i)'_+(F_i^{-1}(1-\gamma_0-\gamma_i))}, & (f_i)'_+(F_i^{-1}(1-\gamma_0-\gamma_i)) \neq 0 \\ F_i^{-1}(1-\gamma_0-\gamma_i), & (f_i)'_+(F_i^{-1}(1-\gamma_0-\gamma_i)) = 0 \end{cases},$$

$b_i = f_i(F_i^{-1}(1-\gamma_0-\gamma_i))$, $c_i = (f_i)'_+(F_i^{-1}(1-\gamma_0-\gamma_i))$, and $d_i \in [0, 1-c_i]$ be the solution of $g_i(s) = 0$, where

$$\begin{aligned}
g_i(s) := &\int_{1-\gamma_0-\gamma_i}^1 (s+c_i)(F_i^{-1}(t) - F_i^{-1}(1-\gamma_0-\gamma_i)) + F_i^{-1}(1-\gamma_0-\gamma_i)\mathrm{d}t \\
&- \int_{1-\gamma_0-\gamma_i}^1 f_i(F_i^{-1}(t))\mathrm{d}t.
\end{aligned}$$

Note that g_i is continuous on $[0,1]$ with $g_i(1-c_i) \geqslant 0$ and $g_i(0) \leqslant 0$. Hence the existence of

d_i is guaranteed. One can easily check that for the above chosen a_i, b_i, c_i, d_i,

$$\int_s^1 h_{a_i,b_i,c_i,d_i}(F_i^{-1}(t))\mathrm{d}t \leqslant \int_s^1 f_i(F_i^{-1}(t))\mathrm{d}t, \tag{2.11}$$

for $s \in (0,1)$. For the rest of the proof, using the same argument as in the proof of (i), we can show that (2.10) holds. This completes the proof of (ii).

Finally, we focus on case (iii). Note that for $\mathbf{f} \in \mathscr{D}_\mathbf{2}^\mathbf{n}(\mathbf{P})$, we have $T_{f_i}(0) = 0$, and $T_{f_i}(x)$ is a convex, continuous, and increasing function on $[0,\infty)$. Note that the distribution of $T_{f_i}(X_i)$ is concave beyond its α-quantile if $T_{f_i} = 0$. We next suppose $T_{f_i} \neq 0$. Direct computation gives:

$$\mathbb{P}(T_{f_i}(X_i) \leqslant x) = \mathbb{P}\left(X_i \leqslant T_{f_i}^{-1}(x)\right) = F_i(T_{f_i}^{-1}(x)),\ x \geqslant 0,$$

where $T_{f_i}^{-1}(x) = \sup\{t \geqslant 0 : T_{f_i}(t) \leqslant x\}$ for $x \geqslant 0$. Note that $T_{f_i}^{-1}$ is concave and continuous on $[0,\infty)$. Hence, $\frac{(F_i(T_{f_i}^{-1}(x))-\alpha)_+}{1-\alpha}$ is concave on its support. Hence in light of Proposition 2.1,

$$\begin{aligned}
&\inf_{\mathbf{f}\in\mathscr{D}_\mathbf{2}^\mathbf{n}(\mathbf{P})} \sup_{(X_1,\ldots,X_n)\in\mathscr{E}_n(\mathbf{F})} \mathrm{VaR}_\alpha(S_n^\mathbf{f}(X_1,\ldots,X_n)) \\
&= \inf_{\mathbf{f}\in\mathscr{D}_\mathbf{2}^\mathbf{n}(\mathbf{P})} \inf_{\boldsymbol{\gamma}\in(1-\alpha)\Delta_n} \sum_{i=1}^n \left\{R_{\gamma_i,\gamma_0}(X_i - f_i(X_i)) + \pi_i(f_i(X_i))\right\} \\
&= \inf_{\boldsymbol{\gamma}\in(1-\alpha)\Delta_n} \inf_{\mathbf{f}\in\mathscr{D}_\mathbf{2}^\mathbf{n}(\mathbf{P})} \sum_{i=1}^n \left\{R_{\gamma_i,\gamma_0}(X_i - f_i(X_i)) + \pi_i(f_i(X_i))\right\}.
\end{aligned}$$

Next, we show that $\mathscr{D}_2^n(P)$ can be replaced by $\mathscr{G}^n(P)$. Analogously to case (ii), we will only show

$$\inf_{\mathbf{f}\in\mathscr{D}_\mathbf{2}^\mathbf{n}(\mathbf{P})} \sum_{i=1}^n \left\{R_{\gamma_i,\gamma_0}(X_i - f_i(X_i)) + \pi_i(f_i(X_i))\right\} \geqslant \inf_{(\mathbf{a},\mathbf{b})\in\mathscr{A}_\mathbf{2}(\mathbf{P})} G(\mathbf{a},\mathbf{b},\boldsymbol{\gamma}). \tag{2.12}$$

Let

$$a_i = \begin{cases} \frac{f_i(F_i^{-1}(1-\gamma_0-\gamma_i))}{F_i^{-1}(1-\gamma_0-\gamma_i)}, & F_i^{-1}(1-\gamma_0-\gamma_i) \neq 0 \\ 1, & F_i^{-1}(1-\gamma_0-\gamma_i) = 0 \end{cases},$$

and $b_i \in [F_i^{-1}(1-\gamma_0-\gamma_i),\infty]$ be the solution of $g_i(s)=0$, where:

$$g_i(s) := \int_{1-\gamma_0-\gamma_i}^{1} g_{a_i,s}(F_i^{-1}(t))\mathrm{d}t - \int_{1-\gamma_0-\gamma_i}^{1} f_i(F_i^{-1}(t))\mathrm{d}t.$$

For the rest of the proof, using the same argument as in the proof of (i), we can show that (2.12) holds. The inverse inequality of (2.12) is trivial. This establishes the claim of case (iii). □

Proof of Proposition 2.3. It is clear that $\mathbf{l}_{\mathbf{a}^*,\mathbf{b}^*}$, $\mathbf{h}_{\mathbf{a}^*,\mathbf{b}^*,\mathbf{c}^*,\mathbf{d}^*}$ and $\mathbf{g}_{\mathbf{a}^*,\mathbf{b}^*}$ are the optimal ceded loss functions for cases (i)-(iii), respectively. We next focus on the existence of these optimal parameters under the continuity of π_1. We only prove the existence of the optimal parameters for case (i), as the other two cases follow similarly. First note that:

$$R_{\gamma_i,\gamma_0}(X_i - l_{a_i,b_i}(X_i)) + \pi_i(l_{a_i,b_i}(X_i)) = R_{\gamma_i,\gamma_0}(X_i) - R_{\gamma_i,\gamma_0}(l_{a_i,b_i}(X_i)) + \pi_i(l_{a_i,b_i}(X_i)).$$

For $0 \leqslant \varepsilon \leqslant T \leqslant 1$, define

$$\overline{\Delta}_n^{\varepsilon,T} = \left\{ (\gamma_0,\gamma_1,\ldots,\gamma_n) \in [\varepsilon,T]\times[0,T]^n : \sum_{i=0}^{n}\gamma_i = T \right\}. \tag{2.13}$$

Clearly, for $0<\varepsilon<1-\alpha$, $R_{\gamma_i,\gamma_0}(X_i)$ is continuous with respect to $\boldsymbol{\gamma}$ over $\overline{\Delta}_n^{\varepsilon,1-\alpha}$, and

$$R_{\gamma_i,\gamma_0}(l_{a_i,b_i}(X_i)) = \frac{1}{\gamma_0}\int_{\gamma_i}^{\gamma_i+\gamma_0} l_{a_i,b_i}(F_i^{-1}(t))\mathrm{d}t$$

is continuous with respect to $(\mathbf{a},\mathbf{b},\boldsymbol{\gamma})$ over $\mathscr{A}(P)\times\overline{\Delta}_n^{\varepsilon,1-\alpha}$. Hence, by the continuity of π_i, $L(\mathbf{a},\mathbf{b},\boldsymbol{\gamma})$ is a continuous function of $(\mathbf{a},\mathbf{b},\boldsymbol{\gamma})$ over $\mathscr{A}(P)\times\overline{\Delta}_n^{\varepsilon,1-\alpha}$ for $\varepsilon>0$, and $\mathscr{A}(P)$ is a closed set. Therefore there exists $(\mathbf{a}_\varepsilon^*,\mathbf{b}_\varepsilon^*,\boldsymbol{\gamma}_\varepsilon^*)\in\mathscr{A}(P)\times\overline{\Delta}_n^{\varepsilon,1-\alpha}$ such that:

$$(\mathbf{a}_\varepsilon^*,\mathbf{b}_\varepsilon^*,\boldsymbol{\gamma}_\varepsilon^*) \in \arg\inf_{(\mathbf{a},\mathbf{b},\boldsymbol{\gamma})\in\mathscr{A}(P)\times\overline{\Delta}_n^{\varepsilon,1-\alpha}} L(\mathbf{a},\mathbf{b},\boldsymbol{\gamma}).$$

Let $n_0\in\mathbb{N}$ such that $n_0\geqslant\frac{1}{1-\alpha}$. We next discuss the limit of $(\mathbf{a}_{1/k}^*,\mathbf{b}_{1/k}^*,\boldsymbol{\gamma}_{1/k}^*), k\geqslant n_0$. To this end, we need to extend the domain of L. Note that:

$$\begin{aligned}\lim_{\gamma_0\downarrow 0} R_{\gamma_i,\gamma_0}(X_i - l_{a_i,b_i}(X_i)) &= \lim_{\gamma_0\downarrow 0}\frac{1}{\gamma_0}\int_{\gamma_i}^{\gamma_i+\gamma_0}\mathrm{VaR}_{1-t}(X_i - l_{a_i,b_i}(X_i))\mathrm{d}t\\ &= \mathrm{VaR}_{1-\gamma_i}\left(X_i - l_{a_i,b_i}(X_i)\right).\end{aligned}$$

We define $R_{\gamma_i,0}(X_i - l_{a_i,b_i}(X_i) = \mathrm{VaR}_{1-\gamma_i}\left(X_i - l_{a_i,b_i}(X_i)\right)$. Notice that $R_{0,0}(X_i - l_{a_i,b_i}(X_i)) = \infty$ for $b_i < \infty$ and ess-sup$X_i = \infty$. Then $R_{\gamma_i,\gamma_0}(X_i - l_{a_i,b_i}(X_i)$ is a continuous function of $(\mathbf{a},\mathbf{b},\boldsymbol{\gamma})$ over $\mathscr{A}(P)\times\overline{\Delta}_n^{0,1-\alpha}$ if we allow $R_{0,0}(X_i - l_{a_i,b_i}(X_i) = \infty$ in the above scenario. It further implies that $L(\mathbf{a},\mathbf{b},\boldsymbol{\gamma})$ is a continuous function of $(\mathbf{a},\mathbf{b},\boldsymbol{\gamma})$ over $\mathscr{A}(P)\times\overline{\Delta}_n^{0,1-\alpha}$ if we allow $L(\mathbf{a},\mathbf{b},\boldsymbol{\gamma}) = \infty$ in the above mentioned scenario. There exists a subsequence $(\mathbf{a}^*_{1/k_l},\mathbf{b}^*_{1/k_l},\boldsymbol{\gamma}^*_{1/k_l}), l \geqslant n_0$ such that $(\mathbf{a}^*_{1/k_l},\mathbf{b}^*_{1/k_l},\boldsymbol{\gamma}^*_{1/k_l})$ converges to $(\mathbf{a}^*,\mathbf{b}^*,\boldsymbol{\gamma}^*)\in\mathscr{A}(P)\times\overline{\Delta}_n^{0,1-\alpha}$ as $l\to\infty$. By the continuity of L,

$$L(\mathbf{a}^*,\mathbf{b}^*,\boldsymbol{\gamma}^*) = \lim_{l\to\infty} L(\mathbf{a}^*_{1/k_l},\mathbf{b}^*_{1/k_l},\boldsymbol{\gamma}^*_{1/k_l}) = \inf_{(\mathbf{a},\mathbf{b})\in\mathscr{A},\boldsymbol{\gamma}\in(1-\alpha)\Delta_n} L(\mathbf{a},\mathbf{b},\boldsymbol{\gamma}) = \inf_{\boldsymbol{\gamma}\in(1-\alpha)\Delta_n} L(\mathbf{a}^*,\mathbf{b}^*,\boldsymbol{\gamma}).$$

This implies that:

$$(\mathbf{a}^*,\mathbf{b}^*)\in\arg\inf_{(\mathbf{a},\mathbf{b})\in\mathscr{A}(\mathbf{P})}\left\{\inf_{\boldsymbol{\gamma}\in(1-\alpha)\Delta_n} L(\mathbf{a},\mathbf{b},\boldsymbol{\gamma})\right\}.$$

Hence we establish the existence of the optimal parameters for case (i). □

Proof of Proposition 2.4. By the result in Makarov (1981) and the continuity of F_1^{-1} and F_2^{-1}, we have:

$$\begin{aligned}\sup_{(X_1,X_2)\in\mathscr{E}_2(\mathbf{F})}\mathrm{VaR}_\alpha(S_2^{\mathbf{f}}(X_1,X_2)) = \inf_{t\in[0,1-\alpha]} & R_{f_1}(\mathrm{VaR}_{\alpha+t}(X_1)) + R_{f_2}(\mathrm{VaR}_{1-t}(X_2))\\ &+\pi_1(f_1(X_1)) + \pi_2(f_2(X_2)).\end{aligned}$$

Hence

$$
\begin{aligned}
&\inf_{(f_1,f_2)\in\mathscr{D}^2(P)} \sup_{(X_1,X_2)\in\mathscr{E}_2(\mathbf{F})} \mathrm{VaR}_\alpha(S_2^{\mathbf{f}}(X_1,X_2)) \\
&= \inf_{t\in[0,1-\alpha]} \inf_{(f_1,f_2)\in\mathscr{D}^2(P)} \mathrm{VaR}_{\alpha+t}(X_1-f_1(X_1))+\mathrm{VaR}_{1-t}(X_2-f_2(X_2)) \\
&\qquad\qquad\qquad\qquad +\pi_1(f_1(X_1))+\pi_2(f_2(X_2)).
\end{aligned}
$$

Let $a_1=F_1^{-1}(\alpha+t)-f_1(F_1^{-1}(\alpha+t))$ and $a_2=F_2^{-1}(1-t)-f_2(F_2^{-1}(1-t))$. Moreover, let $b_1\in[F_1^{-1}(\alpha+t),\infty]$ be the solution of $g_1(s)=0$, where

$$g_1(s)=\int_{\alpha+t}^1 l_{a_1,s}(F_1^{-1}(u))\mathrm{d}u-\int_{\alpha+t}^1 f_1(F_1^{-1}(u))\mathrm{d}u,$$

and $b_2\in[F_2^{-1}(1-t),\infty]$ be the solution of $g_2(s)=0$, where

$$g_2(s)=\int_{1-t}^1 l_{a_2,s}(F_2^{-1}(u))\mathrm{d}u-\int_{1-t}^1 f_2(F_2^{-1}(u))\mathrm{d}u.$$

Using the above a_i,b_i we observe that:

$$\mathrm{VaR}_{\alpha+t}(X_1-f_1(X_1))=\mathrm{VaR}_{\alpha+t}(X_1-l_{a_1,b_1}(X_1)),$$

and

$$\mathrm{VaR}_{1-t}(X_2-f_2(X_2))=\mathrm{VaR}_{1-t}(X_2-l_{a_2,b_2}(X_2)),$$

and for $s\in(0,1)$,

$$\int_s^1 l_{a_i,b_i}(F_i^{-1}(t))\mathrm{d}t\leqslant\int_s^1 f_i(F_i^{-1}(t))\mathrm{d}t,\ i=1,2,$$

which, with the help of Theorem 2.57 in Föllmer and Schied (2016) and property (iii) of π_i, implies that for $i=1,2$, $\pi_i(l_{a_i,b_i}(X_i))\leqslant\pi_i(f_i(X_i))$. Using the similar argument as in the proof of Theorem 2.2, we can establish the claim.

2.7.2 Proofs of results in Section 2.4.

To prove Proposition 2.5, we need the following lemma.

Lemma 2.16. *For $\mathbf{F}^{(k)} = (F_1^{(k)}, \ldots, F_n^{(k)}), k \geqslant 1$ and $\mathbf{F} = (F_1, \ldots, F_n)$, if for each i, $(F_i^{(k)})^{-1}(t) \to F_i^{-1}(t)$ uniformly for $t \in (1-\beta-\alpha, 1)$ with $0 \leqslant \beta < \beta+\alpha \leqslant 1$, then*

$$\lim_{k\to\infty} \sup_{(X_1,\ldots,X_n)\in\mathscr{E}_n(\mathbf{F}^{(k)})} R_{\beta,\alpha}\left(\sum_{i=1}^n X_i\right) = \sup_{(X_1,\ldots,X_n)\in\mathscr{E}_n(\mathbf{F})} R_{\beta,\alpha}\left(\sum_{i=1}^n X_i\right).$$

Proof. First note that:

$$\begin{aligned}
&\sup_{(X_1,\ldots,X_n)\in\mathscr{E}_n(\mathbf{F}^{(k)})} R_{\beta,\alpha}\left(\sum_{i=1}^n X_i\right) \\
&= \sup_{U_1,\ldots,U_n\sim U[0,1]} R_{\beta/(\beta+\alpha),\alpha/(\beta+\alpha)}\left(\sum_{i=1}^n (F_i^{(k)})^{-1}(1-\beta-\alpha+(\beta+\alpha)U_i)\right),
\end{aligned}$$

and

$$\begin{aligned}
&\sup_{(X_1,\ldots,X_n)\in\mathscr{E}_n(\mathbf{F})} R_{\beta,\alpha}\left(\sum_{i=1}^n X_i\right) \\
&= \sup_{U_1,\ldots,U_n\sim U[0,1]} R_{\beta/(\beta+\alpha),\alpha/(\beta+\alpha)}\left(\sum_{i=1}^n (F_i)^{-1}(1-\beta-\alpha+(\beta+\alpha)U_i)\right).
\end{aligned}$$

Using the fact that:

$$\begin{aligned}
u_k := &\sup_{x_1,\ldots,x_n\in(0,1)} \left|\sum_{i=1}^n (F_i^{(k)})^{-1}(1-\beta-\alpha+(\beta+\alpha)x_i) - \sum_{i=1}^n (F_i)^{-1}(1-\beta-\alpha+(\beta+\alpha)x_i)\right| \\
\leqslant &\sup_{x_1,\ldots,x_n\in(0,1)} \sum_{i=1}^n \left|(F_i^{(k)})^{-1}(1-\beta-\alpha+(\beta+\alpha)x_i) - (F_i)^{-1}(1-\beta-\alpha+(\beta+\alpha)x_i)\right| \\
\to &\, 0 \text{ as } k\to\infty,
\end{aligned}$$

we have:

$$\lim_{k\to\infty}\left|\sup_{(X_1,\ldots,X_n)\in\mathscr{E}_n(\mathbf{F}^{(k)})} R_{\beta,\alpha}\left(\sum_{i=1}^n X_i\right) - \sup_{(X_1,\ldots,X_n)\in\mathscr{E}_n(\mathbf{F})} R_{\beta,\alpha}\left(\sum_{i=1}^n X_i\right)\right| \leqslant \lim_{k\to\infty} u_k = 0.$$

This completes the proof. □

Proof of Proposition 2.5.

For $i=1,\ldots,n$ and $k\geqslant 1$, if $q_i := F_i((F_i)_+^{-1}(1-\beta-\alpha))-(1-\beta-\alpha)>0$, let

$$F_i^{(k)}(x)=\begin{cases} F_i(x), & x<(F_i)_+^{-1}(1-\beta-\alpha)\\ F_i(x)-q_i+kq_i(x-(F_i)_+^{-1}(1-\beta-\alpha)), & (F_i)_+^{-1}(1-\beta-\alpha)\leqslant x\\ & \quad <(F_i)_+^{-1}(1-\beta-\alpha)+1/k\\ F_i(x), & x\geqslant (F_i)_+^{-1}(1-\beta-\alpha)+1/k \end{cases}$$

and if $q_i=0$, let $F_i^{(k)}=F_i$. Clearly, each $F_i^{(k)}$ admits a decreasing density beyond its $1-\beta-\alpha$-quantile. Hence, in light of Theorem 1 of Blanchet et al. (2023), we have:

$$\sup_{(X_1,\ldots,X_n)\in\mathscr{E}_n(\mathbf{F}^{(k)})} R_{\beta,\alpha}\left(\sum_{i=1}^n X_i\right)=\inf_{\boldsymbol{\gamma}\in(\beta+\alpha)\Delta_n,\gamma_0\geqslant\alpha}\sum_{i=1}^n R_{\gamma_i,\gamma_0}(F_i^{(k)}).$$

Moreover, by the construction of $F_i^{(k)}$, it follows that: $\sup_{t\in(1-\beta-\alpha,1)}|(F_i^{(k)})^{-1}(t)-F_i^{-1}(t)|\leqslant 1/k\to 0$ as $k\to\infty$. Hence, by Lemma 2.16, we have:

$$\lim_{k\to\infty}\sup_{(X_1,\ldots,X_n)\in\mathscr{E}_n(\mathbf{F}^{(k)})} R_{\beta,\alpha}\left(\sum_{i=1}^n X_i\right)=\sup_{(X_1,\ldots,X_n)\in\mathscr{E}_n(\mathbf{F})} R_{\beta,\alpha}\left(\sum_{i=1}^n X_i\right).$$

One can easily check

$$\lim_{k\to\infty}\inf_{\boldsymbol{\gamma}\in(\beta+\alpha)\Delta_n,\gamma_0\geqslant\alpha}\sum_{i=1}^n R_{\gamma_i,\gamma_0}(F_i^{(k)})=\inf_{\boldsymbol{\gamma}\in(\beta+\alpha)\Delta_n,\gamma_0\geqslant\alpha}\sum_{i=1}^n R_{\gamma_i,\gamma_0}(F_i).$$

A combination of the above results establishes the claim. □

Proof of Theorem 2.6. We first showcase (i). By Theorem 1 of Blanchet et al. (2023), we have:

$$
\begin{aligned}
&\inf_{\mathbf{f}\in\mathscr{D}^{\mathbf{n}}(\mathbf{P})}\sup_{(X_1,\ldots,X_n)\in\mathscr{E}_n(\mathbf{F})} R_{\beta,\alpha}\left(S_n^{\mathbf{f}}(X_1,\ldots,X_n)\right)\\
&=\inf_{\mathbf{f}\in\mathscr{D}^{\mathbf{n}}(\mathbf{P})}\inf_{\boldsymbol{\gamma}\in(\beta+\alpha)\Delta_n,\gamma_0\geqslant\alpha}\sum_{i=1}^{n}\left\{R_{\gamma_i,\gamma_0}(X_i-f_i(X_i))+\pi_i(f_i(X_i))\right\}\\
&=\inf_{\boldsymbol{\gamma}\in(\beta+\alpha)\Delta_n,\gamma_0\geqslant\alpha}\inf_{\mathbf{f}\in\mathscr{D}^{\mathbf{n}}(\mathbf{P})}\sum_{i=1}^{n}\left\{R_{\gamma_i,\gamma_0}(X_i-f_i(X_i))+\pi_i(f_i(X_i))\right\}.
\end{aligned}
$$

Analogously as in the proof of Theorem 2.2, we can show that:

$$
\inf_{\mathbf{f}\in\mathscr{D}^{\mathbf{n}}(\mathbf{P})}\sum_{i=1}^{n}\left\{R_{\gamma_i,\gamma_0}(X_i-f_i(X_i))+\pi_i(f_i(X_i))\right\}=\inf_{(\mathbf{a},\mathbf{b})\in\mathscr{A}(\mathbf{P})}L(\mathbf{a},\mathbf{b},\boldsymbol{\gamma}).
$$

This completes the proof of (i).

For the proof of case (ii), it follows from Proposition 2.5 that:

$$
\begin{aligned}
&\inf_{\mathbf{f}\in\mathscr{D}_{\mathbf{2}}^{\mathbf{n}}(\mathbf{P})}\sup_{(X_1,\ldots,X_n)\in\mathscr{E}_n(\mathbf{F})} R_{\beta,\alpha}\left(S_n^{\mathbf{f}}(X_1,\ldots,X_n)\right)\\
&=\inf_{\mathbf{f}\in\mathscr{D}_{\mathbf{2}}^{\mathbf{n}}(\mathbf{P})}\inf_{\boldsymbol{\gamma}\in(\beta+\alpha)\Delta_n,\gamma_0\geqslant\alpha}\sum_{i=1}^{n}\left\{R_{\gamma_i,\gamma_0}(X_i-f_i(X_i))+\pi_i(f_i(X_i))\right\}\\
&=\inf_{\boldsymbol{\gamma}\in(\beta+\alpha)\Delta_n,\gamma_0\geqslant\alpha}\inf_{\mathbf{f}\in\mathscr{D}_{\mathbf{2}}^{\mathbf{n}}(\mathbf{P})}\sum_{i=1}^{n}\left\{R_{\gamma_i,\gamma_0}(X_i-f_i(X_i))+\pi_i(f_i(X_i))\right\}.
\end{aligned}
$$

Analogously as in the proof of Theorem 2.2, we have:

$$
\inf_{\mathbf{f}\in\mathscr{D}_{\mathbf{2}}^{\mathbf{n}}(\mathbf{P})}\sum_{i=1}^{n}\left\{R_{\gamma_i,\gamma_0}(X_i-f_i(X_i))+\pi_i(f_i(X_i))\right\}=\inf_{(\mathbf{a},\mathbf{b})\in\mathscr{A}_{\mathbf{2}}(\mathbf{P})}G(\mathbf{a},\mathbf{b},\boldsymbol{\gamma}),
$$

which completes the proof of case (ii). □

Proof of Proposition 2.7 It is clear that $\mathbf{l}_{\mathbf{a}^*,\mathbf{b}^*}$ and $\mathbf{g}_{\mathbf{a}^*,\mathbf{b}^*}$ are the optimal ceded loss functions for cases (i)-(ii), respectively. We next focus on the existence of the optimal parameters for case (i). First note that:

$$
R_{\gamma_i,\gamma_0}(X_i-l_{a_i,b_i}(X_i))+\pi_i(l_{a_i,b_i}(X_i))=R_{\gamma_i,\gamma_0}(X_i)-R_{\gamma_i,\gamma_0}(l_{a_i,b_i}(X_i))+\pi_i(l_{a_i,b_i}(X_i)).
$$

Clearly, $R_{\gamma_i,\gamma_0}(X_i)$ is continuous with respect to $\boldsymbol{\gamma}$ over $\overline{\Delta}_n^{\beta+\alpha}$ and

$$R_{\gamma_i,\gamma_0}(l_{a_i,b_i}(X_i)) = \frac{1}{\gamma_0}\int_{\gamma_i}^{\gamma_i+\gamma_0} l_{a_i,b_i}(\mathrm{VaR}_{1-t}(X_i))\mathrm{d}t$$

is continuous with respect to $(\mathbf{a},\mathbf{b},\boldsymbol{\gamma})$ over $\mathscr{A}(P)\times\overline{\Delta}_n^{\alpha,\beta+\alpha}$, where the definition of $\overline{\Delta}_n^{\alpha,\beta+\alpha}$ is given in (2.13). Hence $L(\mathbf{a},\mathbf{b},\boldsymbol{\gamma})$ is continuous over $\mathscr{A}(P)\times\overline{\Delta}_n^{\alpha,\beta+\alpha}$. Therefore, there exists $(\mathbf{a}^*,\mathbf{b}^*,\boldsymbol{\gamma}^*)\in\mathscr{A}(P)\times\overline{\Delta}_n^{\alpha,\beta+\alpha}$ such that

$$L(\mathbf{a}^*,\mathbf{b}^*,\boldsymbol{\gamma}^*) = \inf_{(\mathbf{a},\mathbf{b},\boldsymbol{\gamma})\in\mathscr{A}(P)\times\overline{\Delta}_n^{\beta+\alpha}} L(\mathbf{a},\mathbf{b},\boldsymbol{\gamma}) = \inf_{\boldsymbol{\gamma}\in\overline{\Delta}_n^{\alpha,\beta+\alpha}} L(\mathbf{a}^*,\mathbf{b}^*,\boldsymbol{\gamma}).$$

Hence the existence of $(\mathbf{a}^*,\mathbf{b}^*)$ for case (i) is established. The proof for case (ii) is similar to case (i) and hence we omit it. $\square$

2.7.3 Proofs of results in Section 2.5.

Proof of Propositions 2.8-2.9. By Proposition 2.1, we have:

$$\begin{aligned}&\inf_{\mathbf{w}\in\Lambda_n(P)}\sup_{(X_1,\ldots,X_n)\in\mathscr{E}_n(\mathbf{F})}\mathrm{VaR}_\alpha\left(S^{\mathbf{w}}(X_1,\ldots,X_n)\right)\\ &=\inf_{\boldsymbol{\gamma}\in(1-\alpha)\Delta_n}\inf_{\mathbf{w}\in\Lambda_n(P)}\sum_{i=1}^n\left((1-w_i)R_{\gamma_i,\gamma_0}(X_i)+w_i\pi_i(X_i)\right).\end{aligned}$$

Note that $\sum_{i=1}^n\left((1-w_i)R_{\gamma_i,\gamma_0}(X_i)+w_i\pi_i(X_i)\right)$ is a linear function for each w_i. Hence for each $\boldsymbol{\gamma}\in(1-\alpha)\Delta_n$, we have:

$$\inf_{\mathbf{w}\in\Lambda_n(P)}\sum_{i=1}^n\left((1-w_i)R_{\gamma_i,\gamma_0}(X_i)+w_i\pi_i(X_i)\right) = \inf_{\mathbf{w}\in\partial\Lambda_n(P)}\sum_{i=1}^n\left((1-w_i)R_{\gamma_i,\gamma_0}(X_i)+w_i\pi_i(X_i)\right),$$

which implies:

$$
\begin{aligned}
&\inf_{\mathbf{w}\in\Lambda_n(P)} \sup_{(X_1,\ldots,X_n)\in\mathscr{E}_n(\mathbf{F})} \mathrm{VaR}_\alpha\left(S^{\mathbf{w}}(X_1,\ldots,X_n)\right)\\
&= \inf_{\boldsymbol{\gamma}\in(1-\alpha)\Delta_n} \inf_{\mathbf{w}\in\partial\Lambda_n(P)} \sum_{i=1}^n \big((1-w_i)R_{\gamma_i,\gamma_0}(X_i)+w_i\pi_i(X_i)\big)\\
&= \inf_{\mathbf{w}\in\partial\Lambda_n(P)} \inf_{\boldsymbol{\gamma}\in(1-\alpha)\Delta_n} \sum_{i=1}^n \big((1-w_i)R_{\gamma_i,\gamma_0}(X_i)+w_i\pi_i(X_i)\big).
\end{aligned}
$$

Note that $\inf_{\boldsymbol{\gamma}\in(1-\alpha)\Delta_n}\sum_{i=1}^n \big((1-w_i)R_{\gamma_i,\gamma_0}(X_i)+w_i\pi_i(X_i)\big)$ is a continuous function with respect to $\mathbf{w}$ and $\partial\Lambda_n(P)$ is compact. Hence there exists optimal quota-share levels $\mathbf{w}^*\in\partial\Lambda_n(P)$. This completes the proof of Proposition 2.8. Analogously, we can prove Proposition 2.9. □

Proof of Corollary 2.10. By Proposition 2.8, we have:

$$
\begin{aligned}
&\inf_{\mathbf{w}\in[0,1]^n} \sup_{(X_1,\ldots,X_n)\in\mathscr{E}_n(\mathbf{F})} \mathrm{VaR}_\alpha\left(S^{\mathbf{w}}(X_1,\ldots,X_n)\right)\\
&= \inf_{\mathbf{w}\in[0,1]^n} \inf_{\boldsymbol{\gamma}\in(1-\alpha)\Delta_n} \sum_{i=1}^n \big((1-w_i)R_{\gamma_i,\gamma_0}(X_i)+w_i\pi_i(X_i)\big)\\
&= \inf_{\boldsymbol{\gamma}\in(1-\alpha)\Delta_n} \inf_{\mathbf{w}\in[0,1]^n} \sum_{i=1}^n \big((1-w_i)R_{\gamma_i,\gamma_0}(X_i)+w_i\pi_i(X_i)\big)\\
&= \inf_{\boldsymbol{\gamma}\in(1-\alpha)\Delta_n} \inf_{\mathbf{w}\in\Lambda_n^0} \sum_{i=1}^n \big((1-w_i)R_{\gamma_i,\gamma_0}(X_i)+w_i\pi_i(X_i)\big)\\
&= \inf_{\mathbf{w}\in\Lambda_n^0} \left\{ \inf_{\boldsymbol{\gamma}\in(1-\alpha)\Delta_n} \sum_{i=1}^n \big((1-w_i)R_{\gamma_i,\gamma_0}(X_i)+w_i\pi_i(X_i)\big) \right\}.
\end{aligned}
$$

Hence there exist optimal quota-share levels $\mathbf{w}^*\in\Lambda_n^0$. We can similarly show the same conclusion for Proposition 2.9. □

2.7.4 Proofs of results in Section 2.6.

Proof of Proposition 2.11. By applying the Theorem 3.1 in Chi and Tan (2013), there exists admissible limited stop-loss function $l_{u_1,v_1}(X_1)$ and $l_{u_2,v_2}(X_2)$ such that, for any value of $t \in [0, 1-\alpha]$,

$$\mathrm{VaR}_{\alpha+t}\big(X_1 - l_{u_1,v_1}(X_1)\big) + \pi_1\big(l_{u_1,v_1}(X_1)\big) \leqslant \mathrm{VaR}_{\alpha+t}\big(X_1 - l_{a_1,b_1}(X_1)\big) + \pi_1\big(l_{a_1,b_1}(X_1)\big),$$
$$\forall (a_1, b_1) \in \mathscr{A}(P),$$

and,

$$\mathrm{VaR}_{1-t}\big(X_1 - l_{u_2,v_2}(X_2)\big) + \pi_2\big(l_{u_2,v_2}(X_2)\big) \leqslant \mathrm{VaR}_{1-t}\big(X_2 - l_{a_2,b_2}(X_2)\big) + \pi_2\big(l_{a_2,b_2}(X_2)\big),$$
$$\forall (a_2, b_2) \in \mathscr{A}(P).$$

As a result, for any $t \in [0, 1-\alpha]$,

$$L_1(u_1, u_2, v_1, v_2, t) \leqslant L_1(a_1, a_2, b_1, b_2, t), \quad \forall (a_1, a_2, b_1, b_2) \in \mathscr{A}(P). \tag{2.14}$$

Furthermore, for any given $\hat{t} \in [0, 1-\alpha]$, the Theorem 3.1 in Chi and Tan (2013) also implies that:

$$\begin{aligned}
\min_{\substack{(a_1,a_2,b_1,b_2)\\ \in \mathscr{A}(P)}} L_1(a_1, a_2, b_1, b_2, \hat{t}) &= \min_{\substack{(u_1,u_2,v_1,v_2)\\ \in \mathscr{A}(P)}} L_1(u_1, u_2, v_1, v_2, \hat{t}) \\
&= \min_{\substack{(u_1,u_2)\\ \in \mathscr{A}_v(P,v_1,v_2)}} L_2(u_1, u_2, \hat{t}).
\end{aligned}$$

By building on the above, we are now ready to show that:

$$\min_{\substack{(a_1,a_2,b_1,b_2)\\ \in \mathscr{A}(P)}} \min_{t\in[0,1-\alpha]} L_1(a_1, a_2, b_1, b_2, t) = \min_{\substack{(u_1,u_2)\\ \in \mathscr{A}_v(P,v_1,v_2)}} \min_{t\in[0,1-\alpha]} L_2(u_1, u_2, t)$$

in two steps. In the first step, we show that:

$$\min_{\substack{(a_1,a_2,b_1,b_2)\\ \in \mathscr{A}(P)}} \min_{t\in[0,1-\alpha]} L_1(a_1, a_2, b_1, b_2, t) = \min_{\substack{(u_1,u_2,v_1,v_2)\\ \in \mathscr{A}(P)}} \min_{t\in[0,1-\alpha]} L_1(u_1, u_2, v_1, v_2, t).$$

Let us denote S^*_{ab} to be the set of optimal solutions that solve the optimisation problem

$$\min_{\substack{(a_1,a_2,b_1,b_2)\\ \in\mathscr{A}(P)}} \min_{t\in[0,1-\alpha]} L_1\left(a_1,a_2,b_1,b_2,t\right),$$

while S^*_{uv} denotes the set of optimal solutions that solve

$$\min_{\substack{(u_1,u_2,v_1,v_2)\\ \in\mathscr{A}(P)}} \min_{t\in[0,1-\alpha]} L_1\left(u_1,u_2,v_1,v_2,t\right).$$

Let us assume there exists a solution $(u_1^*,u_2^*,v_1^*,v_2^*,t^*)\in S^*_{uv}$ but $(u_1^*,u_2^*,v_1^*,v_2^*,t^*)\notin S^*_{ab}$, there must exist a solution $\left(\hat{a}_1,\hat{a}_2,\hat{b}_1,\hat{b}_2,\hat{t}\right)\in S^*_{ab}$, such that:

$$L_1\left(\hat{a}_1,\hat{a}_2,\hat{b}_1,\hat{b}_2,\hat{t}\right) < L_1(u_1^*,u_2^*,v_1^*,v_2^*,t^*). \tag{2.15}$$

If $\hat{t}=t^*$, (2.15) contradicts with (2.14). If $\hat{t}\neq t^*$, (2.14) implies that there must exist another solution $\left(\hat{u}_1,\hat{u}_2,\hat{v}_1,\hat{v}_2,\hat{t}\right)$ that is admissible to the optimisation problem

$$\min_{\substack{(u_1,u_2,v_1,v_2)\\ \in\mathscr{A}(P)}} \min_{t\in[0,1-\alpha]} L_1\left(u_1,u_2,v_1,v_2,t\right),$$

such that

$$L_1\left(\hat{u}_1,\hat{u}_2,\hat{v}_1,\hat{v}_2,\hat{t}\right)\leqslant L_1\left(\hat{a}_1,\hat{a}_2,\hat{b}_1,\hat{b}_2,\hat{t}\right) < L_1(u_1^*,u_2^*,v_1^*,v_2^*,t^*),$$

which contradicts with the assumption that $(u_1^*,u_2^*,v_1^*,v_2^*,t^*)\in S^*_{uv}$.

Therefore, if $(u_1^*,u_2^*,v_1^*,v_2^*,t^*)\in S^*_{uv}$, $(u_1^*,u_2^*,v_1^*,v_2^*,t^*)\in S^*_{ab}$ must also hold.

On the other hand, let us assume that there exists a solution $(a_1^*,a_2^*,b_1^*,b_2^*,t^*)\in S^*_{ab}$, but $(a_1^*,a_2^*,b_1^*,b_2^*,t^*)\notin S^*_{uv}$. Then, there must exist a solution $(u_1',u_2',v_1',v_2',t')\in S^*_{uv}$, such that

$$L_1(u_1',u_2',v_1',v_2',t') < L_1(a_1^*,a_2^*,b_1^*,b_2^*,t^*).$$

Since the solution $(u_1', u_2', v_1', v_2', t') \in S_{uv}^*$ is also admissible to the optimisation problem:

$$\min_{\substack{(a_1,a_2,b_1,b_2)\\ \in \mathscr{A}(P)}} \min_{t\in[0,1-\alpha]} L_1\left(a_1, a_2, b_1, b_2, t\right),$$

it contradicts with the assumption that $(a_1^*, a_2^*, b_1^*, b_2^*, t^*) \in S_{ab}^*$. Therefore, if $(a_1^*, a_2^*, b_1^*, b_2^*, t^*) \in S_{ab}^*$, $(a_1^*, a_2^*, b_1^*, b_2^*, t^*) \in S_{uv}^*$ must also be true. By combining the above, the first part of the proof is complete.

Now let us move to the second step of the proof and show that:

$$\min_{\substack{(u_1,u_2,v_1,v_2)\\ \in \mathscr{A}(P)}} \min_{t\in[0,1-\alpha]} L_1\left(u_1, u_2, v_1, v_2, t\right) = \min_{\substack{(u_1,u_2)\\ \in \mathscr{A}_v(P,v_1,v_2)}} \min_{t\in[0,1-\alpha]} L_2\left(u_1, u_2, t\right).$$

We now denote S_u^* to be the set of optimal solutions that solve

$$\min_{\substack{(u_1,u_2)\\ \in \mathscr{A}_v(P,v_1,v_2)}} \min_{t\in[0,1-\alpha]} L_2(u_1, u_2, t).$$

Assume the optimal solution $(u_1^*, u_2^*, v_1^*, v_2^*, t^*) \in S_{uv}^*$ but $(u_1^*, u_2^*, t^*) \notin S_u^*$, there must exist a solution $(\hat{u}_1, \hat{u}_2, \hat{t}) \in S_u^*$, such that:

$$L_2\left(\hat{u}_1, \hat{u}_2, \hat{t}\right) < L_2(u_1^*, u_2^*, t^*). \tag{2.16}$$

Because of the translation invariance property of VaR, it is not difficult to see that:

$$L_1\left(\hat{u}_1, \hat{u}_2, \hat{v}_1, \hat{v}_2, \hat{t}\right) = L_2\left(\hat{u}_1, \hat{u}_2, \hat{t}\right) \tag{2.17}$$

where, by definition, $\hat{v}_1 = \mathrm{VaR}_{\alpha+\hat{t}}(X_1)$ and $\hat{v}_2 = \mathrm{VaR}_{1-\hat{t}}(X_2)$. Similarly, it also holds that:

$$L_1(u_1^*, u_2^*, v_1^*, v_2^*, t^*) = L_2(u_1^*, u_2^*, t^*), \tag{2.18}$$

where, by definition, $v_1^* = \mathrm{VaR}_{\alpha+t^*}(X_1)$ and $v_2^* = \mathrm{VaR}_{1-t^*}(X_2)$. By combining (2.16), (2.17) and (2.18), we obtain:

$$L_1\left(\hat{u}_1, \hat{u}_2, \hat{v}_1, \hat{v}_2, \hat{t}\right) < L_1(u_1^*, u_2^*, v_1^*, v_2^*, t^*). \tag{2.19}$$

Note that $(\hat{u}_1, \hat{u}_2, \hat{v}_1, \hat{v}_2, \hat{t})$ is admissible to the optimisation problem

$$\min_{\substack{(u_1,u_2,v_1,v_2)\\ \in \mathscr{A}(P)}} \min_{t\in[0,1-\alpha]} L_1\left(u_1, u_2, v_1, v_2, t\right),$$

and thus, (2.19) contradicts with the assumption that $(u_1^*, u_2^*, v_1^*, v_2^*, t^*) \in S_{uv}^*$. Therefore, if $(u_1^*, u_2^*, v_1^*, v_2^*, t^*) \in S_{uv}^*$, it must also holds that $(u_1^*, u_2^*, t^*) \in S_u^*$.

On the other hand, let us now show that if $(u_1^*, u_2^*, t^*) \in S_u^*$, $(u_1^*, u_2^*, v_1^*, v_2^*, t^*) \in S_{uv}^*$ must also be true. Assume that $(u_1^*, u_2^*, v_1^*, v_2^*, t^*) \notin S_{uv}^*$, then there must exist a solution $\left(\tilde{u}_1, \tilde{u}_2, \tilde{v}_1, \tilde{v}_2, \tilde{t}\right) \in S_{uv}^*$ such that

$$L_1\left(\tilde{u}_1, \tilde{u}_2, \tilde{v}_1, \tilde{v}_2, \tilde{t}\right) < L_1(u_1^*, u_2^*, v_1^*, v_2^*, t^*), \tag{2.20}$$

where by definition $\tilde{v}_1 = \mathrm{VaR}_{\alpha+\tilde{t}}(X_1)$ and $\tilde{v}_2 = \mathrm{VaR}_{1-\tilde{t}}(X_2)$. If we apply again the translation invariance property of VaR, (2.20) implies the following:

$$L_2\left(\tilde{u}_1, \tilde{u}_2, \tilde{t}\right) < L_2(u_1^*, u_2^*, t^*). \tag{2.21}$$

Since $\left(\tilde{u}_1, \tilde{u}_2, \tilde{v}_1, \tilde{v}_2, \tilde{t}\right)$ is admissible to the optimisation problem

$$\min_{\substack{(u_1,u_2)\\ \in \mathscr{A}_v(P,v_1,v_2)}} \min_{t\in[0,1-\alpha]} L_2(u_1, u_2, t),$$

(2.21) contradicts with the assumption that $(u_1^*, u_2^*, t^*) \in S_u^*$. That is, we have shown that if $(u_1^*, u_2^*, t^*) \in S_u^*$, $(u_1^*, u_2^*, v_1^*, v_2^*, t^*) \in S_{uv}^*$ must also be true, and the second part of the proof is also complete.

Overall, we have shown that:

$$\min_{\substack{(a_1,a_2,b_1,b_2)\\ \in\mathscr{A}(P)}} \min_{t\in[0,1-\alpha]} L_1\left(a_1,a_2,b_1,b_2,t\right) = \min_{\substack{(u_1,u_2,v_1,v_2)\\ \in\mathscr{A}(P)}} \min_{t\in[0,1-\alpha]} L_1\left(u_1,u_2,v_1,v_2,t\right),$$

and

$$\min_{\substack{(u_1,u_2,v_1,v_2)\\ \in\mathscr{A}(P)}} \min_{t\in[0,1-\alpha]} L_1\left(u_1,u_2,v_1,v_2,t\right) = \min_{\substack{(u_1,u_2)\\ \in\mathscr{A}_v(P,v_1,v_2)}} \min_{t\in[0,1-\alpha]} L_2\left(u_1,u_2,t\right),$$

which lead to the conclusion that

$$\min_{\substack{(a_1,a_2,b_1,b_2)\\ \in\mathscr{A}(P)}} \min_{t\in[0,1-\alpha]} L_1\left(a_1,a_2,b_1,b_2,t\right) = \min_{\substack{(u_1,u_2)\\ \in\mathscr{A}_v(P,v_1,v_2)}} \min_{t\in[0,1-\alpha]} L_2\left(u_1,u_2,t\right).$$

Proof of Proposition 2.13 It is not difficult to show that

$$\begin{aligned}\mathbb{E}\big(l_{u_i,v_i}(X_i)\big) &= \int_{u_i}^{v_i}(x-u_i)f_i(x)dx+\int_{v_i}^{\infty}(v_i-u_i)f_i(x)dx\\ &= \lambda_i e^{-u_i/\lambda_i}-\lambda_i e^{-v_i/\lambda_i}.\end{aligned}$$

Moreover, v_i is given by

$$v_1 = F_1^{-1}(\alpha+t) = -\lambda_1\ln(1-\alpha-t), \quad \text{for } i=1,$$

and

$$v_2 = F_2^{-1}(1-t) = -\lambda_2\ln t, \quad \text{for } i=2.$$

As a result, $L_2(u_1,u_2,t)$ becomes

$$\begin{aligned}L_3(u_1,u_2,t) &= \sum_{i=1}^{2}\Big(u_i+w_i+(1+\delta_i)\mathbb{E}\big(l_{u_i,v_i}(X_i)\big)\Big)\\ &= \sum_{i=1}^{2}\left(u_i+w_i+(1+\delta_i)\lambda_i e^{-\frac{u_i}{\lambda_i}}\right)-(1+\delta_1)\lambda_1(1-\alpha)\\ &\quad+\big((1+\delta_1)\lambda_1-(1+\delta_2)\lambda_2\big)t. \end{aligned} \tag{2.22}$$

We are now ready to apply Proposition 2.11 and solve the optimisation problem (2.4) in two steps. The first step considers the inner minimiser and denote

$$\tilde{t} := \arg\min_{t\in[0,1-\alpha]} L_3(u_1,u_2,t),$$

while the second step solves the outer minimiser

$$\min_{\substack{(u_1,u_2)\\ \in\mathscr{A}_v(P,v_1,v_2)}} L_3(u_1,u_2,\tilde{t}).$$

As we can see in (2.22), $L_3(u_1,u_2,t)$ is linear in t. Thus, we will discuss the solution in the following three cases.

(i) If $(1+\delta_1)\lambda_1-(1+\delta_2)\lambda_2=0$, the solution of the inner minimiser ($\tilde{t}$) is not unique and does not affect the value of $L_3(u_1,u_2,t)$. Furthermore, the objective function of the outer minimiser $L_3(u_1,u_2,\tilde{t})$ becomes

$$L_3^{(\mathrm{i})}\left(u_1,u_2,\tilde{t}\right) = \sum_{i=1}^{2}\left(u_i+w_i+(1+\delta_i)\lambda_i e^{-u_i/\lambda_i}\right)-(1+\delta_1)\lambda_1(1-\alpha).$$

Differentiating $L_3^{(\mathrm{i})}(u_1,u_2,\tilde{t})$ with respect to u_1 and u_2 respectively, we obtain:

$$\frac{dL_3^{(\mathrm{i})}\left(u_1,u_2,\tilde{t}\right)}{du_i} = 1-(1+\delta_i)e^{-u_i/\lambda_i}, \quad \text{for } i=1,2. \tag{2.23}$$

Setting (2.23) to 0 gives

$$u_i = -\lambda_i \ln\left(\frac{1}{1+\delta_i}\right) > 0, \quad \text{for } i = 1,2. \tag{2.24}$$

Recall that the feasible region for u_i is $(u_1,u_2) \in \mathscr{A}_v(P,v_1,v_2)$, which lead to the following budget constraint and feasibility set $\mathscr{T}$ for t such that

$$\sum_{i=1}^{2}(w_i+\lambda_i) - (1+\delta_1)(1-\alpha)\lambda_1 \leqslant P, \tag{2.25}$$

and

$$\mathscr{T} := \left\{ t \in [0,1-\alpha] : 1-\alpha-\frac{1}{1+\delta_1} \leqslant t \leqslant \frac{1}{1+\delta_2} \right\}.$$

Note that $\mathscr{T}$ is non-empty, because $1-\alpha-\frac{1}{1+\delta_1} < 1-\alpha$ and $0 < \frac{1}{1+\delta_2}$ hold. Therefore, if (2.25) holds, the feasibility set is not empty and the optimal solution

$$(u_1^*, u_2^*, t^*) := \arg \min_{\substack{(u_1,u_2) \\ \in \mathscr{A}_v(P,v_1,v_2)}} \min_{t\in[0,1-\alpha]} L_3(u_1,u_2,t)$$

is given by

$$\begin{cases} u_i^* = -\lambda_i \ln\left(\frac{1}{1+\delta_i}\right), & \text{for } i = 1,2, \\ t^* = t_0, & \text{where } t_0 \in \mathscr{T}. \end{cases}$$

Otherwise, a numerical approach shall be employed to solve the optimisation problem.

(ii) If $(1+\delta_1)\lambda_1-(1+\delta_2)\lambda_2 > 0$, $L_3(u_1,u_2,t)$ is an increasing function in t. Thus, $\tilde{t} = 0$, and $L_3(u_1,u_2,\tilde{t})$ becomes

$$L_3^{(\mathrm{ii})}\left(u_1,u_2,\tilde{t}\right) = \sum_{i=1}^{2}\left(u_i+w_i+(1+\delta_i)\lambda_i e^{-u_i/\lambda_i}\right) - (1+\delta_1)\lambda_1(1-\alpha).$$

Therefore, differentiating $L_3^{(\mathrm{ii})}\left(u_1,u_2,\tilde{t}\right)$ with respect to u_1 and u_2 respectively will lead to the same first order conditions and solutions as in (2.23) and (2.24). Note that $v_2 = \infty$

when $t = \tilde{t} = 0$. That is, the feasibility constraint of u_2, $u_2 \in [0, v_2)$, will always hold. Provided that $\mathscr{A}_v(P, v_1, v_2) \neq \emptyset$,

$$u_2^* = -\lambda_2 \ln\left(\frac{1}{1+\delta_2}\right) \quad \text{and} \quad u_2^* \in (0, v_2).$$

On the other hand, $v_1 = -\lambda_1 \ln(1-\alpha) < \infty$ when $t = \tilde{t} = 0$. Thus, it is necessary to consider the following two sub-cases and check the budget constraint in each sub-case:

(iia) If $(1-\alpha)(1+\delta_1) < 1$, it holds that

$$-\lambda_1 \ln\left(\frac{1}{1+\delta_1}\right) < -\lambda_1 \ln(1-\alpha) = v_1$$

and thus,

$$u_1^* = -\lambda_1 \ln\left(\frac{1}{1+\delta_1}\right) \in (0, v_1),$$

which leads to the same budget constraint as in (2.25). That is, $\mathscr{A}_v(P, v_1, v_2) \neq \emptyset$ if (2.25) holds and the optimal solution (u_1^*, u_2^*, t^*) is given by:

$$\begin{cases} u_i^* = -\lambda_i \ln\left(\frac{1}{1+\delta_i}\right), & \text{for } i = 1, 2, \\ t^* = 0. \end{cases}$$

Otherwise, a numerical approach will be sought to solve the optimisation problem.

(iib) If $(1-\alpha)(1+\delta_1) \geqslant 1$, we have

$$0 \leqslant u_1 \leqslant v_1 = -\lambda_1 \ln(1-\alpha) \leqslant -\lambda_1 \ln\left(\frac{1}{1+\delta_1}\right).$$

One may notice that the first order condition (2.23) is an increasing function in u_i and is equal to 0 when $u_i = -\lambda_i \ln\left(\frac{1}{1+\delta_i}\right)$. Therefore,

$$\frac{dL_3^{(\mathrm{ii})}\left(u_1, u_2, \tilde{t}\right)}{du_1} < 0, \quad \forall\, u_1 \in \left[0, -\lambda_1 \ln\left(\frac{1}{1+\delta_1}\right)\right).$$

Consequently, $u_1^* = v_1$, i.e. there will be no reinsurance purchased for X_1, which leads to the budget constraint of

$$w_2+\lambda_2 \leqslant P. \tag{2.26}$$

Therefore, if (2.26) holds, the optimal solution (u_1^*, u_2^*, t^*) is given by:

$$\begin{cases} u_1^* = -\lambda_1 \ln(1-\alpha), \\ u_2^* = -\lambda_2 \ln\left(\frac{1}{1+\delta_2}\right), \\ t^* = 0. \end{cases}$$

Otherwise, a numerical approach will be sought to solve the optimisation problem.

(iii) Finally, if $(1+\delta_1)\lambda_1-(1+\delta_2)\lambda_2 < 0$, $L_3(u_1,u_2,t)$ is now a decreasing function in t. Thus, $\tilde{t} = 1-\alpha$, and $L_3\left(u_1,u_2,\tilde{t}\right)$ becomes

$$L_3^{(\mathrm{iii})}\left(u_1,u_2,\tilde{t}\right) = \sum_{i=1}^{2}\left(u_i+w_i+(1+\delta_i)\lambda_i e^{-u_i/\lambda_i}\right)-(1+\delta_2)\lambda_2(1-\alpha).$$

Again, as in case (ii), differentiating $L_3^{(\mathrm{iii})}\left(u_1,u_2,\tilde{t}\right)$ with respect to u_1 and u_2 respectively will lead to the same first order conditions and solutions as in (2.23) and (2.24). Note that $v_1=\infty$ when $t=\tilde{t}=1-\alpha$. That is, the feasibility constraint of u_1, $u_1 \in [0,\infty)$, will always hold. Provided that $\mathscr{A}_v(P,u_1,u_2) \neq \emptyset$,

$$u_1^* = -\lambda_1 \ln\left(\frac{1}{1+\delta_1}\right).$$

On the other hand, $v_2 = -\lambda_2 \ln(1-\alpha) < \infty$ when $t=\tilde{t}=1-\alpha$. Again, it is necessary to discuss two sub-cases and check the budget constraint:

(iiia) If $(1-\alpha)(1+\delta_2) < 1$, it holds that

$$-\lambda_2 \ln\left(\frac{1}{1+\delta_2}\right) < -\lambda_2 \ln(1-\alpha) = v_2,$$

and thus,

$$u_2^* = -\lambda_2 \ln\left(\frac{1}{1+\delta_2}\right) \in (0, v_2),$$

which leads to the following budget constraint

$$\sum_{i=1}^{2}(w_i + \lambda_i) - (1+\delta_2)\lambda_2(1-\alpha) \leqslant P. \tag{2.27}$$

That is, $\mathscr{A}_v(P, v_1, v_2) \neq \emptyset$ if (2.27) holds and the optimal solution (u_1^*, u_2^*, t^*) is given by:

$$\begin{cases} u_i^* = -\lambda_i \ln\left(\frac{1}{1+\delta_i}\right), & \text{for } i = 1, 2, \\ t^* = 1-\alpha. \end{cases}$$

Otherwise, a numerical approach will be sought to solve the optimisation problem.

(iiib) If $(1-\alpha)(1+\delta_2) \geqslant 1$, we have

$$0 \leqslant u_2 \leqslant v_2 = -\lambda_2 \ln(1-\alpha) \leqslant -\lambda_2 \ln\left(\frac{1}{1+\delta_2}\right).$$

Similar to what we have mentioned in sub-case (iib), the first order condition (2.23) is an increasing function in u_2 and is equal to 0 when $u_2 = -\lambda_2 \ln\left(\frac{1}{1+\delta_2}\right)$. Therefore,

$$\frac{dL_3^{(\mathrm{iii})}\left(u_1, u_2, \tilde{t}\right)}{du_2} < 0, \quad \forall\, u_2 \in \left[0, -\lambda_2 \ln\left(\frac{1}{1+\delta_2}\right)\right).$$

Consequently, $u_2^* = v_2$, i.e. there will be no reinsurance purchased for X_2, which leads to the budget constraint of

$$w_1 + \lambda_1 \leqslant P. \tag{2.28}$$

Therefore, if (2.28) holds, the optimal solution (u_1^*, u_2^*, t^*) is given by

$$\begin{cases} u_1^* = -\lambda_1 \ln\left(\frac{1}{1+\delta_1}\right), \\ u_2^* = -\lambda_2 \ln(1-\alpha), \\ t^* = 1-\alpha. \end{cases}$$

Otherwise, a numerical approach will be sought to solve the optimisation problem.

CHAPTER 3

A homotopy-based approach for multi-risk reinsurance problems: Numerical optimisation with perturbations and ensembles

3.1 Introduction

An optimal reinsurance problem involves two key parties, specifically the insurer (i.e., the policy buyer) and the reinsurer (i.e., the policy seller). The insurer transfers a portion of its risk to the reinsurer in exchange for a reinsurance premium. Suppose the buyer operates across n business lines, with X_i denoting the random risk from the ith line for $i = 1, 2, \ldots, n$. A proportion $f_i(X_i)$ of each risk is ceded to the reinsurer as part of the buyer's risk management strategy. The reinsurer charges a premium denoted by $\pi_i(f_i(X_i))$. Consequently, the post-reinsurance risk profiles for the buyer (S_B) and the reinsurer (S_R) are:

$$S_B = \sum_{i=1}^{n} \left(X_i - f_i(X_i) + \pi_i(f_i(X_i))\right), \quad S_R = \sum_{i=1}^{n} \left(f_i(X_i) - \pi_i(f_i(X_i))\right).$$

The objective of optimal reinsurance design is to strike a balance between risk retention and profit generation. A general formulation of the problem is given by the following:

$$\begin{cases} \min\limits_{f_1(X_1),\ldots,f_n(X_n)} & \delta\rho_B(S_B)+(1-\delta)\rho_R(S_R), \\ \text{s.t.} & g_k(f_1(X_1),\ldots,f_n(X_n)) \leqslant 0, \quad \forall k \in \mathcal{N}_g, \\ & h_j(f_1(X_1),\ldots,f_n(X_n)) = 0, \quad \forall j \in \mathcal{N}_h. \end{cases} \tag{3.1}$$

Here, $\mathcal{N}_g = \{1,2,\ldots,n_g\}$ and $\mathcal{N}_h = \{1,2,\ldots,n_h\}$ index the inequality and equality constraints, with $n_g, n_h \geqslant 1$, where g_k are inequality constraints and h_j are equality constraints.

Seminal studies by Borch (1960) and Arrow (1963) demonstrated that the optimal stop-loss reinsurance contract minimises either the variance of retained risk or maximises the expected utility of the insurer's terminal wealth. Subsequently, further contributions explored variance minimisation under various premium principles, including the standard deviation and mean-variance principles (e.g., Gajek and Zagrodny (2000), Kaluszka (2001), Kaluszka (2004)). Recent research trends have focused on risk measure-based models, notably involving Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR), with contributions from Cai and Tan (2007), Cui et al. (2013), amongst others. Additional studies incorporate multi-risk environments and counterparty risk (e.g., Carole and Ludkovski (2012), Cai and Chi (2020), Asimit et al. (2013b)).

The studies mentioned above primarily addressed scenarios that involve two parties in the risk transfer process. However, in practice, multiple reinsurers may share risk with a single insurer or a group of insurers, as discussed in detail in Asimit et al. (2013a). Closed-form solutions become elusive under such complexity, although efficient numerical solutions are possible in convex settings (e.g. Asimit et al. (2018)). In contrast, non-convex formulations present substantial difficulties due to solution non-uniqueness, making even numerical solution identification a challenging task. To address this, we apply a homotopy optimisation approach to tackle such non-convex reinsurance problems and demonstrate the efficacy and efficiency of the algorithm.

The homotopy analysis method, first introduced in Liao (1992), originally addressed highly non-linear partial differential equations. Inspired by parallels between nonlinearity and non-convexity, it was later adapted for non-convex optimisation (Watson and Haftka (1989);

Dunlavy and O'Leary (2005)), giving rise to the Homotopy Optimisation Method (HOM). Building on this, Dunlavy and O'Leary (2005) proposed an enhanced version for global optimisation, termed Homotopy Optimisation with Perturbations and Ensembles (HOPE). The HOPE algorithm is both efficient and computationally viable for tackling complex non-convex problems. Its key strength lies in the ability to effectively handle strong non-linearity and non-convexity, outperforming conventional heuristics. This chapter explores non-convex optimal reinsurance problems using the HOPE algorithm, an approach yet to be widely addressed in existing actuarial literature.

The structure of this chapter is as follows. Section 3.2 presents the HOPE algorithm. Section 3.3 demonstrates its efficiency through a benchmark test. A complete numerical example is provided in Section 3.4, followed by concluding remarks in Section 3.5.

3.2 Formulation

This section describes the general idea and procedure of solving a non-convex optimal reinsurance problem using a homotopy algorithm. We consider a general non-convex objective function $f_{\text{obj}} : \mathbb{D} \to \mathbb{R}$ subject to some constraints, where $\mathbb{D}$ defines the domain and $\mathscr{D}$ is the associated feasible region of the control variables denoted by $\mathbf{x}$. To simplify the notation, we define $\mathscr{L}(f_{\text{obj}}, v)$ as the operation to minimise f_{obj} in $\mathscr{D}$ by an iterative solver with v selected as the initial guess. It is practical to decompose the objective function into two parts as follows:

$$f_{\text{obj}}(\mathbf{x}) = \underbrace{\Phi(\mathbf{x})}_{\text{non-convex}} + \underbrace{\Psi(\mathbf{x})}_{\text{convex}} . \tag{3.2}$$

The homotopy algorithm starts by constructing a continuous homotopy function H, given by:

$$H(\mathbf{x}, \lambda) = \lambda \Phi(\mathbf{x}) + \Psi(\mathbf{x}), \tag{3.3}$$

where λ is the homotopy parameter that varies over the interval $[0, 1]$. When decomposition (3.3) is not readily available, it is possible to define an artificial convex function ϕ and instead consider $H(\mathbf{x},\lambda) = \lambda f_{\text{obj}}(\mathbf{x}) + (1-\lambda)\phi(\mathbf{x})$. However, this alternative approach lies outside the scope of this paper and will not be discussed further.

From equation (3.3), we observe that:

$$\begin{cases} H(\mathbf{x},0) = \Psi(\mathbf{x}), \\ H(\mathbf{x},1) = f_{\text{obj}}(\mathbf{x}). \end{cases} \tag{3.4}$$

Due to the convexity of Ψ, minimising $H(\mathbf{x},0)$ using a local solver yields a unique minimiser, denoted by v. To proceed, the homotopy variable λ is discretised into $\mathscr{M}$ equally spaced steps as follows:

$$\lambda_m = \frac{m}{\mathscr{M}}, \qquad m = 0,1,2,...,\mathscr{M}, \tag{3.5}$$

allowing a continuation procedure along the path of λ. For every $m = j$ with $j \geqslant 1$, the solution obtained at $m = j-1$ is used as the initial guess for the optimisation solver. When $m = \mathscr{M}$, the homotopy function H coincides with the objective function f_{obj}, implying that the resulting solution corresponds to the optimum of the reinsurance problem.

The algorithm 1 summarises the standard homotopy optimisation procedure. As noted previously, this technique does not ensure convergence to the global optimum.

Input v_0 ; `/* initial guess */`
$f_{\text{obj}} = \Psi$; $v \leftarrow \mathscr{L}(f_{\text{obj}}, v_0)$; `/* optimal solution that minimises Ψ */`
for $m = 1 : \mathscr{M}$ **do**
  $\lambda = \frac{m}{\mathscr{M}}$, $f_{\text{obj}} = \lambda\Phi + \Psi$; `/* objective function */`
  $v \leftarrow \mathscr{L}(f_{\text{obj}}, v)$; `/* solution update */`
end
Output v.

Algorithm 1: Classic Homotopy Optimisation Method

In the extended version of the homotopy method, which includes perturbations and ensembles, the initial guesses are randomly perturbed several times, which allows a broader exploration of the feasible solution space. Since we search for a larger area, Algorithm 2 is more likely to find the global solution compared to Algorithm 1. After each iteration for a given λ_m, only a subset of the best-performing solutions, based on their values of the objective function, is retained within an ensemble. This ensemble serves as the source of initial guesses for the subsequent iteration.

The following notation is used:

$\mathscr{M}_1$: the number of perturbations applied in a loop;

$\mathscr{M}_2$: the maximal number of solutions retained at the end of a loop.

In summary, the homotopy optimisation method with perturbations and ensembles involves three key parameters: $\mathscr{M}$, $\mathscr{M}_1$, and $\mathscr{M}_2$. The computations in the remainder of this work rely on this extended framework, whose pseudocode is outlined in Algorithm 2. Further discussion of the choice and configuration of these parameters is provided in Section 3.3.

3.3 Examinations of algorithm 2

This section presents a numerical investigation of a non-convex optimal reinsurance problem using Algorithm 2. We look into both the admissibility of the numerical solution and the computational efficiency of the algorithm. The former assesses whether the solution identified by Algorithm 2 adequately fulfils the decision-maker's objective, while the latter compares the algorithm's computational cost with that of other global optimisation techniques (such as the grid search method). Our focus is on solving the following optimisation model using Algorithm 2:

$$
\begin{cases}
\min_{\substack{a_1,a_2\\ b_1,b_2}} \min_t & \mathrm{VaR}_{\alpha+t}\big(X_1 - l_1(X_1)\big) + \mathrm{VaR}_{1-t}\big(X_2 - l_2(X_2)\big) + \sum_{i=1}^{2} \pi_i\big(l_i(X_i)\big) \\
\text{s.t.} & \sum_{i=1}^{2} \pi_i\big(l_i(X_i)\big) \leqslant P, \\
& 0 \leqslant t \leqslant 1-\alpha, \\
& 0 \leqslant a_i \leqslant b_i \leqslant \infty, \quad \forall\, i \in \{1,2\},
\end{cases}
\tag{3.6}
$$

where $(z)_+ = \max\{z,0\}$, and $l_i(X_i) = (X_i - a_i)_+ - (X_i - b_i)_+, \quad \forall i \in \{1,2\}$.

If the premium function π_i for each $i \in \{1,2\}$ follows the expected premium principle,

Input v_0 ; `/* initial guess */`
$f_{\text{obj}} = \Psi$; $v \leftarrow \mathscr{L}(f_{\text{obj}}, v_0)$;
$V \leftarrow v$; `/* initialisation of the solution matrix */`
$m_{\text{sol}} \leftarrow 1$; `/* number of elements in V */`
for $m = 1 : \mathscr{M}$ **do**
  $\lambda = \dfrac{m}{\mathscr{M}}, f_{\text{obj}} = \lambda\Phi + \Psi$; `/* objective function */`
  $S \leftarrow \{\}$; `/* S the auxiliary solution set */`
  for $j = 1 : m_{\text{sol}}$ **do**
    $v \leftarrow V(:, j)$; `/* new initial guess */`
    $u \leftarrow \mathscr{L}(f_{\text{obj}}, v)$, and add u to the solution set S;
    for $k = 1 : \mathscr{M}_1$ **do**
      $v \leftarrow v + r$; `/* perturbed by a random vector */`
      $u \leftarrow \mathscr{L}(f_{\text{obj}}, v)$;
      if *u is different from the existing elements in S* **then**
        add u to the solution set S ;
      end
    end
  end
  $n \leftarrow |S|$; `/* the number of the elements in S */`
  if $n \leqslant \mathscr{M}_2$ **then**
    $V \leftarrow S, m_{\text{sol}} \leftarrow n$;
  else
    keep the best $\mathscr{M}_2$ solutions in S and delete the others.
    $V \leftarrow S, m_{\text{sol}} \leftarrow \mathscr{M}_2$;
  end
  end
end
Select the best solution, denoted by $v^{\ddagger}$, among the elements from V.
Output $v^{\ddagger}$

Algorithm 2: Homotopy Method with Perturbations and Ensembles

then an analytical solution to (3.6) is available in Chapter 2, where the authors investigate an optimal reinsurance model under dependence uncertainty (see Example 1 in Chapter 2). This analytical solution is adopted as a benchmark against which we validate the numerical results produced by our algorithm.

Moreover, Chapter 2 also provides a numerical solution to (3.6) by employing a grid search approach. We use their implementation as a reference to compare the computational performance of Algorithm 2.

The process of solving optimisation problem (3.6) using Algorithm 2 involves the following numerical steps, labelled as (**NS**):

NS1: We denote N empirical observations of X_i by:

$$\mathbf{x}_i := \{x_{i1}, x_{i2}, \ldots, x_{iN}\}, \quad \forall i \in \{1,2\}. \tag{3.7}$$

These observations may be generated using Monte Carlo simulation or taken from historical data provided by insurance firms. For this example, we adopt the model setup in Chapter 2, with $N = 200$, and assume that X_i is exponentially distributed with

$$E(X_1) = 8000 \quad \text{and} \quad E(X_2) = 3000. \tag{3.8}$$

NS2: The premium function π_i is specified via the expected premium principle:

$$\pi_i\big(l_i(X_i)\big) = w_i + (1+\theta_i)E\big(l_i(X_i)\big), \quad \forall i \in \{1,2\}. \tag{3.9}$$

NS3: The parameters $w_1, w_2, \theta_1, \theta_2$ and P are assumed to be pre-determined. We select values that correspond to case (ii) in the benchmark result from Example 1 of Chapter 2. The corresponding benchmark solution is denoted as:

$$v^* := (a_1^*, a_2^*, b_1^*, b_2^*, t^*), \tag{3.10}$$

and defines the optimal reinsurance functions:

$$l_i^*(X_i) = (X_i - a_i^*)_+ - (X_i - b_i^{\ddagger})_+, \quad \forall i \in \{1,2\}. \tag{3.11}$$

These values are summarised in Table 3.1.

α	a_1^*	a_2^*	b_1^*	b_2^*	t^*
$0 \leqslant \alpha < 0.45$	b_1^*	$E(X_2)\ln(1+\theta_2)$	a_1^*	∞	0
$0.45 \leqslant \alpha \leqslant 1$	$E(X_1)\ln(1+\theta_1)$	$E(X_2)\ln(1+\theta_2)$	$\mathrm{VaR}_{\alpha+t^*}(X_1)$	∞	0

Table 3.1: Relevant section of the benchmark result from Example 1 in Chapter 2 under our parameter assumptions.

NS4: Algorithm 2 is applied to solve problem (3.6), producing an optimal solution denoted

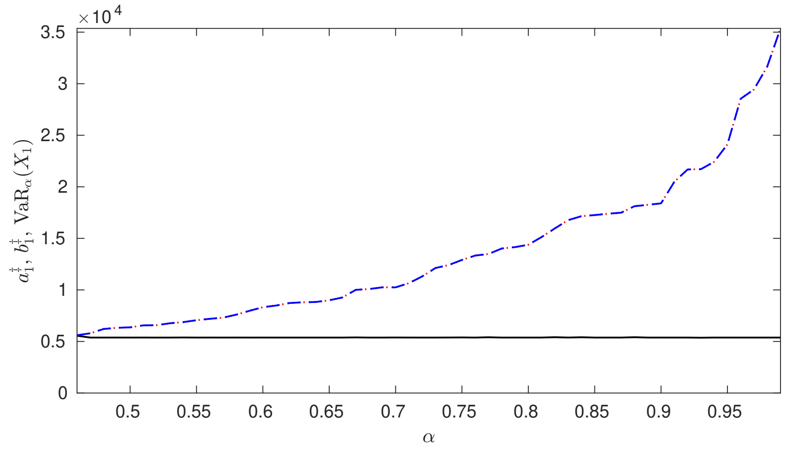

Figure 3.1: Plot of $a_1^\ddagger$ (black solid), $b_1^\ddagger$ (blue dashed), and $\mathrm{VaR}_\alpha(X_1)$ (red dotted) for $\alpha \in [0.45, 1]$.

by:

$$v^\ddagger = \left(a_1^\ddagger, a_2^\ddagger, b_1^\ddagger, b_2^\ddagger, t^\ddagger\right).$$

The corresponding reinsurance strategies are:

$$l_i^\ddagger(X_i) = (X_i - a_i^\ddagger)_+ - (X_i - b_i^\ddagger)_+, \quad \forall i \in \{1,2\}. \tag{3.12}$$

The resulting solution $v^\ddagger$ is illustrated in Figures 3.1 to 3.3. We compare it to the benchmark result from Chapter 2 to validate accuracy. Figure 3.1 plots $a_1^\ddagger$, $b_1^\ddagger$, and $\mathrm{VaR}_\alpha(X_1)$ for $\alpha \in [0.45, 1]$. The following relation holds for all such α:

$$a_1^\ddagger = \bar{x}_1 \ln(1+\theta_1), \quad b_1^\ddagger = \mathrm{VaR}_{\alpha+t^\ddagger}(X_1), \quad t^\ddagger = 0, \tag{3.13}$$

where

$$\bar{x}_1 := \frac{1}{N}\sum_{j=1}^{N} x_{1j} \tag{3.14}$$

is the empirical mean of X_1. These results align with those presented in Table 3.1. For $\alpha \in [0, 0.45)$, we omit plotting $a_1^\ddagger$ and $b_1^\ddagger$ in Figure 3.1 since the benchmark solution is not unique. In fact, the condition $a_1^* = b_1^*$ implies $l_1^\ddagger(X_1) = 0$, as depicted in Figure 3.2.

As indicated in Table 3.1, $l_2^*(X_2)$ is a stop-loss reinsurance with a constant retention a_2^*

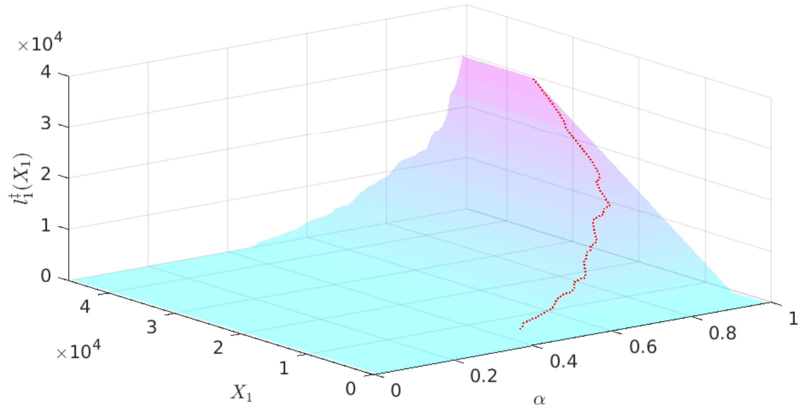

Figure 3.2: 3D view of $(X_1, l_1^{\ddagger}(X_1), \alpha)$; $\mathrm{VaR}_\alpha(X_1)$ shown in red dotted.

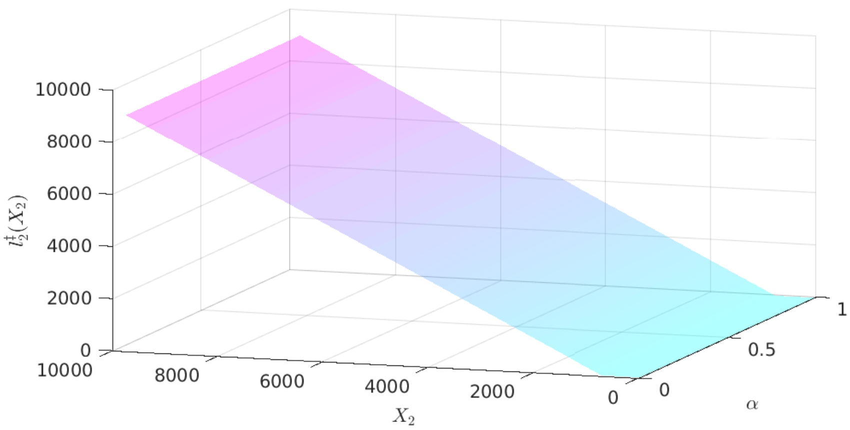

Figure 3.3: 3D plot of $(X_2, l_2^{\ddagger}(X_2), \alpha)$.

across all $\alpha \in [0,1]$. This observation is supported by our numerical result $v^{\ddagger}$. Specifically, we find:

$$a_2^{\ddagger} = \bar{x}_2 \ln(1+\theta_2), \quad b_2^{\ddagger} = \mathrm{VaR}_{1-t^{\ddagger}}(X_1), \quad t^{\ddagger} = 0, \tag{3.15}$$

where

$$\bar{x}_2 := \frac{1}{N} \sum_{j=1}^{N} x_{2j} \tag{3.16}$$

denotes the empirical mean of X_2. Figure 3.3 confirms this structure by visualising $l_2^{\ddagger}(X_2)$ as a stop-loss contract with a fixed retention level across all values of α.

3.3.1 Accuracy

This section presents a detailed evaluation of the accuracy of the numerical solution $v^{\ddagger}$ obtained using Algorithm 2, with a particular focus on the likelihood that $v^{\ddagger}$ is admissible. Let G^{*} denote the objective function value corresponding to the benchmark solution v^{*}, and $G^{\ddagger}$ the value associated with the computed solution $v^{\ddagger}$. The solution $v^{\ddagger}$ is considered admissible if and only if the following criterion is satisfied:

$$\left|\frac{G^{\ddagger}-G^{*}}{G^{*}}\right| < 1\%\,. \tag{3.17}$$

In other words, a numerical solution is deemed admissible if its objective value deviates from the benchmark value by no more than 1%. This is a stringent condition that allows only minor numerical discrepancies. It is important to note that comparing $v^{\ddagger}$ directly with v^{*} is not always feasible due to the potential existence of multiple solutions, such as when $a_1^{*} = b_1^{*}$ for $\alpha \in [0, 0.45)$.

The numerical experiment to assess the accuracy of $v^{\ddagger}$ proceeds according to the following steps, denoted by **ES**:

ES1: Generate random samples of $\mathbf{x}_1$ and $\mathbf{x}_2$ as described in **NS1** of the numerical procedure.

ES2: Solve the optimal reinsurance problem (3.6) using Algorithm 2 with an arbitrary initial guess. In practice, v_0 may be chosen as a Gaussian random vector or a zero vector. The resulting solution is denoted by $v^{\ddagger}$.

ES3: Assess whether $v^{\ddagger}$ is admissible according to the criterion given in equation (3.17).

ES4: Repeat steps **ES1–ES3** a total of 100 times. Let n_{a} denote the number of times an admissible solution is obtained. Therefore, the proportion of admissibility is then given by: $\eta = \frac{n_{\mathrm{a}}}{100} \times 100\%\,.$

This analysis aims to address the following questions:

Q1: How do the algorithmic parameters (i.e., $\mathscr{M}$, $\mathscr{M}_1$, and $\mathscr{M}_2$) influence the convergence behaviour of the algorithm?

Q2: What effect does the choice of initial guess v_0 have on the algorithm's convergence?

To address **Q1**, we conduct experiments using the following six sets of parameter values:

1. $\mathscr{M}_2 = 4$, $\mathscr{M}_1 = 4$, $\mathscr{M} = 10$.
2. $\mathscr{M}_2 = 8$, $\mathscr{M}_1 = 8$, $\mathscr{M} = 10$.
3. $\mathscr{M}_2 = 8$, $\mathscr{M}_1 = 8$, $\mathscr{M} = 20$.
4. $\mathscr{M}_2 = 8$, $\mathscr{M}_1 = 16$, $\mathscr{M} = 10$.
5. $\mathscr{M}_2 = 16$, $\mathscr{M}_1 = 16$, $\mathscr{M} = 10$.
6. $\mathscr{M}_2 = 32$, $\mathscr{M}_1 = 32$, $\mathscr{M} = 10$.

Naturally, increasing the parameter values leads to a greater computational burden. The sets are arranged in ascending order of computational cost, with set (1) being the least demanding and set (6) the most. Furthermore, since the form of the optimal solution is influenced by the value of α, each parameter set is tested at five values of α: 0.1, 0.3, 0.5, 0.7, and 0.9.

	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.7$	$\alpha = 0.9$
Set 1	100%	100%	100%	62%	93%
Set 2	100%	100%	100%	96%	100%
Set 3	100%	100%	100%	99%	100%
Set 4	100%	100%	100%	100%	100%
Set 5	100%	100%	100%	100%	100%
Set 6	100%	100%	100%	100%	100%

Table 3.2: Likelihood of obtaining an admissible solution, η, for various values of α and algorithmic parameter sets.

Table 3.2 reports the values of η in different parameter sets and the values of α. Under parameter set (1), convergence is excellent for $\alpha = 0.1$, 0.3, and 0.5, but deteriorates at $\alpha = 0.7$ and 0.9. This decline is attributed to the presence of multiple local optima in close proximity. However, this issue can be mitigated by increasing computational effort via larger parameter values. In particular, sets (4) to (6) achieve perfect convergence for all tested values of α.

Effect of Initial Guess

We now turn to the effect of the initial guess v_0. Up to this point, the algorithm has used arbitrary initial guesses. To investigate whether incorporating prior knowledge into the choice of v_0 can enhance convergence, we conduct further experiments with $\alpha = 0.7$ under parameter sets (1) and (2). Instead of arbitrary choices, the initial guess is set to $v_0 = v^{\ddagger}_{0.69}$, where $v^{\ddagger}_{0.69}$ denotes the numerical solution previously obtained for $\alpha = 0.69$. The value 0.69 is chosen because it is the nearest α value below 0.7.

The results show that, under set (1), the likelihood of admissibility improves from 62% to 65%, while under set (2), it increases from 96% to 98%. These marginal gains suggest that the influence of the initial guess is minor, due to the robustness of Algorithm 2, which benefits from the perturbation and ensemble strategies. Therefore, the algorithm remains effective even when little or no prior information is available, reinforcing its suitability for solving complex optimisation problems.

3.3.2 The computational efficiency

All simulations are carried out on a desktop computer equipped with an Intel Quad-Core i7-3770 CPU, running a 64-bit Linux operating system. The computation times required to solve the optimal reinsurance problem 3.6 using the grid search method introduced in Chapter 2, for various values of α, are reported in Table 3.3.

α	Computing time (seconds)
0.9	977.97
0.7	4364.30
0.5	7748.17
0.3	9371.02
0.1	12390.15

Table 3.3: Computation times of the global grid search method for different values of α.

It can be observed that the computation time increases considerably as α decreases. This trend is primarily due to the rapid growth in the number of grid points, denoted by τ, required to be evaluated. In fact, $\tau = N(1-\alpha)$. In contrast, the computing time for solving problem (3.6) using Algorithm 2 remains relatively stable across different values of α. For this reason, we only present the computation times for $\alpha = 0.9$ in Table 3.4.

$\mathscr{M}_2$	$\mathscr{M}_1$	$\mathscr{M}$	Computing time (seconds)
4	4	10	25.65
		20	51.29
	8	10	63.80
		20	76.75
8	8	10	104.32
		20	158.70
	16	10	142.10
		20	332.82
16	16	10	357.90
		20	743.58
32	32	10	1461.14

Table 3.4: Computation times of Algorithm 2 for different parameter settings. Results correspond to the case $\alpha = 0.9$ (similar times are observed for other values of α).

As noted previously, increasing the parameter values leads to a higher computational burden and, consequently, longer run times. However, even when using parameter set (4), i.e., $\mathscr{M}_2 = 8$, $\mathscr{M}_1 = 16$, and $\mathscr{M} = 10$, the algorithm completes in 142.10 seconds, which is about only 14.5% of the time required by the grid search method (977.97 seconds). This efficiency advantage becomes even more significant as α decreases.

Further, since the loop from line 5 to line 15 in Algorithm 2 involves independent computations, it can be readily adapted for parallel computing, which would result in substantial gains in efficiency. However, to keep the analysis straightforward, no parallelisation has been implemented in the present study. It is also worth mentioning that modern computers, with significantly more powerful CPUs and GPUs than the demonstration setup, are well-suited for employing larger parameter settings than those used in our experiments. This makes it practically feasible for practitioners to further improve both the efficiency and the accuracy of the algorithm.

3.4 Numerical example

In Section 3.3, both risk variables X_1 and X_2 were assumed to follow exponential distributions. While convenient, this assumption may not reflect realistic risk profiles, especially in heavy-tail settings. Therefore, in this section, we consider the more representative Pareto distribution as the underlying risk distribution. As closed-form analytical solutions are not available for this case, a numerical method is employed. Specifically, we apply Algorithm 2 with parameters $\mathscr{M}_1 = 32$, $\mathscr{M}_2 = 32$, and $\mathscr{M} = 10$, which produced the best performance to solve the optimal reinsurance problem defined in (3.6).

We now assume that both X_1 and X_2 follow Pareto distributions with density functions given by:

$$f_{X_1}(x) = \frac{\gamma_1 \beta_1^{\gamma_1}}{(\beta_1 + x)^{\gamma_1+1}}, \qquad f_{X_2}(x) = \frac{\gamma_2 \beta_2^{\gamma_2}}{(\beta_2 + x)^{\gamma_2+1}}. \tag{3.18}$$

In this example, the parameters $\gamma_1 = \gamma_2 = 3$ while β_1 and β_2 are selected such that $\mathbb{E}[X_1] = 8000$ and $\mathbb{E}[X_2] = 3000$, consistent with the values used in (3.8).

Let the optimal solution returned by the numerical solver be denoted by:

$$\tilde{v} := (\tilde{a}_1, \tilde{a}_2, \tilde{b}_1, \tilde{b}_2, \tilde{t}),$$

which yields the corresponding optimal reinsurance functions:

$$\tilde{l}_i(X_i) = (X_i - \tilde{a}_i)_+ - (X_i - \tilde{b}_i)_+, \quad \text{for } i \in \{1,2\}.$$

Figures 3.4 and 3.5 illustrate the gradient profiles $d\tilde{l}_1(X_1)/dX_1$ and $d\tilde{l}_2(X_2)/dX_2$, respectively, using a black and white colour scheme. Black indicates regions where the gradient is zero, white white denotes regions where it is one.

From the figures, we observe that $\tilde{l}_1$ changes form at three key values of α: approximately 0.34, 0.375, and 0.4, as indicated by vertical dashed lines in the zoomed plot of Figure 3.4.

A zero reinsurance function $\tilde{l}_1$ appears for $\alpha < 0.34$ and $0.375 < \alpha < 0.4$, corresponding to the horizontal black bands. For other values of α, the classical two-layer stop-loss form emerges. Selected solution profiles for various α values are shown in Figures 3.6 and 3.7.

The first turning points $\tilde{a}_1$ and $\tilde{a}_2$ are constants equal to the 44.5%-quantile of X_1 and the 23.5%-quantile of X_2, respectively, corresponding to vertical boundaries between colour bands in Figures 3.4 and 3.5.

The second turning points $\tilde{b}_1$ and $\tilde{b}_2$ coincide with $\mathrm{VaR}_{\alpha+\tilde{t}}(X_1)$ and $\mathrm{VaR}_{1-\tilde{t}}(X_2)$, respectively, as shown in Figure 3.8.

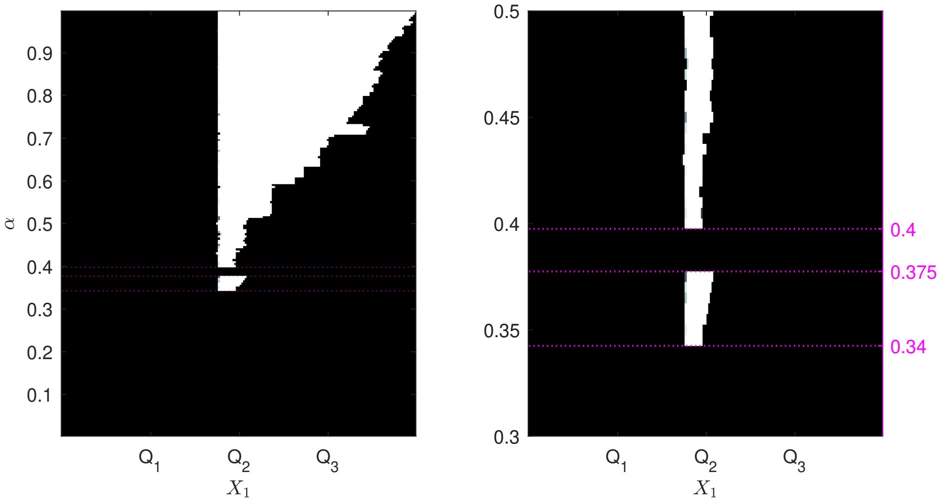

Figure 3.4: (Left) Gradient plots of solution profiles: black regions represent zero gradient, while white regions represent unit gradient. Q_1, Q_2, and Q_3 denote the first, second, and third quantiles of X_1. (Right) Zoom in around $\alpha \approx 0.4$.

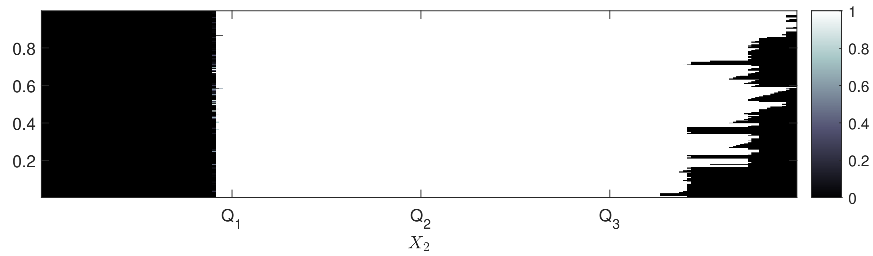

Figure 3.5: Same plot as Figure 3.4, but for X_2. A colour bar is shown on the right.

In particular, unlike in Section 3.3, the parameter $\tilde{t}$ is not always zero (Figure 3.9). This leads to non-smooth transitions in the gradient plots, and causes $\tilde{b}_1$ to be non-monotonic in α, and $\tilde{b}_2$ to deviate from the maximum of X_2, capping $\tilde{l}_2$ with a second layer.

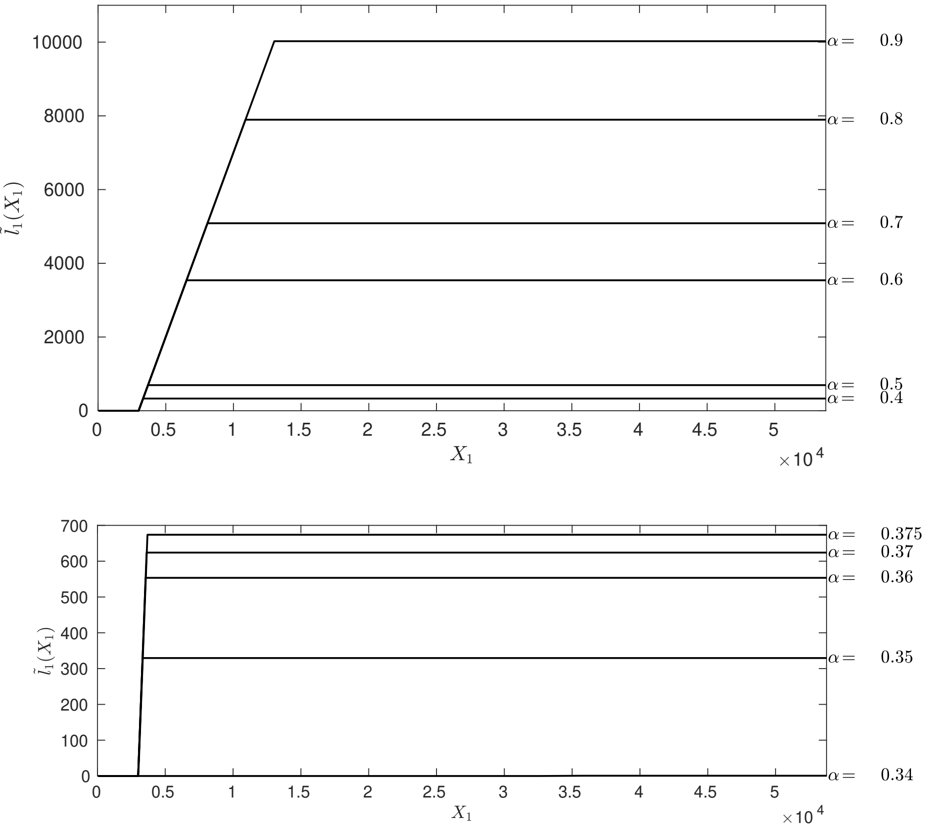

Figure 3.6: (Top) Plots of $\tilde{l}_1(X_1)$ for $\alpha = 0.9$, 0.8, 0.7, 0.6, 0.5, and 0.4. (Bottom) $\alpha = 0.375$, 0.37, 0.36, 0.35, and 0.34. $\tilde{a}_1 = 2995.335$. $\tilde{l}_1$ is identically zero for $\alpha \leqslant 0.34$.

Remark 3.1. Numerical findings for problem (3.6) with Pareto densities in (3.18) are summarised below:

- For $\alpha \in (0, 0.339] \cup [0.376, 0.399]$, $\tilde{l}_1(X_1)$ is identically zero.

- For $\alpha \in [0.34, 0.375] \cup [0.4, 1)$, $\tilde{l}_1(X_1)$ takes the form of a two-layered stop-loss:

$$\tilde{a}_1 = 2995.335, \quad \tilde{b}_1 = \mathrm{VaR}_{\alpha+\tilde{t}}(X_1).$$

- Structural changes in $\tilde{l}_1$ occur around $\alpha \approx 0.34$, 0.375, and 0.4.

- For all $\alpha \in (0, 1)$, $\tilde{l}_2(X_2)$ is a two-layered stop-loss with:

$$\tilde{a}_2 = 577.374, \quad \tilde{b}_2 = \mathrm{VaR}_{1-\tilde{t}}(X_2).$$

- The auxiliary variable $\tilde{t}$, computed numerically, is often non-zero.

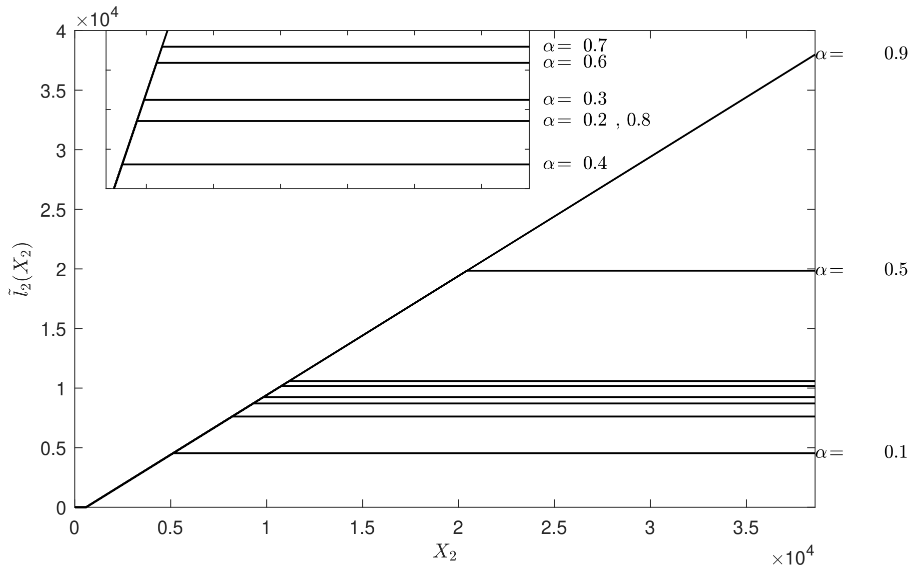

Figure 3.7: Plots of $\tilde{l}_2(X_2)$ for various α values. Top: zoomed-in view. $\tilde{a}_2 = 577.374$.

Compared to $l_1^{\ddagger}$ in Section 3.3, the function $\tilde{l}_1$ exhibits more frequent transitions between zero and stop-loss forms as α varies. This pattern appears novel in the literature and currently lacks a theoretical explanation, possibly due to the existence of multiple optima. To validate these findings, we compare them with the results obtained from a global optimisation routine. As shown in Figure 3.10, the two sets of results align closely, supporting the presence of an unusual region around $[0.34, 0.375]$.

Further theoretical investigation is needed to fully understand the origin and nature of this behaviour.

3.5 Conclusion

Reinsurance is a widely adopted risk management strategy among insurers. By purchasing reinsurance contracts, insurers can effectively transfer a portion of their risk exposure to one or more reinsurers. The fundamental objective of an optimal reinsurance problem is to determine the best trade-off between risk mitigation and cost efficiency for the insurer. However, the existing literature often relies on simplifying assumptions to maintain the convexity and tractability of such models, as incorporating more realistic objectives or constraints usually compromises the tractability of the optimisation problem.

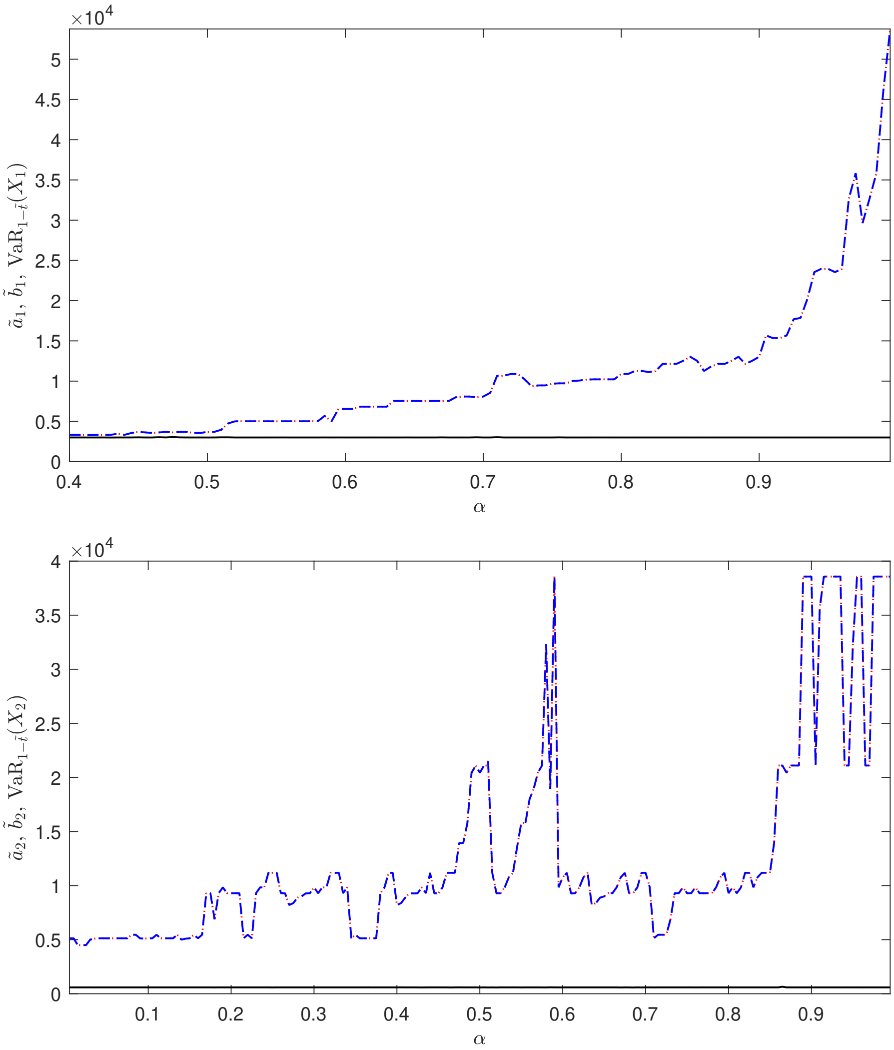

Figure 3.8: (Top) Plots of $\tilde{a}_1$ (black), $\tilde{b}_1$ (blue dashed), and $\mathrm{VaR}_{\alpha+\tilde{t}}(X_1)$ (red dotted). (Bottom) Same for X_2.

To tackle this issue, the present work proposes a homotopy-based algorithm, one that integrates perturbation and ensemble techniques to address non-convex optimal reinsurance problems. A benchmark model is solved using Algorithm 2, and the resulting numerical solution is validated against an analytical solution previously established in the literature. Notably, the numerical solution obtained by Algorithm 2 exhibits the same functional structure as its analytical counterpart.

Further experiments have been conducted to evaluate the robustness of the numerical solution. Specifically, the influence of algorithmic parameter settings and initial guess selection on solution stability is investigated, providing deeper insight into algorithmic performance. The computational efficiency of Algorithm 2 is also assessed by comparing its

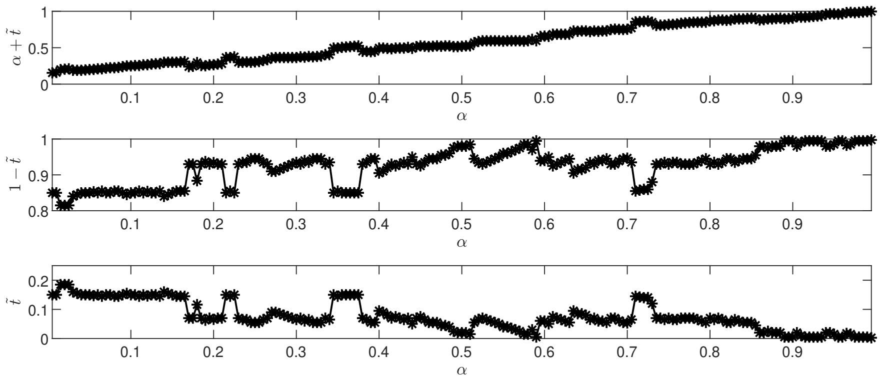

Figure 3.9: (Top) Plot of $\alpha+\tilde{t}$ versus α. (Middle) $1-\tilde{t}$ vs α. (Bottom) $\tilde{t}$ vs α.

run times to those obtained using the traditional grid search approach.

Subsequently, Algorithm 2 is applied to a second non-convex actuarial problem, for which no analytical solution exists, involving risk variables that follow a Pareto distribution. The numerical results, summarised in Remark 3.1, reveal novel features not previously reported in the literature, as illustrated in Figure 3.4. These insights, unattainable via classical analytical or numerical methods, underscore the novelty of the approach.

Overall, the homotopy-based algorithm enhanced with perturbations and ensembles demonstrates strong potential as a robust and efficient numerical method for solving non-convex optimal reinsurance problems. Its success suggests promising avenues for future applications.

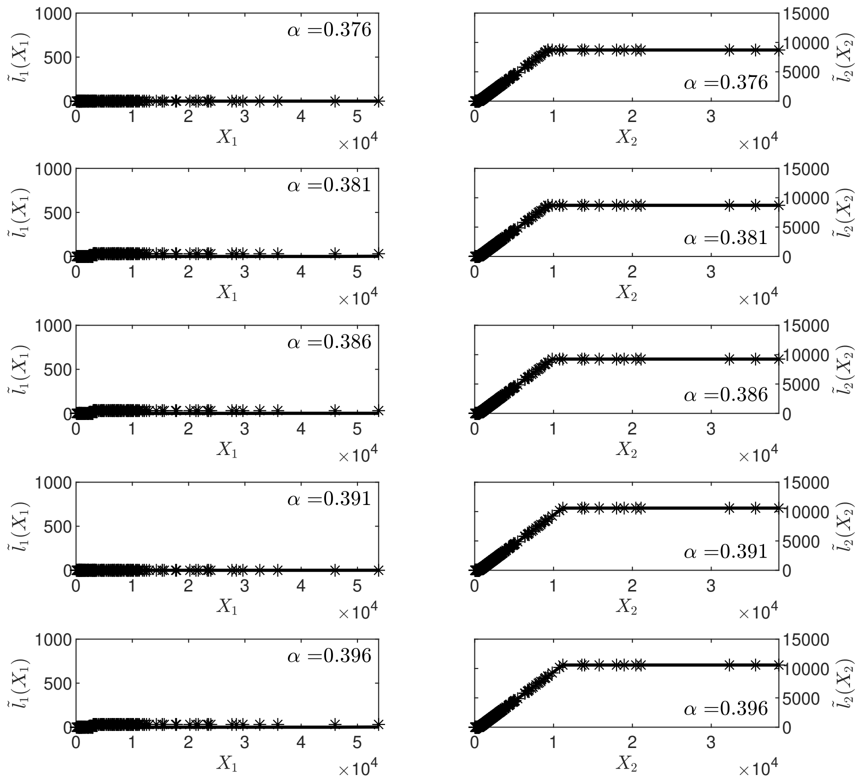

Figure 3.10: Solution profiles for $\alpha \in [0.375, 0.4]$: solid curves (homotopy method), asterisks (global search).

CHAPTER 4

Robust distortion risk metrics and portfolio optimisation

4.1 Introduction

In classical decision-making models, the decision-maker typically assumes that the random variable X follows a known distribution F, and then evaluates risk through a law-invariant functional $\mathscr{V}$, such as variance, expected (dis)utility, or value-at-risk (VaR), so that $V(X) := V(F)$. This methodology, however, relies critically on the accuracy of a single probabilistic model. When the assumed distribution is uncertain or misspecified, such computed decisions may turn out to be poor. Consequently, how we incorporate distributional uncertainty into decision-making has become a fundamental issue in economics, finance, engineering, and operations research. One particularly influential modelling approach is Distributionally Robust Optimisation (DRO). In its standard form, DRO amounts to dealing with a problem of the form:

$$\min_{\theta \in \Theta} \max_{F_{\mathbf{X}} \in \mathscr{F}} \mathscr{V}(f(\theta, \mathbf{X})). \tag{4.1}$$

Here, f is a loss function, θ reflects a decision vector (e.g., weights), $F_{\mathbf{X}}$ is an admissible distribution function for a given risk vector $\mathbf{X}$, and $\mathscr{F}$ is the uncertainty set containing all admissible distribution functions. A DRO thus reflects the basic idea that one aims to make decisions that perform optimally under worst-case scenarios. In this chapter, we contribute to the literature by solving for uncertainty sets that are reflective of real-world ambiguity in relation to the inner max problem for a very broad class of decision functionals, $\mathscr{V}$, that have been used in real-world decision making. Here, we apply the results to address novel robust portfolio selection problems of far-reaching practical interest.

The evolution of the min–max framework for decision making under uncertainty originates with Scarf (1958), who studied the news vendor problem under distributional uncertainty of demand. The theoretical justification for the min–max framework is based on the axiomatic foundation of Gilboa and Schmeidler (1989), who analysed expected utility under conditions of uncertainty by considering the worst-case expected utility over a family of distributions (see also Hansen and Sargent (2001), who introduced the min–max principle in the context of robust control). The development of modern robust optimisation techniques over the past decade, such as Ben-Tal et al. (2009), has further contributed to the successful application of DROs in various areas, including engineering, finance, operations research, and economics.

As for the choice of the decision functional, $\mathscr{V}$, adopting the expected (dis)utility framework appears to be the most natural choice given its prominent place in the academic literature. There are, however, a series of limitations to its use, as stated in Starmer (2000). First, it is not at all obvious for decision-makers to specify their utility function. In the context of optimal portfolio strategies, Brennan and Solanki (1981) points out that, “from a practical point of view, it may well prove easier for the investor to choose directly his optimal quantile function than it would be for him to communicate his utility function to a portfolio manager.” The same observation has led Goldstein et al. (2008) to introduce a tool, called the distribution builder, which makes it possible for investors to analyse their desired pay-off function and to elicit a utility to explain their choice. Second, there is ample empirical evidence that real-world decision-making cannot be reconciled with the use of utility functions (e.g., Allais (1953)). In response to this criticism, numerous alternatives, including Yaari’s dual theory (Yaari (1987)), rank-dependent utility theory (Quiggin (1982)), and cumulative prospect theory (Tversky and Kahneman (1992)), have been proposed.

All these theories have been justified by proposing axioms that are considered “sensible". While providing a prescriptive foundation for a decision theory, as a practical guide to making choices, is important, the real issue lies in understanding how people actually make choices. As such, observed real-world behaviour should not be dismissed simply because it violates conventional choice axioms as formulated in Starmer (2000). In this context, Yaari’s dual theory holds considerable appeal because it aligns more closely with observed decision-making behaviours. Indeed, this theory gives rise to quantile-based functionals, called distortion risk measures, such as value-at-risk (VaR), range value at risk (RVaR), and tail value at risk (TVaR). All of these functionals are actually used in real-world decision-making, with the central tenet being that they reflect the human tendency to ask such questions as “What if this happens?” or “What would I lose under this scenario?"[1] Our setup is, however, broader in that we do not require the distortion function to be non-decreasing. This makes it possible to extend the scope and also include the inter-quartile range (IQR), the Gini deviation (GD), and the mean-median deviation (or the difference between two distortion risk measures) within our framework. Specifically, the functionals that we consider are labelled distortion risk metrics[2] and will be denoted by ρ_g, where g reflects the underlying distortion function. They were first proposed in Wang et al. (2020a).

We are not the first to pursue DRO using distortion risk measures and their associated generalisation distortion risk metrics. This was first performed by Ben-Tal et al. (2009) for the case of VaR, which was further extended in Chen et al. (2011) to include TVaR, and later significantly generalised by Li (2018) to extend to the entire class of convex distortion risk measures, and by Cai et al. (2025) to general distortion risk measures. In all these works, the ambiguity set $\mathscr{F}$ is characterised by the mean and covariance matrix of the random vector $\mathbf{X}$.

[1]Roese and Olson (1995) attribute this to mental simulation and counterfactual thinking. That is, people naturally engage in mental simulations of (extreme) events, imagining what could happen in those scenarios; see also Tversky and Kahneman Tversky and Kahneman (1973) who explain this behaviour as rational, given that extreme events are more memorable.

[2]In Wang et al. (2020b) they are called distortion risk metrics, which is a slightly different terminology from ours.

To deal with a DRO, one needs to first solve an inner problem of the form:

$$\max_{F_Y \in \widetilde{\mathscr{F}}} \rho_g(Y). \tag{4.2}$$

Problem 4.2 specifically deals with the extent to which measurements can be affected by model mis-specification. This problem is of particular relevance in statistics, where it has been studied under the preconceived notion of robust statistics, and also in the financial industry, where the assessment of model uncertainty became a regulatory priority in the aftermath of the 2008–2009 Global Financial Crisis (GFC). For instance, in February 2017, the European Central Bank (ECB) published its Guide to the Targeted Review of Internal Models, in which it declared that every institution "should have a model risk management framework in place that allows it to identify, understand, and manage its model risk" European Central Bank (2017). An early contribution in this regard is the seminal Cantelli bounds on tail risk (survival probabilities); by inversion, this result yields a sharp bound on VaR. An explicit formula can be traced back to El Ghaoui et al. (2003). Interestingly, the bound on VaR is achieved by a two-point distribution, and it follows that the same bound also applies to TVaR, which can be viewed as the concave distortion risk measure closest to VaR. Since then, it has become apparent that this correspondence between VaR and its concavation TVaR carries over to general distortion risk measures. Indeed, Li (2018), Cai et al. (2025), and Pesenti et al. (2024) show that, for suitable uncertainty sets, Problem (4.2) is equivalent to the case in which g is replaced by the smallest concave function g^* that dominates g. This is a very relevant result, as when considering the DRO problem, if (4.1) f is concave in θ, it becomes a convex-concave optimisation problem for which powerful computational methods are available. As shown in Pesenti et al. (2024), the stated equivalence does not hold if the Wasserstein distance is involved, as is the case in our study. For more studies of worst-case problems in operations research and its associated applications (see e.g., Chen et al. (2011) and Zymler et al. (2013)).

The ambiguity set is a key component of any distributionally robust optimisation model. Rather than letting mathematical convenience drive the choice of the set, it should instead be primarily guided by the available data, possibly complemented by expert opinion. It should be sufficiently large to reasonably include the true distribution, but not so broad that it admits implausible distributions, as this could lead to overly conservative decision-making. For example, the distribution function that maximises VaR and TVaR (i.e., the Cantelli bound), under the sole knowledge of mean and variance, has only two mass points, although this is not plausible in practice. This serves to highlight that more (possibly qualitative) information is required.

In this paper, we study bounds on distortion risk metrics for ambiguity sets that are arguably highly relevant from a practical standpoint, including the case in which the ambiguity set, in addition to containing distributions with given first two moments, is also restricted to unimodal distributions that are close to some reference distribution. The assumption of unimodality is very reasonable in that it usually complies with data, which also explains why unimodal distributions are routinely used in engineering, operations research, insurance, and financial risk modelling. For instance, Pareto, Gamma, Normal, Log-Normal, Beta, Weibull, and Student t-distributions are all unimodal. The literature on risk bounds for unimodal distribution functions is, however, limited. Popescu (2005) proposes semi-definite programming to determine best-possible bounds on tail (survival) probabilities under mean, variance, and unimodality constraints, from which, by numerical inversion, VaR bounds can be obtained. Li et al. (2018) and Bernard et al. (2023) derive explicit bounds on RVaR, although results for general distortion risk metrics appear to be missing.

In this chapter, we show that bounds on distortion risk metrics obtained under the unimodality assumption offer significant improvements over bounds obtained without this assumption. To address the natural requirement that admissible distributions are "close enough" to a reference distribution, we use the 2-Wasserstein distance. This choice, aside from offering mathematical convenience, allows us to handle distributions with differing supports and to be consistent with mean-variance constraint.

4.1.1 Our contributions

For uncertainty sets defined by fixed mean, variance, and 2-Wasserstein distance, we establish worst-case values for general distortion metrics under minimal conditions in which the distortion function only needs to have finite variation on $[0,1]$, meaning that its total variation is finite:

$$\sup_{0=t_0<\ldots<t_n=1} \sum_{i=1}^{n} |g(t_i) - g(t_{i-1})| < \infty.$$

This includes cases where g is non-monotone or discontinuous. When g is upper semi-continuous, we also identify the worst-case distribution. Our findings extend those of Bernard et al. (2024), who required g to be increasing and absolutely continuous. Since the projection method in Bernard et al. (2024) does not apply to discontinuous distortions, we introduce an alternative approach, which is a variant of the concave envelope method. Unlike Cai et al. (2025) and Pesenti et al. (2024), who constructed envelopes directly on g, we build the concave envelope on a linear combination of g and a functional of the reference distribution, with a combination parameter. We then optimise over this parameter, with the continuity of the concave envelope with respect to the parameter being crucial, and its proof is non-trivial. Using this method, we derive sharp bounds for measures such as GlueVaR[3], inter-quantile difference, and the discrepancy between expected shortfall (ES) and VaR. These are results which were not previously available.

When the uncertainty set also incorporates unimodality, alongside mean, variance, and Wasserstein distance, we develop bounds for absolutely continuous distortion functions. We begin by considering sets with fixed mean, variance, and unimodality at a known inflection point. By projecting onto the class of increasing and concave-convex functions with a fixed inflection point on $(0,1)$, we derive worst-case values and distributions. While prior work focused on specific measures such as VaR or RVaR (without fixing the inflection point), we develop a general theory, one that incorporates unimodality. Explicit projection formulas are often unavailable, so we have designed an efficient approximation algorithm, one that attains any desired degree of accuracy. Extending this, we obtain bounds when all four constraints—the mean, variance, unimodality, and Wasserstein distance—are imposed. Finally, we consider unimodality with an unknown inflection point, deriving bounds valid over an interval.

[3]Proposed by Belles-Sampera et al. (2013) as a compromise between VaR and ES: the former often underestimates, while the latter is seen as overly conservative. Cai et al. (2025) stress the practical demand for such measures in industry.

The results are then applied to portfolio optimisation and risk quantification under conditions of model uncertainty.

4.2 Preliminary

Let $(\Omega, \mathscr{A}, \mathbb{P})$ be an atomless probability space. Denote by L^2 the collection of real-valued random variables with finite second moment, and by $\mathscr{M}^2$, the set of their corresponding distribution functions. A positive value of a random variable is interpreted as a financial loss. All random variables and distribution functions that appear below are assumed to lie in L^2 and $\mathscr{M}^2$, respectively. For a distribution function G, the left and right quantiles are defined by:

$$G^{-1}(p) = \inf\{x \in \mathbb{R} : G(x) \geqslant p\}, \qquad p \in (0,1],$$

and

$$G_+^{-1}(p) = \inf\{x \in \mathbb{R} : G(x) > p\}, \qquad p \in [0,1),$$

respectively, with the convention $\inf \emptyset = \infty$. The left quantile is often called the VaR. We use the notation VaR for the left quantiles and VaR^+ for the right quantiles. The expected shortfall (ES), another widely used regulatory risk measure in banking and finance, is given by:

$$\mathrm{ES}_\alpha(G) = \frac{1}{1-\alpha} \int_\alpha^1 G^{-1}(t)\,\mathrm{d}t, \qquad 0 \leqslant \alpha < 1.$$

We denote by $\mathscr{H}$ being the set of functions $g : [0,1] \to \mathbb{R}$ with finite variation that satisfy $g(0) = g(0+) = 0$ and $g(1) = g(1-)$. For $g \in \mathscr{H}$ set $\hat{g}(x) = \max\{g(x-), g(x), g(x+)\}$ for $x \in (0,1)$ and $\hat{g}(x) = g(x)$ for $x \in \{0,1\}$. Thus, $\hat{g}$ is the upper semi-continuous representative of g. For $g \in \mathscr{H}$, the associated distortion risk metric $\rho_g : \mathscr{M}^2 \to \mathbb{R}$ is defined by:

$$\rho_g(G) = \int_0^\infty g(1-G(x))\,\mathrm{d}x + \int_{-\infty}^0 \big(g(1-G(x)) - g(1)\big)\,\mathrm{d}x. \tag{4.3}$$

The objective of this paper is to identify the *worst-case* and *best-case* values of a distortion risk metric, ρ_g, over specified distributional uncertainty sets $\mathscr{F} \subseteq \mathscr{M}^2$. That is, we consider optimisation problems of the form:

$$\sup_{G\in\mathscr{F}} \rho_g(G) \qquad (4.4a) \qquad\qquad \inf_{G\in\mathscr{F}} \rho_g(G)\,. \qquad (4.4b)$$

The sets $\mathscr{F}$ will contain all distribution functions with prescribed mean and variance that lie within a Wasserstein ball around a given reference distribution F and/or are unimodal. By $\mathscr{F}^{-1}$, we denote the collection of quantile functions corresponding to the cdfs in $\mathscr{F}$.

Besides, the worst-case and best-case values, we also investigate the *worst-case* and *best-case distribution functions* if they exist, i.e. the distributions that attain (4.4a) and (4.4b), respectively.

Note that

$$\inf_{G\in\mathscr{F}} \rho_g(G) = -\sup_{G\in\mathscr{F}} -\rho_g(G) = -\sup_{G\in\mathscr{F}} \rho_{-g}(G), \tag{4.5}$$

where ρ_{-g} is a distortion risk metric as well. Therefore, G^* is a worst-case distribution for $\sup_{G\in\mathscr{F}} \rho_{-g}(G)$, if and only if it is a best-case distribution for $\inf_{G\in\mathscr{F}} \rho_g(G)$. Therefore, the minimisation problem (4.4b) can be transferred to a maximisation problem of the same form (4.4a). This observation is one of the motivations for studying distortion risk metrics (rather than only distortion risk measures), as also noted in Pesenti et al. (2024). In this paper, we concentrate primarily on (4.4a).

We conclude this section by describing several distortion risk metrics that will be discussed in the following sections. We begin with three variability distortion risk metrics: the Gini deviation (GD), the mean–median deviation (MMD), and the inter-quantile difference (IQD).

The Gini deviation of a distribution G is given by:

$$\mathrm{GD}(G) = \frac{1}{2}\mathbb{E}(|X-Y|) = \rho_{g_{\mathrm{GD}}}(G),$$

where $X \sim G$ and $Y \sim G$ are independent, and $g_{\mathrm{GD}}(t) = t - t^2$ for $t \in [0,1]$. Thus, GD measures the average absolute difference between two independent draws from G. After suitable normalisation, it yields the Gini coefficient, a standard measure of income inequality. In finance, Shalit and Yitzhaki (1984) proposed the Gini deviation as an alternative to variance in Markowitz's portfolio model. Specifically, these authors develop a portfolio selection approach based on the mean and the Gini deviation as measures of return and risk, respectively. Apart from being more robust, the use of the Gini deviation also enables the derivation of necessary conditions for stochastic dominance, allowing agents to eliminate any feasible solutions from the efficient set that are stochastically dominated by others.

The mean–median deviation of G is defined as:

$$\mathrm{MMD}(G) = \min_{x\in\mathbb{R}} \mathbb{E}(|X-x|) = \mathbb{E}\big(|X-G^{-1}(1/2)|\big) = \rho_{g_{\mathrm{MMD}}}(G),$$

where $X \sim G$ and $g_{\mathrm{MMD}}(t) = t \wedge (1-t)$ for $t \in [0,1]$.

For the inter-quantile difference, we define

$$\mathrm{IQD}^{+}_{\alpha}(G) = G_{+}^{-1}(1-\alpha) - G^{-1}(\alpha) = \rho_{g_{\mathrm{IQD}^{+}}}(G), \qquad \alpha \in (0,1/2],$$

and

$$\mathrm{IQD}^{-}_{\alpha}(G) = G^{-1}(1-\alpha) - G_{+}^{-1}(\alpha) = \rho_{g_{\mathrm{IQD}^{-}}}(G), \qquad \alpha \in (0,1/2),$$

where $g_{\mathrm{IQD}^{+}}(t) = \mathbf{1}_{[\alpha,1-\alpha]}(t)$ and $g_{\mathrm{IQD}^{-}}(t) = \mathbf{1}_{(\alpha,1-\alpha)}(t)$ for $t \in [0,1]$. Definitions of $\mathrm{IQD}^{+}_{\alpha}$ and $\mathrm{IQD}^{-}_{\alpha}$ appear in Bellini et al. (2022) and Lauzier et al. (2023), respectively. The measures MMD and IQD are important in robust statistics and have applications in portfolio selection and risk management, wherein robustness to outliers is desirable (see Lauzier et al. (2023) for applications to risk sharing).

The GlueVaR of a distribution G was introduced in Belles-Sampera et al. (2013). For parameters α,β,h_1,h_2 with $\alpha,\beta \in [0,1]$, $\alpha \leqslant \beta$, $h_1 \in [0,1]$ and $h_2 \in [h_1,1]$, the associated distortion $g := g^{h_1,h_2}_{\beta,\alpha}$ is given by:

$$g^{h_1,h_2}_{\beta,\alpha}(t) = \begin{cases} \frac{h_1}{1-\beta}t, & 0 \leqslant t < 1-\beta, \\ h_1 + \frac{h_2-h_1}{\beta-\alpha}[t-(1-\beta)], & 1-\beta \leqslant t \leqslant 1-\alpha, \\ 1, & 1-\alpha < t \leqslant 1. \end{cases} \tag{4.6}$$

Notably, VaR_{α}, ES_{α} and the RVaR arise as special cases of this family of risk measures with the corresponding distortion functions $g^{0,0}_{\alpha,\alpha}$, $g^{1,1}_{\alpha,\alpha}$ and $g^{0,1}_{\beta,\alpha}$ with $\alpha < \beta$, respectively.

The RVaR introduced in Cont et al. (2010) as a family of robust risk measures is defined by:

$$\mathrm{RVaR}_{\alpha,\beta}(G) = \frac{1}{\beta-\alpha}\int_{\alpha}^{\beta} G^{-1}(t)\,\mathrm{d}t, \qquad 0<\alpha<\beta<1.$$

Under appropriate conditions on h_1 and h_2, the GlueVaR distortion can produce measures that lie between VaR and ES. In particular, if $\frac{h_1}{1-\beta} \geqslant \frac{h_2-h_1}{\beta-\alpha}$ then $\mathrm{GlueVaR}_{\beta,\alpha}^{h_1,h_2}$ can be expressed as a convex combination $w_1\mathrm{ES}_{\alpha}+w_2\mathrm{ES}_{\beta}+w_3\mathrm{VaR}_{\alpha}$ with non-negative weights w_1,w_2,w_3 summing to one (see Belles-Sampera et al. (2013) for further details).

Finally, for $0<\alpha_1<\alpha_2<1$, one may consider the discrepancy between ES and VaR as being defined by:

$$\rho_{g_{\alpha_1,\alpha_2}} = \mathrm{ES}_{\alpha_1} - \mathrm{VaR}_{\alpha_2},$$

where $g_{\alpha_1,\alpha_2}(t) = \frac{t}{1-\alpha_1}\wedge 1 - \mathbf{1}_{(1-\alpha_2,1]}(t)$. In practice, typical values for parameter α can be $\alpha_1 = 0.975$ and $\alpha_2 = 0.99$.

Below, we use V to denote a standard uniform random variable.

4.3 Bounds for distortion risk metrics under Wasserstein distance constraints

A widely used notion in mass transportation and distributionally robust optimisation is the Wasserstein metric (Esfahani and Kuhn (2018) and Blanchet and Murthy (2019)).The connection between optimal transport and portfolio optimisation arises naturally. In DRO, the decision-maker hedges against model misspecification by optimising over a neighbourhood of distributions around a reference model. The Wasserstein distance provides a geometrically meaningful notion of "neighbourhood" in which two distributions are close if the cost of transporting probability mass from one to the other is small. Unlike divergence-based ambiguity sets, the Wasserstein distance can compare distributions with differing supports, which is important in financial applications where the support of the true return distribution is unknown. In portfolio selection, the reference distribution is typically estimated from historical return data, and the Wasserstein ball captures the uncertainty inherent in this estimation. The radius $\sqrt{\varepsilon}$ quantifies the confidence of the decision-maker in the estimated return distribution.

For two random variables, X and Y, with distribution functions F and G, respectively, the one-dimensional Wasserstein distance of order 2 is defined by:

$$d_W(X,Y) = d_W(F,G) = d_W(F^{-1},G^{-1}) = \left(\int_0^1 |F^{-1}(x) - G^{-1}(x)|^2 \,\mathrm{d}x\right)^{1/2}.$$

In this section, we study the problem (4.4a) when the uncertainty set $\mathscr{F}$ is taken to be:

$$\mathscr{F} := \mathscr{M}_\varepsilon(\mu,\sigma) = \left\{G \in \mathscr{M}^2 : \int_{\mathbb{R}} x \,\mathrm{d}G = \mu,\ \int_{\mathbb{R}} x^2 \,\mathrm{d}G = \mu^2 + \sigma^2,\ d_W(G,F) \leqslant \sqrt{\varepsilon}\right\},$$

where $\mu \in \mathbb{R}$, $\sigma > 0$, $\varepsilon > 0$ and $F \in \mathscr{M}^2$. Here, F is the centre of a Wasserstein ball; we denote its mean and standard deviation by μ_F and $\sigma_F > 0$, respectively. Note that:

$$\mathscr{M}_\infty(\mu,\sigma) = \left\{G \in \mathscr{M}^2 : \int_{\mathbb{R}} x \,\mathrm{d}G = \mu,\ \int_{\mathbb{R}} x^2 \,\mathrm{d}G = \mu^2 + \sigma^2\right\}.$$

For $g \in \mathscr{H}$, let g^* and g_* be the concave and convex envelopes of g, i.e.,

$$g^* = \inf\{h \in \mathscr{H} : h \text{ is concave on } [0,1] \text{ and } h(x) \geqslant g(x),\ \text{for all } x \in [0,1]\},$$

and

$$g_* = \sup\{h \in \mathscr{H} : h \text{ is convex on } [0,1] \text{ and } h(x) \leqslant g(x),\ \text{for all } x \in [0,1]\}.$$

For any concave or convex value of $h \in \mathscr{H}$ we write $h'(t) := \partial_+ h(t)$ for the right derivative. Set

$$c_0 = Corr(F^{-1}(V), (g^*)'(1-V)),$$

with the convention $c_0 = 0$ when $(g^*)'$ is constant; note that $c_0 \geqslant 0$. Moreover, for $\lambda \geqslant 0$, define

$$g_\lambda(t) = g(t) + \lambda \int_{1-t}^1 F^{-1}(s) \,\mathrm{d}s, \qquad t \in [0,1],$$

and introduce

$$h_\lambda(t) = \mu + \sigma \frac{(g_\lambda^*)'(1-t) - a_\lambda}{b_\lambda}, \tag{4.7}$$

where $a_\lambda = \mathbb{E}\big((g_\lambda^*)'(V)\big) = g(1) + \lambda \mu_F$ and $b_\lambda = \sqrt{\mathrm{VaR}\big((g_\lambda^*)'(V)\big)}$. We denote the distribution function corresponding to h_λ by H_λ, so that $H_\lambda^{-1} = h_\lambda$. To guarantee that h_λ is well-defined, we impose the following standing assumption:

Assumption A: $\int_0^1 \big((g^*)'(t)\big)^2 \,\mathrm{d}t < \infty$ and $(g_\lambda^*)'$ is not constant for all values of $\lambda > 0$.

Under **Assumption A**, for any $G \in \mathscr{M}_\varepsilon(\mu,\sigma)$ and $g \in \mathscr{H}$ we have:

$$\rho_g(G) \leqslant \rho_{g^*}(G) = \int_0^1 (g^*)'(1-t) G^{-1}(t)\,\mathrm{d}t \leqslant \sigma \left(\int_0^1 \big((g^*)'(t)\big)^2 \,\mathrm{d}t \right)^{1/2} < \infty,$$

so **Assumption A** ensures $\rho_g(G) < \infty$ for every $G \in \mathscr{M}_\varepsilon(\mu,\sigma)$.

The next lemma, whose proof is technical and deferred to the Appendix, is central to our main results (see Theorem 4.2 below) and will be used in the proof of Corollary 4.5.

Lemma 4.1. *The map $\lambda \mapsto Corr(F^{-1}(V), (g_\lambda^*)'(1-V))$ is continuous on $[0,\infty)$ and*

$$\lim_{\lambda \to \infty} Corr(F^{-1}(V), (g_\lambda^*)'(1-V)) = 1.$$

We now present our first main theorem, which identifies the worst-case distribution and the worst-case value of distortion risk metrics over the Wasserstein uncertainty set $\mathscr{M}_\varepsilon(\mu,\sigma)$.

Theorem 4.2. *Suppose $g \in \mathscr{H}$ and $g = \hat{g}$.*

(i) If

$$(\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 < \varepsilon < (\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 + 2\sigma_F \sigma (1 - c_0),$$

then

$$\sup_{G \in \mathscr{M}_\varepsilon(\mu,\sigma)} \rho_g(G) = \rho_g(H_{\lambda_\varepsilon}),$$

where H_{λ_ε} is the unique worst-case distribution function determined by $d_W(F, H_{\lambda_\varepsilon}) = \sqrt{\varepsilon}$ for some values of $\lambda_\varepsilon > 0$.

(ii) If

$$\varepsilon \geqslant (\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 + 2\sigma_F \sigma (1 - c_0),$$

then, provided $(g^*)'$ is not constant, the conclusion of (i) holds with $\lambda_\varepsilon = 0$. If $(g^*)'$ is constant, then:

$$\sup_{G \in \mathscr{M}_\varepsilon(\mu,\sigma)} \rho_g(G) = g(1)\mu.$$

Theorem 4.2 applies to a broad class of distortion functions. It subsumes several results from the literature, including Bernard et al. (2024) (increasing, absolutely continuous g), Shao and Zhang (2023) (increasing g, $\varepsilon = \infty$), Li et al. (2018) (Range-Value-at-Risk, $\varepsilon = \infty$), Li (2018) (concave, increasing g, $\varepsilon = \infty$), and Pesenti et al. (2024) (general g, $\varepsilon = \infty$). The novelty of Theorem 4.2 is that it permits g to be non-monotone and discontinuous, thereby covering distortion functions such as those for GD, MMD, VaR^+ and IQD^+ as special cases. In particular, Theorem 4.2 exactly extends the results of Bernard et al. (2024) from increasing, absolutely continuous distortion functions to the more general class of upper semi-continuous distortions with finite variation. The projection method used in Bernard et al. (2024) requires absolute continuity and cannot be applied to our more general setting.

Our proof employs a variant of the concave-envelope technique. Unlike the constructions in Cai et al. (2025) and Pesenti et al. (2024), where the envelope is taken directly on the distortion function, we form the concave envelope of a linear combination of the distortion function and a functional of the reference distribution, F, with combination parameter λ, and then optimise over λ. The existence of an optimal λ relies on the continuity property established in Lemma 4.1, whose proof is highly non-trivial.

To characterise the worst-case distribution, it is important to obtain an explicit form for $(g_\lambda^*)'(1-t)$. When g is concave, one has:

$$(g_\lambda^*)'(1-t) = g'(1-t) + \lambda F^{-1}(t), \qquad t \in (0,1),$$

which covers the cases of GD and MMD. For non-concave g the computation of $(g_\lambda^*)'(1-t)$ is more involved, but Corollary 4.3 below provides explicit expressions for VaR_α^+, IQD_α^+ and the discrepancy $\rho_{g_{\alpha_1,\alpha_2}}$ between ES and VaR.

For $\alpha \in (0,1)$ and $\lambda \geqslant 0$ define

$$t_{\alpha,\lambda} = \inf\left\{ t \in [0,\alpha) : \frac{1+\lambda \int_{1-\alpha}^{1-t} F^{-1}(s)\,\mathrm{d}s}{\alpha - t} \geqslant \lambda F^{-1}(1-t) \right\}, \tag{4.8}$$

and

$$\hat{t}_{\alpha,\lambda} = \sup\left\{ t \in (1-\alpha,1] : \frac{\lambda \int_{1-t}^{\alpha} F^{-1}(s)\,\mathrm{d}s - 1}{t-1+\alpha} \leqslant \lambda F^{-1}(1-t) \right\}. \tag{4.9}$$

For $0 < \alpha_1 < \alpha_2 < 1$ and $\lambda \geqslant 0$ set

$$\begin{aligned} u_{\alpha_1,\alpha_2,\lambda} = \sup\Bigg\{ t \in (1-\alpha_2,1] : &\frac{\frac{(t-1+\alpha_2)\wedge(\alpha_2-\alpha_1)}{1-\alpha_1} + \lambda \int_{1-t}^{\alpha_2} F^{-1}(s)\,\mathrm{d}s - 1}{t-1+\alpha_2} \\ &\leqslant \frac{1}{1-\alpha_1}\mathbf{1}_{(0,1-\alpha_1)}(t) + \lambda F^{-1}(1-t) \Bigg\}. \end{aligned} \tag{4.10}$$

The following corollary collects explicit worst-case values and quantiles in these special cases.

Corollary 4.3. *Suppose*

$$(\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 < \varepsilon < (\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 + 2\sigma_F\sigma(1-c_0).$$

(i) For $\alpha \in (0,1)$,

$$\sup_{G \in \mathscr{M}_\varepsilon(\mu,\sigma)} \mathrm{VaR}_\alpha^+(G) = \mu + \sigma \frac{\frac{1+\lambda_\varepsilon \int_\alpha^{1-t_{1-\alpha,\lambda_\varepsilon}} F^{-1}(s)\,\mathrm{d}s}{1-\alpha-t_{1-\alpha,\lambda_\varepsilon}} - a_{\lambda_\varepsilon}}{b_{\lambda_\varepsilon}},$$

and the worst-case quantile equals h_{λ_ε} from (4.7) *with*

$$\begin{aligned} (g_\lambda^*)'(1-t) =& \lambda F^{-1}(t)\mathbf{1}_{(0,\alpha]\cup(1-t_{1-\alpha,\lambda},1)}(t) \\ &+ \frac{1+\lambda \int_\alpha^{1-t_{1-\alpha,\lambda}} F^{-1}(s)\,\mathrm{d}s}{1-\alpha-t_{1-\alpha,\lambda}}\mathbf{1}_{(\alpha,1-t_{1-\alpha,\lambda}]}(t), \end{aligned}$$

for $t \in (0,1)$, where λ_ε solves $d_W(F,H_\lambda) = \sqrt{\varepsilon}$.

(ii) For $\alpha \in (0,1/2)$,

$$\sup_{G \in \mathscr{M}_\varepsilon(\mu,\sigma)} \mathrm{IQD}_\alpha^+(G) = \left(\frac{1+\lambda_\varepsilon \int_{1-\alpha}^{1-t_{\alpha,\lambda_\varepsilon}} F^{-1}(s)\,\mathrm{d}s}{\alpha - t_{\alpha,\lambda_\varepsilon}} - \frac{\lambda_\varepsilon \int_{1-\hat{t}_{\alpha,\lambda_\varepsilon}}^{\alpha} F^{-1}(s)\,\mathrm{d}s - 1}{\hat{t}_{\alpha,\lambda_\varepsilon} - 1 + \alpha} \right) \frac{\sigma}{b_{\lambda_\varepsilon}},$$

and the worst-case quantile is h_{λ_ε} from (4.7) *with*

$$\begin{aligned}(g_\lambda^*)'(1-t) = & \frac{1+\lambda \int_{1-\alpha}^{1-t_{\alpha,\lambda}} F^{-1}(s)\,\mathrm{d}s}{\alpha - t_{\alpha,\lambda}} \mathbf{1}_{(1-\alpha,1-t_{\alpha,\lambda})}(t) \\ & + \frac{\lambda \int_{1-\hat{t}_{\alpha,\lambda}}^{\alpha} F^{-1}(s)\,\mathrm{d}s - 1}{\hat{t}_{\alpha,\lambda} - 1 + \alpha} \mathbf{1}_{(1-\hat{t}_{\alpha,\lambda},\alpha)}(t) \\ & + \lambda F^{-1}(t) \mathbf{1}_{(0,1-\hat{t}_{\alpha,\lambda}) \cup (\alpha,1-\alpha) \cup (1-t_{\alpha,\lambda},1)}(t),\end{aligned}$$

for $t \in (0,1)$, where λ_ε solves $d_W(F,H_\lambda) = \sqrt{\varepsilon}$.

(iii) For $0 < \alpha_1 < \alpha_2 < 1$,

$$\begin{aligned}& \sup_{G \in \mathscr{M}_\varepsilon(\mu,\sigma)} \rho_{g_{\alpha_1,\alpha_2}}(G) \\ & = \frac{\sigma \lambda_\varepsilon}{b_{\lambda_\varepsilon}(1-\alpha_1)} \left(\int_{\alpha_1}^{\alpha_1 \vee (1-u_{\alpha_1,\alpha_2,\lambda_\varepsilon})} F^{-1}(s)\,\mathrm{d}s + \int_{\alpha_2}^{1} F^{-1}(s)\,\mathrm{d}s \right) \\ & + \frac{\sigma}{b_{\lambda_\varepsilon}} \frac{1-\alpha_2 + (1-\alpha_1 - u_{\alpha_1,\alpha_2,\lambda_\varepsilon})_+}{(1-\alpha_1)^2} \\ & + \frac{\sigma c_{\alpha_1,\alpha_2,\lambda_\varepsilon}}{b_{\lambda_\varepsilon}} \left(\frac{(\alpha_2-\alpha_1) \wedge (u_{\alpha_1,\alpha_2,\lambda_\varepsilon} - 1 + \alpha_2)}{1-\alpha_1} - 1 \right),\end{aligned}$$

and the worst-case quantile is h_{λ_ε} from (4.7) *with*

$$\begin{aligned}(g_\lambda^*)'(1-t) = & \left(\frac{1}{1-\alpha_1} \mathbf{1}_{(\alpha_1,1)}(t) + \lambda F^{-1}(t) \right) \mathbf{1}_{(0,1-u_{\alpha_1,\alpha_2,\lambda}) \cup (\alpha_2,1)}(t) \\ & + c_{\alpha_1,\alpha_2,\lambda} \mathbf{1}_{(1-u_{\alpha_1,\alpha_2,\lambda},\alpha_2)}(t),\end{aligned}$$

for $t \in (0,1)$, where λ_ε solves $d_W(F,H_\lambda) = \sqrt{\varepsilon}$ and

$$c_{\alpha_1,\alpha_2,\lambda} = \frac{\frac{(u_{\alpha_1,\alpha_2,\lambda} - 1 + \alpha_2) \wedge (\alpha_2 - \alpha_1)}{1-\alpha_1} + \lambda \int_{1-u_{\alpha_1,\alpha_2,\lambda}}^{\alpha_2} F^{-1}(s)\,\mathrm{d}s - 1}{u_{\alpha_1,\alpha_2,\lambda} - 1 + \alpha_2}.$$

Part (i) of Corollary 4.3 reproduces Proposition 4.6 of Bernard et al. (2024), where the derivation is rather involved; here it follows readily from Theorem 4.2. The worst-case quantile for VaR^+ is linear in the reference quantile on the tail regions, which is not a constant in general, in contrast to the step-function (two-point) form arising without a Wasserstein constraint (Cantelli bound). The case with only a Wasserstein constraint is studied in Liu et al. (2022). Part (ii) of Corollary 4.3 is, to our knowledge, new in that it provides the worst-case value of a variability measure (statistical dispersion) for distributions with fixed mean and variance inside a Wasserstein ball. The corresponding worst-case quantile is linear for both the central and tail regions. Part (iii) gives the worst-case discrepancy between ES and VaR, which can be interpreted as the maximal additional capital requirement when replacing VaR by ES compared with the Basel framework in Basel Committee on Banking Supervision (2019).

Theorem 4.2 assumes upper semi-continuity of g, which excludes some distortion functions such as those for VaR, IQD^- and GlueVaR. Using (4.5), to obtain the best-case value of ρ_g is equivalent to finding the worst-case value of ρ_{-g}; however, if g is upper semi-continuous, then $-g$ is lower semi-continuous and may not be covered by Theorem 4.2 (examples include VaR^+, IQD^-+ and $\rho_{g_{\alpha_1,\alpha_2}}$). We, therefore, remove the upper semi-continuity restriction and treat general $g \in \mathscr{H}$ in the next result.

Theorem 4.4. *Suppose $g \in \mathscr{H}$.*

(i) If

$$(\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 < \varepsilon < (\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 + 2\sigma_F \sigma (1 - c_0)$$

and the map $\lambda \mapsto \rho_{\hat{g}}(H_\lambda)$ is continuous on $(0, \infty)$, then

$$\sup_{G \in \mathscr{M}_\varepsilon(\mu,\sigma)} \rho_g(G) = \sup_{G \in \mathscr{M}_\varepsilon(\mu,\sigma)} \rho_{\hat{g}}(G) = \rho_{\hat{g}}(H_{\lambda_\varepsilon}),$$

where λ_ε solves $d_W(F, H_\lambda) = \sqrt{\varepsilon}$.

(ii) If

$$\varepsilon \geqslant (\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 + 2\sigma_F \sigma (1 - c_0)$$

and $(g^*)'$ is not constant, then

$$\sup_{G\in\mathscr{M}_\varepsilon(\mu,\sigma)} \rho_g(G) = \sup_{G\in\mathscr{M}_\varepsilon(\mu,\sigma)} \rho_{\hat{g}}(G) = \rho_{\hat{g}}(H_0).$$

If $(g^*)'$ is constant, then $\sup_{G\in\mathscr{M}_\varepsilon(\mu,\sigma)} \rho_g(G) = g(1)\mu$.

In part (i) of Theorem 4.4, we impose the additional assumption that $\rho_{\hat{g}}(H_\lambda)$ is continuous in λ over $(0,\infty)$. This assumption arises from the Wasserstein constraint and cannot be removed from our technical proof of Theorem 4.4 (see Section 4.7.1 in the Appendix). The assumption is satisfied for the distortion risk metrics considered in Corollary 4.5 below. The arguments used in the proof of Lemma 4.1 play a key role in verifying this continuity property.

For $0<\alpha<\beta<1$, $0<h_1<h_2<1$ and $\lambda \geqslant 0$ define

$$u^{h_1,h_2}_{\alpha,\beta,\lambda} = \inf\Bigg\{ t\in[0,1-\alpha) : \frac{1-g^{h_1,h_2}_{\alpha,\beta}(t)+\lambda\int_\alpha^{1-t} F^{-1}(s)\,\mathrm{d}s}{1-\alpha-t} \geqslant \frac{h_1}{1-\beta}\mathbf{1}_{(0,1-\beta)}(t) + \frac{h_2-h_1}{\beta-\alpha}\mathbf{1}_{(1-\beta,1-\alpha)}(t) + \lambda F^{-1}(1-t) \Bigg\}. \tag{4.11}$$

Corollary 4.5. *Suppose*

$$(\mu_F-\mu)^2+(\sigma_F-\sigma)^2 < \varepsilon < (\mu_F-\mu)^2+(\sigma_F-\sigma)^2+2\sigma_F\sigma(1-c_0).$$

(i) For $\alpha\in(0,1)$, the map $\lambda\mapsto \mathrm{VaR}^+_\alpha(H_\lambda)$ is continuous on $(0,\infty)$, and

$$\sup_{G\in\mathscr{M}_\varepsilon(\mu,\sigma)} \mathrm{VaR}_\alpha(G) = \sup_{G\in\mathscr{M}_\varepsilon(\mu,\sigma)} \mathrm{VaR}^+_\alpha(G) = \mu+\sigma\frac{\frac{1+\lambda_\varepsilon\int_\alpha^{1-t_{1-\alpha,\lambda_\varepsilon}} F^{-1}(s)\,\mathrm{d}s}{1-\alpha-t_{1-\alpha,\lambda_\varepsilon}} - a_{\lambda_\varepsilon}}{b_{\lambda_\varepsilon}},$$

where λ_ε solves $d_W(F,H_\lambda)=\sqrt{\varepsilon}$.

(ii) For $\alpha\in(0,1/2)$, the map $\lambda\mapsto \mathrm{IQD}^+_\alpha(H_\lambda)$ is continuous on $(0,\infty)$ and

$$\sup_{G\in\mathscr{M}_\varepsilon(\mu,\sigma)} \mathrm{IQD}^-_\alpha(G) = \sup_{G\in\mathscr{M}_\varepsilon(\mu,\sigma)} \mathrm{IQD}^+_\alpha(G) = \left(\frac{1+\lambda_\varepsilon\int_{1-\alpha}^{1-t_{\alpha,\lambda_\varepsilon}} F^{-1}(s)\,\mathrm{d}s}{\alpha-t_{\alpha,\lambda_\varepsilon}} - \frac{\lambda_\varepsilon\int_{1-\hat{t}_{\alpha,\lambda_\varepsilon}}^{\alpha} F^{-1}(s)\,\mathrm{d}s-1}{\hat{t}_{\alpha,\lambda_\varepsilon}-1+\alpha}\right)\frac{\sigma}{b_{\lambda_\varepsilon}},$$

with λ_ε solving $d_W(F,H_\lambda)=\sqrt{\varepsilon}$.

(iii) For $0<\alpha<\beta<1$ and $0<h_1<h_2<1$ satisfying $\frac{h_1}{1-\beta}\geqslant\frac{h_2-h_1}{\beta-\alpha}$, the map $\lambda\mapsto\rho_{\hat{g}^{h_1,h_2}_{\beta,\alpha}}(H_\lambda)$ is continuous on $(0,\infty)$ and

$$
\begin{aligned}
&\sup_{G\in\mathscr{M}_\varepsilon(\mu,\sigma)}\mathrm{GlueVaR}^{h_1,h_2}_{\beta,\alpha}(G)=\sup_{G\in\mathscr{M}_\varepsilon(\mu,\sigma)}\rho_{\hat{g}^{h_1,h_2}_{\beta,\alpha}}(G)\\
=&\mu-\frac{\sigma(1+\lambda_\varepsilon\mu_F)}{b_{\lambda_\varepsilon}}+\frac{\sigma(1-h_2)}{b_{\lambda_\varepsilon}}c^{h_1,h_2}_{\alpha,\beta,\lambda_\varepsilon}\\
&+\frac{\sigma h_1}{b_{\lambda_\varepsilon}(1-\beta)}\Bigg(c^{h_1,h_2}_{\alpha,\beta,\lambda_\varepsilon}(1-u^{h_1,h_2}_{\alpha,\beta,\lambda_\varepsilon}-\beta)_++\frac{h_1((1-\beta)\wedge u^{h_1,h_2}_{\alpha,\beta,\lambda_\varepsilon})}{1-\beta}\\
&\qquad\qquad+\lambda_\varepsilon\int_{\beta\vee(1-u^{h_1,h_2}_{\alpha,\beta,\lambda_\varepsilon})}^{1}F^{-1}(s)\,\mathrm{d}s\Bigg)\\
&+\frac{\sigma(h_2-h_1)}{b_{\lambda_\varepsilon}(\beta-\alpha)}\Bigg(c^{h_1,h_2}_{\alpha,\beta,\lambda_\varepsilon}(\beta\wedge(1-u^{h_1,h_2}_{\alpha,\beta,\lambda_\varepsilon})-\alpha)+\frac{(h_2-h_1)(\beta-1+u^{h_1,h_2}_{\alpha,\beta,\lambda_\varepsilon})_+}{\beta-\alpha}\\
&\qquad\qquad+\lambda_\varepsilon\int_{\beta\wedge(1-u^{h_1,h_2}_{\alpha,\beta,\lambda_\varepsilon})}^{\beta}F^{-1}(s)\,\mathrm{d}s\Bigg),
\end{aligned}
$$

and the worst-case quantile for this supremum is h_{λ_ε} from (4.7) *with*

$$
\begin{aligned}
(g^*_\lambda)'(1-t)=&\,c^{h_1,h_2}_{\alpha,\beta,\lambda}\mathbf{1}_{(\alpha,1-u^{h_1,h_2}_{\alpha,\beta,\lambda})}(t)+\frac{h_1}{1-\beta}\mathbf{1}_{(\beta\vee(1-u^{h_1,h_2}_{\alpha,\beta,\lambda}),1)}(t)\\
&+\frac{h_2-h_1}{\beta-\alpha}\mathbf{1}_{(\beta\wedge(1-u^{h_1,h_2}_{\alpha,\beta,\lambda}),\beta)}(t)+\lambda F^{-1}(t)\mathbf{1}_{(0,\alpha)\cup(1-u^{h_1,h_2}_{\alpha,\beta,\lambda},1)},
\end{aligned}
$$

for $t\in(0,1)$, where λ_ε solves $d_W(F,H_\lambda)=\sqrt{\varepsilon}$ and

$$
c^{h_1,h_2}_{\alpha,\beta,\lambda}=\frac{1-g^{h_1,h_2}_{\alpha,\beta}(u^{h_1,h_2}_{\alpha,\beta,\lambda})+\lambda\int_\alpha^{1-u^{h_1,h_2}_{\alpha,\beta,\lambda}}F^{-1}(s)\,\mathrm{d}s}{1-\alpha-u^{h_1,h_2}_{\alpha,\beta,\lambda}}.
$$

When $\frac{h_1}{1-\beta}\geqslant\frac{h_2-h_1}{\beta-\alpha}$ one can write $\mathrm{GlueVaR}^{h_1,h_2}_{\beta,\alpha}=w_1\mathrm{ES}_\alpha+w_2\mathrm{ES}_\beta+w_3\mathrm{VaR}_\alpha$ for non-negative weights w_1,w_2,w_3 with $w_1+w_2+w_3=1$. By selecting suitable parameters, GlueVaR lies between VaR and ES, representing a risk attitude that is more conservative than VaR, yet less conservative than ES. Thus, part (iii) of Corollary 4.5 yields the worst-case value for convex combinations of VaR and ES. Using (4.5) and Theorem 4.4, one can also obtain best-case expressions for VaR^+, IQD^+ and $\rho_{g_{\alpha_1,\alpha_2}}$ analogous to Corollary 4.5; the best-case value for GlueVaR follows from (4.5) and Theorem 4.2.

We remark, in general, that Theorem 4.4 does not assert the existence of a worst-case distribution, because any such existence may fail when g is not upper semi-continuous. Indeed, suppose a worst-case distribution G_0 exists for $\sup_{G\in\mathscr{M}_\varepsilon(\mu,\sigma)}\rho_g(G)$. Then, by Theorem 4.4 G_0 must also be a worst-case distribution for $\sup_{G\in\mathscr{M}_\varepsilon(\mu,\sigma)}\rho_{\hat{g}}(G)$, and uniqueness from Theorem 4.2 would force $G_0 = H_{\lambda_\varepsilon}$. However, direct computation using the expressions in Corollaries 4.3 and 4.5 yields contradictions, e.g.,

$$\mathrm{VaR}_\alpha(H_{\lambda_\varepsilon}) = \mu + \sigma\frac{\lambda F^{-1}(\alpha) - a_{\lambda_\varepsilon}}{b_{\lambda_\varepsilon}}, \qquad \mathrm{IQD}^-_\alpha(H_{\lambda_\varepsilon}) = \sigma\frac{\lambda F^{-1}(1-\alpha) - \lambda F_+^{-1}(\alpha)}{b_{\lambda_\varepsilon}},$$

and

$$\mathrm{GlueVaR}^{h_1,h_2}_{\beta,\alpha}(H_{\lambda_\varepsilon}) = \rho_{\hat{g}^{h_1,h_2}_{\beta,\alpha}}(H_{\lambda_\varepsilon}) + \frac{\sigma(1-h_2)}{b_{\lambda_\varepsilon}}(\lambda F^{-1}(\alpha) - c^{h_1,h_2}_{\alpha,\beta,\lambda_\varepsilon}),$$

which demonstrate that worst-case distributions need not exist for VaR_α, IQD^-_α and $\mathrm{GlueVaR}^{h_1,h_2}_{\beta,\alpha}$. Similar phenomena for VaR with $\varepsilon = \infty$ are discussed in Corollary 4.1 of Bernard et al. (2024) and Example 17 of Pesenti et al. (2024).

Although obtaining a closed form for λ_ε is challenging for a general $g \in \mathscr{H}$, an explicit formula is available in the concave case.

Proposition 4.6. *If $g \in \mathscr{H}$ is concave and $\int_0^1 (g'(t))^2\,\mathrm{d}t < \infty$, then for*

$$(\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 < \varepsilon < (\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 + 2\sigma_F\sigma(1 - c_0),$$

the solution of $d_W(F, H_\lambda) = \sqrt{\varepsilon}$ is:

$$\lambda_\varepsilon = \frac{-C_{g,F} + \sqrt{C^2_{g,F} - \sigma^2_F\frac{V_g C^2_{\varepsilon,F} - \sigma^2 C^2_{g,F}}{C^2_{\varepsilon,F} - \sigma^2\sigma^2_F}}}{\sigma^2_F},$$

where

$$C_{\varepsilon,F} = \frac{\mu_F^2 + \sigma_F^2 + \mu^2 + \sigma^2 - 2\mu\mu_F - \varepsilon}{2} \geqslant 0, \qquad V_g = \mathrm{VaR}(g'(V)),$$

and

$$C_{g,F} = Cov(F^{-1}(V), g'(1-V)) \geqslant 0.$$

4.4 Bounds for unimodal distribution functions with Wasserstein constraint

In this section, we suppose $g \in \mathscr{H}$ is such that the associated distortion risk metric admits the representation

$$\rho_g(G) = \int_0^1 \gamma(u) G^{-1}(u)\, \mathrm{d}u, \tag{4.12}$$

where the *weight function* is given by $\gamma(u) = \partial_- g(x)\big|_{x=1-u}$ for $0 < u < 1$, and ∂_- denotes the left derivative. We also assume $\int_0^1 |\gamma(u)|^2\, \mathrm{d}u < \infty$. Equivalently, the distortion risk metric may be written in terms of the quantile function, and we will sometimes write $\rho_g(G^{-1})$ instead of $\rho_g(G)$.

Definition 4.7. *A cdf $G \in \mathscr{M}^2$ is called unimodal if G is convex–concave, i.e. there exists $x_m \in \mathbb{R}$ (the* mode*) such that G is convex on $(-\infty, x_m)$ and concave on $(x_m, +\infty)$.*

We say that a (left) quantile function G^{-1} is *concave–convex* if there exists an inflection point $\xi \in [0,1]$ such that G^{-1} is concave on $(0,\xi)$ and convex on $(\xi,1)$. In Bernard et al. (2023) characterises the unimodality of a cumulative distribution function in terms of its quantile function: a cdf G is unimodal if and only if G^{-1} is continuous on $(0,1)$ and is one of: concave, convex, or concave–convex.

Throughout, we refer to quantile functions whose cdfs are unimodal as *unimodal quantile functions*. Let us define the sets:

$$\mathscr{F}_U = \{G \in \mathscr{M}^2 :\ G \text{ is unimodal}\}, \qquad \mathscr{F}_{U,\xi} = \{G \in \mathscr{F}_U :\ \text{the inflection point is } \xi\},$$

and

$$\mathscr{F}_{U,\xi}(\mu,\sigma) = \left\{G \in \mathscr{F}_{U,\xi} : \int_{\mathbb{R}} x\, \mathrm{d}G = \mu,\ \int_{\mathbb{R}} x^2\, \mathrm{d}G = \mu^2 + \sigma^2\right\},$$

$$\mathscr{F}_{U,\xi}(\mu,\sigma,\varepsilon) = \left\{G \in \mathscr{F}_{U,\xi} : \int_{\mathbb{R}} x\, \mathrm{d}G = \mu,\ \int_{\mathbb{R}} x^2\, \mathrm{d}G = \mu^2 + \sigma^2,\ d_W(F,G) \leqslant \sqrt{\varepsilon}\right\},$$

where $F \in \mathscr{M}^2$, $\mu \in \mathbb{R}$, $\sigma > 0$ and $\varepsilon > 0$. For inflection points varying in an interval $[\xi_1, \xi_2]$

with $0 \leqslant \xi_1 < \xi_2 \leqslant 1$ we write $\mathscr{F}_{U,[\xi_1,\xi_2]}$, $\mathscr{F}_{U,[\xi_1,\xi_2]}(\mu,\sigma)$ and $\mathscr{F}_{U,[\xi_1,\xi_2]}(\mu,\sigma,\varepsilon)$, noting that $\mathscr{F}_{U,[\xi_1,\xi_2]} = \bigcup_{\xi\in[\xi_1,\xi_2]} \mathscr{F}_{U,\xi}$ and similarly for the sets with moment or Wasserstein constraints.

In this section, we investigate upper bounds for distortion risk metrics when the uncertainty set consists of unimodal distributions with prescribed mean and variance, with either a fixed inflection point or an inflection point known only to lie on an interval. Concretely, we study the optimisation problems:

$$\sup_{G\in\mathscr{F}_{U,\xi}(\mu,\sigma)} \rho_g(G) \qquad (4.13a) \qquad\qquad \sup_{G\in\mathscr{F}_{U,\xi}(\mu,\sigma,\varepsilon)} \rho_g(G), \qquad (4.13b)$$

and,

$$\sup_{G\in\mathscr{F}_{U,[\xi_1,\xi_2]}(\mu,\sigma)} \rho_g(G) \qquad (4.14a) \qquad\qquad \sup_{G\in\mathscr{F}_{U,[\xi_1,\xi_2]}(\mu,\sigma,\varepsilon)} \rho_g(G). \qquad (4.14b)$$

4.4.1 Fixed inflection point

Recall that $\mathscr{F}_{U,\xi}^{-1}$ denotes the collection of quantile functions associated with $\mathscr{F}_{U,\xi}$. The set $\mathscr{F}_{U,\xi}^{-1}$ is a closed convex cone, hence the L_2-projection of any function on $(0,1)$ onto this set exists and is unique (see Theorem 2.1 of Brunk (1965)). Let $\gamma_\xi^\uparrow$ denote the L_2-projection of γ onto $\mathscr{F}_{U,\xi}^{-1}$, and set $\hat{a}_\xi = \mathbb{E}(\gamma_\xi^\uparrow(V))$ and $\hat{b}_\xi = \sqrt{\mathrm{VaR}(\gamma_\xi^\uparrow(V))}$ for $V \sim U(0,1)$. By Corollary 2.3 of Brunk (1965) we have $\hat{a}_\xi = \int_0^1 \gamma_\xi^\uparrow(u)\,\mathrm{d}u = g(1)$ and $\hat{b}_\xi = \sqrt{\int_0^1 \left(\gamma_\xi^\uparrow(u) - g(1)\right)^2 \mathrm{d}u}$.

Proposition 4.8 (Bounds for unimodal distribution functions with a given inflection point)**.** *Suppose $\gamma_\xi^\uparrow$ is not constant. Then*

$$\sup_{G\in\mathscr{F}_{U,\xi}(\mu,\sigma)} \rho_g(G) = \mu g(1) + \sigma\sqrt{\int_0^1 \left(\gamma_\xi^\uparrow(u) - g(1)\right)^2 \mathrm{d}u},$$

and the unique worst-case quantile is

$$h_\xi^\uparrow(u) := \mu + \sigma\frac{\gamma_\xi^\uparrow(u) - \hat{a}_\xi}{\hat{b}_\xi}.$$

If unimodality is not imposed, then the corresponding uncertainty set is $\mathscr{M}_\infty(\mu,\sigma)$. Denote by $\gamma^\uparrow$ the projection of γ on $\mathscr{M}_\infty^{-1}(\mu,\sigma)$. Then, Corollary 3.9 of Bernard et al. (2024) and

Theorem 5 together with Remark 2 of Pesenti et al. (2024) yield:

$$\sup_{G \in \mathscr{M}_\infty(\mu,\sigma)} \rho_g(G) = \mu g(1) + \sigma \sqrt{\int_0^1 \left(\gamma^{\uparrow}(u) - g(1)\right)^2 \mathrm{d}u},$$

provided $\gamma^{\uparrow}$ is not constant, and the worst-case quantile equals $\mu + \frac{\sigma\left(\gamma^{\uparrow} - g(1)\right)}{\sqrt{\int_0^1 \left(\gamma^{\uparrow}(u) - g(1)\right)^2 \mathrm{d}u}}$.

By Theorem 2.8 of Brunk (1965), we have the Pythagorean-type inequality

$$\int_0^1 \left(\gamma^{\uparrow}(u) - g(1)\right)^2 \mathrm{d}u \geqslant \int_0^1 \left(\gamma_\xi^{\uparrow}(u) - g(1)\right)^2 \mathrm{d}u + \int_0^1 \left(\gamma^{\uparrow}(u) - \gamma_\xi^{\uparrow}(u)\right)^2 \mathrm{d}u,$$

so if $\gamma^{\uparrow} \notin \mathscr{F}_{U,\xi}^{-1}(\mu,\sigma)$ then enforcing unimodality with a fixed inflection point can strictly reduce the worst-case value.

Furthermore, for functionals such as ES and RVaR, the worst-case distributions over the unrestricted set $\mathscr{M}_\infty(\mu,\sigma)$ are two-point distributions, which are often undesirable in applications (see Bernard et al. (2024) and Pesenti et al. (2024)). Imposing unimodality typically yields worst-case distributions that are not discrete.

In Li et al. (2018), the RVaR, with $\gamma(u) = \frac{1}{\beta-\alpha}\mathbf{1}_{(\alpha,\beta)}(u)$ for $0 \leqslant \alpha < \beta \leqslant 1$, was considered under mean, variance, and unimodality constraints without fixing the inflection point. Proposition 4.8 is valid for all distortion risk metrics with absolutely continuous distortion functions g, and it is more precise when the inflection point is fixed. By maximising the worst-case values over different inflection points, we can immediately derive the results for unknown inflection points (see Section 4.4.2). Optimal solutions to the problem (4.13a) are thus obtained once the projection $\gamma_\xi^{\uparrow}$ of the function γ on the set $F_{U,\xi}^{-1}$ is established.

Whilst computing such a projection, L_2-projection $\gamma_\xi^{\uparrow}$, in closed form is, in general, very difficult. It can be numerically obtained to any desired degree of precision. The next result shows that when γ is a step function, its projection on $F_{U,\xi}^{-1}$ is piecewise linear.

Proposition 4.9. *Let γ be a step function with n steps, i.e. $\gamma(u) = \sum_{i=1}^n y_i \mathbf{1}_{(x_{i-1},x_i)}(u)$ with $0 = x_0 < x_1 < \cdots < x_n = 1$ and $y_i \in \mathbb{R}$. Then $\gamma_\xi^{\uparrow}$ is piecewise linear with at most $2n+2$ pieces.*

More precisely, if $\xi = x_{i_0}$ for some $i_0 \in \{0,1,\dots,n\}$ then

$$\frac{\partial \gamma_\xi^\uparrow(u)}{\partial u} = \sum_{i=1}^{n}\Big(c_i^-\mathbf{1}_{(x_{i-1},a_i)}(u) + c_i^+\mathbf{1}_{(a_i,x_i)}(u)\Big), \tag{4.15}$$

and

$$\gamma_\xi^\uparrow(0) = g(1) - \sum_{i=1}^{n}\big(c_i^-(a_i - x_{i-1}) + c_i^+(x_i - a_i)\big), \tag{4.16}$$

with parameters constrained to the set

$$\begin{aligned}\mathscr{D}_n = \Big\{(\mathbf{a},\mathbf{c}) :& a_i \in [x_{i-1},x_i],\ i = 1,\dots,n,\ c_1^- \geqslant c_1^+ \geqslant \dots \geqslant c_{i_0}^- \geqslant c_{i_0}^+ \geqslant 0,\\ & 0 \leqslant c_{i_0+1}^- \leqslant c_{i_0+1}^+ \leqslant \dots \leqslant c_n^- \leqslant c_n^+\Big\},\end{aligned}$$

where $\mathbf{a} = (a_1,\dots,a_n)$ and $\mathbf{c} = (c_1^-,c_1^+,\dots,c_n^-,c_n^+)$. The optimal parameters $(\mathbf{a}^*,\mathbf{c}^*)$ solve

$$\begin{aligned}\arg\min_{(\mathbf{a},\mathbf{c})\in\mathscr{D}_n}\sum_{i=1}^{n}\Big\{&(a_i - x_{i-1})\big[(e_i^+)^2 + e_i^+e_i^- + (e_i^-)^2\big]\\ &+ (x_i - a_i)\big[(e_i^+ + c_i^+(x_i - a_i))^2 + (e_i^+ + c_i^+(x_i - a_i))e_i^+ + (e_i^+)^2\big]\Big\},\end{aligned} \tag{4.17}$$

where $e_i^- = g(1) - \sum_{j=i}^{n}\big(c_j^-(a_j - x_{j-1}) + c_j^+(x_j - a_j)\big) - y_i$ and $e_i^+ = e_i^- + c_i^-(a_i - x_{i-1})$ for $i = 1,\dots,n$. If $\xi \in (x_{i_0-1},x_{i_0})$ for some $i_0 \in \{1,\dots,n\}$, rewrite γ as a step function with a break point at ξ and apply the same construction.

Thus, for a step-function γ with n steps, finding the projection reduces to solving the finite-dimensional optimisation (4.17) in $3n$ (or $3n+3$) parameters, depending on the position of ξ. In that parametrisation, one obtains the representation:

$$\begin{aligned}\gamma_\xi^\uparrow(u) &= e_i^- + y_i + c_i^-(u - x_{i-1}), \qquad u \in [x_{i-1},a_i),\\ \gamma_\xi^\uparrow(u) &= e_i^+ + y_i + c_i^+(u - a_i), \qquad u \in [a_i,x_i), \quad i = 1,\dots,n.\end{aligned}$$

A general γ can be approximated arbitrarily well by step functions γ_n. Let $\gamma_{\xi,n}^\uparrow$ be the projection of γ_n on $\mathscr{F}_{U,\xi}^{-1}$, and set $h_{\xi,n}^\uparrow = \mu g(1) + \sigma\frac{\gamma_{\xi,n}^\uparrow - \hat{a}_{\xi,n}}{\hat{b}_{\xi,n}}$, where $\hat{a}_{\xi,n} = \mathbb{E}(\gamma_{\xi,n}^\uparrow(V))$ and

$\hat{b}_{\xi,n} = \sqrt{\mathrm{VaR}(\gamma^{\uparrow}_{\xi,n}(V))}$. Denote by $\|\cdot\|_2$ the L_2-norm on $(0,1)$.

Proposition 4.10. *If $\gamma^{\uparrow}_{\xi}$ and $\gamma^{\uparrow}_{\xi,n}$ are not constant, then*

$$\|\gamma^{\uparrow}_{\xi,n} - \gamma^{\uparrow}_{\xi}\|_2 \leqslant \|\gamma_n - \gamma\|_2, \qquad \|h^{\uparrow}_{\xi,n} - h^{\uparrow}_{\xi}\|_2 \leqslant \Big(2 + \frac{2\|\gamma\|_2 + \|\gamma_n\|_2}{\hat{b}_{\xi}}\Big)\frac{\sigma}{\hat{b}_{\xi}}\|\gamma_n - \gamma\|_2,$$

and

$$|\rho_g(h^{\uparrow}_{\xi}) - \rho_g(h^{\uparrow}_{\xi,n})| \leqslant \frac{\sigma(2\|\gamma\|_2 + \|\gamma_n\|_2)\|\gamma_n - \gamma\|_2}{\hat{b}_{\xi}}.$$

These bounds show that approximation errors are controlled by the L_2-distance between γ and its stepwise approximation γ_n, validating the practical approach of approximating the projection numerically.

We give a few illustrative computations.

Example 4.11. (i) For the Gini deviation (GD), we have $\gamma(u) = 1 - 2u$ for $u \in (0,1)$. Hence $\gamma^{\uparrow}_{\xi}(u) = 1 - 2u$ for all $u \in (0,1)$.

(ii) For the mean–median deviation (MMD), we have $g(t) = t \wedge (1-t)$ on $[0,1]$. Therefore, $\gamma(u) = -\mathbf{1}_{(0,1/2)}(u) + \mathbf{1}_{(1/2,1)}(u)$. If $\xi = 1/2$ then one obtains $\gamma^{\uparrow}_{\xi}(u) = 3(u - 1/2)$ for $u \in (0,1)$.

Next, we look into the Wasserstein-constrained set $\mathscr{F}_{U,\xi}(\mu,\sigma,\varepsilon)$.

For $\lambda > 0$, we set $k_\lambda(u) = \gamma(u) + \lambda F^{-1}(u)$ and let $k^{\uparrow}_{\lambda,\xi}$ denote the L_2-projection of k_λ onto $\mathscr{F}^{-1}_{U,\xi}$. Define for $\lambda > 0$,

$$h_{\lambda,\xi} = \mu + \frac{k^{\uparrow}_{\lambda,\xi} - a_{\lambda,\xi}}{b_{\lambda,\xi}}\sigma, \tag{4.18}$$

where, $a_{\lambda,\xi} = \mathbb{E}(k^{\uparrow}_{\lambda,\xi}(V)) = g(1) + \lambda\mu_F$ and $b_{\lambda,\xi} = \sqrt{\mathrm{VaR}(k^{\uparrow}_{\lambda,\xi}(V))}$. Accordingly, c_1 is written as $c_1 = Corr(F^{-1}(V), \gamma^{\uparrow}_{\xi}(V))$.

We first identify the range of ε for which $\mathscr{F}_{U,\xi}(\mu,\sigma,\varepsilon)$ is non-empty. Let $F^{-1,\uparrow}_{\xi}$ be the L_2-projection of F^{-1} on $\mathscr{F}^{-1}_{U,\xi}$ and set $\hat{c}_0 = Corr(F^{-1}(V), F^{-1,\uparrow}_{\xi}(V))$ (when $F^{-1,\uparrow}_{\xi}$ is not constant). Clearly $\mathbb{E}(F^{-1,\uparrow}_{\xi}(V)) = \mu_F$.

Lemma 4.12. *Suppose $F_\xi^{-1,\uparrow}$ is not constant. If*

$$\varepsilon < (\mu_F-\mu)^2+(\sigma_F-\sigma)^2+2\sigma_F\sigma(1-\hat{c}_0),$$

then $\mathscr{F}_{U,\xi}^{-1}(\mu,\sigma,\varepsilon)=\varnothing$. If

$$\varepsilon = (\mu_F-\mu)^2+(\sigma_F-\sigma)^2+2\sigma_F\sigma(1-\hat{c}_0),$$

then $\mathscr{F}_{U,\xi}^{-1}(\mu,\sigma)=\left\{\dfrac{F_\xi^{-1,\uparrow}-\mu_F}{\sigma_F^\uparrow}\sigma+\mu\right\}$, where $\sigma_F^\uparrow=\sqrt{\mathrm{VaR}(F_\xi^{-1,\uparrow}(V))}$.

Henceforth we restrict attention to the regime $\varepsilon>(\mu_F-\mu)^2+(\sigma_F-\sigma)^2+2\sigma_F\sigma(1-\hat{c}_0)$, for which $\mathscr{F}_{U,\xi}$ contains infinitely many distributions.

The following continuity result for the correlation between $F^{-1}(V)$ and the projected functions $k_{\lambda,\xi}^\uparrow(V)$ is a key technical ingredient.

Lemma 4.13. *Suppose $F^{-1}\in\mathscr{M}^2$, $F_\xi^{-1,\uparrow}$ is not constant, and $k_{\lambda,\xi}^\uparrow$ is not constant for all values of $\lambda>0$. Then the map $\lambda\mapsto Corr(F^{-1}(V),k_{\lambda,\xi}^\uparrow(V))$ is continuous on $[0,\infty)$ and*

$$\lim_{\lambda\to\infty} Corr(F^{-1}(V),k_{\lambda,\xi}^\uparrow(V))=Corr(F^{-1}(V),F_\xi^{-1,\uparrow}(V)).$$

Note that $\hat{c}_0$ equals 1 when $F^{-1}\in\mathscr{F}_{U,\xi}^{-1}$, while if $F^{-1}\notin\mathscr{F}_{U,\xi}^{-1}$ one has $\hat{c}_0<1$.

We can now state the main result for the unimodal class with a fixed inflection point.

Theorem 4.14. *Assume $F^{-1}\in\mathscr{M}^2$, $F_\xi^{-1,\uparrow}$ is not constant, and $k_{\lambda,\xi}^\uparrow$ is not constant for all values of $\lambda>0$.*

(i) If

$$(\mu_F-\mu)^2+(\sigma_F-\sigma)^2+2\sigma_F\sigma(1-\hat{c}_0)<\varepsilon<(\mu_F-\mu)^2+(\sigma_F-\sigma)^2+2\sigma_F\sigma(1-c_1),$$

then $h_{\lambda_\varepsilon,\xi}(u)$ defined in (4.18) *is the unique worst-case quantile for problem* (4.13b)*, and*

$$\sup_{G\in\mathscr{F}_{U,\xi}(\mu,\sigma,\varepsilon)}\rho_g(G)=\mu g(1)+\frac{\sigma}{b_{\lambda_\varepsilon,\xi}}\Big(\int_0^1 k_{\lambda_\varepsilon,\xi}^\uparrow(u)\gamma(u)\,\mathrm{d}u-g(1)(g(1)+\lambda_\varepsilon\mu_F)\Big),$$

where $\lambda_\varepsilon > 0$ solves $d_W(h_{\lambda_\varepsilon,\xi}, F^{-1}) = \sqrt{\varepsilon}$.

(ii) If

$$\varepsilon \geqslant (\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 + 2\sigma_F\sigma(1 - c_1),$$

and $\gamma_\xi^\uparrow$ is not constant, then the unique worst-case quantile for (4.13b) *is $h_\xi^\uparrow(u) = \mu + \sigma\frac{\gamma_\xi^\uparrow(u) - \hat{a}_\xi}{\hat{b}_\xi}$. If $\gamma_\xi^\uparrow$ is constant, then $\sup_{G\in\mathscr{F}_{U,\xi}(\mu,\sigma)} \rho_g(G) = g(1)\mu$.*

When g is absolutely continuous, Theorem 4.14 refines Theorem 4.2 by incorporating unimodality with a fixed inflection point into the uncertainty set, and imposing this information typically reduces the worst-case value. The practical computation of the bounds relies on numerically obtaining the projections, for which Propositions 4.9 and 4.10 are crucial.

4.4.2 Unknown inflection points

Diverging from $\mathscr{F}_{U,\xi}^{-1}$, the set $\mathscr{F}_{U,[\xi_1,\xi_2]}^{-1}$ with $0 \leqslant \xi_1 < \xi_2 \leqslant 1$ is not convex. Hence, Theorem 2.1 of Brunk (1965) cannot guarantee the existence and uniqueness of the projection. To compute the worst-case value of the distortion risk metrics, one can apply Proposition 4.8, Theorem 4.14, and the following relations, which reduce the optimisation over an unknown inflection point to a collection of fixed-inflection-point problems:

$$\sup_{G\in\mathscr{F}_{U,[\xi_1,\xi_2]}(\mu,\sigma)} \rho_g(G) = \sup_{\xi\in[\xi_1,\xi_2]} \sup_{G\in\mathscr{F}_{U,\xi}(\mu,\sigma)} \rho_g(G), \tag{4.19}$$

$$\sup_{G\in\mathscr{F}_{U,[\xi_1,\xi_2]}(\mu,\sigma,\varepsilon)} \rho_g(G) = \sup_{\xi\in[\xi_1,\xi_2]} \sup_{G\in\mathscr{F}_{U,\xi}(\mu,\sigma,\varepsilon)} \rho_g(G). \tag{4.20}$$

Although these formulae may appear cumbersome, they are amenable to numerical implementation as one can scan over ξ and apply the fixed-ξ solution method to problem (4.14a).

The existence of an L_2-projection of γ onto the non-convex set $\mathscr{F}_{U,[\xi_1,\xi_2]}^{-1}$ is shown below:

Lemma 4.15. *For any γ with $\int_0^1 |\gamma(u)|^2\,\mathrm{d}u < \infty$ there exists $\gamma_{\xi_1,\xi_2}^\uparrow \in \mathscr{F}_{U,[\xi_1,\xi_2]}^{-1}$ such that:*

$$\gamma_{\xi_1,\xi_2}^\uparrow \in \arg\min_{h\in\mathscr{F}_{U,[\xi_1,\xi_2]}^{-1}} \|\gamma - h\|_2.$$

Uniqueness need not hold because of non-convexity. Let $\gamma^{\uparrow}_{\xi_1,\xi_2}$ be one such projection and write $a_{\xi_1,\xi_2} = \mathbb{E}(\gamma^{\uparrow}_{\xi_1,\xi_2}(V))$ and $b_{\xi_1,\xi_2} = \mathrm{Stdev}(\gamma^{\uparrow}_{\xi_1,\xi_2}(V))$.

Proposition 4.16 (Bounds for distortion risk measures for unimodal distribution functions)**.** *Suppose $\gamma^{\uparrow}_{\xi_1,\xi_2}$ is not constant. Then,*

$$h^{\uparrow}_{\xi_1,\xi_2}(u) := \mu + \sigma \frac{\gamma^{\uparrow}_{\xi_1,\xi_2}(u) - a_{\xi_1,\xi_2}}{b_{\xi_1,\xi_2}}$$

is a worst-case quantile for problem (4.14a).

The worst-case quantile of Problem (4.14a), given in Proposition 4.16, may not be unique, since different choices of $\xi \in [\xi_1, \xi_2]$ can yield the same maximum in (4.19). Proposition 4.16 also encompasses the result of Li et al. (2018) for the special case $\gamma(u) = \frac{1}{\beta-\alpha}\mathbf{1}_{(\alpha,\beta)}(u)$.

Because $\mathscr{F}^{-1}_{U,[\xi_1,\xi_2]}$ is non-convex, establishing continuity properties analogous to Lemma 4.13 is more difficult, so problem (4.14b) cannot be solved by the same method as Theorem 4.14. Nevertheless, Theorem 4.14 together with (4.20) provides a practical route to compute the supremum in (4.14b).

4.5 Robust portfolio optimisation

In this section, we apply the preceding results to robust portfolio optimisation under two types of uncertainty sets: (i) a mean–variance constraint together with a probabilistic constraint given by a Wasserstein metric on the multivariate return vector; and (ii) the same constraints plus an additional unimodality constraint on the portfolio return.

4.5.1 Mean–variance and Wasserstein distance constraints

Let X_i, $i = 1, \ldots, n$, denote the negative return (loss) from investing one unit in asset i, and let $\mathbf{w} = (w_1, \ldots, w_n) \in \mathscr{A} \subseteq \delta_n$ be the portfolio weights, where

$$\delta_n = \{(w_1, \ldots, w_n) : w_i \geqslant 0, \sum_{i=1}^{n} w_i = 1\}.$$

Hence, the negative return of the portfolio is $\sum_{i=1}^n w_i X_i$. Suppose only partial information is available about the return vector $\mathbf{X} = (X_1, \ldots, X_n)$: its mean vector $\mu = (\mu_1, \ldots, \mu_n)$, its covariance matrix Σ_0, and that the distribution $F_{\mathbf{X}}$ lies within a Wasserstein ball of radius $\sqrt{\varepsilon}$ around a reference distribution $F_{\mathbf{X}_0}$, i.e.,

$$d_W^{(n)}(F_{\mathbf{X}}, F_{\mathbf{X}_0}) \leqslant \sqrt{\varepsilon}, \qquad \varepsilon > 0.$$

The n-dimensional Wasserstein distance (order two) between two distributions F and G is defined by:

$$d_W^{(n)}(F,G) = \inf_{\mathbf{X}\sim F,\ \mathbf{Y}\sim G} \left(\mathbb{E}\|\mathbf{X}-\mathbf{Y}\|_2^2\right)^{1/2}$$

(e.g., Blanchet et al. (2022)). For the purposes of simplification and practicality, let us suppose that the reference distribution $F_{\mathbf{X}_0}$ shares the same mean and covariance as the (unknown) true law, i.e. $\mathbb{E}(\mathbf{X}_0) = \mu$ and $Cov(\mathbf{X}_0) = \Sigma_0$, and that Σ_0 is positive definite.

For a given weight vector $\mathbf{w}$ and radius $\varepsilon > 0$ define the univariate uncertainty set for the portfolio negative return by:

$$\mathscr{M}_{\mathbf{w},\varepsilon} = \left\{F_{\sum_{i=1}^n w_i X_i} : E(\mathbf{X}) = \mu, Cov(\mathbf{X}) = \Sigma_0\right\} \cap \left\{F_{\sum_{i=1}^n w_i X_i} : d_W^{(n)}(F_{\mathbf{X}}, F_{\mathbf{X}_0}) \leqslant \sqrt{\varepsilon}\right\}.$$

The robust portfolio problem considered is:

$$\arg\min_{\mathbf{w}\in\mathscr{A}} \sup_{G\in\mathscr{M}_{\mathbf{w},\varepsilon}} \rho_g(G). \tag{4.21}$$

As an example, one may take $\mathscr{A} = \{\mathbf{w} \in \Delta_n :\ -\mathbf{w}^\top \mu \geqslant a\}$ for some $a > 0$, imposing a lower bound on the (positive) portfolio return.

For robust portfolio optimisation, the uncertainty set $\mathscr{M}_{\mathbf{w},\varepsilon}$ is novel and differs from those previously considered in the literature (e.g. Blanchet et al. (2022), Bernard et al. (2024), Pesenti et al. (2024), Mao et al. (2025)). In $\mathscr{M}_{\mathbf{w},\varepsilon}$, the Wasserstein constraint is imposed on the multivariate distribution of the negative-return vector $\mathbf{X}$, whereas in Bernard et al. (2024), it is placed on the univariate distribution of the portfolio negative return $\mathbf{w}^\top\mathbf{X}$.

The following proposition shows that $\mathscr{M}_{\mathbf{w},\varepsilon}$ can be expressed as a univariate uncertainty set for $F_{\sum_{i=1}^n w_i X_i}$, which plays an important role in enabling the application of our results in Section 4.3.

Proposition 4.17. *We have*

$$\mathscr{M}_{\mathbf{w},\varepsilon} = \mathscr{M}_{\varepsilon\|\mathbf{w}\|_2^2}\left(\mathbf{w}^\top \mu, \sqrt{\mathbf{w}^\top \Sigma_0 \mathbf{w}}\right),$$

with reference distribution $F = F_{\mathbf{w}^\top \mathbf{X}_0}$.

Proposition 4.17 implies that problem (4.21) is equivalent to

$$\arg\min_{\mathbf{w}\in\mathscr{A}} \sup_{G\in\mathscr{M}_{\varepsilon\|\mathbf{w}\|_2^2}(\mathbf{w}^\top\mu,\sqrt{\mathbf{w}^\top\Sigma_0\mathbf{w}})} \rho_g(G),$$

which allows us to apply Theorems 4.2 and 4.4 to evaluate the inner supremum. The following proposition summarises the solution approach.

Proposition 4.18. *Problem* (4.21) *can be addressed as follows:*

(i) If g is concave and $(g^*)'$ is not constant, then the robust optimisation reduces to

$$\arg\min_{\mathbf{w}\in\mathscr{A}} \left\{ \mathbf{w}^\top \mu\, g(1) + \frac{\sqrt{\mathbf{w}^\top\Sigma_0\mathbf{w}}}{b_\mathbf{w}} \left(V_g + \lambda_\mathbf{w} C_{g,F_{\mathbf{w}^\top\mathbf{X}_0}}\right) \right\},$$

where $V_g = \mathrm{VaR}(g'(1-V))$, $C_{g,F_{\mathbf{w}^\top\mathbf{X}_0}} = Cov(F^{-1}_{\mathbf{w}^\top\mathbf{X}_0}(V), g'(1-V))$, and

$$b_\mathbf{w} = \sqrt{V_g + 2\lambda_\mathbf{w} C_{g,F_{\mathbf{w}^\top\mathbf{X}_0}} + \lambda_\mathbf{w}^2 \mathbf{w}^\top\Sigma_0\mathbf{w}},$$

with

$$\lambda_\mathbf{w} = \frac{-C_{g,F_{\mathbf{w}^\top\mathbf{X}_0}} + \sqrt{\frac{\left(C^2_{g,F_{\mathbf{w}^\top\mathbf{X}_0}} - V_g\, \mathbf{w}^\top\Sigma_0\mathbf{w}\right) A_\mathbf{w}^2}{A_\mathbf{w}^2 - (\mathbf{w}^\top\Sigma_0\mathbf{w})^2}}}{\mathbf{w}^\top\Sigma_0\mathbf{w}},$$

and

$$A_\mathbf{w} = \left(\mathbf{w}^\top\Sigma_0\mathbf{w} - \frac{\varepsilon\|\mathbf{w}\|_2^2}{2}\right) \vee \left(\sqrt{\mathbf{w}^\top\Sigma_0\mathbf{w}/V_g}\, C_{g,F_{\mathbf{w}^\top\mathbf{X}_0}}\right).$$

(ii) For $\rho = \mathrm{IQD}_\alpha^+$ *or* IQD_α^- *with* $\alpha \in (0,1/2)$*, the optimal robust portfolio is*

$$\arg\min_{\mathbf{w}\in\mathscr{A}} \left\{ \frac{\sqrt{\mathbf{w}^\top\Sigma_0\mathbf{w}}}{\sqrt{V_{\mathbf{w},\lambda_\mathbf{w}}}} \left(\frac{1+\lambda_\mathbf{w}\int_{1-\alpha}^{1-t_{\alpha,\lambda_\mathbf{w}}} F^{-1}_{\mathbf{w}^\top\mathbf{X}_0}(s)\,\mathrm{d}s}{\alpha - t_{\alpha,\lambda_\mathbf{w}}} - \frac{\lambda_\mathbf{w}\int_{1-\hat{t}_{\alpha,\lambda_\mathbf{w}}}^{\alpha} F^{-1}_{\mathbf{w}^\top\mathbf{X}_0}(s)\,\mathrm{d}s - 1}{\hat{t}_{\alpha,\lambda_\mathbf{w}} - 1 + \alpha} \right) \right\},$$

where $V_{\mathbf{w},\lambda} = \mathrm{VaR}((g_\lambda^*)'(V))$ *and* $\lambda_\mathbf{w}$ *equals the solution of* $d_W(F_{\mathbf{w}^\top\mathbf{X}_0}, H_\lambda) = \sqrt{\varepsilon}\|\mathbf{w}\|_2$ *when* $d_W(F_{\mathbf{w}^\top\mathbf{X}_0}, H_0) > \sqrt{\varepsilon}\|\mathbf{w}\|_2$*, and* $\lambda_\mathbf{w} = 0$ *otherwise.*

In the proposition above, the optimal portfolio position $\mathbf{w}$ depends on the reference distribution $F_{\mathbf{X}_0}$. Next, let us assume that $F_{\mathbf{X}_0}$ is elliptical, i.e., $F_{\mathbf{X}_0} \sim E_n(\mu, \Sigma_0, \psi)$, where μ is the mean, Σ_0 the covariance matrix, and ψ the characteristic generator. Note that elliptical distributions include the family of multivariate normal distributions and multivariate t-distributions as special cases. For any portfolio weight vector $\mathbf{w}$, the marginal distribution satisfies $F_{\mathbf{w}^\top\mathbf{X}_0} \sim E_1(\mathbf{w}^\top\mu,\, \mathbf{w}^\top\Sigma_0\mathbf{w},\, \psi)$. Define the standardised marginal distribution $F_0 = F_{\frac{\mathbf{w}^\top\mathbf{X}_0 - \mathbf{w}^\top\mu}{\sqrt{\mathbf{w}^\top\Sigma_0\mathbf{w}}}} \sim E_1(0,1,\psi)$. Hence,

$$C_{g,F_{\mathbf{w}^\top\mathbf{X}_0}} = \sqrt{\mathbf{w}^\top\Sigma_0\mathbf{w}}\,C_0,\ C_0 = \mathrm{Cov}(F_0^{-1}(V), g'(1-V)).$$

For this special $F_{\mathbf{X}_0}$, we can simplify Proposition 4.18 as the following corollaries.

Corollary 4.19. *Suppose* $F_{\mathbf{X}_0} \sim E_n(\mu, \Sigma_0, \psi)$*. If* g *is concave and* $(g^*)'$ *is not constant, the robust optimisation* (4.21) *is equivalent to:*

$$\arg\min_{\mathbf{w}\in\mathscr{A}} \left\{ \mathbf{w}^\top\mu\, g(1) + \frac{\sqrt{\mathbf{w}^\top\Sigma_0\mathbf{w}}}{\sqrt{V_g + 2C_0B_\mathbf{w} + B_\mathbf{w}^2}} \left(V_g + C_0B_\mathbf{w}\right) \right\}, \tag{4.22}$$

where

$$B_\mathbf{w} = -C_0 + \sqrt{\frac{(C_0^2 - V_g)A_\mathbf{w}^2}{A_\mathbf{w}^2 - (\mathbf{w}^\top\Sigma_0\mathbf{w})^2}}, \quad A_\mathbf{w} = \left(\mathbf{w}^\top\Sigma_0\mathbf{w} - \frac{\varepsilon\|\mathbf{w}\|_2^2}{2}\right) \vee \frac{C_0\mathbf{w}^\top\Sigma_0\mathbf{w}}{\sqrt{V_g}}.$$

If $g(1) = 0$ the linear term $\mathbf{w}^\top\mu\, g(1)$ vanishes from the objective in (4.22), though the mean vector μ may still enter implicitly via constraints in $\mathscr{A}$.

For the IQD-based optimisation, it is convenient to define, for $\alpha \in (0,1)$, $\lambda \geqslant 0$ and

$\mathbf{w} \in \mathscr{A}$,

$$t_{\alpha,\mathbf{w},\lambda} = \inf\left\{t \in [0,\alpha) : \frac{1/\sqrt{\mathbf{w}^\top \Sigma_0 \mathbf{w}} + \lambda \int_{1-\alpha}^{1-t} F_0^{-1}(s)\,\mathrm{d}s}{\alpha - t} \geqslant \lambda F_0^{-1}(1-t)\right\},$$

and

$$\hat{t}_{\alpha,\mathbf{w},\lambda} = \sup\left\{t \in (1-\alpha, 1] : \frac{\lambda \int_{1-t}^{\alpha} F_0^{-1}(s)\,\mathrm{d}s - 1/\sqrt{\mathbf{w}^\top \Sigma_0 \mathbf{w}}}{t - 1 + \alpha} \leqslant \lambda F_0^{-1}(1-t)\right\}.$$

Corollary 4.20. *Suppose $F_{\mathbf{X}_0} \sim E_n(\mu, \Sigma_0, \psi)$. If $\rho = \mathrm{IQD}_\alpha^+$ or IQD_α^- with $\alpha \in (0, 1/2)$, then the robust portfolio problem* (4.21) *is equivalent to*

$$\begin{aligned}\underset{\mathbf{w}\in\mathscr{A}}{\arg\min}\Bigg\{\sqrt{\frac{\mathbf{w}^\top\Sigma_0\mathbf{w}}{V_{\mathbf{w},\lambda_\mathbf{w}}}}\Bigg(&\frac{1+\lambda_\mathbf{w}\sqrt{\mathbf{w}^\top\Sigma_0\mathbf{w}}\int_{1-\alpha}^{1-t_{\alpha,\mathbf{w},\lambda_\mathbf{w}}}F_0^{-1}(s)\,\mathrm{d}s}{\alpha-t_{\alpha,\mathbf{w},\lambda_\mathbf{w}}}\\ &-\frac{\lambda_\mathbf{w}\sqrt{\mathbf{w}^\top\Sigma_0\mathbf{w}}\int_{1-\hat{t}_{\alpha,\mathbf{w},\lambda_\mathbf{w}}}^{\alpha}F_0^{-1}(s)\,\mathrm{d}s-1}{\hat{t}_{\alpha,\mathbf{w},\lambda_\mathbf{w}}-1+\alpha}\Bigg)\Bigg\},\end{aligned}$$

where

$$\lambda_\mathbf{w} = \eta_\mathbf{w} \mathbb{1}_{\left\{2\mathbf{w}^\top\Sigma_0\mathbf{w}\left(1-\frac{\int_{1-\alpha}^{1}F_0^{-1}(t)\mathrm{d}t-\int_0^{\alpha}F_0^{-1}(t)\mathrm{d}t}{\sqrt{2\alpha}}\right)>\varepsilon\|\mathbf{w}\|_2^2\right\}}$$

with

$$\delta_\mathbf{w} = \frac{\int_{1-\alpha}^{1}F_0^{-1}(t)\,\mathrm{d}t - \int_0^{\alpha}F_0^{-1}(t)\,\mathrm{d}t}{\sqrt{2\alpha}},$$

and $\eta_\mathbf{w} > 0$ is the solution of the equation

$$\begin{aligned}\left(1-\frac{\varepsilon\|\mathbf{w}\|_2^2}{2\mathbf{w}^\top\Sigma_0\mathbf{w}}\right)\sqrt{V_{\mathbf{w},\lambda}} = {}&\frac{1+\lambda\sqrt{\mathbf{w}^\top\Sigma_0\mathbf{w}}\int_{1-\alpha}^{1-t_{\alpha,\mathbf{w},\lambda}}F_0^{-1}(s)\,\mathrm{d}s}{\alpha-t_{\alpha,\mathbf{w},\lambda}}\int_{1-\alpha}^{1-t_{\alpha,\mathbf{w},\lambda}}F_0^{-1}(t)\,\mathrm{d}t\\ &+\frac{\lambda\sqrt{\mathbf{w}^\top\Sigma_0\mathbf{w}}\int_{1-\hat{t}_{\alpha,\mathbf{w},\lambda}}^{\alpha}F_0^{-1}(s)\,\mathrm{d}s-1}{\hat{t}_{\alpha,\mathbf{w},\lambda}-1+\alpha}\int_{1-\hat{t}_{\alpha,\mathbf{w},\lambda}}^{\alpha}F_0^{-1}(t)\,\mathrm{d}t\\ &+\lambda\sqrt{\mathbf{w}^\top\Sigma_0\mathbf{w}}\int_{(0,1-\hat{t}_{\alpha,\mathbf{w},\lambda})\cup(\alpha,1-\alpha)\cup(1-t_{\alpha,\mathbf{w},\lambda},1)}(F_0^{-1}(t))^2\,\mathrm{d}t.\end{aligned}$$

For numerical computation, $V_{\mathbf{w},\lambda}$ admits the explicit representation

$$\begin{aligned} V_{\mathbf{w},\lambda} = &\Big(\frac{1+\lambda\sqrt{\mathbf{w}^\top\Sigma_0\mathbf{w}}\int_{1-\alpha}^{1-t_{\alpha,\mathbf{w},\lambda}} F_0^{-1}(s)\,\mathrm{d}s}{\alpha - t_{\alpha,\mathbf{w},\lambda}}\Big)^2(\alpha - t_{\alpha,\mathbf{w},\lambda}) \\ &+\Big(\frac{\lambda\sqrt{\mathbf{w}^\top\Sigma_0\mathbf{w}}\int_{1-\hat{t}_{\alpha,\mathbf{w},\lambda}}^{\alpha} F_0^{-1}(s)\,\mathrm{d}s - 1}{\hat{t}_{\alpha,\mathbf{w},\lambda}-1+\alpha}\Big)^2(\hat{t}_{\alpha,\mathbf{w},\lambda}-1+\alpha) \\ &+\lambda^2\mathbf{w}^\top\Sigma_0\mathbf{w}\int_{(0,1-\hat{t}_{\alpha,\mathbf{w},\lambda})\cup(\alpha,1-\alpha)\cup(1-t_{\alpha,\mathbf{w},\lambda},1)} (F_0^{-1}(t))^2\,\mathrm{d}t. \end{aligned}$$

Corollary 4.21. *Suppose $F_{\mathbf{X}_0} \sim E_n(\mu,\Sigma_0,\psi)$. If $\rho = \mathrm{VaR}_\alpha$ or VaR_α^+ with $\alpha \in (0,1)$, then the solution of problem* (4.21) *is given by:*

$$\operatorname*{arg\,min}_{\mathbf{w}\in\mathscr{A}}\left\{\mathbf{w}^\top\mu + \sqrt{\frac{\mathbf{w}^\top\Sigma_0\mathbf{w}}{V_{\mathbf{w},\lambda_\mathbf{w}}}}\left(\frac{1+\lambda_\mathbf{w}\sqrt{\mathbf{w}^\top\Sigma_0\mathbf{w}}\int_\alpha^{1-t_{1-\alpha,\mathbf{w},\lambda_\mathbf{w}}} F_0^{-1}(s)\mathrm{d}s}{1-\alpha-t_{1-\alpha,\mathbf{w},\lambda_\mathbf{w}}} - 1\right)\right\},$$

where

$$\lambda_\mathbf{w} = \eta_\mathbf{w}\mathbb{1}_{\left\{2\mathbf{w}^\top\Sigma_0\mathbf{w}\left(1-\frac{\int_\alpha^1 F_0^{-1}(t)\mathrm{d}t}{\sqrt{\alpha(1-\alpha)}}\right)>\varepsilon\|\mathbf{w}\|_2^2\right\}}$$

with $\eta_\mathbf{w} \in (0,\infty)$ being the solution of

$$\begin{aligned} \left(1-\frac{\varepsilon\|\mathbf{w}\|_2^2}{2\mathbf{w}^\top\Sigma_0\mathbf{w}}\right)\sqrt{V_{\mathbf{w},\lambda}} = &\frac{1+\lambda\sqrt{\mathbf{w}^\top\Sigma_0\mathbf{w}}\int_\alpha^{1-t_{1-\alpha,\mathbf{w},\lambda}} F_0^{-1}(s)\mathrm{d}s}{1-\alpha-t_{1-\alpha,\mathbf{w},\lambda}}\int_\alpha^{1-t_{1-\alpha,\mathbf{w},\lambda}} F_0^{-1}(t)\mathrm{d}t \\ &+\lambda\sqrt{\mathbf{w}^\top\Sigma_0\mathbf{w}}\int_{(0,\alpha)\cup(1-t_{1-\alpha,\mathbf{w},\lambda},1)} (F_0^{-1}(t))^2\mathrm{d}t. \end{aligned}$$

Note that in Corollary 4.21, $V_{\mathbf{w},\lambda}$ can be expressed in a more explicit way for computation as below:

$$\begin{aligned} V_{\mathbf{w},\lambda} = &\frac{\left(1+\lambda\sqrt{\mathbf{w}^\top\Sigma_0\mathbf{w}}\int_\alpha^{1-t_{1-\alpha,\mathbf{w},\lambda}} F_0^{-1}(s)\mathrm{d}s\right)^2}{1-\alpha-t_{1-\alpha,\mathbf{w},\lambda}} \\ &+\lambda^2\mathbf{w}^\top\Sigma_0\mathbf{w}\int_{(0,\alpha)\cup(1-t_{1-\alpha,\mathbf{w},\lambda},1)} (F_0^{-1}(t))^2\mathrm{d}t - 1. \end{aligned}$$

For the portfolio optimization using GlueVaR, we need the following notation: For

$0 < \alpha < \beta < 1$, $0 < h_1 < h_2 < 1$, $\lambda \geqslant 0$ and $\mathbf{w} \in \mathscr{A}$, let

$$u^{h_1,h_2}_{\alpha,\beta,\mathbf{w},\lambda} = \inf \Bigg\{ t \in [0, 1-\alpha) : \frac{1 - g^{h_1,h_2}_{\alpha,\beta}(t) + \lambda \sqrt{\mathbf{w}^\top \Sigma_0 \mathbf{w}} \int_\alpha^{1-t} F_0^{-1}(s)\mathrm{d}s}{1-\alpha-t}$$
$$\geqslant \frac{h_1}{1-\beta} \mathbb{1}_{(0,1-\beta)}(t) + \frac{h_2-h_1}{\beta-\alpha} \mathbb{1}_{[1-\beta,1-\alpha)}(t) + \lambda \sqrt{\mathbf{w}^\top \Sigma_0 \mathbf{w}} F_0^{-1}(1-t) \Bigg\}$$

and

$$d_{\mathbf{w},\lambda} = 1 - h_2 + \frac{h_1}{1-\beta}(1 - u^{h_1,h_2}_{\alpha,\beta,\mathbf{w},\lambda} - \beta)_+ + \frac{h_2-h_1}{\beta-\alpha}(\beta \wedge (1 - u^{h_1,h_2}_{\alpha,\beta,\mathbf{w},\lambda}) - \alpha).$$

Corollary 4.22. *Suppose $F_{\mathbf{X}_0} \sim E_n(\mu, \Sigma_0, \psi)$. If $\rho = \mathrm{GlueVaR}^{h_1,h_2}_{\beta,\alpha}$ with $0 < \alpha < \beta < 1$ and $0 < h_1 < h_2 < 1$ satisfying $\frac{h_1}{1-\beta} \geqslant \frac{h_2-h_1}{\beta-\alpha}$, then the solution of problem* (4.21) *is given by:*

$$\begin{aligned}
\operatorname*{arg\,min}_{\mathbf{w}\in\mathscr{A}} \Bigg\{ & \mathbf{w}^\top \mu - \frac{\sqrt{\mathbf{w}^\top \Sigma_0 \mathbf{w}}}{\sqrt{V_{\mathbf{w},\lambda_\mathbf{w}}}} \left(1 - \frac{h_1^2((1-\beta) \wedge u^{h_1,h_2}_{\alpha,\beta,\mathbf{w},\lambda_\mathbf{w}})}{(1-\beta)^2} - \frac{(h_2-h_1)^2(\beta - 1 + u^{h_1,h_2}_{\alpha,\beta,\mathbf{w},\lambda_\mathbf{w}})_+}{(\beta-\alpha)^2} \right) \\
& + \frac{\sqrt{\mathbf{w}^\top \Sigma_0 \mathbf{w}}}{\sqrt{V_{\mathbf{w},\lambda_\mathbf{w}}}} \frac{1 - g^{h_1,h_2}_{\alpha,\beta}(u^{h_1,h_2}_{\alpha,\beta,\mathbf{w},\lambda})}{1-\alpha-u^{h_1,h_2}_{\alpha,\beta,\mathbf{w},\lambda}} d_{\mathbf{w},\lambda_\mathbf{w}} + \frac{\lambda_\mathbf{w} \mathbf{w}^\top \Sigma_0 \mathbf{w}}{\sqrt{V_{\mathbf{w},\lambda_\mathbf{w}}}} \frac{\int_\alpha^{1-u^{h_1,h_2}_{\alpha,\beta,\mathbf{w},\lambda_\mathbf{w}}} F_0^{-1}(s)\mathrm{d}s}{1-\alpha-u^{h_1,h_2}_{\alpha,\beta,\mathbf{w},\lambda_\mathbf{w}}} d_{\mathbf{w},\lambda_\mathbf{w}} \\
& + \frac{\lambda_\mathbf{w} \mathbf{w}^\top \Sigma_0 \mathbf{w}}{\sqrt{V_{\mathbf{w},\lambda_\mathbf{w}}}} \left(\frac{h_1}{1-\beta} \int_{\beta \vee (1-u^{h_1,h_2}_{\alpha,\beta,\mathbf{w},\lambda_\mathbf{w}})}^1 F_0^{-1}(s)\mathrm{d}s + \frac{h_2-h_1}{\beta-\alpha} \int_{\beta \wedge (1-u^{h_1,h_2}_{\alpha,\beta,\mathbf{w},\lambda_\mathbf{w}})}^\beta F_0^{-1}(s)\mathrm{d}s \right) \Bigg\},
\end{aligned}$$

where

$$\lambda_\mathbf{w} = \eta_\mathbf{w} \mathbb{1}_{\left\{ 2\mathbf{w}^\top \Sigma_0 \mathbf{w} \left(1 - \frac{\left(\frac{1}{1-\alpha} \vee \frac{h_1}{1-\beta}\right) \int_\beta^1 F_0^{-1}(t)\mathrm{d}t + \left(\frac{1}{1-\alpha} \wedge \frac{1-h_1}{\beta-\alpha}\right) \int_\alpha^\beta F_0^{-1}(t)\mathrm{d}t}{\sqrt{\frac{1-\beta}{(1-\alpha)^2} \vee \frac{h_1^2}{1-\beta} + \frac{\beta-\alpha}{(1-\alpha)^2} \wedge \frac{(1-h_1)^2}{\beta-\alpha} - 1}} \right) > \varepsilon \|\mathbf{w}\|_2^2 \right\}}$$

with $\eta_{\mathbf{w}} \in (0,\infty)$ being the solution of

$$
\begin{aligned}
&\left(1-\frac{\varepsilon\|\mathbf{w}\|_2^2}{2\mathbf{w}^\top\Sigma_0\mathbf{w}}\right)\sqrt{V_{\mathbf{w},\lambda}} \\
&= \frac{1-g_{\alpha,\beta}^{h_1,h_2}(u_{\alpha,\beta,\mathbf{w},\lambda}^{h_1,h_2})}{1-\alpha-u_{\alpha,\beta,\mathbf{w},\lambda}^{h_1,h_2}}\int_\alpha^{1-u_{\alpha,\beta,\mathbf{w},\lambda}^{h_1,h_2}} F_0^{-1}(t)\mathrm{d}t + \frac{h_1}{1-\beta}\int_{\beta\vee(1-u_{\alpha,\beta,\mathbf{w},\lambda}^{h_1,h_2})}^1 F_0^{-1}(t)\mathrm{d}t \\
&\quad + \frac{h_2-h_1}{\beta-\alpha}\int_{\beta\vee(1-u_{\alpha,\beta,\mathbf{w},\lambda}^{h_1,h_2})}^{\beta} F_0^{-1}(t)\mathrm{d}t + \lambda\sqrt{\mathbf{w}^\top\Sigma_0\mathbf{w}}\int_{(0,\alpha)\cup(1-u_{\alpha,\beta,\mathbf{w},\lambda}^{h_1,h_2},1)} (F_0^{-1}(t))^2\mathrm{d}t \\
&\quad + \frac{\lambda\sqrt{\mathbf{w}^\top\Sigma_0\mathbf{w}}}{1-\alpha-u_{\alpha,\beta,\mathbf{w},\lambda}^{h_1,h_2}}\left(\int_\alpha^{1-u_{\alpha,\beta,\mathbf{w},\lambda}^{h_1,h_2}} F_0^{-1}(s)\mathrm{d}s\right)^2.
\end{aligned}
$$

Note that in Corollary 4.22, an explicit expression for $V_{\mathbf{w},\lambda}$ for computation is given by:

$$
\begin{aligned}
V_{\mathbf{w},\lambda} &= \frac{1}{1-\alpha-u_{\alpha,\beta,\mathbf{w},\lambda}^{h_1,h_2}}\left(1-g_{\alpha,\beta}^{h_1,h_2}(u_{\alpha,\beta,\mathbf{w},\lambda}^{h_1,h_2})+\lambda\sqrt{\mathbf{w}^\top\Sigma_0\mathbf{w}}\int_\alpha^{1-u_{\alpha,\beta,\mathbf{w},\lambda}^{h_1,h_2}} F_0^{-1}(t)\mathrm{d}t\right)^2 \\
&\quad + \frac{h_1^2((1-\beta)\wedge u_{\alpha,\beta,\mathbf{w},\lambda}^{h_1,h_2})}{(1-\beta)^2} + \frac{(h_2-h_1)^2(\beta-1+u_{\alpha,\beta,\mathbf{w},\lambda}^{h_1,h_2})_+}{(\beta-\alpha)^2} \\
&\quad + \lambda^2\mathbf{w}^\top\Sigma_0\mathbf{w}\int_{(0,\alpha)\cup(1-u_{\alpha,\beta,\mathbf{w},\lambda}^{h_1,h_2},1)} (F_0^{-1}(t))^2\mathrm{d}t - 1.
\end{aligned}
$$

4.5.2 Unimodal constraints

In this subsection, we additionally suppose that the portfolio negative return $\sum_{i=1}^n w_i X_i$ is known to be unimodal with inflection point $\xi \in (0,1)$. Then the uncertainty set becomes

$$
\mathscr{M}_{\mathbf{w},\xi,\varepsilon} = \mathscr{M}_{\mathbf{w},\varepsilon} \cap \mathscr{F}_{U,\xi},
$$

and by Proposition 4.17, we have

$$
\mathscr{M}_{\mathbf{w},\xi,\varepsilon} = \mathscr{F}_{U,\xi}\big(\mathbf{w}^\top\mu, \sqrt{\mathbf{w}^\top\Sigma_0\mathbf{w}}, \varepsilon\|\mathbf{w}\|_2^2\big).
$$

The robust optimisation problem is to solve

$$
\mathop{\arg\min}_{\mathbf{w}\in\mathscr{A}} \sup_{G\in\mathscr{F}_{U,\xi}(\mathbf{w}^\top\mu,\sqrt{\mathbf{w}^\top\Sigma_0\mathbf{w}},\varepsilon\|\mathbf{w}\|_2^2)} \rho_g(G). \tag{4.23}
$$

Assume $g \in \mathscr{H}$ has a density $\gamma(u) = \partial_{-}g(x)\big|_{x=1-u}$ for $0 < u < 1$. First, consider the case $\varepsilon = \infty$ i.e., there is no probability constraint. Recall the following definition, $\hat{b}_\xi = \sqrt{\int_0^1 (\gamma_\xi^\uparrow(u) - g(1))^2 \, \mathrm{d}u}$.

Proposition 4.23. *When $\varepsilon = \infty$ the robust problem* (4.23) *is equivalent to*

$$\arg\min_{\mathbf{w}\in\mathscr{A}} \left(\mathbf{w}^\top \mu\, g(1) + \hat{b}_\xi \sqrt{\mathbf{w}^\top \Sigma_0 \mathbf{w}}\right).$$

Note that for GD and MMD, they have $g(1) = 0$. Therefore, if $\mathscr{A} = \delta_n$ the unimodality information does not alter the optimisation problem, although it reduces the worst-case value of the distortion risk metrics through the constant $\hat{b}_\xi$. For RVaR, where $g(1) = 1$, unimodality may affect the optimal portfolio via $\hat{b}_\xi$.

Let $(F_{\mathbf{w}^\top \mathbf{X}_0})_\xi^{-1,\uparrow}$ denote the projection of $F^{-1}_{\mathbf{w}^\top \mathbf{X}_0}$ onto $\mathscr{F}^{-1}_{U,\xi}$. The following result treats the case $0 < \varepsilon < \infty$.

Proposition 4.24. *Suppose $k^\uparrow_{\lambda,\xi}$ is non-constant for all $\lambda \geqslant 0$, and $(F_{\mathbf{w}^\top \mathbf{X}_0})_\xi^{-1,\uparrow}$ is not constant for all values of $\mathbf{w} \in \delta_n$. For $0 < \varepsilon < \infty$ the problem* (4.23) *is equivalent to:*

$$\arg\min_{\mathbf{w}\in\delta_{n,\varepsilon}} \left(\mathbf{w}^\top \mu\, g(1) + \frac{\sqrt{\mathbf{w}^\top \Sigma_0 \mathbf{w}} \left(\int_0^1 \gamma(u) k^\uparrow_{\lambda_\mathbf{w},\xi}(u)\, \mathrm{d}u - g(1)(g(1) + \lambda_\mathbf{w} \mathbf{w}^\top \mu) \right)}{\sqrt{\int_0^1 \left(k^\uparrow_{\lambda_\mathbf{w},\xi}(u) - g(1) - \lambda_\mathbf{w} \mathbf{w}^\top \mu\right)^2 \mathrm{d}u}} \right),$$

where

$$\delta_{n,\varepsilon} = \mathscr{A} \cap \left\{ 2\mathbf{w}^\top \Sigma_0 \mathbf{w} - 2\sqrt{\mathbf{w}^\top \Sigma_0 \mathbf{w}} \sqrt{\int_0^1 \left((F_{\mathbf{w}^\top \mathbf{X}_0})_\xi^{-1,\uparrow}(u) - \mathbf{w}^\top \mu\right)^2 \mathrm{d}u} < \varepsilon \|\mathbf{w}\|_2^2 \right\},$$

and

$$\lambda_\mathbf{w} = \eta_\mathbf{w} \mathbf{1}_{\{2\mathbf{w}^\top \Sigma_0 \mathbf{w} - 2\sqrt{\mathbf{w}^\top \Sigma_0 \mathbf{w}}\, \delta'_\mathbf{w} > \varepsilon \|\mathbf{w}\|_2^2\}},$$

with

$$\delta'_\mathbf{w} = \frac{\int_0^1 F^{-1}_{\mathbf{w}^\top \mathbf{X}_0}(u) \gamma_\xi^\uparrow(u)\, \mathrm{d}u - g(1)\mathbf{w}^\top \mu}{\sqrt{\int_0^1 (\gamma_\xi^\uparrow(u) - g(1))^2\, \mathrm{d}u}},$$

and $\eta_{\mathbf{w}} > 0$ *solves*

$$\int_0^1 F^{-1}_{\mathbf{w}^\top \mathbf{X}_0}(u) k^{\uparrow}_{\lambda,\xi}(u)\,\mathrm{d}u = \frac{2\mathbf{w}^\top \Sigma_0 \mathbf{w} - \varepsilon \|\mathbf{w}\|_2^2}{2\sqrt{\mathbf{w}^\top \Sigma_0 \mathbf{w}}} \sqrt{\int_0^1 \left(k^{\uparrow}_{\lambda,\xi}(u) - g(1) - \lambda \mathbf{w}^\top \mu\right)^2 \mathrm{d}u} \\ + \mathbf{w}^\top \mu \left(g(1) + \lambda \mathbf{w}^\top \mu\right).$$

Next, assume that the reference distribution is elliptical, i.e., $F_{\mathbf{X}_0} \sim E_n(\mu, \Sigma_0, \psi)$. It follows that for any $\mathbf{w}$, the marginal satisfies $F_{\mathbf{w}^\top \mathbf{X}_0} \sim E_1(\mathbf{w}^\top \mu,\, \mathbf{w}^\top \Sigma_0 \mathbf{w},\, \psi)$. Define the standardised marginal

$$F_0 = F_{\frac{\mathbf{w}^\top \mathbf{X}_0 - \mathbf{w}^\top \mu}{\sqrt{\mathbf{w}^\top \Sigma_0 \mathbf{w}}}} \;\sim\; E_1(0, 1, \psi).$$

We denote by $(F_0)^{-1,\uparrow}_{\xi}$ the projection of F_0^{-1} onto $\mathscr{F}_{U,\xi}$. This yields the following corollary.

Corollary 4.25. *Suppose* $k^{\uparrow}_{\lambda,\xi}$ *is non-constant for all* $\lambda \geqslant 0$ *and* $(F_0)^{-1,\uparrow}_{\xi}$ *is not constant. For* $0 < \varepsilon < \infty$ *the robust problem* (4.23) *becomes*

$$\underset{\mathbf{w} \in \delta_{n,\varepsilon}}{\arg\min} \left(\mathbf{w}^\top \mu\, g(1) + \frac{\sqrt{\mathbf{w}^\top \Sigma_0 \mathbf{w}} \left(\int_0^1 \gamma(u) k^{\uparrow}_{\lambda_{\mathbf{w}},\xi}(u)\,\mathrm{d}u - g(1)(g(1) + \lambda_{\mathbf{w}} \mathbf{w}^\top \mu)\right)}{\sqrt{\int_0^1 \left(k^{\uparrow}_{\lambda_{\mathbf{w}},\xi}(u) - g(1) - \lambda_{\mathbf{w}} \mathbf{w}^\top \mu\right)^2 \mathrm{d}u}} \right),$$

where

$$\delta_{n,\varepsilon} = \mathscr{A} \cap \left\{ 2\mathbf{w}^\top \Sigma_0 \mathbf{w} \left(1 - \sqrt{\int_0^1 ((F_0)^{-1,\uparrow}_{\xi}(u))^2\,\mathrm{d}u}\right) < \varepsilon \|\mathbf{w}\|_2^2 \right\},$$

and

$$\lambda_{\mathbf{w}} = \eta_{\mathbf{w}} \mathbf{1}_{\{2\mathbf{w}^\top \Sigma_0 \mathbf{w}(1 - \delta''_{\mathbf{w}}) > \varepsilon \|\mathbf{w}\|_2^2\}}, \quad \delta''_{\mathbf{w}} = \frac{\int_0^1 F_0^{-1}(u) \gamma^{\uparrow}_{\xi}(u)\,\mathrm{d}u}{\sqrt{\int_0^1 (\gamma^{\uparrow}_{\xi}(u) - g(1))^2\,\mathrm{d}u}},$$

with $\eta_{\mathbf{w}} > 0$ *solving*

$$\int_0^1 F_0^{-1}(u) k^{\uparrow}_{\lambda,\xi}(u)\,\mathrm{d}u = \frac{2\mathbf{w}^\top \Sigma_0 \mathbf{w} - \varepsilon \|\mathbf{w}\|_2^2}{2\mathbf{w}^\top \Sigma_0 \mathbf{w}} \sqrt{\int_0^1 \left(k^{\uparrow}_{\lambda,\xi}(u) - g(1) - \lambda \mathbf{w}^\top \mu\right)^2 \mathrm{d}u}.$$

4.6 Numerical Examples

Our main results reduce robust portfolio optimisation under an ambiguity set characterised by mean, variance and a Wasserstein ball to minimising a deterministic objective. In this section, we demonstrate numerical results to illustrate the impact of model uncertainty on portfolio optimisation for different risk metrics: GD, MMD, IQD, VaR, ES and GlueVaR. Specifically, we solve the optimisation problem in Corollary 4.19 for GD, MMD and ES, Corollary 4.20 for IQD, Corollary 4.21 for VaR and Corollary 4.22 for GlueVaR.

We assume that $\mathscr{A} = \Delta_n$ and that the reference distribution $\mathbf{X}_0 \sim N(\mu, \Sigma_0)$ represents the negative returns from investing one dollar in each of n different assets in the market. To simplify the analysis, we consider the case $n = 2$. For the reference mean vector, we set $\mu = (-0.06, -0.03)^\top$. We consider both positively and negatively correlated covariance matrices Σ_0, as follows:

$$\Sigma_0^{(1)} = \begin{bmatrix} 0.15^2, & 0.5 \times 0.15 \times 0.1 \\ 0.5 \times 0.15 \times 0.1 & 0.1^2 \end{bmatrix}, \quad \Sigma_0^{(2)} = \begin{bmatrix} 0.15^2, & -0.5 \times 0.15 \times 0.1 \\ -0.5 \times 0.15 \times 0.1 & 0.1^2 \end{bmatrix}.$$

The uncertainty is controlled by the Wasserstein radius ε.

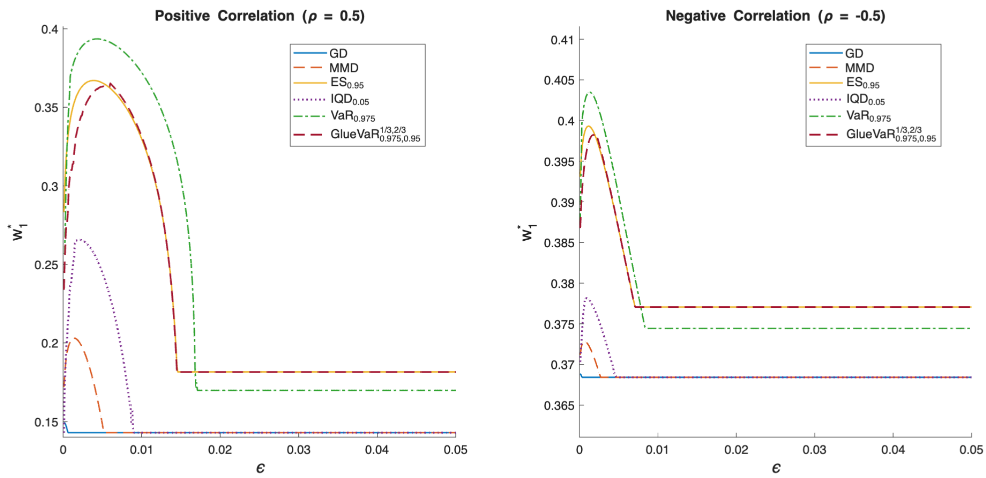

Figure 4.1: Optimal weight w_1^* under different Wasserstein radius ε with positive and negative correlations.

The impact of model uncertainty on portfolio optimisation under these distortion risk metrics can be seen clearly in Figure 4.1. In fact, as ε changes, w_1 follows a similar pattern across all six distortion risk metrics. For small ε (low uncertainty), more weight is allocated to the first asset; for intermediate ε (moderate uncertainty), the weight allocated to the first asset begins to decline; and for large ε, it becomes constant (corresponding to the case in which the Wasserstein constraint becomes inactive). The distortion risk metrics can be naturally divided into two groups in Figure 4.1: the optimal weights of the first asset for GD, MMD, and IQD (variability measures) are always smaller than those for VaR, ES and GlueVaR (tail-risk measures), and the first group is less sensitive than the second group in portfolio optimisation. In particular, the optimal weights for GD remain nearly constant, showing that model uncertainty has only a small impact on portfolio selection under this criterion. These qualitative findings are driven by the fact that $g(1) = 0$ for GD, MMD and IQD, implying that μ does not contribute to the portfolio optimisation; consequently, all these criteria behave similarly to variance. In contrast, for the tail-risk measures VaR, ES and GlueVaR, substantially more weight is allocated to the riskiest asset, and these criteria exhibit more sensitivity to model uncertainty.

For the variability measures, the distortion function satisfies $g(1) = 0$. Inspecting the objective in Corollary 4.19, the term $\mathbf{w}^\top \mu\, g(1)$ vanishes, so the expected return vector μ does not enter the optimisation directly. The objective then depends on $\mathbf{w}$ only through the portfolio variance $\mathbf{w}^\top \Sigma_0 \mathbf{w}$ (and the Wasserstein-related correction), which makes the problem qualitatively similar to minimum-variance optimisation. This explains why GD, MMD, and IQD produce similar optimal weights and exhibit low sensitivity to changes in the Wasserstein radius ε: the optimisation is governed almost entirely by the covariance structure, regardless of whether the underlying functional measures average absolute deviation, median deviation, or inter-quantile spread. Within this group, the optimal weights are nearly identical because the differences among GD, MMD, and IQD are encoded in the shape of g on the interior of $[0,1]$, which affects the worst-case value of the risk metric, yet only has a minor influence on the location of the optimal portfolio. For the tail-risk measures, $g(1) = 1$, and the term $\mathbf{w}^\top \mu$ contributes directly to the objective. This couples the portfolio's expected return with its tail-risk exposure, leading the optimiser to tilt towards assets with higher expected returns (lower expected losses) even at the cost of increased variance. As a result, the optimal weight w_1^* is substantially larger for VaR, ES, and GlueVaR than for the variability group, since asset 1 offers a higher expected return ($\mu_1 = -0.06$ versus $\mu_2 = -0.03$). Moreover, because tail-risk measures focus on the upper quantiles of the loss distribution, the worst-case value is more sensitive to distributional perturbations: widening the Wasserstein ball ε enlarges the set of plausible tail behaviours, causing a more pronounced shift in the optimal allocation. This explains the greater sensitivity of the tail-risk group to the uncertainty parameter ε observed in Figure 4.1. Within this group, VaR, ES, and GlueVaR yield similar optimal weights because they all target the same tail region of the distribution (as controlled by the confidence level α), and GlueVaR is, by construction, a convex combination of VaR and ES, so its optimum naturally lies close to those of its constituents. In summary, the key is whether $g(1) = 0$ or $g(1) = 1$: when $g(1) = 0$, the optimisation is mean-blind and variance-driven; when $g(1) = 1$, the mean-risk trade-off actively shapes the portfolio, thereby amplifying sensitivity to model uncertainty.

We also include the result of the case under Student's t-distribution assuming that $\mathscr{A} = \Delta_n$ and that the reference distribution $\mathbf{X}_0 \sim \mathbf{t}(\nu)$ represents the negative returns with $\nu = 5$. For the reference mean vector, we set $\mu = (-0.06, -0.03)^\top$. We consider both positively and negatively correlated covariance matrices Σ_0 as a normal distribution case :

$$\Sigma_0^{(1)} = \begin{bmatrix} 0.15^2, & 0.5 \times 0.15 \times 0.1 \\ 0.5 \times 0.15 \times 0.1 & 0.1^2 \end{bmatrix}, \quad \Sigma_0^{(2)} = \begin{bmatrix} 0.15^2, & -0.5 \times 0.15 \times 0.1 \\ -0.5 \times 0.15 \times 0.1 & 0.1^2 \end{bmatrix}.$$

The uncertainty is controlled by the Wasserstein radius ε. The impact of model uncertainty can be seen in Figure 4.2 Compared with the normal case in Figure 4.1, the Student's t case

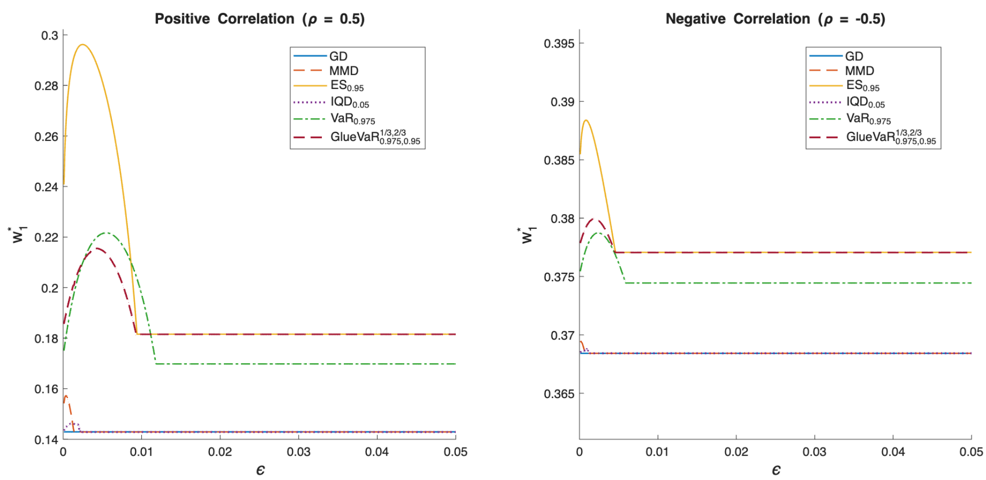

Figure 4.2: Optimal weight w_1^* under different Wasserstein radius ε with positive and negative correlations under Student's t distribution.

in Figure 4.2 leads to noticeably more conservative allocations to the first asset under the tail–risk criteria VaR, ES and GlueVaR, especially for small and intermediate values of ε. This effect is most pronounced under positive correlation, where the heavier tails of the Student's t distribution reduce the optimal weight assigned to the riskier asset relative to the Gaussian benchmark. By contrast, the allocations obtained under GD, MMD and IQD are only mildly affected by the change of distribution and remain close to those in the normal setting, which is consistent with the fact that these criteria are driven primarily by variability rather than extreme tail behaviour. For larger ε, the curves again flatten, and the differences between the normal and Student's t cases become much smaller, indicating that the distributional effect is strongest when the Wasserstein ambiguity radius is not too large.

4.6.1 Future work

The numerical examples above demonstrate the tractability of our framework. While a full empirical back-test using historical return data lies beyond the scope of this chapter, we briefly discuss the expected practical benefits and potential future work.

By construction, the robust portfolio minimises the worst-case distortion risk metric over an ambiguity set. In doing so, it hedges against model misspecification in the tails and, consequently, avoids concentrated positions that arise from over-fitting to a single distributional assumption. A direct consequence of the min–max formulation is that the robust portfolio is guaranteed to achieve a lower worst-case distortion risk metric than any portfolio that ignores distributional uncertainty. This is the defining feature of the DRO approach: by accounting for all plausible distributions within the Wasserstein ball, the resulting allocation is inherently more resilient to adverse distributional shifts.

Whether the robust portfolio also improves the Sharpe ratio is not confirmed. The Sharpe ratio is defined as the ratio of expected excess return to the standard deviation, yet the distortion risk metrics considered here, such as VaR, ES, GD, and IQD, measure risk in fundamentally different ways from the standard deviation. In particular, the notion of risk in our framework is not the standard deviation. It is a quantile-based or dispersion-based functional that may penalise tail losses or variability quite differently from variance. Consequently, a portfolio that is optimal in the worst-case distortion risk metric sense need not be optimal in the mean-variance sense. A systematic numerical comparison between robust distortion-based portfolios and Sharpe ratio-optimal portfolios, including an analysis of whether the former yields competitive or superior Sharpe ratios, is an important direction for future research.

Several extensions merit investigation. First, an empirical study using historical equity return data would allow a direct comparison of realised risk, return, and the Sharpe ratio arising between robust distortion-based portfolios and classical benchmarks. Second, the unimodality constraint developed in Section 4.4 could be tested empirically to quantify the extent to which it tightens the ambiguity set and reduces conservatism in practice. Finally, extending the framework to accommodate transaction costs and dynamic rebalancing would further enhance its applicability.

4.7 Proofs of results

All the proofs of the results in Sections 4.3-4.4 are presented in this section.

4.7.1 Proof of results in Section 4.3

In this section we present the proofs of the results stated in Section 4.3.

Proof of Lemma 4.1. By definition,

$$\begin{aligned} Corr(F^{-1}(V),(g_\lambda^*)'(1-V)) &= \frac{\mathbb{E}\big(F^{-1}(V)(g_\lambda^*)'(1-V)\big)-\mathbb{E}(F^{-1}(V))\mathbb{E}((g_\lambda^*)'(1-V))}{\sigma_F\sqrt{\mathrm{VaR}((g_\lambda^*)'(V))}} \\ &= \frac{\mathbb{E}\big(F^{-1}(V)(g_\lambda^*)'(1-V)\big)-\mu_F\big(g(1)+\lambda\mu_F\big)}{\sigma_F\sqrt{\mathbb{E}\big((g_\lambda^*)'(V)\big)^2-(g(1)+\lambda\mu_F)^2}}. \end{aligned}$$

Fix $\lambda_0\in[0,\infty)$. We first show that, if $(g_{\lambda_0}^*)'(t)$ is continuous at $t\in(0,1)$, then $(g_\lambda^*)'(t)$ depends continuously on λ at λ_0. For $\lambda_1,\lambda_2\in[0,\infty)$ one can easily check:

$$|g_{\lambda_2}(t)-g_{\lambda_1}(t)|\leqslant C|\lambda_2-\lambda_1|, \qquad C=\int_0^1|F^{-1}(s)|\,\mathrm{d}s,$$

and hence also $|g_{\lambda_2}^*(t)-g_{\lambda_1}^*(t)|\leqslant C|\lambda_2-\lambda_1|$. Therefore, $\sup_{t\in(0,1)}|g_{\lambda_2}^*(t)-g_{\lambda_1}^*(t)|\leqslant C|\lambda_2-\lambda_1|\to 0$ as $|\lambda_2-\lambda_1|\to 0$.

Suppose, by contradiction, that there exists a point $t \in (0,1)$ of continuity of $(g^*_{\lambda_0})'$, for which $(g^*_\lambda)'(t) \not\to (g^*_{\lambda_0})'(t)$ as $\lambda \to \lambda_0$. Without any loss of generality, assume there is a sequence $\lambda_n \to \lambda_0$ such that: $\lim_{n\to\infty}(g^*_{\lambda_n})'(t) = c > (g^*_{\lambda_0})'(t)$. Set $d = c - (g^*_{\lambda_0})'(t) > 0$. By continuity of $(g^*_{\lambda_0})'$ at t there exists $\varepsilon > 0$ with $(g^*_{\lambda_0})'(s) \leqslant (g^*_{\lambda_0})'(t) + d/3$ for all $s \in (t-\varepsilon, t]$. Moreover, for large values of n, we have $(g^*_{\lambda_n})'(t) > c - d/3$. Since $(g^*_{\lambda_n})'$ is non-increasing, it follows that $(g^*_{\lambda_n})'(s) \geqslant c - d/3$ on $(t-\varepsilon, t]$ for all large n. Hence,

$$\int_{t-\varepsilon}^{t} (g^*_{\lambda_n})'(s)\,\mathrm{d}s \geqslant \int_{t-\varepsilon}^{t} (g^*_{\lambda_0})'(s)\,\mathrm{d}s + \frac{d\varepsilon}{3},$$

or equivalently

$$g^*_{\lambda_n}(t) - g^*_{\lambda_n}(t-\varepsilon) \geqslant g^*_{\lambda_0}(t) - g^*_{\lambda_0}(t-\varepsilon) + \frac{d\varepsilon}{3},$$

for all large n. This contradicts the uniform convergence $\sup_{t\in(0,1)} |g^*_{\lambda_n}(t) - g^*_{\lambda_0}(t)| \to 0$. Thus $(g^*_\lambda)'(t) \to (g^*_{\lambda_0})'(t)$ whenever $(g^*_{\lambda_0})'$ is continuous at t. Since $(g^*_{\lambda_0})'$ has at most countably many discontinuities on $(0,1)$, we obtain $(g^*_\lambda)'(t) \to (g^*_{\lambda_0})'(t)$ for almost every $t \in (0,1)$, as $\lambda \to \lambda_0$.

Set $X_\lambda = (g^*_\lambda)'(1-V)$ and $Y_\lambda = (g^*)'(1-V) + \lambda F^{-1}(V)$. Elementary computations give:

$$\mathrm{ES}_\alpha(X_\lambda) = \frac{g^*_\lambda(1-\alpha)}{1-\alpha}, \qquad \mathrm{ES}_\alpha(Y_\lambda) = \frac{g^*(1-\alpha) + \lambda \int_\alpha^1 F^{-1}(t)\,\mathrm{d}t}{1-\alpha}.$$

Since $g^*_\lambda(t) \leqslant g^*(t) + \lambda \int_{1-t}^1 F^{-1}(s)\,\mathrm{d}s$ we have: $\mathrm{ES}_\alpha(X_\lambda) \leqslant \mathrm{ES}_\alpha(Y_\lambda)$ for all $\alpha \in (0,1)$, and obviously $\mathbb{E}(X_\lambda) = \mathbb{E}(Y_\lambda) = g(1) + \lambda\mu_F$. By Corollary 2.61 and Theorem 2.57 in Föllmer and Schied (2016), this implies $X_\lambda \leqslant_{\mathrm{cx}} Y_\lambda$ (convex order), hence,

$$\mathbb{E}(X_\lambda^2) \leqslant \mathbb{E}(Y_\lambda^2),$$

and, therefore,

$$\begin{aligned}\int_0^1 \left((g^*_\lambda)'(t)\right)^2 \mathrm{d}t &\leqslant \int_0^1 \left((g^*)'(t) + \lambda F^{-1}(1-t)\right)^2 \mathrm{d}t \\ &\leqslant 2\int_0^1 \left((g^*)'(t)\right)^2 \mathrm{d}t + 2\lambda^2 \int_0^1 \left(F^{-1}(t)\right)^2 \mathrm{d}t < \infty.\end{aligned}$$

Let us fix $\lambda_0 > 0$ and let $\varepsilon \in (0,1)$ be arbitrary. Define the concave continuous functions as:

$$f_\lambda(t) = g_\lambda^*(t)\mathbf{1}_{\{0\leqslant t\leqslant\varepsilon\}}(t) + \Big(g_\lambda^*(\varepsilon) + \frac{g_\lambda^*(1) - g_\lambda^*(\varepsilon)}{1-\varepsilon}(t-\varepsilon)\Big)\mathbf{1}_{\{\varepsilon<t\leqslant 1\}}(t),$$
$$k_\lambda(t) = r_\lambda(t)\mathbf{1}_{\{0\leqslant t\leqslant\varepsilon\}}(t) + \Big(r_\lambda(\varepsilon) + \frac{r_\lambda(1) - r_\lambda(\varepsilon)}{1-\varepsilon}(t-\varepsilon)\Big)\mathbf{1}_{\{\varepsilon<t\leqslant 1\}}(t),$$

where $r_\lambda(t) = g^*(t) + \lambda\int_{1-t}^1 F^{-1}(s)\,\mathrm{d}s$. Note that $f_\lambda \leqslant k_\lambda$, $f_\lambda(0) = k_\lambda(0) = 0$, $f_\lambda(1) = k_\lambda(1)$, and both are continuous concave functions on $[0,1]$.

Arguing as above, we obtain $f_\lambda'(1-V) \leqslant_{\mathrm{cx}} k_\lambda'(1-V)$, so

$$\int_0^1 \left(f_\lambda'(t)\right)^2 \mathrm{d}t \leqslant \int_0^1 \left(k_\lambda'(t)\right)^2 \mathrm{d}t.$$

Rewriting the left-hand side gives

$$\begin{aligned}\int_0^\varepsilon \left((g_\lambda^*)'(t)\right)^2 \mathrm{d}t \leqslant &\int_0^\varepsilon \left((g^*)'(t) + \lambda F^{-1}(1-t)\right)^2 \mathrm{d}t \\ &+ \frac{\left(r_\lambda(1) - r_\lambda(\varepsilon)\right)^2 - \left(g_\lambda^*(1) - g_\lambda^*(\varepsilon)\right)^2}{1-\varepsilon} \\ \leqslant &2\int_0^\varepsilon \left((g^*)'(t)\right)^2 \mathrm{d}t + 2\lambda^2\int_0^\varepsilon \left(F^{-1}(1-t)\right)^2 \mathrm{d}t \\ &+ \frac{\left|r_\lambda(1) - r_\lambda(\varepsilon) + g_\lambda^*(1) - g_\lambda^*(\varepsilon)\right| \cdot \left|r_\lambda(\varepsilon) - g_\lambda^*(\varepsilon)\right|}{1-\varepsilon}.\end{aligned}$$

Since $g_\lambda^*(1) = r_\lambda(1) = g_\lambda(1)$ and $g_\lambda(\varepsilon) \leqslant g_\lambda^*(\varepsilon) \leqslant r_\lambda(\varepsilon)$, we deduce that,

$$\begin{aligned}&\frac{\left|r_\lambda(1) - r_\lambda(\varepsilon) + g_\lambda^*(1) - g_\lambda^*(\varepsilon)\right| \cdot \left|r_\lambda(\varepsilon) - g_\lambda^*(\varepsilon)\right|}{1-\varepsilon} \\ \leqslant &\frac{2\left(|g_\lambda(1)| + |r_\lambda(\varepsilon)| + |g_\lambda(\varepsilon)|\right)|g^*(\varepsilon) - g(\varepsilon)|}{1-\varepsilon}.\end{aligned}$$

Hence, for any value of $\eta > 0$, there exists $\varepsilon_0 > 0$ such that for all $0 < \varepsilon < \varepsilon_0$,

$$\begin{aligned}&\sup_{0\leqslant\lambda\leqslant\lambda_0}\int_0^\varepsilon \left((g_\lambda^*)'(t)\right)^2 \mathrm{d}t \\ \leqslant &2\int_0^\varepsilon \left((g^*)'(t)\right)^2 \mathrm{d}t + 2(\lambda_0+1)^2\int_0^\varepsilon \left(F^{-1}(1-t)\right)^2 \mathrm{d}t + M|g^*(\varepsilon) - g(\varepsilon)| < \eta,\end{aligned}$$

for some constant M. A similar argument on the upper tail yields that, for sufficiently small values of ε,

$$\sup_{0\leqslant\lambda\leqslant\lambda_0}\int_{1-\varepsilon}^{1}\big((g_\lambda^*)'(t)\big)^2\,\mathrm{d}t<\eta.$$

Since each $(g_\lambda^*)'$ is monotone on $(0,1)$, we conclude that the family $\{((g_\lambda^*)'(t))^2 : 0\leqslant\lambda\leqslant\lambda_0\}$ is uniformly integrable for every instance where $\lambda_0>0$.

Using Hölder's inequality and the a.e. convergence established earlier, we obtain:

$$\begin{aligned}
&\big|\mathbb{E}\big(F^{-1}(V)(g_\lambda^*)'(1-V)\big)-\mathbb{E}\big(F^{-1}(V)(g_{\lambda_0}^*)'(1-V)\big)\big|\\
&\quad=\Big|\int_0^1 F^{-1}(t)\big((g_\lambda^*)'(1-t)-(g_{\lambda_0}^*)'(1-t)\big)\,\mathrm{d}t\Big|\\
&\quad\leqslant\Big(\int_0^1\big(F^{-1}(t)\big)^2\,\mathrm{d}t\Big)^{1/2}\Big(\int_0^1\big((g_\lambda^*)'(t)-(g_{\lambda_0}^*)'(t)\big)^2\,\mathrm{d}t\Big)^{1/2}\to 0,
\end{aligned}$$

as $\lambda\to\lambda_0$. Similarly,

$$\begin{aligned}
&\big|\mathbb{E}\big((g_\lambda^*)'(V)\big)^2-\mathbb{E}\big((g_{\lambda_0}^*)'(V)\big)^2\big|\\
&\quad=\Big|\int_0^1\big((g_\lambda^*)'(t)\big)^2-\big((g_{\lambda_0}^*)'(t)\big)^2\,\mathrm{d}t\Big|\\
&\quad\leqslant\Big(\int_0^1\big((g_\lambda^*)'(t)+(g_{\lambda_0}^*)'(t)\big)^2\,\mathrm{d}t\Big)^{1/2}\Big(\int_0^1\big((g_\lambda^*)'(t)-(g_{\lambda_0}^*)'(t)\big)^2\,\mathrm{d}t\Big)^{1/2}\to 0,
\end{aligned}$$

as $\lambda\to\lambda_0$. Combining these limits shows that the correlation $Corr(F^{-1}(V),(g_\lambda^*)'(1-V))$ is continuous in λ on $[0,\infty)$.

It remains to prove that $\lim_{\lambda\to\infty}Corr(F^{-1}(V),(g_\lambda^*)'(1-V))=1$. Put $l_\lambda=g_\lambda/\lambda$ for $\lambda>0$, so $l_\lambda^*=g_\lambda^*/\lambda$. Then,

$$Corr(F^{-1}(V),(g_\lambda^*)'(1-V))=Corr(F^{-1}(V),(l_\lambda^*)'(1-V))$$

and

$$Corr(F^{-1}(V),(l_\lambda^*)'(1-V))=\frac{\mathbb{E}(F^{-1}(V)(l_\lambda^*)'(1-V))-\mu_F\big(g(1)/\lambda+\mu_F\big)}{\sigma_F\sqrt{\mathbb{E}\big((l_\lambda^*)'(V)\big)^2-\big(g(1)/\lambda+\mu_F\big)^2}}.$$

Denote $l_\infty(t)=\int_{1-t}^1F^{-1}(s)\,\mathrm{d}s$. Then, $\sup_{t\in[0,1]}|l_\lambda(t)-l_\infty(t)|\leqslant\sup_{t\in[0,1]}|g(t)|/\lambda$ and, consequently, $\sup_{t\in[0,1]}|l_\lambda^*(t)-l_\infty^*(t)|\leqslant \sup_{t\in[0,1]}|g(t)|/\lambda$. Arguing as above yields $(l_\lambda^*)'(t)\to (l_\infty^*)'(t)$ for almost every value of t and the family $\{(l_\lambda^*)'(t):\lambda\geqslant 1\}$ is uniformly integrable. Hence,

$$\lim_{\lambda\to\infty}\mathbb{E}\big(F^{-1}(V)(l_\lambda^*)'(1-V)\big)=\mathbb{E}\big(F^{-1}(V)(l_\infty^*)'(1-V)\big)=\mathbb{E}\big((F^{-1}(V))^2\big),$$

and

$$\lim_{\lambda\to\infty}\mathbb{E}\big((l_\lambda^*)'(V)\big)^2=\mathbb{E}\big((l_\infty^*)'(V)\big)^2=\mathbb{E}\big((F^{-1}(V))^2\big).$$

Therefore, the correlation tends to 1 as $\lambda\to\infty$. □

Combining Theorem 5 and Remark 2 of Pesenti et al. (2024), we obtain the next lemma, which is used in the proof of Theorem 4.2.

Lemma 4.26. *For $g\in\mathscr{H}$,*

$$\sup_{G\in\mathscr{M}_\infty(\mu,\sigma)}\rho_g(G)=\rho_{\hat{g}}(H_0),$$

and if $g=\hat{g}$ then the supremum is uniquely attained at H_0.

Proof of Theorem 4.2. Observe first that for any $G\in\mathscr{M}_\varepsilon(\mu,\sigma)$ one has:

$$(\mu_F-\mu)^2+(\sigma_F-\sigma)^2\leqslant d_W^2(F,G)\leqslant\varepsilon.$$

If $d_W^2(F,G)=(\mu_F-\mu)^2+(\sigma_F-\sigma)^2$ then necessarily $G^{-1}(t)=\mu+\sigma\frac{F^{-1}(t)-\mu_F}{\sigma_F}$, whose law we denote by H_∞.

Now consider G with $(\mu_F-\mu)^2+(\sigma_F-\sigma)^2<d_W^2(F,G)\leqslant\varepsilon$. By Lemma 4.1, there exists a condition $\lambda\geqslant 0$ such that $d_W(F,H_\lambda)=d_W(F,G)$. This equality of Wasserstein distances is equivalent to:

$$\int_0^1 F^{-1}(t)h_\lambda(t)\,\mathrm{d}t=\int_0^1 F^{-1}(t)G^{-1}(t)\,\mathrm{d}t,$$

which in turn is equivalent to

$$\rho_{g_\lambda-g}(H_\lambda)=\rho_{g_\lambda-g}(G). \tag{4.24}$$

By Lemma 4.26, we know $\sup_{G\in\mathscr{M}_\infty(\mu,\sigma)}\rho_{g_\lambda}(G)=\rho_{g_\lambda}(H_\lambda)$ and H_λ is the unique maximiser. Therefore, for any value of $G\in\mathscr{M}_\infty(\mu,\sigma)$ with $G\neq H_\lambda$, we have $\rho_{g_\lambda}(G)<\rho_{g_\lambda}(H_\lambda)$, i.e.,

$$\rho_g(G)+\rho_{g_\lambda-g}(G)<\rho_g(H_\lambda)+\rho_{g_\lambda-g}(H_\lambda).$$

Using (4.24) this yields $\rho_g(G)<\rho_g(H_\lambda)$ whenever $d_W(F,H_\lambda)=d_W(F,G)$ and $G\neq H_\lambda$. Thus, the maximiser over the Wasserstein-sphere of given radius must be of the form H_λ for some value(s) of $\lambda\in(0,\infty]$.

If $d_W(F,H_{\lambda_1})<d_W(F,H_{\lambda_2})$, then one should check (using the same envelope arguments) that: $\rho_g(H_{\lambda_1})<\rho_g(H_{\lambda_2})$. Consequently, for a given radius, there is a unique $\lambda_\varepsilon>0$ with $d_W(F,H_{\lambda_\varepsilon})=\sqrt{\varepsilon}$ and H_{λ_ε} attains the maximum. This proves (i).

For (ii), suppose $\varepsilon\geqslant(\mu_F-\mu)^2+(\sigma_F-\sigma)^2+2\sigma_F\sigma(1-c_0)$ and $(g^*)'$ is not constant. By Lemma 4.26,

$$\sup_{G\in\mathscr{M}_\varepsilon(\mu,\sigma)}\rho_g(G)\leqslant\sup_{G\in\mathscr{M}_\infty(\mu,\sigma)}\rho_g(G)=\rho_g(H_0).$$

Since $H_0\in\mathscr{M}_\varepsilon(\mu,\sigma)$, the reverse inequality holds as well, hence $\sup_{G\in\mathscr{M}_\varepsilon(\mu,\sigma)}\rho_g(G)=\rho_g(H_0)$ and H_0 is the unique maximiser.

If $(g^*)'$ is constant then $c_0=0$ and $\varepsilon\geqslant(\mu_F-\mu)^2+(\sigma_F-\sigma)^2+2\sigma_F\sigma$, which implies that $\mathscr{M}_\varepsilon(\mu,\sigma)=\mathscr{M}_\infty(\mu,\sigma)$. In this case,

$$\sup_{G\in\mathscr{M}_\infty(\mu,\sigma)}\rho_g(G)\leqslant\sup_{G\in\mathscr{M}_\infty(\mu,\sigma)}\rho_{g^*}(G)=g(1)\mu.$$

For the reverse bound define, for $n\geqslant1$,

$$G_n=(1-1/n)\delta_\mu+\frac{1}{2n}\delta_{\mu-\sqrt{n}\sigma}+\frac{1}{2n}\delta_{\mu+\sqrt{n}\sigma},$$

which belongs to $\mathscr{M}_\infty(\mu,\sigma)$ and converges weakly to δ_μ. A direct calculation shows that:

$$\rho_g(G_n)=g(1)\mu+\big(g(1-1/(2n))-g(1)+g(1/(2n))\big)\sqrt{n}\sigma\to g(1)\mu,$$

as $n\to\infty$. Hence $\sup_{G\in\mathscr{M}_\infty(\mu,\sigma)}\rho_g(G)=g(1)\mu$. This completes the proof. $\square$

Proof of Corollary 4.3.

(i) For VaR_α^+ take $g(t)=\mathbf{1}_{[1-\alpha,1]}(t)$ and $g_\lambda(t)=\mathbf{1}_{[1-\alpha,1]}(t)+\lambda\int_{1-t}^1 F^{-1}(s)\,\mathrm{d}s$. Using the definition of $t_{1-\alpha,\lambda}$ one obtains:

$$
\begin{aligned}
g_\lambda^*(t) = {} & \lambda\int_{1-t}^1 F^{-1}(s)\,\mathrm{d}s\,\mathbf{1}_{[0,t_{1-\alpha,\lambda})}(t) \\
& + \Big(\frac{g_\lambda(1-\alpha)-g_\lambda(t_{1-\alpha,\lambda})}{1-\alpha-t_{1-\alpha,\lambda}}(t-t_{1-\alpha,\lambda})+g_\lambda(t_{1-\alpha,\lambda})\Big)\mathbf{1}_{[t_{1-\alpha,\lambda},1-\alpha]}(t) \\
& + \big(1+\lambda\int_{1-t}^1 F^{-1}(s)\,\mathrm{d}s\big)\mathbf{1}_{(1-\alpha,1]}(t).
\end{aligned}
$$

Differentiating yields, for $t\in(0,1)$,

$$
(g_\lambda^*)'(1-t) = \lambda F^{-1}(t)\,\mathbf{1}_{(0,\alpha]\cup(1-t_{1-\alpha,\lambda},1)}(t) + \frac{1+\lambda\int_\alpha^{1-t_{1-\alpha,\lambda}}F^{-1}(s)\,\mathrm{d}s}{1-\alpha-t_{1-\alpha,\lambda}}\mathbf{1}_{(\alpha,1-t_{1-\alpha,\lambda}]}(t).
$$

Hence,

$$
\mathrm{VaR}_\alpha^+(H_\lambda) = \mu+\sigma\frac{\frac{1+\lambda\int_\alpha^{1-t_{1-\alpha,\lambda}}F^{-1}(s)\,\mathrm{d}s}{1-\alpha-t_{1-\alpha,\lambda}}-a_\lambda}{b_\lambda}.
$$

(ii) For IQD_α^+ take $g(t)=\mathbf{1}_{[\alpha,1-\alpha]}(t)$ and $g_\lambda(t)=\mathbf{1}_{[\alpha,1-\alpha]}(t)+\lambda\int_{1-t}^1 F^{-1}(s)\,\mathrm{d}s$. By construction,

$$
\begin{aligned}
g_\lambda^*(t) = {} & \Big(\frac{g_\lambda(\alpha)-g_\lambda(t_{\alpha,\lambda})}{\alpha-t_{\alpha,\lambda}}(t-t_{\alpha,\lambda})+g_\lambda(t_{\alpha,\lambda})\Big)\mathbf{1}_{[t_{\alpha,\lambda},\alpha]}(t) \\
& + g_\lambda(t)\mathbf{1}_{(0,t_{\alpha,\lambda})\cup(\alpha,1-\alpha]\cup(\hat{t}_{\alpha,\lambda},1]}(t) \\
& + \Big(\frac{g_\lambda(\hat{t}_{\alpha,\lambda})-g_\lambda(1-\alpha)}{\hat{t}_{\alpha,\lambda}-1+\alpha}(t-\hat{t}_{\alpha,\lambda})+g_\lambda(\hat{t}_{\alpha,\lambda})\Big)\mathbf{1}_{(1-\alpha,\hat{t}_{\alpha,\lambda}]}(t).
\end{aligned}
$$

Thus for $t\in(0,1)$,

$$
\begin{aligned}
(g_\lambda^*)'(1-t) = {} & \frac{g_\lambda(\alpha)-g_\lambda(t_{\alpha,\lambda})}{\alpha-t_{\alpha,\lambda}}\mathbf{1}_{(1-\alpha,1-t_{\alpha,\lambda})}(t) + \frac{g_\lambda(\hat{t}_{\alpha,\lambda})-g_\lambda(1-\alpha)}{\hat{t}_{\alpha,\lambda}-1+\alpha}\mathbf{1}_{(1-\hat{t}_{\alpha,\lambda},\alpha)}(t) \\
& + \lambda F^{-1}(t)\,\mathbf{1}_{(0,1-\hat{t}_{\alpha,\lambda})\cup(\alpha,1-\alpha)\cup(1-t_{\alpha,\lambda},1)}(t).
\end{aligned}
$$

Theorem 4.2 then gives:

$$\mathrm{IQD}^+_\alpha(H_\lambda) = \frac{\frac{g_\lambda(\alpha)-g_\lambda(t_{\alpha,\lambda})}{\alpha-t_{\alpha,\lambda}} - \frac{g_\lambda(\hat{t}_{\alpha,\lambda})-g_\lambda(1-\alpha)}{\hat{t}_{\alpha,\lambda}-1+\alpha}}{b_\lambda}\,\sigma.$$

(iii) For $g(t) = \frac{t}{1-\alpha_1} \wedge 1 - \mathbf{1}_{(1-\alpha_2,1]}(t)$ one computes similarly that:

$$g^*_\lambda(t) = \Big(g_\lambda(1-\alpha_2) + \frac{g_\lambda(u_{\alpha_1,\alpha_2,\lambda}) - g_\lambda(1-\alpha_2)}{u_{\alpha_1,\alpha_2,\lambda} - 1 + \alpha_2}(t-1+\alpha_2)\Big)\mathbf{1}_{[1-\alpha_2,u_{\alpha_1,\alpha_2,\lambda}]}(t)$$
$$+ g_\lambda(t)\mathbf{1}_{(0,1-\alpha_2)\cup(u_{\alpha_1,\alpha_2,\lambda},1)}(t).$$

Hence, for $t \in (0,1)$,

$$(g^*_\lambda)'(1-t) = \Big(\frac{1}{1-\alpha_1}\mathbf{1}_{(\alpha_1,1)}(t) + \lambda F^{-1}(t)\Big)\mathbf{1}_{(0,1-u_{\alpha_1,\alpha_2,\lambda})\cup(\alpha_2,1)}(t) + c_{\alpha_1,\alpha_2,\lambda}\mathbf{1}_{(1-u_{\alpha_1,\alpha_2,\lambda},\alpha_2)}(t).$$

Applying Theorem 4.2 yields the claimed expression for the supremum. □

Proof of Theorem 4.4. By Theorem 4.2 it suffices to prove

$$\sup_{G\in\mathscr{M}_\varepsilon(\mu,\sigma)} \rho_g(G) = \sup_{G\in\mathscr{M}_\varepsilon(\mu,\sigma)} \rho_{\hat{g}}(G).$$

Since $g \leqslant \hat{g}$, we trivially have

$$\sup_{G\in\mathscr{M}_\varepsilon(\mu,\sigma)} \rho_g(G) \leqslant \sup_{G\in\mathscr{M}_\varepsilon(\mu,\sigma)} \rho_{\hat{g}}(G).$$

We now show the reverse inequality.

First assume that $g(t) \neq \hat{g}(t)$ only at finitely many points $\{t_1,\dots,t_m\}$ with $t_1 < \cdots < t_m$. For each $n \geqslant 1$ and $t \in (0,1)$ define the quantile

$$G_n^{-1}(t) = G^{-1}(t)\mathbf{1}_{\left((0,1)\setminus\bigcup_{i\in\mathscr{D}_1}(1-t_i-1/n^2,1-t_i+1/n)\right)\setminus\bigcup_{i\in\mathscr{D}_2}(1-t_i-1/n,1-t_i+1/n^2)}(t)$$
$$+ \sum_{i\in\mathscr{D}_1} \frac{n^2}{n+1}\int_{1-t_i-1/n^2}^{1-t_i+1/n} G^{-1}(s)\,\mathrm{d}s\,\mathbf{1}_{(1-t_i-1/n^2,1-t_i+1/n)}(t)$$
$$+ \sum_{i\in\mathscr{D}_2} \frac{n^2}{n+1}\int_{1-t_i-1/n}^{1-t_i+1/n^2} G^{-1}(s)\,\mathrm{d}s\,\mathbf{1}_{(1-t_i-1/n,1-t_i+1/n^2)}(t),$$

where $\mathscr{D}_1 = \{i : \hat{g}(t_i) = \lim_{s\downarrow t_i} g(s)\}$ and $\mathscr{D}_2 = \{i : \hat{g}(t_i) = \lim_{s\uparrow t_i} g(s)\} \setminus \mathscr{D}_1$. For sufficiently large values of n, the small intervals appearing above are disjointed. Let σ_n be the standard deviation of G_n and set

$$\widehat{G}_n^{-1}(t) = \mu + \frac{G_n^{-1}(t) - \mu}{\sigma_n}\sigma. \tag{4.25}$$

One checks that $\lim_{n\to\infty} d_W(\widehat{G}_n, F) \leqslant \sqrt{\varepsilon}$, hence for every $\eta > 0$ there exists n_0 with $\widehat{G}_n \in \mathscr{M}_{\varepsilon+\eta}(\mu,\sigma)$ for all $n \geqslant n_0$.

A direct decomposition yields

$$\begin{aligned}\rho_{\hat{g}}(G) = \rho_{\hat{g}}(G_n) &+ \sum_{i\in\mathscr{D}_1}\Big(\int_{G^{-1}(1-t_i-1/n^2)}^{G^{-1}(1-t_i+1/n)} \hat{g}(1-G(x))\,\mathrm{d}x - \int_{G^{-1}(1-t_i-1/n^2)}^{G^{-1}(1-t_i+1/n)} \hat{g}(1-G_n(x))\,\mathrm{d}x\Big)\\ &+ \sum_{i\in\mathscr{D}_2}\Big(\int_{G^{-1}(1-t_i-1/n)}^{G^{-1}(1-t_i+1/n^2)} \hat{g}(1-G(x))\,\mathrm{d}x - \int_{G^{-1}(1-t_i-1/n)}^{G^{-1}(1-t_i+1/n^2)} \hat{g}(1-G_n(x))\,\mathrm{d}x\Big).\end{aligned}$$

Elementary estimates and the definitions of the G_n show that, as $n \to \infty$, the correction terms vanish and, therefore, $\rho_{\hat{g}}(G) = \lim_{n\to\infty} \rho_{\hat{g}}(G_n)$.

Using (4.25) together with the homogeneity properties of ρ_g one obtains

$$\rho_{\hat{g}}(G_n) = \frac{\rho_{\hat{g}}(\widehat{G}_n) - \mu g(1)}{\sigma}\sigma_n + \mu g(1)$$

and since $\rho_{\hat{g}}(\widehat{G}_n) = \rho_g(\widehat{G}_n)$ we derive

$$\rho_{\hat{g}}(G) = \lim_{n\to\infty} \rho_{\hat{g}}(G_n) = \lim_{n\to\infty} \frac{\rho_g(\widehat{G}_n) - \mu g(1)}{\sigma}\sigma_n + \mu g(1) = \lim_{n\to\infty} \rho_g(\widehat{G}_n) \leqslant \sup_{H\in\mathscr{M}_{\varepsilon+\eta}(\mu,\sigma)} \rho_g(H). \tag{4.26}$$

Thus, $\rho_{\hat{g}}(G) \leqslant \sup_{H\in\mathscr{M}_{\varepsilon+\eta}(\mu,\sigma)} \rho_g(H)$ for every $\eta > 0$.

Next, treat the case where $g \neq \hat{g}$ on infinitely many points $\{t_i : i \geqslant 1\}$. Define approximations

$$g_m(t) = \hat{g}(t)\mathbf{1}_{\{t_1,\dots,t_m\}}(t) + g(t)\mathbf{1}_{[0,1]\setminus\{t_1,\dots,t_m\}}(t).$$

Each value of g_m differs from g only on a finite set and obeys the previous argument, so for every $G \in \mathscr{M}_\varepsilon(\mu,\sigma)$ and every $\eta > 0$ we have:

$$\rho_{g_m}(G) \leqslant \sup_{H \in \mathscr{M}_{\varepsilon+\eta}(\mu,\sigma)} \rho_g(H).$$

Since $g_m \uparrow \hat{g}$ pointwise and $\rho_{g_m}(G) > -\infty$ for each point, m, the monotone convergence theorem yields $\rho_{g_m}(G) \uparrow \rho_{\hat{g}}(G)$. Hence,

$$\rho_{\hat{g}}(G) \leqslant \sup_{H \in \mathscr{M}_{\varepsilon+\eta}(\mu,\sigma)} \rho_g(H).$$

If $\rho_{\hat{g}}(G) = -\infty$, the inequality is trivial. If $\rho_{\hat{g}}(G) > -\infty$ but $\rho_g(G) = -\infty$, one may truncate G from below, renormalise, and argue as above to obtain the same bound in the limit. Combining the cases we obtain, for every instance where $\eta > 0$,

$$\sup_{G \in \mathscr{M}_\varepsilon(\mu,\sigma)} \rho_{\hat{g}}(G) \leqslant \sup_{G \in \mathscr{M}_{\varepsilon+\eta}(\mu,\sigma)} \rho_g(G).$$

Hence,

$$\sup_{G \in \mathscr{M}_\varepsilon(\mu,\sigma)} \rho_g(G) \leqslant \sup_{G \in \mathscr{M}_\varepsilon(\mu,\sigma)} \rho_{\hat{g}}(G) \leqslant \lim_{\eta \downarrow 0} \sup_{G \in \mathscr{M}_{\varepsilon+\eta}(\mu,\sigma)} \rho_g(G).$$

Let $l(\varepsilon) = \sup_{G \in \mathscr{M}_\varepsilon(\mu,\sigma)} \rho_{\hat{g}}(G)$. Then l is non-decreasing and

$$\lim_{\eta \downarrow 0} l(\varepsilon - \eta) \leqslant \sup_{G \in \mathscr{M}_\varepsilon(\mu,\sigma)} \rho_g(G) \leqslant l(\varepsilon).$$

The assumed continuity of $\rho_{\hat{g}}(H_\lambda)$ in λ on the relevant range of radii implies continuity of $l(\varepsilon)$ for $(\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 < \varepsilon < (\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 + 2\sigma_F\sigma(1 - c_0)$, and, therefore, equality holds there:

$$\sup_{G \in \mathscr{M}_\varepsilon(\mu,\sigma)} \rho_g(G) = \sup_{G \in \mathscr{M}_\varepsilon(\mu,\sigma)} \rho_{\hat{g}}(G).$$

Applying Theorem 4.2 yields the conclusion of part (i).

For $\varepsilon > (\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 + 2\sigma_F\sigma(1 - c_0)$ the function $l(\varepsilon)$ is constant, so the same identity holds; the boundary case $\varepsilon = (\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 + 2\sigma_F\sigma(1 - c_0)$ follows from Theorem 5 and Remark 2 of Pesenti et al. (2024). Combining these observations with Theorem 4.2 gives part (ii). □

Proof of Corollary 4.5.

(i) From Corollary 4.3 we have the explicit formula:

$$\mathrm{VaR}_{\alpha}^{+}(H_{\lambda}) = \mu + \sigma \frac{\frac{1+\lambda\int_{\alpha}^{1-t_{1-\alpha,\lambda}} F^{-1}(s)\,\mathrm{d}s}{1-\alpha-t_{1-\alpha,\lambda}} - a_{\lambda}}{b_{\lambda}},$$

where $a_{\lambda} = g(1)+\lambda\mu_F$ and $b_{\lambda} = \sqrt{\int_0^1 \left((g_{\lambda}^{*})'(t)\right)^2 \mathrm{d}t - (g(1)+\lambda\mu_F)^2}$. The continuity of a_{λ} in λ is immediate. The continuity of b_{λ} follows from the uniform integrability of the family $\{((g_{\lambda}^{*})'(t))^2 : 0 \leqslant \lambda \leqslant \lambda_0\}$ for any fixed $\lambda_0 > 0$ and the a.e. convergence $(g_{\lambda}^{*})'(t) \to (g_{\lambda_0}^{*})'(t)$ as $\lambda \to \lambda_0$ established in the proof of Lemma 4.1. Using the explicit form of $(g_{\lambda}^{*})'(t)$ in Corollary 4.3(i), one also checks that the ratio

$$\frac{1+\lambda\int_{\alpha}^{1-t_{1-\alpha,\lambda}} F^{-1}(s)\,\mathrm{d}s}{1-\alpha-t_{1-\alpha,\lambda}}$$

depends continuously on λ. Hence, $\mathrm{VaR}_{\alpha}^{+}(H_{\lambda})$ is continuous in $\lambda \in (0,\infty)$, and Theorem 4.4 applies to give the claimed result.

(ii) The expression for $\mathrm{IQD}_{\alpha}^{+}(H_{\lambda})$ is given in Corollary 4.3(ii). The continuity of b_{λ} was discussed in (i), and the continuity of the numerators $\frac{g_{\lambda}(\alpha)-g_{\lambda}(t_{\alpha,\lambda})}{\alpha-t_{\alpha,\lambda}}$ and $\frac{g_{\lambda}(\hat{t}_{\alpha,\lambda})-g_{\lambda}(1-\alpha)}{\hat{t}_{\alpha,\lambda}-1+\alpha}$ follows from the a.e. convergence $(g_{\lambda}^{*})'(t) \to (g_{\lambda_0}^{*})'(t)$ and the uniform integrability arguments used in Lemma 4.1. Therefore, $\mathrm{IQD}_{\alpha}^{+}(H_{\lambda})$ is continuous in λ on $(0,\infty)$ and Theorem 4.4 yields the conclusion.

(iii) For $g = \hat{g}_{\beta,\alpha}^{h_1,h_2}$ note that, under the condition $\frac{h_1}{1-\beta} \geqslant \frac{h_2-h_1}{\beta-\alpha}$, the function $g_{\beta,\alpha}^{h_1,h_2} \wedge h_2$ is concave on $(0,1)$. One then computes, as in Corollary 4.3, that:

$$g_{\lambda}^{*}(t) = \left(g_{\lambda}(u_{\alpha,\beta,\lambda}^{h_1,h_2}) + \frac{g_{\lambda}(1-\alpha)-g_{\lambda}(u_{\alpha,\beta,\lambda}^{h_1,h_2})}{1-\alpha-u_{\alpha,\beta,\lambda}^{h_1,h_2}}(t-u_{\alpha,\beta,\lambda}^{h_1,h_2})\right)\mathbf{1}_{[u_{\alpha,\beta,\lambda}^{h_1,h_2},1-\alpha]}(t) + g_{\lambda}(t)\mathbf{1}_{(0,u_{\alpha,\beta,\lambda}^{h_1,h_2})\cup(1-\alpha,1)}(t).$$

Differentiating gives the formula for $(g_{\lambda}^{*})'(1-t)$ appearing in Corollary 4.5(iii). Using the decomposition $\rho_{\hat{g}_{\beta,\alpha}^{h_1,h_2}} = w_1\mathrm{ES}_{\alpha} + w_2\mathrm{ES}_{\beta} + w_3\mathrm{VaR}_{\alpha}^{+}$ with $w_1, w_2, w_3 \geqslant 0$, $w_1+w_2+w_3 = 1$, together with the uniform integrability and a.e. convergence arguments used above, one obtains the continuity in λ of all terms in the expression for the supremum. Then Theorem 4.4 yields the asserted identity. $\square$

Proof of Proposition 4.6. If g is concave, then $g_\lambda^* = g_\lambda = g + \lambda \int_{1-\cdot}^1 F^{-1}(s)\,\mathrm{d}s$, hence,

$$(g_\lambda^*)'(t) = g'(t) + \lambda F^{-1}(1-t), \qquad h_\lambda(t) = \mu + \sigma \frac{g'(1-t) + \lambda F^{-1}(t) - a_\lambda}{b_\lambda},$$

with $a_\lambda = g(1) + \lambda \mu_F$ and

$$b_\lambda = \sqrt{\int_0^1 \left((g_\lambda^*)'(t)\right)^2 \mathrm{d}t - (g(1)+\lambda\mu_F)^2}.$$

The equation $d_W(F, H_\lambda) = \sqrt{\varepsilon}$ is equivalent to:

$$\varepsilon = \mu_F^2 + \sigma_F^2 + \mu^2 + \sigma^2 - 2Cov(F^{-1}(V), h_\lambda(V)) - 2\mu\mu_F,$$

hence,

$$Cov(F^{-1}(V), h_\lambda(V)) = C_{\varepsilon,F} := \frac{\mu_F^2 + \sigma_F^2 + \mu^2 + \sigma^2 - 2\mu\mu_F - \varepsilon}{2} \geqslant 0.$$

Computing the covariance using the expression of h_λ gives:

$$Cov(F^{-1}(V), h_\lambda(V)) = \frac{\sigma}{b_\lambda}\left(C_{g,F} + \lambda\sigma_F^2\right),$$

where $C_{g,F} = Cov(F^{-1}(V), g'(1-V))$. Thus,

$$b_\lambda = \frac{\sigma\left(C_{g,F} + \lambda\sigma_F^2\right)}{C_{\varepsilon,F}}.$$

On the other hand, by definition $b_\lambda^2 = V_g + 2\lambda C_{g,F} + \lambda^2\sigma_F^2$, with $V_g = \mathrm{VaR}(g'(1-V))$. Equating the two expressions for b_λ^2 and simplifying yields the quadratic equation in λ

$$\sigma_F^2\lambda^2 + 2C_{g,F}\lambda + \frac{V_g C_{\varepsilon,F}^2 - \sigma^2 C_{g,F}^2}{C_{\varepsilon,F}^2 - \sigma^2\sigma_F^2} = 0.$$

Solving this quadratic produces the claimed formula

$$\lambda_\varepsilon = \frac{-C_{g,F} + \sqrt{C_{g,F}^2 - \sigma_F^2 \frac{V_g C_{\varepsilon,F}^2 - \sigma^2 C_{g,F}^2}{C_{\varepsilon,F}^2 - \sigma^2 \sigma_F^2}}}{\sigma_F^2}.$$

$\square$

4.7.2 Proofs of results in Section 4.4

In this section, we give the proofs of all statements from Section 4.4. Denote by $\langle \cdot, \cdot \rangle$ the inner product on the space $\mathscr{F}_{U,\xi}^{-1}$.

Proof of Proposition 4.8. By construction $\gamma_\xi^\uparrow \in \mathscr{F}_{U,\xi}^{-1}$. Since $\gamma_\xi^\uparrow$ is non-constant, the function

$$h_\xi^\uparrow(u) = \mu + \sigma \frac{\gamma_\xi^\uparrow(u) - \hat{a}_\xi}{\hat{b}_\xi}$$

belongs to $\mathscr{F}_{U,\xi}^{-1}(\mu, \sigma)$. Take an arbitrary $h \in \mathscr{F}_{U,\xi}^{-1}(\mu, \sigma)$ and set

$$k(u) = \hat{b}_\xi \frac{h(u) - \mu}{\sigma} + \hat{a}_\xi \in \mathscr{F}_{U,\xi}^{-1}.$$

Note that $\|\gamma_\xi^\uparrow\|_2 = \|k\|_2$. The claim of the proposition follows from the chain of equivalent inequalities

$$\begin{aligned} \|\gamma - \gamma_\xi^\uparrow\|_2 \leqslant \|\gamma - k\|_2 &\iff \langle \gamma, \gamma_\xi^\uparrow \rangle \geqslant \langle \gamma, k \rangle \\ &\iff \langle \gamma, h_\xi^\uparrow \rangle \geqslant \langle \gamma, h \rangle \\ &\iff \rho_g(h_\xi^\uparrow) \geqslant \rho_g(h). \end{aligned}$$

Strict inequality holds in the above relations unless $\gamma_\xi^\uparrow = k$, which yields the uniqueness of the maximiser.

Using the above identity we obtain:

$$\sup_{G^{-1}\in\mathscr{F}^{-1}_{U,\xi}(\mu,\sigma)} \rho_g(G^{-1}) = \langle \gamma, \mu + \sigma \frac{\gamma^{\uparrow}_{\xi} - \hat{a}_{\xi}}{\hat{b}_{\xi}} \rangle$$
$$= \mu g(1) + \frac{\sigma}{\hat{b}_{\xi}} \langle \gamma, \gamma^{\uparrow}_{\xi} - \hat{a}_{\xi} \rangle.$$

By Corollary 2.3 in Brunk (1965) we have the orthogonality relations $\langle \gamma - \gamma^{\uparrow}_{\xi}, \gamma^{\uparrow}_{\xi} \rangle = 0$ and $\langle \gamma - \gamma^{\uparrow}_{\xi}, c \rangle = 0$ for any constant $c \in \mathbb{R}$. Hence,

$$\langle \gamma, \gamma^{\uparrow}_{\xi} - \hat{a}_{\xi} \rangle = \langle \gamma^{\uparrow}_{\xi}, \gamma^{\uparrow}_{\xi} - \hat{a}_{\xi} \rangle = \langle \gamma^{\uparrow}_{\xi} - \hat{a}_{\xi}, \gamma^{\uparrow}_{\xi} - \hat{a}_{\xi} \rangle = \hat{b}^2_{\xi},$$

and, therefore,

$$\sup_{G^{-1}\in\mathscr{F}^{-1}_{U,\xi}(\mu,\sigma)} \rho_g(G^{-1}) = \mu g(1) + \sigma \hat{b}_{\xi},$$

as claimed. □

Proof of Proposition 4.9. Let

$$\gamma(u) = \sum_{i=1}^{n} y_i \mathbf{1}_{(x_{i-1},x_i)}(u), \qquad 0 = x_0 < x_1 < \cdots < x_n = 1,\ y_i \in \mathbb{R}.$$

Without loss of generality, assume $\xi \in (x_{i_0-1}, x_{i_0})$. Fix one interval in the collection: $\{(x_{i-1},x_i) : i \neq i_0\} \cup \{(x_{i_0-1},\xi),(\xi,x_{i_0})\}$ and denote it by (a,b), and write the constant value of γ on (a,b) by y. We treat the case $b \leqslant \xi$ (so $\gamma^{\uparrow}_{\xi}$ must be concave on (a,b)); the complementary case $a \geqslant \xi$ is analogous (convex on (a,b)).

(1) If $\gamma^{\uparrow}_{\xi}(a) \geqslant y$, define on (a,b) the affine function

$$\gamma_1(u) = \frac{\gamma^{\uparrow}_{\xi}(b) - \gamma^{\uparrow}_{\xi}(a)}{b-a}(u-a) + \gamma^{\uparrow}_{\xi}(a),$$

and set $\gamma_1 = \gamma^{\uparrow}_{\xi}$ outside (a,b). Then $\gamma_1 \in \mathscr{F}^{-1}_{U,\xi}$ and $\|\gamma_1 - \gamma\|_2 \leqslant \|\gamma^{\uparrow}_{\xi} - \gamma\|_2$. By the uniqueness of the projection, we must have $\gamma^{\uparrow}_{\xi} = \gamma_1$, hence $\gamma^{\uparrow}_{\xi}$ is linear on (a,b).

(2) If $\gamma_\xi^\uparrow(a) < y < \gamma_\xi^\uparrow(b)$, there exists $c \in (a,b)$ with $\gamma_\xi^\uparrow(c) = y$. Define for $u \in (a,b)$

$$\gamma_2(u) = \Big(\frac{\gamma_\xi^\uparrow(b) - \gamma_\xi^\uparrow(c)}{b-c}(u-b) + \gamma_\xi^\uparrow(b)\Big) \wedge \Big((\gamma_\xi^\uparrow)'_+(a)(u-a) + \gamma_\xi^\uparrow(a)\Big),$$

and set $\gamma_2 = \gamma_\xi^\uparrow$ outside (a,b). Again $\gamma_2 \in \mathscr{F}_{U,\xi}^{-1}$ and $\|\gamma_2 - \gamma\|_2 \leqslant \|\gamma_\xi^\uparrow - \gamma\|_2$, so uniqueness yields $\gamma_\xi^\uparrow = \gamma_2$, which is linear on (a,b).

(3) If $\gamma_\xi^\uparrow(b) \leqslant y$, put for $u \in (a,b)$

$$\gamma_3(u) = \Big((\gamma_\xi^\uparrow)'_+(a)(u-a) + \gamma_\xi^\uparrow(a)\Big) \wedge \Big((\gamma_\xi^\uparrow)'_-(b)(u-b) + \gamma_\xi^\uparrow(b)\Big),$$

and $\gamma_3 = \gamma_\xi^\uparrow$ off (a,b). Again $\gamma_3 \in \mathscr{F}_{U,\xi}^{-1}$ and $\|\gamma_3 - \gamma\|_2 \leqslant \|\gamma_\xi^\uparrow - \gamma\|_2$, so uniqueness forces $\gamma_\xi^\uparrow = \gamma_3$, hence linearity on (a,b).

The preceding arguments show that $\gamma_\xi^\uparrow$ is piecewise linear on each step interval (or its subdivision when the interval contains ξ). If $\xi = x_{i_0}$ for some i_0, the representation (4.15) follows. By Corollary 2.3 in Brunk (1965), we have the moment constraint $\langle \gamma - \gamma_\xi^\uparrow, 1\rangle = 0$, which yields (4.16).

A direct calculation (carried out by integrating the quadratic deviations piecewise) yields:

$$\begin{aligned}\int_0^1 (\gamma(u) - \gamma_\xi^\uparrow(u))^2\,\mathrm{d}u &= \sum_{i=1}^n \Big(\frac{(e_i^+)^3 - (e_i^-)^3}{3c_i^-} + \frac{(e_i^+ + c_i^+(x_i - a_i))^3 - (e_i^+)^3}{3c_i^+}\Big)\\ &= \sum_{i=1}^n \Big(\frac{a_i - x_{i-1}}{3}\big((e_i^+)^2 + e_i^+e_i^- + (e_i^-)^2\big)\\ &\quad + \frac{x_i - a_i}{3}\big((e_i^+ + c_i^+(x_i - a_i))^2 + (e_i^+ + c_i^+(x_i - a_i))e_i^+ + (e_i^+)^2\big)\Big),\end{aligned}$$

where

$$e_i^- = g(1) - \sum_{j=i}^n \big(c_j^-(a_j - x_{j-1}) + c_j^+(x_j - a_j)\big) - y_i, \qquad e_i^+ = e_i^- + c_i^-(a_i - x_{i-1}),$$

for $i = 1,\dots,n$. Therefore, the optimal parameters $(\mathbf{a},\mathbf{c})$ are obtained by minimising the displayed expression over the domain $\mathscr{D}_n$, which yields (4.17). This completes the proof. □

Proof of Proposition 4.10. The L^2-projection operator is contractive (Theorem 2.3 in Brunk

(1965)), hence

$$\|\gamma^{\uparrow}_{\xi,n} - \gamma^{\uparrow}_{\xi}\|_2 \leqslant \|\gamma_n - \gamma\|_2.$$

Write

$$\hat{b}^2_{\xi} - \hat{b}^2_{\xi,n} = \langle \gamma^{\uparrow}_{\xi}, \gamma^{\uparrow}_{\xi}\rangle - \langle \gamma^{\uparrow}_{\xi,n}, \gamma^{\uparrow}_{\xi,n}\rangle - (\hat{a}^2_{\xi} - \hat{a}^2_{\xi,n}).$$

Using Corollary 2.3 of Brunk (1965) we obtain:

$$\big|\langle \gamma^{\uparrow}_{\xi}, \gamma^{\uparrow}_{\xi}\rangle - \langle \gamma^{\uparrow}_{\xi,n}, \gamma^{\uparrow}_{\xi,n}\rangle\big| = |\langle \gamma, \gamma^{\uparrow}_{\xi}\rangle - \langle \gamma, \gamma^{\uparrow}_{\xi,n}\rangle| \leqslant \|\gamma\|_2\|\gamma_n - \gamma\|_2.$$

A direct estimate yields

$$|\hat{a}^2_{\xi} - \hat{a}^2_{\xi,n}| \leqslant (\|\gamma\|_2 + \|\gamma_n\|_2)\|\gamma_n - \gamma\|_2.$$

Hence,

$$|\hat{b}_{\xi} - \hat{b}_{\xi,n}| = \frac{|\hat{b}^2_{\xi} - \hat{b}^2_{\xi,n}|}{\hat{b}_{\xi} + \hat{b}_{\xi,n}} \leqslant \frac{(2\|\gamma\|_2 + \|\gamma_n\|_2)\|\gamma_n - \gamma\|_2}{\hat{b}_{\xi}}.$$

Using these bounds, we obtain:

$$\begin{aligned}
\|h^{\uparrow}_{\xi,n} - h^{\uparrow}_{\xi}\|_2 &= \sigma\Big\|\frac{\gamma^{\uparrow}_{\xi,n} - \hat{a}_{\xi,n}}{\hat{b}_{\xi,n}} - \frac{\gamma^{\uparrow}_{\xi} - \hat{a}_{\xi}}{\hat{b}_{\xi}}\Big\|_2 \\
&\leqslant \sigma\frac{\|\gamma^{\uparrow}_{\xi,n} - \gamma^{\uparrow}_{\xi}\|_2 + |\hat{a}_{\xi,n} - \hat{a}_{\xi}|}{\hat{b}_{\xi}} + \sigma\frac{\hat{b}_{\xi,n}}{\hat{b}_{\xi,n}\hat{b}_{\xi}}|\hat{b}_{\xi,n} - \hat{b}_{\xi}| \\
&\leqslant \Big(2 + \frac{2\|\gamma\|_2 + \|\gamma_n\|_2}{\hat{b}_{\xi}}\Big)\frac{\sigma}{\hat{b}_{\xi}}\|\gamma_n - \gamma\|_2,
\end{aligned}$$

and, consequently,

$$|\rho_g(h^{\uparrow}_{\xi}) - \rho_g(h^{\uparrow}_{\xi,n})| = \sigma|\hat{b}_{\xi} - \hat{b}_{\xi,n}| \leqslant \frac{\sigma(2\|\gamma\|_2 + \|\gamma_n\|_2)\|\gamma_n - \gamma\|_2}{\hat{b}_{\xi}}.$$

This proves the proposition. □

Proof of Example 4.11. Part (i) is immediate. For (ii) assume $\xi = 1/2$. Then any candidate projection takes the form:

$$\gamma_\xi^{\uparrow}(u) = \big(c_1(u-1/2)+a-1\big)\mathbf{1}_{(0,1/2]}(u) + \big(c_2(u-1/2)+a-1\big)\mathbf{1}_{(1/2,1)}(u),$$

with parameters $a \in [0,2]$, $c_1 \geqslant 2a$ and $c_2 \geqslant 2(2-a)$. A direct computation gives:

$$\begin{aligned}\int_0^1 \big(\gamma_\xi^{\uparrow}(u)-\gamma(u)\big)^2\,\mathrm{d}u &= c_1^2\Big(\int_0^{a/c_1} u^2\,\mathrm{d}u + \int_0^{1/2-a/c_1} u^2\,\mathrm{d}u\Big)\\ &\quad + c_2^2\Big(\int_0^{(2-a)/c_2} u^2\,\mathrm{d}u + \int_0^{1/2-(2-a)/c_2} u^2\,\mathrm{d}u\Big)\\ &=: G(c_1,c_2,a).\end{aligned}$$

To minimise expressions of the form

$$f(c) = c^2\Big(\Big(\frac{a}{c}\Big)^3 + \Big(b-\frac{a}{c}\Big)^3\Big),$$

with $a>0$, $b>0$ and $c \geqslant a/b$, one observes that: $\frac{f(c)}{a^3} = \frac{1+(\frac{b}{a}c-1)^3}{c}$ and differentiating shows the unique minimiser is $c = \frac{3}{2}\frac{a}{b}$. Applying this to both terms in G (with appropriate identifications) yields that, for a fixed a, the minimum is attained at $c_1 = 3a$ and $c_2 = 3(2-a)$, with $G(3a,3(2-a),a) = \frac{1}{8}(a^2+(2-a)^2)$, and this is minimised at $a=1$. Thus,

$$\gamma_\xi^{\uparrow}(u) = 3\big(u-\tfrac{1}{2}\big), \qquad u \in (0,1),$$

as stated. □

Proof of Lemma 4.12. For any $G^{-1} \in \mathscr{F}_{U,\xi}^{-1}(\mu,\sigma)$ a straightforward computation yields:

$$d_W^2(F^{-1},G^{-1}) = (\mu-\mu_F)^2 + (\sigma-\sigma_F)^2 + 2\sigma\sigma_F + 2\mu\mu_F - 2\langle F^{-1},G^{-1}\rangle.$$

If $G^{-1} \in \mathscr{F}_{U,\xi}^{-1}(\mu_F,\sigma_F^{\uparrow})$ then by Theorem 2.2 of Brunk (1965) we have

$$\langle F^{-1} - F_\xi^{-1,\uparrow}, F_\xi^{-1,\uparrow} - G^{-1}\rangle \geqslant 0,$$

which implies

$$\langle F^{-1}, G^{-1}\rangle \leqslant \langle F^{-1}, F_\xi^{-1,\uparrow}\rangle + \langle F_\xi^{-1,\uparrow}, G^{-1}\rangle - \langle F_\xi^{-1,\uparrow}, F_\xi^{-1,\uparrow}\rangle.$$

By Cauchy–Schwarz, we can also have $\langle F_\xi^{-1,\uparrow}, G^{-1}\rangle \leqslant \langle F_\xi^{-1,\uparrow}, F_\xi^{-1,\uparrow}\rangle$.
Therefore, for $G^{-1} \in \mathscr{F}_{U,\xi}^{-1}(\mu_F, \sigma_F^{\uparrow})$,

$$\langle F^{-1}, G^{-1}\rangle \leqslant \langle F^{-1}, F_\xi^{-1,\uparrow}\rangle.$$

Consequently, for any value of $G^{-1} \in \mathscr{F}_{U,\xi}^{-1}(\mu, \sigma)$,

$$\begin{aligned} d_W^2(F^{-1}, G^{-1}) &\geqslant (\mu-\mu_F)^2 + (\sigma-\sigma_F)^2 + 2\sigma\sigma_F + 2\mu\mu_F - 2\Big\langle F^{-1}, \frac{F_\xi^{-1,\uparrow}-\mu_F}{\sigma_F^{\uparrow}}\sigma + \mu\Big\rangle \\ &= (\mu-\mu_F)^2 + (\sigma-\sigma_F)^2 + 2\sigma\sigma_F\big(1 - Corr(F^{-1}(V), F_\xi^{-1,\uparrow}(V))\big). \end{aligned} \tag{4.27}$$

Let $\hat{c}_0 = Corr(F^{-1}(V), F_\xi^{-1,\uparrow}(V))$. The bound (4.27) shows that if:

$$\varepsilon < (\mu_F-\mu)^2 + (\sigma_F-\sigma)^2 + 2\sigma_F\sigma(1-\hat{c}_0),$$

then $\mathscr{F}_{U,\xi}^{-1}(\mu, \sigma, \varepsilon) = \varnothing$. If equality holds, then the lower bound in (4.27) is attained and one obtains the singleton

$$\mathscr{F}_{U,\xi}^{-1}(\mu, \sigma) = \Big\{\frac{F_\xi^{-1,\uparrow}-\mu_F}{\sigma_F^{\uparrow}}\sigma + \mu\Big\}.$$

This proves the lemma. □

Proof of Lemma 4.13. Theorem 2.3 of Brunk (1965) gives the L^2-continuity estimate

$$\|k_{\lambda_2,\xi}^{\uparrow} - k_{\lambda_1,\xi}^{\uparrow}\|_2^2 \leqslant |\lambda_2-\lambda_1|^2\big(\mu_F^2 + \sigma_F^2\big),$$

which immediately implies continuity of the map $\lambda \mapsto Corr(F^{-1}(V), k_{\lambda,\xi}^{\uparrow}(V))$ on $[0,\infty)$. Let

$$\hat{k}_{\lambda,\xi} = \lambda\gamma + F^{-1}, \qquad \hat{k}_{\lambda,\xi}^{\uparrow} = L^2\text{-projection of } \hat{k}_{\lambda,\xi} \text{ on } \mathscr{F}_{U,\xi}^{-1}.$$

Then $\hat{k}^{\uparrow}_{0,\xi} = F_{\xi}^{-1,\uparrow}$ and, again by Theorem 2.3 of Brunk (1965),

$$\|\hat{k}^{\uparrow}_{\lambda,\xi} - F_{\xi}^{-1,\uparrow}\|_2^2 \leqslant \lambda^2 \|\gamma\|_2^2.$$

Hence $\hat{k}^{\uparrow}_{\lambda,\xi} \to F_{\xi}^{-1,\uparrow}$ in L^2 as $\lambda \downarrow 0$, and, therefore,

$$\lim_{\lambda \downarrow 0} Corr(F^{-1}(V), \hat{k}^{\uparrow}_{\lambda,\xi}(V)) = Corr(F^{-1}(V), F_{\xi}^{-1,\uparrow}(V)).$$

Combining with the L^2-continuity argument above yields

$$\lim_{\lambda \to \infty} Corr(F^{-1}(V), k^{\uparrow}_{\lambda,\xi}(V)) = Corr(F^{-1}(V), F_{\xi}^{-1,\uparrow}(V)).$$

This completes the proof. □

Proof of Theorem 4.14.

(i) By Lemma 4.13, for any $G^{-1} \in \mathscr{F}^{-1}_{U,\xi}(\mu, \sigma, \varepsilon)$ there exists $\lambda > 0$ such that:

$$d_W(h^{\uparrow}_{\lambda,\xi}, F^{-1}) = d_W(G^{-1}, F^{-1}).$$

Hence $\langle F^{-1}, h^{\uparrow}_{\lambda,\xi}\rangle = \langle F^{-1}, G^{-1}\rangle$. Proposition 4.8 applied to the distortion $\gamma + \lambda F^{-1}$ yields:

$$\sup_{H^{-1} \in \mathscr{F}^{-1}_{U,\xi}(\mu,\sigma)} \langle \gamma + \lambda F^{-1}, H^{-1}\rangle = \langle \gamma + \lambda F^{-1}, h^{\uparrow}_{\lambda,\xi}\rangle,$$

with $h^{\uparrow}_{\lambda,\xi}$ the unique maximiser. Therefore, for any $G^{-1} \neq h^{\uparrow}_{\lambda,\xi}$ having the same Wasserstein distance to F^{-1}, we obtain:

$$\langle \gamma + \lambda F^{-1}, h^{\uparrow}_{\lambda,\xi}\rangle > \langle \gamma + \lambda F^{-1}, G^{-1}\rangle,$$

which, using the equality of inner products with F^{-1}, implies $\rho_g(h^{\uparrow}_{\lambda,\xi}) > \rho_g(G^{-1})$. Thus, $h^{\uparrow}_{\lambda,\xi}$ is the unique maximiser on the Wasserstein sphere determined by that radius.

If $d_W(h^{\uparrow}_{\lambda_1,\xi}, F^{-1}) < d_W(h^{\uparrow}_{\lambda_2,\xi}, F^{-1})$ then $\langle F^{-1}, h^{\uparrow}_{\lambda_2,\xi}\rangle < \langle F^{-1}, h^{\uparrow}_{\lambda_1,\xi}\rangle$ and by Proposition 4.8 one gets $\rho_g(h^{\uparrow}_{\lambda_2,\xi}) > \rho_g(h^{\uparrow}_{\lambda_1,\xi})$. Hence, the maximiser increases with the radius, and part

(i) follows.,

(ii) If $\gamma_\xi^\uparrow$ is not constant, then $h_\xi^\uparrow \in \mathscr{F}_{U,\xi}^{-1}(\mu,\sigma,\varepsilon)$ and Proposition 4.8 gives the stated conclusion. If $\gamma_\xi^\uparrow$ is constant, Corollary 2.3 of Brunk (1965) yields $\langle \gamma, k\rangle \leqslant \langle \gamma_\xi^\uparrow, k\rangle$ for all $k \in \mathscr{F}_{U,\xi}^{-1}$. Taking $k = \pm 1$ shows that $\gamma_\xi^\uparrow = g(1)$ and, therefore,

$$\sup_{G^{-1}\in\mathscr{F}_{U,\xi}^{-1}(\mu,\sigma)} \rho_g(G^{-1}) \leqslant \sup_{G^{-1}\in\mathscr{F}_{U,\xi}^{-1}(\mu,\sigma)} \langle \gamma_\xi^\uparrow, G^{-1}\rangle \leqslant g(1)\mu.$$

To see the reverse inequality construct for $n \geqslant 1$, the distribution G_n of the random variable $\mu + \sigma\sqrt{3}nU[-1/n, 1/n]$ (uniform on a shrinking interval). Then $G_n^{-1} \in \mathscr{F}_{U,\xi}^{-1}(\mu,\sigma)$ and $G_n^{-1}(t) \to \mu$ for all $t \in (0,1)$, which implies $\rho_g(G_n^{-1}) \to g(1)\mu$. Hence,

$$\sup_{G^{-1}\in\mathscr{F}_{U,\xi}^{-1}(\mu,\sigma)} \rho_g(G^{-1}) = g(1)\mu,$$

completing the proof. $\square$

Proof of Lemma 4.15. Let

$$\delta := \inf_{h\in\mathscr{F}_{U,[\xi_1,\xi_2]}^{-1}} \|\gamma - h\|_2$$

and choose a sequence $(h_n)_{n\geqslant 1} \subset \mathscr{F}_{U,[\xi_1,\xi_2]}^{-1}$ with $\|\gamma - h_n\|_2 \downarrow \delta$. For each rational $u \in (0,1) \cap \mathbb{Q}$ the sequence $(h_n(u))_{n\geqslant 1}$ is bounded, hence by a diagonal argument we may extract a subsequence $(h_{n_m})_{m\geqslant 1}$ such that the pointwise limit

$$h^*(u) := \lim_{m\to\infty} h_{n_m}(u)$$

exists for every $u \in (0,1)\cap\mathbb{Q}$. The function h^* is monotone increasing on the rationals. Define for $u \in (0,1)$

$$\gamma_{\xi_1,\xi_2}^\uparrow(u) := \inf_{u'\in(u,1)\cap\mathbb{Q}} h^*(u'),$$

which is increasing and right-continuous on $(0,1)$.

We next show $\gamma_{\xi_1,\xi_2}^\uparrow \in \mathscr{F}_{U,[\xi_1,\xi_2]}^{-1}$. Let C denote the set of continuity points of $\gamma_{\xi_1,\xi_2}^\uparrow$; for any value of $u \in C$ there are rational sequences $u_\ell \uparrow u$ and $u'_\ell \downarrow u$ with $h^*(u_\ell) \to \gamma_{\xi_1,\xi_2}^\uparrow(u)$ and

$h^*(u'_\ell) \to \gamma^{\uparrow}_{\xi_1,\xi_2}(u)$. By definition of h^* and the subsequence,

$$\lim_{m\to\infty} h_{n_m}(u) = \gamma^{\uparrow}_{\xi_1,\xi_2}(u) \qquad \text{for all } u \in C.$$

Let ξ_m denote the inflection point of h_{n_m}. The sequence $(\xi_m)_{m\geqslant 1}$ is bounded, so after passing to a further subsequence, we may assume $\xi_m \to \xi \in [\xi_1,\xi_2]$. We show $\gamma^{\uparrow}_{\xi_1,\xi_2}$ is convex on $(\xi,1)$ and concave on $(0,\xi)$. Fix $u_1 < u_2 \leqslant u_3 < u_4$ in $(\xi,1)$. For large m we have $\xi_m < u_1$, and convexity of h_{n_m} on $(\xi_m,1)$ yields:

$$\frac{h_{n_m}(u_{2,\ell}) - h_{n_m}(u_{1,\ell})}{u_{2,\ell}-u_{1,\ell}} \leqslant \frac{h_{n_m}(u_{4,\ell}) - h_{n_m}(u_{3,\ell})}{u_{4,\ell}-u_{3,\ell}},$$

for appropriately chosen rational approximations $u_{i,\ell} \downarrow u_i$. Letting $m\to\infty$ and then $\ell\to\infty$ and using right-continuity gives the corresponding inequality for $\gamma^{\uparrow}_{\xi_1,\xi_2}$, hence convexity on $(\xi,1)$. The concavity on $(0,\xi)$ is analogous.

If $\xi \in (0,1)$ we verify continuity at ξ by a local Lipschitz-type estimate: for u_1,u_2 close to ξ, one obtains an m-uniform bound on increments of h_{n_m} from unimodality, and passing to the limit gives $\gamma^{\uparrow}_{\xi_1,\xi_2}(\xi+) = \gamma^{\uparrow}_{\xi_1,\xi_2}(\xi-)$. Thus, $\gamma^{\uparrow}_{\xi_1,\xi_2}$ is continuous on $(0,1)$, concave on $(0,\xi)$ and convex on $(\xi,1)$, with $\xi\in[\xi_1,\xi_2]$, so $\gamma^{\uparrow}_{\xi_1,\xi_2} \in \mathscr{F}^{-1}_{U,[\xi_1,\xi_2]}$.

Finally, Fatou's lemma yields@

$$\|\gamma - \gamma^{\uparrow}_{\xi_1,\xi_2}\|_2 \leqslant \liminf_{m\to\infty} \|\gamma - h_{n_m}\|_2 = \delta,$$

hence $\gamma^{\uparrow}_{\xi_1,\xi_2}$ attains the infimum and the proof is complete. □

Proof of Proposition 4.16. The argument is identical to that of Proposition 4.6 (replace the projection on the full unimodal class by the projection on $\mathscr{F}^{-1}_{U,[\xi_1,\xi_2]}$). We therefore omit the details. □

Proof of Proposition 4.17. By Theorem 1 of Popescu (2007) (see also the discussion there) the set of marginal laws of the portfolio return under the mean–covariance constraint equals

$$\left\{F_{\sum_{i=1}^n w_i X_i} : \mathbb{E}(X_i) = \mu_i,\ Cov(\mathbf{X}) = \Sigma_0\right\} = \mathscr{M}_\infty\left(\mathbf{w}^\top\mu, \sqrt{\mathbf{w}^\top\Sigma_0\mathbf{w}}\right).$$

Moreover, Theorem 5 of Mao et al. (2025) shows that the multivariate Wasserstein constraint on the vector $\mathbf{X}$ translates to a univariate Wasserstein constraint on the portfolio $\mathbf{w}^\top \mathbf{X}$, namely

$$\left\{F_{\sum_{i=1}^n w_i X_i} : d_W^{(n,2)}(F_{\mathbf{X}}, F_{\mathbf{X}_0}) \leqslant \sqrt{\varepsilon}\right\} = \{G : d_W(F_{\mathbf{w}^\top \mathbf{X}_0}, G) \leqslant \sqrt{\varepsilon \, \|\mathbf{w}\|_2^2}\}.$$

Combining these equalities yields:

$$\mathscr{M}_{\mathbf{w},\varepsilon} = \mathscr{M}_{\varepsilon \|\mathbf{w}\|_2^2}\big(\mathbf{w}^\top \mu, \sqrt{\mathbf{w}^\top \Sigma_0 \mathbf{w}}\big),$$

as asserted. $\square$

CHAPTER 5

Conclusion

This thesis has investigated the design and implementation of robust reinsurance and portfolio selection under several complementary forms of uncertainty. The three main chapters, Chapters 2-3 address distinct, yet related aspects of uncertainty that insurers face in practice, specifically, ambiguous dependence across business lines, computational difficulties arising from non-convex optimisation problems, while Chapter 4 addresses distributional ambiguity affecting risk metrics and portfolio optimisation. Together, these contributions form a coherent framework that blends analytical characterisation, computational methodology, and distributionally robust bounds to produce implementable guidance for both theory and practice.

Chapter 2 studies the worst-case reinsurance design when only marginal loss distributions are known and the dependence between risks is uncertain. Using VaR and related criteria, the chapter provides tractable characterisations of optimal ceded-loss functions under mild assumptions on the marginals. A key result is that, under realistic tail conditions, the complex infinite-dimensional search for optimal contracts reduces to a search within a few interpretable families. These include limited stop-loss, piecewise-linear (i.e., generalised stop-loss), and capped quota-share arrangements. The chapter also explores special cases, such as two lines of business and quota-share extremes, to demonstrate how worst-case dependence can be constructed and to clarify the practical meaning of the solutions. Overall, the results offer insurers explicit and robust contract forms, supported by theory, that remain valid under dependence ambiguity.

Chapter 3 tackles the numerical challenges arising from the non-convex, high-dimensional reinsurance optimisation problems. The chapter introduces a homotopy-based computational strategy, Homotopy with Perturbations and Ensembles (HOPE), which has been specifically adapted to these robust formulations. Through careful algorithmic design and extensive numerical experiments, HOPE reliably finds high-quality solutions and is thus far more efficient than traditional global optimisation methods (e.g., grid search optimisation). By rendering previously intractable formulations computationally tractable, this chapter closes the gap between theoretical reduction and practical implementation.

Chapter 4 develops distributionally robust bounds for a broad class of distortion risk metrics under Wasserstein transport constraints, with extensions that incorporate unimodality and moment constraints. The chapter derives sharp worst-case bounds for several practically relevant measures, including both discontinuous and non-monotone distortions that interpolate between VaR and expected shortfall. It also identifies the form of the worst-case distributions across different parameter regimes. By relaxing restrictive regularity assumptions and adding economically informed constraints to the ambiguity set, the analysis generates risk bounds that are theoretically precise and less conservative than many classical alternatives. These bounds are suitable for embedding into capital allocation and portfolio optimisation routines, thereby enabling insurers and regulators to quantify model uncertainty in a principled way.

Collectively, this thesis delivers three contributions. First, it provides analytical insight into the structure of robust reinsurance contracts under conditions of dependence uncertainty. Second, it supplies a practical and scalable computational toolbox for solving non-convex robust optimisation problems. Third, it establishes sharp and distributionally robust risk bounds that are adaptable to real-world constraints. Together, these contributions form a clear pipeline that strengthens both the theoretical foundations and the applied feasibility of robust reinsurance and portfolio design. Beyond these theoretical and computational contributions, the robust frameworks developed in this thesis carry direct practical implications for actuarial decision-making. The robust optimisation approach can be applied across a range of actuarial settings: when designing reinsurance programmes under Solvency II or analogous risk-based capital regimes, where Value-at-Risk and expected shortfall constraints must be satisfied under regulatory scrutiny; when pricing and reserving for multi-line portfolios in which the dependence structure between business lines is unknown or otherwise difficult to estimate from limited historical data; and when stress-testing existing contracts against adverse scenarios. In particular, the limited stop-loss and capped quota-share structures identified in Chapter 2 are not merely theoretical constructs but correspond to contract forms that are widely used in the reinsurance market. Limited stop-loss reinsurance, which caps both the insurer's retained loss and the reinsurer's liability, provides actuaries with a principled basis for selecting retention levels and coverage limits that remain effective even under dependence misspecification, while also reducing solvency capital requirements and stabilising underwriting results. Moreover, the HOPE algorithm of Chapter 3 equips practitioners with a computationally feasible means of solving these otherwise intractable optimisation problems at portfolio scale, and the distributionally robust risk bounds of Chapter 4 can be embedded directly into internal capital models to produce risk assessments that account explicitly for model uncertainty. Collectively, these tools enable actuaries to move beyond point-estimate-based reinsurance design towards strategies that are formally resilient to the ambiguities that pervade real-world insurance operations.

This research, however, has limitations that lead to clear avenues for future work. Relaxing the convexity or concavity assumptions used in some analytical reductions for multi-risk settings would broaden applicability, but would also introduce substantial technical challenges. The empirical calibration of ambiguity sets (e.g., principled choices of Wasserstein radii and validation on industrial multi-line portfolios) would serve to strengthen the potential practical impact. Extending the framework to dynamic contracts, including counterparty default, market frictions, and modelling strategic interactions between insurers and reinsurers, is a promising direction for future research. These extensions would align the models more closely with regulatory and market realities and thereby increase their operational usefulness. Addressing these gaps will require a combination of theoretical advances, computational innovation, and industrial collaboration.

In conclusion, this thesis advances the design of reinsurance and risk management strategies under conditions of uncertainty by combining rigorous analysis, efficient computation, and distributionally informed robustness. The results presented in this thesis thus offer clear theoretical insight and usable tools.

References

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

[2] Allais, M. (1953). "Le Comportement de l'Homme Rationnel Devant le Risque: Critique des Axiomes et Postulats de l'Ecole Américaine". *Econometrica* 21.4, pp. 503–546.

[3] Arrow, K. J. (1963). "Uncertainty and the Welfare Economics of Medical Care". *American Economic Review* 53.5, pp. 941–973.

[4] Asimit, A. V., Badescu, A. M., and Tsanakas, A. (2013a). "Optimal Risk Transfers in Insurance Groups". *European Actuarial Journal* 3, pp. 159–190.

[5] 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, pp. 720–732.

[6] Asimit, A. V., Gao, T., Hu, J., and Kim, E.-S. (2018). "Optimal Risk Transfer: A Numerical Optimization Approach". *North American Actuarial Journal* 22.3, pp. 341–364.

[7] Asimit, A. V., M., B. A., and Cheung, K. C. (2013b). "Optimal Reinsurance in the Presence of Counterparty Default Risk". *Insurance: Mathematics and Economics* 53.3, pp. 690–697.

[8] Basel Committee on Banking Supervision (Feb. 2019). *Minimum Capital Requirements for Market Risk*. Tech. rep. Document number d457. Basel: Bank for International Settlements.

[9] Belles-Sampera, J., Guillén, M., and Santolino, M. (2013). "Beyond value-at-risk: GlueVaR distortion risk measures". *Risk Analysis* 34.1, pp. 121–134.

[10] Bellini, F., Fadina, T., Wang, R., and Wei, Y. (2022). "Parametric measures of variability induced by risk measures". *Insurance: Mathematics and Economics* 106, pp. 270–284.

[11] Ben-Tal, A., El Ghaoui, L., and Nemirovski, A. (2009). *Robust Optimization*. Princeton, New Jersey: Princeton University Press.

[12] Ben-Tal, A. and Nemirovski, A. (2008). "Selected topics in robust convex optimization". *Mathematical Programming* 112.1, pp. 125–158.

[13] Benati, S. and Conde, E. (2022). "A relative robust approach on expected returns with bounded CVaR for portfolio selection". *European Journal of Operational Research* 296.1, pp. 332–352.

[14] Bernard, C., Jiang, X., and Wang, R. (2014). "Risk aggregation with dependence uncertainty". *Insurance: Mathematics and Economics* 54, pp. 93–108.

[15] Bernard, C., Kazzi, R., and Vanduffel, S. (2023). "Model uncertainty assessment for symmetric and right-skewed distributions". Available at SSRN 4468467.

[16] Bernard, C., Liu, F., and Vanduffel, S. (2020). "Optimal insurance in the presence of multiple policyholders". *Journal of Economic Behavior and Organization* 180, pp. 638–656.

[17] Bernard, C., Pesenti, S. M., and Vanduffel, S. (2024). "Robust Distortion Risk Measures". *Mathematical Finance* 34.3, pp. 774–818.

[18] Bertsimas, D., Brown, D. B., and Caramanis, C. (2011). "Theory and applications of robust optimization". *SIAM Review* 53.3, pp. 464–501.

[19] Blanchet, J., Chen, L., and Zhou, X. (2022). "Distributionally robust mean-variance portfolio selection with Wasserstein distances". *Management Science* 68.9, pp. 6382–6410.

[20] Blanchet, J., Lam, H., Liu, Y., and Wang, R. (2023). *Convolution bounds on quantile aggregation*. arXiv:2007.09320.

[21] Blanchet, J. and Murthy, K. (2019). "Quantifying distributional model risk via optimal transport". *Mathematics of Operations Research* 44.2, pp. 565–600.

[22] Boonen, T. J. and Jiang, W. (2024). "Robust insurance design with distortion risk measures". *European Journal of Operational Research* 316.2, pp. 694–706.

[23] Borch, K. (1960). “An attempt to determine the optimum amount of stop-loss reinsurance”. In: *Transactions of the 16th International Congress of Actuaries*. Vol. 1, pp. 597–610.

[24] Brennan, M. J. and Solanki, R. (1981). “Optimal portfolio insurance”. *Journal of Financial and Quantitative Analysis* 16, pp. 279–300.

[25] Brunk, H. D. (1965). “Conditional expectation given a σ-lattice and applications”. *The Annals of Mathematical Statistics* 36.5, pp. 1339–1350.

[26] Cai, J. and Chi, Y. (2020). “Optimal reinsurance designs based on risk measures: a review”. *Statistical Theory and Related Fields* 4.1, pp. 1–13.

[27] Cai, J., Li, J. Y.-M., and Mao, T. (2025). “Distributionally robust optimization under distorted expectations”. *Operations Research* 73.2, pp. 969–985.

[28] Cai, J., Liu, F., and Yin, M. (2024). “Worst-case risk measures of stop-loss and limited loss random variables under distribution uncertainty with applications to robust reinsurance”. *European Journal of Operational Research* 318(1), pp. 310–326.

[29] Cai, J. and Tan, K. S. (2007). “Optimal Retention for a Stop-loss Reinsurance under the VaR and CTE Risk Measures”. *ASTIN Bulletin* 37.1, pp. 93–112.

[30] Cai, J., Tan, K. S., Weng, C., and Zhang, Y. (2008). “Optimal reinsurance under VaR and CTE risk measures”. *Insurance: Mathematics and Economics* 43.1, pp. 185–196.

[31] Cai, J. and Wei, W. (2012). “Optimal reinsurance with positively dependent risks”. *Insurance: Mathematics and Economics* 50.1, pp. 57–63.

[32] Carole, B. and Ludkovski, M. (2012). “Impact of Counterparty Risk on the Reinsurance Market”. *North American Actuarial Journal* 16.1, pp. 87–111.

[33] Chen, L., He, S., and Zhang, S. (2011). “Tight bounds for some risk measures, with applications to robust portfolio selection”. *Operations Research* 59.4, pp. 847–865.

[34] Cheung, K. C. (2010). “Optimal reinsurance revisited – a geometric approach”. *ASTIN Bulletin* 40.1, pp. 221–239.

[35] Cheung, K. C., Sung, K. C. J., and Yam, S. C. P. (2014). “Risk-minimizing reinsurance protection for multivariate risks”. *Journal of Risk and Insurance* 81.1, pp. 219–236.

[36] Chi, Y. and Lin, X. S. (2014). "Optimal reinsurance with limited ceded risk: A stochastic dominance approach". *ASTIN Bulletin* 44.1, pp. 103–126.

[37] Chi, Y. and Tan, K. S. (2013). "Optimal reinsurance with general premium principles". *Insurance: Mathematics and Economics* 52.2, pp. 180–189.

[38] Chi, Y., Xu, Z., and Zhang, S. (2022). "Distributionally robust goal-reaching optimization in the presence of background risk". *North American Actuarial Journal* 26.3, pp. 351–382.

[39] Cont, R., Deguest, R., and Scandolo, G. (2010). "Robustness and sensitivity analysis of risk measurement procedures". *Quantitative Finance* 10.6, pp. 593–606.

[40] Cui, W., Yang, J., and Wu, L. (2013). "Optimal Reinsurance Minimising the Distortion Risk Measure under General Reinsurance Premium Principles". *Insurance: Mathematics and Economics* 53.1, pp. 74–85.

[41] Denuit, M. and Vermandele, C. (1998). "Optimal reinsurance and stop-loss order". *Insurance: Mathematics and Economics* 22.3, pp. 229–233.

[42] Dhaene, J. and Denuit, M. (1999). "The safest dependence structure among risks". *Insurance: Mathematics and Economics* 25.1, pp. 11–21.

[43] Dunlavy, D. M. and O'Leary, D. P. (2005). *Homotopy optimization methods for global optimization*. Tech. rep. SAND2005-7495. Sandia National Laboratories.

[44] El Ghaoui, L., Oks, M., and Oustry, F. (2003). "Worst-case value-at-risk and robust portfolio optimization: A conic programming approach". *Operations Research* 51.4, pp. 543–556.

[45] Embrecht, P., Puccetti, G., and Rüschendorf, L. (2013). "Model uncertainty and VaR aggregation". *Journal of Banking and Finance* 37.8, pp. 2750–2764.

[46] Esfahani, P. M. and Kuhn, D. (2018). "Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations". *Mathematical Programming* 171.1, pp. 115–166.

[47] European Central Bank (2017). *Guidance on model risk management*. Tech. rep. European Central Bank.

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

[49] Gabrel, V., Murat, C., and Thiele, A. (2014). "Recent advances in robust optimization: An overview". *European Journal of Operational Research* 235.3, pp. 471–483.

[50] Gajek, L. and Zagrodny, D. (2000). "Insurer's Optimal Reinsurance Strategies". *Insurance: Mathematics and Economics* 27.1, pp. 105–112.

[51] Gilboa, I. and Schmeidler, D. (1989). "Maxmin expected utility with non-unique prior". *Journal of Mathematical Economics* 18.2, pp. 141–153.

[52] Goldstein, D. G., Johnson, E. J., and Sharpe, W. F. (2008). "Choosing outcomes versus choosing products: consumer-focused retirement investment advice". *Journal of Consumer Research* 35, pp. 440–456.

[53] Hansen, L. and Sargent, T. J. (2001). "Robust control and model uncertainty". *American Economic Review* 91.2, pp. 60–66.

[54] Jakobsons, E., Han, X., and Wang, R. (2016). "General convex order on risk aggregation". *Scandinavian Actuarial Journal* 2016.8, pp. 713–740.

[55] Kaluszka, M. (2001). "Optimal Reinsurance under Mean-variance Premium Principles". *Insurance: Mathematics and Economics* 28.1, pp. 61–67.

[56] Kaluszka, M. (2004). "Mean-Variance Optimal Reinsurance Arrangements". *Scandinavian Actuarial Journal* 1, pp. 28–41.

[57] Lauzier, J.-G., Lin, L., and Wang, R. (2023). *Risk sharing, measuring variability, and distortion riskmetrics*. arXiv:2302.04034.

[58] 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, pp. 1533–1541.

[59] Li, L., Shao, H., Wang, R., and Yang, J. (2018). "Worst-case Range Value-at-Risk with partial information". *SIAM Journal on Financial Mathematics* 9.1, pp. 190–218.

[60] Liao, S. (1992). "The proposed homotopy analysis technique for the solution of nonlinear problems". PhD thesis. Shanghai Jiao Tong University.

[61] Lin, X., Yang, Z., Zhang, X., and Zhang, Q. (2023). “Continuation Path Learning for Homotopy Optimization”. In: *International Conference on Machine Learning*, pp. 21288–21311.

[62] Liu, F., Mao, T., Wang, R., and Wei, L. (2022). “Inf-convolution, optimal allocations, and model uncertainty for tail risk measures”. *Mathematics of Operations Research* 47.3, pp. 2494–2519.

[63] Makarov, G. D. (1981). “Estimates for the distribution function of a sum of two random variables when the marginal distributions are fixed”. *Theory of Probability and Its Applications* 26, pp. 803–806.

[64] Mao, T., Wang, R., and Wu, Q. (2025). “Model Aggregation for Risk Evaluation and Robust Optimization”. *Management Science*. forthcoming.

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

[66] Pesenti, S. M., Wang, Q., and Wang, R. (2024). “Optimizing distortion riskmetrics with distributional uncertainty”. *Mathematical Programming*. forthcoming.

[67] Polak, G. G., Rogers, D. F., and Sweeney, D. J. (2010). “Risk management strategies via minimax portfolio optimization”. *European Journal of Operational Research* 207.1, pp. 409–419.

[68] Popescu, I. (2005). “A Semidefinite Programming Approach to Optimal-Moment Bounds for Convex Classes of Distributions”. *Mathematics of Operations Research* 30.3, pp. 632–657.

[69] Popescu, I. (2007). “Robust mean-covariance solutions for stochastic optimization”. *Operations Research* 55.1, pp. 98–112.

[70] Quiggin, J. (1982). “A theory of anticipated utility”. *Journal of Economic Behavior & Organization* 3.4, pp. 323–343.

[71] Roese, N. J. and Olson, J. M. (1995). “Counterfactual thinking: A critical overview”. *Psychological Bulletin* 118.1, pp. 1–19.

[72] Rüschendorf, L. (1982). “Random variables with maximum sums”. *Advances in Applied Probability* 14, pp. 623–632.

[73] Scarf, H. E. (1958). "A Min–Max Solution of an Inventory Problem". In: *Studies in the mathematical theory of inventory and production*. Ed. by Arrow, K. J., Karlin, S., and Scarf, H. E. Stanford: Stanford University Press, pp. 201–209.

[74] Shalit, H. and Yitzhaki, S. (1984). "Mean-Gini, portfolio theory, and the pricing of risky assets". *The Journal of Finance* 39.5, pp. 1449–1468.

[75] Shao, H. and Zhang, Z. (2023). "Distortion risk measure under parametric ambiguity". *European Journal of Operational Research* 311.3, pp. 1159–1172.

[76] Starmer, C. (2000). "Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk". *Journal of Economic Literature* 38.2, pp. 332–382.

[77] Tan, K. S., Wei, P., Wei, W., and Zhuang, S. (2020). "Optimal dynamic reinsurance policies under a generalize Denneberg's absolute deviation principle". *European Journal of Operational Research* 282.1, pp. 345–362.

[78] Tversky, A. and Kahneman, D. (1973). "Availability: A heuristic for judging frequency and probability". *Cognitive Psychology* 5.2, pp. 207–232.

[79] Tversky, A. and Kahneman, D. (1992). "Advances in prospect theory: Cumulative representation of uncertainty". *Journal of Risk and Uncertainty* 5, pp. 297–323.

[80] Van Heerwaarden, A. E., Kaas, R., and Goovaerts, M. J. (1989). "Optimal reinsurance in relation to ordering of risks". *Insurance: Mathematics and Economics* 8.1, pp. 11–17.

[81] Wang, B. and Wang, R. (2016). "Joint mixability". *Mathematics of Operations Research* 41.3, pp. 808–826.

[82] Wang, Q., Wang, R., and Wei, Y. (2020a). "Distortion riskmetrics on general spaces". *ASTIN Bulletin: The Journal of the IAA* 50.3, pp. 827–851.

[83] Wang, R., Peng, L., and Yang, J. (2013). "Bounds for the sum of dependent risks and worst Value-at-Risk with monotone marginal densities". *Finance and Stochastics* 17.2, pp. 395–417.

[84] Wang, R., Wei, Y., and Willmot, G. E. (2020b). "Characterization, robustness, and aggregation of signed Choquet integrals". *Mathematics of Operations Research* 45.3, pp. 993–1015.

[85] Wang, S. S., Young, V. R., and Panjer, H. H. (1997). “Axiomatic Characterization of Insurance Prices”. *Insurance: Mathematics and Economics* 21.2, pp. 173–183.

[86] Watson, L. T. and Haftka, R. T. (1989). “Modern homotopy methods in optimization”. *Computer Methods in Applied Mechanics and Engineering* 74, pp. 289–304.

[87] Yaari, M. E. (1987). “The dual theory of choice under risk”. *Econometrica* 55.1, pp. 95–115.

[88] Young, V. R. (2004). “Premium Principles”. In: *Encyclopedia of Actuarial Science*. Ed. by Teugels, J. and Sundt, B. Vol. 3. New York: John Wiley & Sons.

[89] Zhu, Y., Chi, Y., and Weng, C. (2014). “Multivariate reinsurance designs for minimizing an insurer’s capital requirement”. *Insurance: Mathematics and Economics* 59, pp. 144–155.

[90] Zymler, S., Kuhn, D., and Rustem, B. (2013). “Distributionally robust joint chance constraints with second-order moment information”. *Mathematical Programming* 137, pp. 167–198.