UNIVERSITY OF SOUTHAMPTON

Faculty of Social Sciences

School of Mathematical Sciences

MULTI-ATTRIBUTE PREFERENCE ROBUST UTILITY BASED SHORTFALL RISK OPTIMIZATION AND DISTRIBUTIONALLY ROBUST REWARD RISK OPTIMIZATION

by
Yuan Zhang

Thesis for the Degree of Doctor of Philosophy

September 2020

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL SCIENCES

SCHOOL OF MATHEMATICAL SCIENCES

Doctor of Philosophy

MULTI-ATTRIBUTE PREFERENCE ROBUST UTILITY BASED SHORTFALL RISK OPTIMIZATION AND DISTRIBUTIONALLY ROBUST REWARD RISK OPTIMIZATION

by Yuan Zhang

The study of decision making under uncertainty is important in many areas (e.g. portfolio theory, control theory and utility theory). The exogenous and endogenous uncertainties, such as variations in stock prices, changes in consumer demand and ambiguity about investor's risk attitude, are beyond deciders' control and knowledge and significantly influence the effectiveness of any decision. In this thesis, we concentrate on this issue and propose some efficient models to deal with the uncertainties. Specifically, (a) we introduce a utility-based reward-risk ratio (URR) optimization model and consider a situation where an investor does not have complete information on the probability distribution of the underlying random variables, and we propose a distributionally robust URR optimization model to mitigate the risk arising from ambiguity of the true probability distribution; (b) we introduce a multivariate utility-based shortfall risk measure (MSR) and focus on a case that a decision maker's true loss function in the definition of MSR is unknown but it is possible to elicit a set of plausible loss functions with partial information, and consequently propose a robust formulation of MSR based on the worst case loss function; (c) we investigate an issue that whether a statistical estimator such as the optimal value of a preference robust optimization model based on empirical data is reliable when the empirical data contain some noise, and we derive moderate sufficient conditions under which the optimal value of the model is robust against perturbation of the exogenous uncertainty data.

Contents

List of Figures

List of Tables

DECLARATION OF AUTHORSHIP

I, Yuan Zhang, declare that the thesis entitled “Multi-attribute preference robust utility based shortfall risk optimization and distributionally robust reward risk optimization” and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date: 24/09/2020

Acknowledgements

It is my pleasure to thank who made this thesis possible. I would like to thank my supervisors, Professor Huifu Xu and Professor Houduo Qi for their support, encouragement and thoughtful guidance during this research. Their extensive knowledge and passion for science helped me to develop my understanding and my inspiration in carrying out this research. They have taught me innumerable lessons and insights on the working of academic research in general. Their technical advice was essential to the completion of this thesis.

I would like to thank my examiners, Professor XuanVinh Doan and Professor Stefano Coniglio, who provided many valuable and constructive feedbacks. I also would like to thank Dr Shaoyan Guo and Mr Wei Wang for their helpful discussions. I acknowledge my thanks to the School of Mathematical Sciences, University of Southampton for all the support and its members of staff for their help with administration.

I would like to acknowledge the support of all my friends. Last but not least, I owe my special thanks to my parents. I would not have contemplated this road without their love, patience and support.

To my dear parents

List of Abbreviations

CVaR	Conditional Value-at-Risk
DRO	Distributionally Robust Optimization
DRURR	Distributionally Robust Utility-Based Reward-Risk Ratio
MSR	Multivariate Utility-Based Shortfall Risk Measure
PRO	Preference Robust Optimization
PRMSR	Preference Robust Multivariate Utility-Based Shortfall Risk Measure
PRMSR-Opti	PRMSR Optimization Problem
SAA	Sample Average Approximation
SR	Utility-Based Shortfall Risk Measure
URR	Utility-Based Reward-Risk Ratio
UGC	Uniform Glivenko-Cantelli
VaR	Value-at-Risk

Chapter 1

Introduction

Optimal decision making problems naturally arise in many fields of engineering and management science. In investment and risk management, for example, a decision maker aims to find the optimal allocation of resources among available assets that would meet the expected return and risk preference. However, there are many underlying exogenous and endogenous uncertainties involved in the decision making. For instance, an investor need to select a portfolio to maximize gain with uncertain future stock price, a manager need to decide an inventory level to maximize profit when future demand is unknown, and there is an ambiguity about the decision maker's risk preference. Thus the desired outcome will be affected by these factors outside the decision maker's control and knowledge at the time of decision. In this thesis, we will investigate these issues and propose some efficient models to deal with the uncertainties. The following sections give a brief introduction of the motivations and contributions of this thesis and also give a plan for subsequent chapters.

1.1 Distributionally Robust Optimization

The primary objective of portfolio management is to allocate monetary resources among available assets so as to achieve the highest possible return while controlling the fund's risk exposure. Reward-risk ratio optimization is an important mathematical approach in portfolio management. In previous researches, the reward-risk ratio optimization problem has been studied based on various risk measures. One of the most referenced reward/risk measures used in finance is the Sharpe ratio [76, 77], which is based on mean-variance analysis.

In general, a decision maker's risk preference is important in reward-risk analysis. A decision maker often needs to identify the utility function that characterizes his/her attitude toward risk when making decisions. Thus in Chapter 2, we propose a new reward-risk ratio associated with personal preference which is called utility-based reward-risk ratio (URR). We apply the URR to an optimal decision making problem where the objective is to maximize the ratio.

An accurate prediction of the distribution of asset returns is an essential factor in determining whether the reward-risk ratio can be precisely evaluated. However, in practice, a decision maker does not have complete information on the distribution of the underlying exogenous uncertainty. To deal with this difficulty, the research on worst-case analysis of the portfolio selection problem has increased over the past few years.

In Chapter 2, we consider a robust scheme for the URR optimization model to mitigate the risk arising from ambiguity of the true distribution. We propose a distributionally robust utility-based reward-risk ratio (DRURR) optimization model where the ambiguity set of probability distributions is constructed through prior moment information. We reformulate the DRURR optimization model as a mathematical program with robust inequality constraints and further transform it into a nonlinear semi-infinite programming problem through the Lagrange dualization. We then apply the entropic approximation scheme to deal with the semi-infinite constraints and stability analysis is presented for the approximation scheme, and consequently propose a numerical scheme to solve the approximated optimization problem.

To see how the proposed framework of modelling works, we investigate a specific case that the ambiguity set is determined by the mean and covariance. we consider box constraints for the mean and covariance, which restricts each component of the two quantities to an interval with finite lower and upper bound. We analyse the likelihood of the true distribution to lie in the ambiguity set and the convergence of the optimal value and the optimal solutions obtained on the basis of the ambiguity set. Finally, we apply the DRURR optimization model to a portfolio selection problem and report some numerical test results.

1.2 Preference Robust Optimization

The DRURR optimization model investigated in Chapter 2 is under the assumption that a decision maker's true utility function is known. In many practical applications, however, the true utility function is unavailable because there may not be enough information to specify it or a group of decision makers have difficulty agreeing on which utility function to use. Thus in Chapter 3, we will focus on this issue in the context of the utility-based shortfall risk measure (SR).

The SR model introduced by Föllmer and Schied [32] has received increasing attention over the past few years. For a given loss function (as a risk attitude) and a threshold value (as a prespecified risk level), the SR of a financial position is the minimal capital added to the position such that the new position's risk level is below the prespecified risk level [47]. Previous research on SR was based on the univariate case, whereas in natural applications, many financial positions possess multi-attributes, e.g., an insurance company typically has several business lines each of which has a distinct attribute. This motivates us to consider a multi-attribute SR model.

In Chapter 3, we introduce a multivariate utility-based shortfall risk measure (MSR) and consider a situation where a decision maker's true loss function is unknown but it is possible to elicit a set of plausible loss functions with empirical data or subjective judgements. Consequently, we define a preference robust multivariate utility-based shortfall risk measure (PRMSR) through the worst loss function from the set to mitigate the risk from the ambiguity. We demonstrate that MSR and PRMSR are convex risk measures and discuss the domains of MSR and PRMSR.

Since a risk measure is often associated with some decision making problems, we apply the PRMSR to an optimization problem (denoted by PRMSR-Opti) where the objective is to minimize the PRMSR of a vector-valued cost function. Considering a case that the underlying probability distribution is continuous, we propose a sample average approximation scheme and show it converges to the true problem in terms of the optimal value and optimal solutions as the sample size increases. A tractable formulation is developed for the approximated optimization problem when the ambiguity set of loss functions is defined with some specified characteristics such as convex increasing, Lipschitz continuous and pairwise comparisons. Some numerical studies are also given to examine the performance of the proposed robust model and numerical scheme.

1.3 Statistical Robustness

In preference robust optimization models, e.g., the PRMSR-Opti in Chapter 3, the true probability distribution is often assumed to be either known or can be recovered via empirical data which do not contain any noise. However, it is unclear whether a statistical estimator such as the optimal value of a preference robust optimization model based on empirical data is reliable when the empirical data contain some noise, that is, if we let Q_N denote the empirical distribution based on the data with noise, P_N the empirical distribution based on the data with noise detached and $\vartheta(\cdot)$ the optimal value of a preference robust optimization model, we ask whether $\vartheta(Q_N)$ is close to $\vartheta(P_N)$ under some metric when N is sufficiently large. This issue is also known as statistical robustness which can be traced back to the work of Hampel [41].

In Chapter 4, we investigate the above issue in the context of PRMSR-Opti model. We recall some basic notions and results about ψ-weak topology and derive uniform continuity of $\vartheta(P)$ w.r.t. variation of the true probability distribution P under some metrics. We investigate the uniform Glivenko-Cantelli property and identify appropriate metrics under which $\vartheta(P_N)$ uniformly converges to $\vartheta(P)$. We establish statistical robustness of estimator of $\vartheta(\cdot)$, that is, $\vartheta(Q_N)$ is close to $\vartheta(P_N)$ under the Prokhorov metric when N is sufficiently large as long as Q (the distribution generates the data with noise) is close to P.

Unlike Chapter 2, which considers P is unknown but can be approximated with an ambiguity set of distributions and the optimal decision is based on the worst probability distribution from the ambiguity set, the statistical robustness in Chapter 4 does not consider worst probability distribution, rather it concerns with quality of statistical estimators based on real data. In addition, compared with the convergence analysis in Chapter 3, the convergence analysis in Chapter 4 focuses on uniform convergence of $\vartheta(P_N)$ to $\vartheta(P)$ for all P in a specified set of probability measures rather than the rate of convergence for a fixed P.

1.4 Notation

Throughout this thesis, we will use the following notations. By convention, we write $a^T b$ for the scalar product of two vectors $a, b \in \mathbb{R}^n$, e denotes a vector with all elements being 1, $\|\cdot\|$ denotes the Euclidean norm of a vector, $\|\cdot\|_\infty$ denotes the infinity norm of matrix, $\|\cdot\|_F$ denotes the Frobenius norm of matrix and $(a)_+ = \max\{a, 0\}$. We use 'cl$A$' to denote the closure of a set A, Diam(A) denotes the diameter of a set A, $d(x, A) := \inf_{x' \in A} \|x - x'\|$ denotes the distance from a point x to the set A, $\mathbb{D}(C, A) := \sup_{x \in C} d(x, A)$ denotes the deviation of C from A, and $\mathbb{H}(C, A) := \max\{\mathbb{D}(C, A), \mathbb{D}(A, C)\}$ denotes the Hausdorff distance between C and A.

Chapter 2

Distributionally Robust Utility-Based Reward-Risk Ratio Optimization

Reward-risk ratio optimization is an important mathematical approach in financial portfolio management. It helps decision makers allocate limited resource efficiently so as to maximize their wealth while reduce the fund's risk exposure. Generally, a decision maker's risk preference has a significant influence on his/her investment decision making, and it is important for a decision maker to find an appropriate reward-risk ratio which reflects his/her risk attitude. In practice, a decision maker often needs to identify the utility function that characterizes his/her attitude toward risk when making decisions. Thus in this chapter, we propose a new reward-risk ratio associated with personal preference which is called utility-based reward-risk ratio (URR). In the real world, a decision maker often needs to make decisions with underlying uncertainties (e.g., variations in stock prices and changes in consumer demand), and these uncertainties will significantly influence the effectiveness of decisions. Hence in this chapter, we will consider a situation where there is an ambiguity about the probability distribution of random variable, and use a robust scheme for the URR optimization model to alleviate the risk arising from it.

2.1 Introduction

2.1.1 Literature Review

One of the main issues in financial portfolio management is to develop appropriate measures for hedging risks arising from various uncertain factors. Since the pioneering Markowitz work on mean-variance portfolio selection [58], the reward-risk analysis framework has been widely used in financial portfolio management. Under the reward-risk analysis framework, the portfolio choice is made according to two criteria: the expected return and the risk. One portfolio is preferred to another if it has higher expected return and lower risk.

Related to reward-risk analysis is the reward-risk ratio optimization, which has been studied based on the various performance measures in the literature, see Stoyanov et al. [79] for an overview of this topic. One of the most referenced return/risk measures used in finance is the Sharpe ratio [76, 77], which is based on mean-variance analysis. The ratio measures the excess return per unit of deviation in an investment asset, and it characterizes how well the return of an asset compensates the investor for the risk taken. After the publication of the Sharpe ratio, some new performance measures, like the STARR ratio, Sortino-Satchell ratio and the Rachev ratio, have been proposed, see Biglova et al. [14] for an empirical comparison.

In practice, no matter which performance ratio is adopted, whether the ratio can be accurately evaluated mainly depends on the reliable and exact forecast of the distribution of asset returns. The work of Black and Litterman [16] indicated that when adopting the mean-variance model, the portfolio decision is very sensitive to the mean, and a small error in the estimator of this variable can be amplified into a significant change of the optimal portfolio strategy. It is obvious that the ambiguity in underlying distributions can be observed in situations where data samples are insufficient or unstable. In addition, other cases, such as indetermination of exit time in portfolio selection problem [59] or no consensus on the future markets among decision makers in decentralized investment management problem [60], may lead to the uncertainty of underlying distributions.

This phenomenon motivates an increase in study of robustness of the portfolio selection when dealing with the ambiguity in the underlying distribution. The work of Ben-Tal et al. [12] developed a robust multi-stage asset allocation model using a robust linear programming approach. Lobo and Boyd [56], Goldfarb and Iyengar [37] and Lu [57] studied the robust mean-variance portfolio selection problem. In their work, some uncertainty sets (e.g., box uncertainty set and ellipsoidal uncertainty set) of the problem parameters (e.g., mean and covariance of the random returns) were introduced, and they showed that the robust portfolio selection problems corresponding to these uncertainty structures can be transformed into semidefinite programs or second-order cone programs, which can be solved by interior-point algorithms.

El Ghaoui et al. [35] studied the robust portfolio optimization problem using worst-case value-at-risk, which requires the information of the first and second moments of the distribution only. A study by Natarajan et al. [61] considered the robust value-at-risk optimization problem by incorporating asymmetric distributional information. Zhu and Fukushima [88] explored the robust portfolio selection problem using worst-case conditional value-at-risk with several structures of uncertainty in the underlying distribution. The work of Zhu et al. [89] focused on the robustness of lower-partial moment based portfolio selection problems. Delage and Ye [26] investigated the distributionally robust optimization problem where the ambiguity set is constructed through moment constraints.

The work of Kapsos et al. [50] proposed a robust omega ratio model where the information of the true probability distribution is incomplete, and they computed the omega ratio on the basis of the worst probability distribution in an ambiguity set of distributions. In their work, three types of uncertainty (mixture distribution, box and ellipsoidal uncertainty) were considered. A study by Gorissen [38] extended robust optimization to fractional programming, where both the objective and the constraints contain uncertain parameters, and he proposed tractable formulation for addressing the problem under some conditions.

The majority of prior research on robust reward-risk ratio optimization have not taken into account personal preference. However, in practice, a decision maker needs to identify the utility function that characterizes his/her attitude toward risk when making decisions. Tong and Wu [81] stated that decision maker's risk preference is important in reward-risk analysis. In the standard portfolio analysis, it is often assumed that investors are risk averse and their utility can be expressed as a function of the mean and variance of the portfolios rate of return [51]. The work of De Giorgi [23] characterized reward and risk measures to obtain a link between reward-risk framework for portfolio selection and utility expectation approach.

2.1.2 Contribution

Our research in this chapter proposes a new reward-risk ratio associated with decision maker's risk preference which is called utility-based reward-risk ratio (URR). We apply the URR to an optimization problem and consider a situation where the true probability distribution is unknown. To tackle this issue, we investigate a distributionally robust utility-based reward-risk ratio (DRURR) model, which is varied from the ex ante Sharpe ratio model with the ambiguity set of probability distribution being constructed through prior moment information.

We reformulate the DRURR optimization model as a mathematical program with robust inequality constraints and further transform it into a nonlinear semi-infinite programming problem. We use the entropic approximation scheme to deal with the semi-infinite constraints and give a stability analysis for the approximation scheme. Compared with mainstream approximation methods in the literature of distributionally robust optimization, the approximation scheme we used has no specific requirement (e.g., linear or convex) for the underlying functions w.r.t. the random variables. Thus the framework of our research has a wider range of applications.

We investigate a specific case that the ambiguity set is determined by the mean and covariance. Unlike the work of Popescu [64] assumes complete information of the mean and covariance, we consider some degree of uncertainty for the two quantities. Compared with the work of Delage and Ye [26] which uses ellipsoid constraints for the mean and semi-definite constraints for the covariance, we consider box constraints for the mean and covariance which restricts each component of the two quantities to an interval with finite lower and upper bound. The ambiguity set defined as such can convert the robust formulation to a moment problem which can be solved by a general numerical scheme. We also analyse the likelihood of the true distribution to lie in the ambiguity set and the convergence of the optimal value and the optimal solutions obtained on the basis of the ambiguity set.

We propose a numerical scheme to solve the approximated optimization problem based on the framework of the Dinkelbach method. Compared with the standard Dinkelbach method, our algorithm updates the optimal value automatically at each iteration by solving a nonlinear equation. We apply the DRURR optimization model to a portfolio selection problem and carry out some numerical tests to show the effectiveness and efficiency of the model.

2.1.3 Structure

The rest of this chapter is structured as follows: In section 2.2, we propose a distributionally robust utility-based reward-risk ratio optimization model and reformulate it as a mathematical program with robust inequality constraints. In section 2.3, we transform the robust optimization problem into a semi-infinite programming problem in the case when the ambiguity set is constructed through prior moment information and further use the entropic risk measure to construct an approximation of the semi-infinite constraints. In section 2.4, we consider a case that the ambiguity set is determined by the mean and covariance with some degree of uncertainty and present stability analysis of the robust optimization problem w.r.t. change of sample data. An iterative scheme is developed for solving the approximated optimization problem in section 2.5, and we apply the proposed model to a portfolio optimization problem and report some numerical test results in section 2.6.

2.2 URR and Distributionally Robust URR

We use $u(\cdot)$ to denote a utility function which is increasing and concave and motivate our discuss with the following stochastic program:

$$\max_{x\in X} \mathbb{E}_P[u(f(x,\xi)) - u(Y(\xi))] \tag{2.2.1}$$

where x is a vector of decision variables, X is a nonempty convex and compact subset of $\mathbb{R}^n$, $f : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$ is a continuous return function and it is concave w.r.t. x for every fixed ξ, ξ is a random variable on probability space $(\Xi, \mathcal{F}, P)$ with $\Xi \subset \mathbb{R}^k$, $Y(\xi)$ is a benchmark return, $\mathbb{E}_P[\cdot]$ denotes the expected value w.r.t. the probability distribution of ξ.

Here $\mathbb{E}_P[u(f(x,\xi))]$ measures a decision maker's preference to random returns of the chosen action, likewise, $\mathbb{E}_P[u(Y(\xi))]$ measures a decision maker's preference to a benchmark return. If

$$\mathbb{E}_P[u(f(x,\xi))] \geq \mathbb{E}_P[u(Y(\xi))],$$

we can assert that the random returns of the chosen action $f(x,\xi)$ are preferred to the benchmark return $Y(\xi)$. It means that the decision maker is satisfied with this investment action, and it can be considered as a reward to the investor. Therefore, we can regard (2.2.1) as a reward maximization problem.

Similarly, we have the following stochastic program:

$$\min_{x\in X} \mathbb{E}_P[(u(Y(\xi)) - u(f(x,\xi)))_+], \tag{2.2.2}$$

where $(a)_+ = \max\{a, 0\}$. It considers the situation when

$$\mathbb{E}_P[u(Y(\xi))] \geq \mathbb{E}_P[u(f(x,\xi))].$$

According to the above discussion, it can be regarded as a risk to the investor, and (2.2.2) is a risk minimization problem.

As discussed in the introduction, reward-risk ratio optimization is an important mathematical approach in financial portfolio management which takes into account the return and risk of the investment and helps an investor make the optimal decision to maximise the expected return and minimise the risk. Thus in what follows, we focus on a so-called *utility-based reward-risk ratio (URR)* optimization problem:

$$\sup_{x\in X} \frac{\mathbb{E}_P[u(f(x,\xi)) - u(Y(\xi))]}{\mathbb{E}_P[(u(Y(\xi)) - u(f(x,\xi)))_+]}. \tag{2.2.3}$$

In this setup, as discussed above, the numerator is regarded as a reward and the denominator as a risk. Note that we define $\mathbb{E}_P[((u(Y(\xi)) - u(f(x,\xi)))_+]$ as the risk because we only care about the scenarios where the utility of f (portfolio return) falls below the utility of benchmark return and we look at the expected utility values of these scenarios. This is in accordance with the semi-deviation risk measures in practice where only deviation from the targeted value is regarded as a risk. We assume that the denominator is positive for all $x \in X$ in that if there is an x_0 such that the quantity is zero, it means that in all scenarios (or almost surely) the new portfolio is worse than the benchmark, and we will exclude such strategy from our feasible set.

The main issue relates to the above model is the information on the underlying uncertainty. In practice, the true probability distribution of random variable ξ may be unknown, but it may be possible to obtain some partial information to construct an ambiguity set of distributions which approximate the true distribution. We use $\mathcal{P}$ to denote an ambiguity set and consider a distributionally robust utility-based reward-risk ratio (DRURR) optimization model:

$$\sup_{x\in X} \inf_{P\in\mathcal{P}} \frac{\mathbb{E}_P[u(f(x,\xi)) - u(Y(\xi))]}{\mathbb{E}_P[(u(Y(\xi)) - u(f(x,\xi)))_+]}. \tag{2.2.4}$$

In this formulation, the optimal solution is based on the worst probability distribution from $\mathcal{P}$ and it provides a lower bound for the optimal value of problem (2.2.3) if the true probability distribution is contained in $\mathcal{P}$.

To deal with the fractional form of the objective function, we use a variable $\gamma \in \mathbb{R}$ and reformulate problem (2.2.4) as follows:

$$\begin{aligned} &\sup_{x\in X,\gamma\in\mathbb{R}} \quad \gamma \\ &\text{s.t.} \quad \inf_{P\in\mathcal{P}} \mathbb{E}_P[u(f(x,\xi)) - u(Y(\xi)) - \gamma(u(Y(\xi)) - u(f(x,\xi)))_+] \geq 0 \end{aligned} \tag{2.2.5}$$

Compared to (2.2.4), problem (2.2.5) is relatively easier to handle as both the objective and constraint functions are linear w.r.t. $\mathbb{E}_P[\cdot]$. We show the equivalence of (2.2.4) and (2.2.5) in the following proposition.

Proposition 2.2.1 *Problems (2.2.5) and (2.2.4) are equivalent when both have finite optimal value and optimal solutions, that is, if* $\{\gamma^*, (x^*, \gamma^*)\}$ *is a pair of optimal value and optimal solution of problem (2.2.5), then* $\{\gamma^*, x^*\}$ *is a pair of optimal value and optimal solution of problem (2.2.4), and vice versa.*

Proof. We use the convention $\frac{0}{0} = +\infty$. Let $\{\gamma^*, (x^*, \gamma^*)\}$ and $\{\hat{\gamma}, \hat{x}\}$ be a pair of optimal value and optimal solution of problems (2.2.5) and (2.2.4) respectively. Since (x^*, γ^*) is a feasible solution of problem (2.2.5), we have

$$\inf_{P\in\mathcal{P}} \mathbb{E}_P[u(f(x^*,\xi)) - u(Y(\xi)) - \gamma^*(u(Y(\xi)) - u(f(x^*,\xi)))_+] \geq 0. \tag{2.2.6}$$

Next we show (2.2.6) implies

$$\inf_{P\in\mathcal{P}} \frac{\mathbb{E}_P[u(f(x^*,\xi)) - u(Y(\xi))]}{\mathbb{E}_P[(u(Y(\xi)) - u(f(x^*,\xi)))_+]} \geq \gamma^*. \tag{2.2.7}$$

Assume for the sake of a contradiction that (2.2.7) fails to hold. Then there exists a small positive number ϵ and a sequence $\{P^N\} \subset \mathcal{P}$ such that

$$\frac{\mathbb{E}_{P^N}[u(f(x^*,\xi)) - u(Y(\xi))]}{\mathbb{E}_{P^N}[(u(Y(\xi)) - u(f(x^*,\xi)))_+]} \leq \gamma^* - \epsilon \tag{2.2.8}$$

when N is sufficiently large. For a fixed P^N, we have

$$\begin{aligned} & \mathbb{E}_{P^N}[u(f(x^*,\xi)) - u(Y(\xi)) - \gamma^*(u(Y(\xi)) - u(f(x^*,\xi)))_+] \\ \leq \; & -\epsilon\mathbb{E}_{P^N}[(u(Y(\xi)) - u(f(x^*,\xi)))_+]. \end{aligned}$$

If $\mathbb{E}_{P^N}[(u(Y(\xi)) - u(f(x^*,\xi)))_+] > 0$, then the inequality above contradicts to (2.2.6). On the other hand, if $\mathbb{E}_{P^N}[(u(Y(\xi)) - u(f(x^*,\xi)))_+] = 0$, then $\mathbb{E}_{P^N}[u(f(x^*,\xi)) - u(Y(\xi))] \geq 0$, which entails the left hand side of (2.2.8) to be positive infinity. This is not possible because of the boundedness of γ^*. Thus we get a contradiction as desired.

Since x^* is a feasible solution of problem (2.2.4), then

$$\hat{\gamma} \geq \inf_{P\in\mathcal{P}} \frac{\mathbb{E}_P[u(f(x^*,\xi)) - u(Y(\xi))]}{\mathbb{E}_P[(u(Y(\xi)) - u(f(x^*,\xi)))_+]} \geq \gamma^*. \tag{2.2.9}$$

On the other hand, as $\hat{x}$ is an optimal solution of problem (2.2.4), we have

$$\frac{\mathbb{E}_P[u(f(\hat{x},\xi)) - u(Y(\xi))]}{\mathbb{E}_P[(u(Y(\xi)) - u(f(\hat{x},\xi)))_+]} \geq \hat{\gamma}$$

for all $P \in \mathcal{P}$. Multiplying both sides of the above inequality by $\mathbb{E}_P[(u(Y(\xi)) - u(f(\hat{x},\xi)))_+]$ and taking minimum w.r.t. P over $\mathcal{P}$, we obtain

$$\inf_{P\in\mathcal{P}} \mathbb{E}_P[u(f(\hat{x},\xi)) - u(Y(\xi)) - \hat{\gamma}(u(Y(\xi)) - u(f(\hat{x},\xi)))_+] \geq 0,$$

which means $(\hat{x}, \hat{\gamma})$ is a feasible solution of problem (2.2.5). Thus

$$\gamma^* \geq \inf_{P\in\mathcal{P}} \frac{\mathbb{E}_P[u(f(\hat{x},\xi)) - u(Y(\xi))]}{\mathbb{E}_P[(u(Y(\xi)) - u(f(\hat{x},\xi)))_+]} \geq \hat{\gamma} \tag{2.2.10}$$

Combining (2.2.9) and (2.2.10), we have

$$\begin{aligned}\gamma^* &= \inf_{P\in\mathcal{P}} \frac{\mathbb{E}_P[u(f(x^*,\xi)) - u(Y(\xi))]}{\mathbb{E}_P[(u(Y(\xi)) - u(f(x^*,\xi)))_+]} \\ &= \inf_{P\in\mathcal{P}} \frac{\mathbb{E}_P[u(f(\hat{x},\xi)) - u(Y(\xi))]}{\mathbb{E}_P[(u(Y(\xi)) - u(f(\hat{x},\xi)))_+]} \quad = \quad \hat{\gamma}.\end{aligned}$$

The proof is complete. □

To ensure the optimal value and the optimal solutions of problem (2.2.5) to be bounded, we make the following assumption.

Assumption 2.2.1 *Assume that:*

(a) X is a compact convex set;

(b) $\Xi \subset \mathbb{R}^k$ is a compact set;

(c) $f(\cdot,\cdot), Y(\cdot)$ and $u(\cdot)$ are continuous;

(d) there exists a positive number ϵ such that

$$\min_{x\in X} \inf_{P\in\mathcal{P}} \mathbb{E}_P[(u(Y(\xi)) - u(f(x,\xi)))_+] \geq \epsilon. \tag{2.2.11}$$

Parts (a)-(c) of the Assumption 2.2.1 are standard in the literature, see [30] for example. Part (d) provides a sufficient condition for the well-definedness of the robust formulation (2.2.4) and (2.2.5).

Proposition 2.2.2 *Under Assumption 2.2.1, problem (2.2.5) have a finite optimal value.*

Proof. Condition (2.2.11) ensures that for all $P \in \mathcal{P}$ and all $x \in X$

$$\begin{aligned}\left|\frac{\mathbb{E}_P[u(f(x,\xi)) - u(Y(\xi))]}{\mathbb{E}_P[(u(Y(\xi)) - u(f(x,\xi)))_+]}\right| &\leq \frac{1}{\epsilon}|\mathbb{E}_P[u(f(x,\xi)) - u(Y(\xi))]| \\ &\leq \sup_{x\in X}\sup_{P\in\mathcal{P}} \frac{1}{\epsilon}|\mathbb{E}_P[u(f(x,\xi)) - u(Y(\xi))]| \\ &\leq \frac{1}{\epsilon}\sup_{x\in X}\sup_{\xi\in\Xi}|u(f(x,\xi)) - u(Y(\xi))|.\end{aligned}$$

Under Assumption 2.2.1, Ξ and X are compact sets and functions f, Y and u are continuous. Therefore, $\sup_{x\in X}\sup_{\xi\in\Xi}|u(f(x,\xi)) - u(Y(\xi))|$ is bounded and so is the optimal value of (2.2.4). As we have shown in Proposition 2.2.1 that (2.2.4) and (2.2.5) are equivalent, thus problem (2.2.5) has a finite optimal value. □

For the convenience of exposition, we rewrite problem (2.2.5) as a minimization problem:

$$\begin{array}{ll}\inf\limits_{x\in X,\gamma\in\mathbb{R}} & -\gamma \\ \text{s.t.} & \sup\limits_{P\in\mathcal{P}} \mathbb{E}_P[-u(f(x,\xi))+u(Y(\xi))+\gamma(u(Y(\xi))-u(f(x,\xi)))_+] \leq 0.\end{array} \tag{2.2.12}$$

To simplify the notation, we let

$$H(x,\xi,\gamma) := -u(f(x,\xi))+u(Y(\xi))+\gamma(u(Y(\xi))-u(f(x,\xi)))_+$$

and write problem (2.2.12) in a concise form

$$\begin{array}{ll}\inf\limits_{x\in X,\gamma\in\mathbb{R}} & -\gamma \\ \text{s.t.} & \sup\limits_{P\in\mathcal{P}} \mathbb{E}_P[H(x,\xi,\gamma)] \leq 0.\end{array} \tag{2.2.13}$$

Proposition 2.2.3 *Under Assumption 2.2.1, the following assertions hold.*

(i) For each $P\in\mathcal{P}$ and $x\in X$, $\mathbb{E}_P[H(x,\xi,\gamma)]$ is strictly increasing in γ;

(ii) there exists a finite γ^* such that

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P[H(x,\xi,\gamma^*)] \geq 0, \forall x\in X; \tag{2.2.14}$$

(iii) if given a γ^* satisfying (2.2.14) and there exists $x^*\in X$ such that $\mathbb{E}_P[H(x^*,\xi,\gamma^*)] = 0$, then $-\gamma^*$ and (x^*,γ^*) are the optimal value and optimal solution of problem (2.2.13).

Proof. Part (i). Under Assumption 2.2.1 (d), it is easy to see that $\mathbb{E}_P[H(x,\xi,\cdot)]$ is strictly increasing in γ.

Part (ii). Under Assumption 2.2.1 (a)-(c), we have

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P[|-u(f(x,\xi))+u(Y(\xi))|] \leq \max_{x\in X,\xi\in\Xi} |-u(f(x,\xi))+u(Y(\xi))| < +\infty.$$

By Assumption 2.2.1 (d),

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P[\gamma(u(Y(\xi))-u(f(x,\xi)))_+] \geq \gamma\epsilon.$$

Therefore, there must exist a γ^* sufficiently large such that

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P[H(x,\xi,\gamma^*)] \geq 0, \forall x \in X.$$

Part (iii). The equality of (2.2.14) implies (x^*, γ^*) is a feasible solution of problem (2.2.13). By Part (i) and (ii), for any $x \in X$ and any $\delta > 0$

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P[H(x,\xi,\gamma^*+\delta)] > 0,$$

which implies that the optimal value of problem (2.2.13) cannot be smaller than $-\gamma^*$. □

The assertions in Proposition 2.2.3 ensure the boundedness of the optimal solution of problem (2.2.13) and the sufficient condition for optimality.

2.3 Dual Formulation and Entropic Approximation

In this section, we consider a dual formulation for the problem (2.2.13) when the ambiguity set $\mathcal{P}$ is constructed through moments. Specifically, we construct the ambiguity set $\mathcal{P}$ as follows:

$$\mathcal{P} := \left\{ P \in \mathscr{P} : \begin{array}{ll} \mathbb{E}_P[\psi_i(\xi)] = 0, & i = 1,\ldots,p \\ \mathbb{E}_P[\psi_i(\xi)] \leq 0, & i = p+1,\ldots,q \end{array} \right\}, \tag{2.3.1}$$

where $\psi_i : \Xi \to \mathbb{R}, i = 1,\ldots,q$, are continuous functions, and $\mathscr{P}$ denotes the set of probability measures of random variable ξ. For the simplicity of discussion, we restrict $\psi(\xi)$ to be scalar functions and make the following assumption.

Assumption 2.3.1 *Let $\psi_i(\xi)$ be defined as in (2.3.1) and Ξ be the support set of ξ. Let $\psi := (\psi_1,\ldots,\psi_q)$. We assume:*

$$0_q \in \text{int}\{\mathbb{E}_P[\psi(\xi)] : P \in \mathscr{P}\} - \mathcal{K},$$

where 'int' denotes the interior of a set, $\mathcal{K} := 0_p \times \mathbb{R}_+^{q-p}$, 0_q is a zero vector with q dimensions and 0_p is a zero vector with p dimensions.

The assumption is a standard condition for deriving Lagrange dual of moment problems, see Xu et al. [86, Proposition 2.1]. Under Assumption 2.3.1, we can reformulate problem (2.2.13) as the following semi-infinite programming problem (see Appendix

A):

$$\begin{array}{lll} \inf\limits_{x\in X,\gamma\in\mathbb{R},\lambda\in\mathbb{R}^p\times\mathbb{R}_+^{q-p}} & -\gamma & \\ \text{s.t.} & \sup\limits_{\xi\in\Xi} H(x,\xi,\gamma) - \sum\limits_{i=1}^{q}\lambda_i\psi_i(\xi) \leq 0. \end{array} \tag{2.3.2}$$

If Assumption 2.2.1 holds, then the optimal value of the primal and the dual problems is finite. In addition, by [72, Proposition 3.4], the set of optimal solutions is nonempty and bounded.

For the convenience of exposition, we let

$$T(x,\gamma,\lambda,\xi) := H(x,\xi,\gamma) - \sum_{i=1}^{q}\lambda_i\psi_i(\xi) \tag{2.3.3}$$

and rewrite the semi-infinite programming problem (2.3.2) as follows:

$$\begin{array}{ll} \inf\limits_{x\in X,\gamma\in\mathbb{R},\lambda\in\mathbb{R}^p\times\mathbb{R}_+^{q-p}} & -\gamma \\ \text{s.t.} & \sup\limits_{\xi\in\Xi} T(x,\gamma,\lambda,\xi) \leq 0. \end{array} \tag{2.3.4}$$

In what follows, we consider an approximation of the constraint in (2.3.4) through entropic risk measure. The entropic risk measure is defined as

$$e_\alpha(Z) := \frac{1}{\alpha}\ln\mathbb{E}_P[e^{-\alpha Z}],$$

where $Z \in \mathscr{L}^\infty(\Xi,\mathcal{F},P)$ [1] is a random variable and α is a positive number. It is well known that $e_\alpha(Z)$ is monotonically increasing in α and

$$\lim_{\alpha\to+\infty} e_\alpha(Z) = \operatorname{ess\,sup}(-Z),$$

where 'ess sup' denotes essential supremum of the random variable, see [33] for a thorough treatment of the subject.

The following lemma states the uniform approximation of entropic risk measure for a general class of random functions.

Lemma 2.3.1 *([55, Proposition 2.1]) Let $g : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$ be a continuous function and X be a subset of $\mathbb{R}^n$. Let ξ be a random variable on the probability space $(\Xi,\mathcal{F},P)$ with support set $\Xi \subset \mathbb{R}^k$. Let $G(x)$, $F_x(\cdot)$ and Ξ_x denote the essential supremum, the cumulative distribution function and the support set of $-g(x,\xi)$ respectively, that is, Ξ_x is the smallest set satisfied $P(-g(x,\xi)\in\Xi_x) = 1$. Let* $\mathrm{Diam}(\Xi_x)$ *denote the diameter*

[1] $\mathscr{L}^\infty(\Xi,\mathcal{F},P)$ denotes the set of essentially bounded measurable functions.

of Ξ_x*. Assume: (a)* $X \subset \mathbb{R}^n$ *is a compact set, (b) for each fixed* $x \in X$*,*

$$\inf_{\xi \in \Xi} g(x, \xi) > -\infty.$$

Then for each fixed $x \in X$*,*

$$\lim_{\alpha \to +\infty} e_\alpha(g(x, \xi)) = G(x).$$

Assume in addition: (c)

$$\inf_{x \in X} \inf_{\xi \in \Xi} g(x, \xi) > -\infty,$$

(d) for any fixed small positive number ϵ*, there exists* $\delta(\epsilon) \in (0, 1)$ *such that*

$$F_x(G(x) - \epsilon) \leq 1 - \delta(\epsilon), \ \forall x \in X_\epsilon,$$

where $X_\epsilon := \{x \in X : \mathrm{Diam}(\Xi_x) > 2\epsilon\}$*. Then*

$$|e_\alpha(g(x, \xi)) - G(x)| < 2\epsilon + \left| \frac{1}{\alpha} \ln \delta(\epsilon) \right|. \tag{2.3.5}$$

By Lemma 2.3.1, we consider an approximation of the constraint in problem (2.3.4) as follows:

$$e_\alpha(-T(x, \gamma, \lambda, \xi)) := \frac{1}{\alpha} \ln \mathbb{E}[e^{\alpha T(x, \gamma, \lambda, \xi)}] \leq 0. \tag{2.3.6}$$

Note that the expectation $\mathbb{E}[\cdot]$ in (2.3.6) is different from the expectation $\mathbb{E}_P[\cdot]$ in the preceding section. From Lemma 2.3.1, we can see the conclusion holds for any probability distribution of $\hat{\xi}$ with support set Ξ. However, in problem (2.2.3), we do not assume any knowledge of the true probability distriution P except the support set Ξ.

With (2.3.6), we can construct an approximation of problem (2.3.4) as follows:

$$\begin{aligned} \inf_{x \in X, \gamma \in \mathbb{R}, \lambda \in \mathbb{R}^p \times \mathbb{R}_+^{q-p}} \quad & -\gamma \\ \text{s.t.} \quad & e_\alpha(-T(x, \gamma, \lambda, \xi)) \leq 0 \end{aligned} \tag{2.3.7}$$

Since the set of optimal solutions to problem (2.3.4) is nonempty and bounded under Assumption 2.2.1, we may restrict the variables $\lambda_i, i = 1, \ldots, q$ in problem (2.3.7) to take finite values. Specifically, we assume that there exists a positive constant C_0 such that

$$|\lambda_i| \leq C_0, \ i = 1, \ldots, q.$$

The following proposition states some important properties of the feasible sets and the optimal values of problems (2.3.4) and (2.3.7).

Proposition 2.3.1 *Let $\mathcal{F}$ and ϑ denote the feasible set and the optimal value of problem (2.3.4) respectively. Likewise, define $\mathcal{F}(\alpha)$ and $\vartheta(\alpha)$ for problem (2.3.7). Then the following assertions hold.*

(i) $\mathcal{F} \subset \mathcal{F}(\alpha)$ *for all* $\alpha > 0$*;*

(ii) $\mathcal{F}(\alpha)$ *is monotonically decreasing, that is,* $\alpha_1 < \alpha_2$, $\mathcal{F}(\alpha_2) \subset \mathcal{F}(\alpha_1)$*;*

(iii) $\vartheta(\alpha)$ *is non-decreasing and* $\vartheta(\alpha) \leq \vartheta$*;*

(iv) if Assumptions 2.2.1 and 2.3.1 hold and there exists a positive number ϵ *such that*

$$\min_{x\in X} \mathbb{E}[(u(Y(\xi)) - u(f(x,\xi)))_+] \geq \epsilon, \tag{2.3.8}$$

then both ϑ *and* $\vartheta(\alpha)$ *are finite, and the set of optimal solutions to (2.3.7) is nonempty and bounded.*

Proof. Part (i). Since $e_\alpha(-T(x,\gamma,\lambda,\xi)) \leq \sup_{\xi\in\Xi} T(x,\gamma,\lambda,\xi)$, then $\mathcal{F} \subset \mathcal{F}(\alpha)$.

Part (ii). It follows form the fact that for any fixed x, γ, λ, $e_\alpha(-T(x,\gamma,\lambda,\xi))$ increases in α.

Part (iii). It follows from Part (i) and Part (ii).

Part (iv). Under Assumption 2.2.1, as we have discussed above, ϑ is finite. By Jensen's inequality,

$$\mathbb{E}[e^{\alpha T(x,\gamma,\lambda,\xi)}] \geq e^{\alpha \mathbb{E}[T(x,\gamma,\lambda,\xi)]}.$$

Thus we have

$$e_\alpha(-T(x,\gamma,\lambda,\xi)) = \frac{1}{\alpha}\ln \mathbb{E}[e^{\alpha T(x,\gamma,\lambda,\xi)}] \geq \frac{1}{\alpha}\ln e^{\alpha \mathbb{E}[T(x,\gamma,\lambda,\xi)]}.$$

In addition, we have

$$\mathbb{E}[T(x,\gamma,\lambda,\xi)] = \mathbb{E}[H(x,\xi,\gamma)] - \sum_{i=1}^{q} \lambda_i \mathbb{E}[\psi_i(\xi)]$$

and

$$\begin{aligned} \mathbb{E}[H(x,\xi,\gamma)] &= -\mathbb{E}[u(f(x,\xi))] + \mathbb{E}[u(Y(\xi))] \\ &\quad +\gamma \mathbb{E}[(u(Y(\xi)) - u(f(x,\xi)))_+] \\ &\geq -\mathbb{E}[u(f(x,\xi))] + \mathbb{E}[u(Y(\xi))] + \gamma\epsilon. \end{aligned}$$

Since $|\lambda_i| \leq C_0$ and $\psi_i(\cdot)$ is continuous on Ξ, for $i = 1, \ldots, q$, then the inequality above means $\mathbb{E}[T(x,\gamma,\lambda,\xi)] \to +\infty$ as $\gamma \to +\infty$, and hence

$$e_\alpha(-T(x,\gamma,\lambda,\xi)) = \frac{1}{\alpha}\ln \mathbb{E}[e^{\alpha T(x,\gamma,\lambda,\xi)}] \to +\infty.$$

It shows that a large γ will violate the constraint of problem (2.3.7) regardless of the value of α, which means γ must be bounded at its optimum. The nonemptiness and boundedness of the set of optimal solutions to (2.3.7) follow from the boundedness of γ and the fact that the other variables of the problem are restricted to take a value from a compact set. □

In what follows, we consider the approximation of (2.3.7) to (2.3.4) in terms of the feasible sets and the optimal values as α changes. For the convenience of exposition, we let

$$\mathfrak{M} := X \times \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}_+^{q-p} \text{ and } \mathfrak{m} := (x, \gamma, \lambda) \in \mathfrak{M},$$

then $T(x, \gamma, \lambda, \xi)$ can be written as $T(\mathfrak{m}, \xi)$. The below error bound condition [62] will be used in the following stability analysis.

Assumption 2.3.2 *There exist positive constants C and δ such that*

$$d(\mathfrak{m}, \mathcal{F}) \leq C \left(\sup_{\xi \in \Xi} T(\mathfrak{m}, \xi) \right)_+ \tag{2.3.9}$$

for any $\mathfrak{m} \in \mathfrak{M}$ satisfying $d(\mathfrak{m}, \mathcal{F}) \leq \delta$.

Note that For $\mathfrak{m} \in \mathcal{F}$, $\sup_{\xi \in \Xi} T(\mathfrak{m}, \xi) \leq 0$, the constraint (2.3.9) holds trivially. Thus the quantity $(\sup_{\xi \in \Xi} T(\mathfrak{m}, \xi))_+$ describes the significance of constraint violation when $\mathfrak{m} \notin \mathcal{F}$.

Theorem 2.3.1 *Assume: (a) $\mathcal{F}$ is a compact set; (b) $\psi_i, i = 1, \ldots, q$, is continuous; (c) Assumption 2.2.1 and the conditions of Lemma 2.3.1 hold for function $T(\mathfrak{m}, \xi)$. Then*

(i) for any $\epsilon > 0$, there exists a positive number α_0 such that

$$\mathbb{H}(\mathcal{F}(\alpha), \mathcal{F}) \leq \epsilon, \forall \alpha \in [\alpha_0, +\infty);$$

(ii) if Assumption 2.3.2 holds, then there exist positive constants C and α^* such that

$$\mathbb{H}(\mathcal{F}(\alpha), \mathcal{F}) \leq C \sup_{\mathfrak{m} \in \mathfrak{M}} \Delta_\alpha(\mathfrak{m}), \forall \alpha \in [\alpha^*, +\infty), \tag{2.3.10}$$

where

$$\Delta_\alpha(\mathfrak{m}) := \left(\sup_{\xi \in \Xi} T(\mathfrak{m}, \xi) - e_\alpha(-T(\mathfrak{m}, \xi)) \right);$$

(iii)

$$|\vartheta(\alpha) - \vartheta| \leq C \sup_{\mathfrak{m} \in \mathfrak{M}} \Delta_\alpha(\mathfrak{m}), \forall \alpha \in [\alpha^*, +\infty).$$

Proof. By Proposition 2.3.1 (i), $\mathbb{D}(\mathcal{F}, \mathcal{F}(\alpha)) = 0$, therefore in Parts (i) and (ii), we only need to show the inequalities hold for $\mathbb{D}(\mathcal{F}(\alpha), \mathcal{F})$.

Part (i). Let ϵ be a fixed small positive number. Define

$$F(\epsilon) := \inf_{\substack{\mathfrak{m}\in\mathfrak{M} \\ d(\mathfrak{m},\mathcal{F})\geq\epsilon}} \sup_{\xi\in\Xi} T(\mathfrak{m}, \xi).$$

Then $F(\epsilon) > 0$. Let $\delta := F(\epsilon)/2$. By Lemma 2.3.1, there exists a positive number α_0 such that

$$\sup_{\mathfrak{m}\in\mathfrak{M}} \left[\sup_{\xi\in\Xi} T(\mathfrak{m}, \xi) - e_\alpha(-T(\mathfrak{m}, \xi))\right] \leq \delta,$$

for $\alpha > \alpha_0$. For any $\mathfrak{m} \in \mathfrak{M}$ with $d(\mathfrak{m}, \mathcal{F}) \geq \epsilon$,

$$\begin{array}{rcl} e_\alpha(-T(\mathfrak{m}, \xi)) & = & \sup_{\xi\in\Xi} T(\mathfrak{m}, \xi) + e_\alpha(-T(\mathfrak{m}, \xi)) - \sup_{\xi\in\Xi} T(\mathfrak{m}, \xi) \\ & \geq & F(\epsilon) - F(\epsilon)/2 = F(\epsilon)/2 > 0, \end{array}$$

which implies $\mathfrak{m} \notin \mathcal{F}(\alpha)$. It shows that for every $\mathfrak{m} \in \mathcal{F}(\alpha)$, we have $d(\mathfrak{m}, \mathcal{F}) < \epsilon$, that is, $\mathbb{D}(\mathcal{F}(\alpha), \mathcal{F}) \leq \epsilon$.

Part (ii). Under Assumpion 2.3.2, it follows by Part (i) that there exists a sufficiently large α^* such that

$$d(\mathfrak{m}, \mathcal{F}) \leq C \left(\sup_{\xi\in\Xi} T(\mathfrak{m}, \xi)\right)_+$$

for all $\mathfrak{m} \in \mathcal{F}(\alpha)$ when $\alpha \geq \alpha^*$. Since $\mathfrak{m} \in \mathcal{F}(\alpha)$ is equivalent to $e_\alpha(-T(\mathfrak{m}, \xi)) \leq 0$, then for any $\mathfrak{m} \in \mathcal{F}(\alpha)$,

$$\begin{array}{rcl} d(\mathfrak{m}, \mathcal{F}) & \leq & C\left(\sup_{\xi\in\Xi} T(\mathfrak{m}, \xi)\right)_+ - C\left(e_\alpha(-T(\mathfrak{m}, \xi))\right)_+ \\ & \leq & C\left(\sup_{\xi\in\Xi} T(\mathfrak{m}, \xi) - e_\alpha(-T(\mathfrak{m}, \xi))\right)_+ \\ & = & C\left(\sup_{\xi\in\Xi} T(\mathfrak{m}, \xi) - e_\alpha(-T(\mathfrak{m}, \xi))\right) \\ & \leq & C \sup_{\mathfrak{m}\in\mathfrak{M}} \left(\sup_{\xi\in\Xi} T(\mathfrak{m}, \xi) - e_\alpha(-T(\mathfrak{m}, \xi))\right), \end{array}$$

where the second inequality follows from the fact $(a)_+ - (b)_+ \leq (a - b)_+$, the equality follows from the fact $\sup_{\xi\in\Xi} T(\mathfrak{m}, \xi) \geq e_\alpha(-T(\mathfrak{m}, \xi))$ for any $\mathfrak{m} \in \mathfrak{M}$. It shows

$$\mathbb{D}(\mathcal{F}(\alpha), \mathcal{F}) \leq C \sup_{\mathfrak{m}\in\mathfrak{M}} \left(\sup_{\xi\in\Xi} T(\mathfrak{m}, \xi) - e_\alpha(-T(\mathfrak{m}, \xi))\right).$$

Part (iii). Let $\mathfrak{m}^*$ and $\mathfrak{m}_\alpha$ be an optimal solution of problem (2.3.4) and problem (2.3.7) respectively. Let $\gamma*$ and γ_α be the corresponding second component of $\mathfrak{m}^*$ and $\mathfrak{m}_\alpha$ respectively. Then $\vartheta = -\gamma^*$ and $\vartheta(\alpha) = -\gamma_\alpha$. By Part (ii), there exists $\bar{\mathfrak{m}} \in \mathcal{F}$ such that

$$\|\mathfrak{m}_\alpha - \bar{\mathfrak{m}}\| \leq C \sup_{\mathfrak{m}\in\mathfrak{M}} \Delta_\alpha(\mathfrak{m}).$$

Let $\bar{\gamma}$ be the corresponding second component of $\bar{\mathfrak{m}}$. Then $\vartheta = -\gamma^* \leq -\bar{\gamma}$. Thus we have

$$\vartheta \leq -\bar{\gamma} \leq -\gamma_\alpha + |\gamma_\alpha - \bar{\gamma}| \leq \vartheta(\alpha) + C \sup_{\mathfrak{m}\in\mathfrak{M}} \Delta_\alpha(\mathfrak{m}).$$

Exchange the role of $\mathfrak{m}_\alpha$ and $\mathfrak{m}^*$ under the symmetry of Hausdorff distance between $\mathcal{F}(\alpha)$ and $\mathcal{F}$, we have

$$\vartheta(\alpha) \leq \vartheta + C \sup_{\mathfrak{m}\in\mathfrak{M}} \Delta_\alpha(\mathfrak{m}).$$

The conclusion follows. □

Under condition (c), it follows by Lemma 2.3.1 that $\Delta_\alpha(\mathfrak{m})$ goes to 0 uniformly for all $\mathfrak{m} \in \mathfrak{M}$ as $\alpha \to +\infty$, thus Theorem 2.3.1 states that $|\vartheta(\alpha) - \vartheta| \to 0$.

2.4 Specific Case of Ambiguity Set

In order to see how the proposed framework of modelling works, in this section we investigate a specific case that the ambiguity set is determined by the mean and covariance. Unlike the work of Popescu [64] that assuming complete information of the mean and covariance, we consider some degree of uncertainty for the two quantities. Specifically, based on the study by Tütüncü and Koenig [82], we consider box constraints for the mean and covariance, which restricts each component of the two quantities to an interval with finite lower and upper bound. The ambiguity set is defined as follows:

$$\mathcal{P}^* := \left\{ P \in \mathscr{P} : \begin{array}{l} -\epsilon \leq (\mathbb{E}_P[\xi] - \bar{\mu})_i \leq \epsilon, \ i = 1, \ldots, m \\ \|\mathbb{E}_P[(\xi - \bar{\mu})(\xi - \bar{\mu})^T] - \bar{\Sigma}\|_\infty \leq \sigma \end{array} \right\}, \tag{2.4.1}$$

where $\bar{\mu}$ and $\bar{\Sigma}$ denote the mean and covariance matrix respectively, $\|A\|_\infty = \max |a_{ij}|$.

The main advantage of such construction is that the ambiguity set (2.4.1) can be fitted into the framework of the preceding moment problem so that we can solve the latter by a general numerical scheme (to be detailed in Section 2.5). Specifically, we recast (2.4.1) as follows:

$$\mathcal{P}^* := \left\{ P \in \mathscr{P} : \begin{array}{ll} -\epsilon \leq \mathbb{E}_P[\psi_i(\xi)] \leq \epsilon, & i = 1, \ldots, p \\ -\sigma \leq \mathbb{E}_P[\psi_j(\xi)] \leq \sigma, & j = p+1, \ldots, q \end{array} \right\}. \tag{2.4.2}$$

where p is the dimension of random ξ, $q = \frac{p^2+3p}{2}$, $\psi_i(\xi) = \xi_i - \bar{\mu}_i$ and $\psi_j(\xi)$, $j = p+1, \ldots, q$ are the elements of the upper triangular of matrix $(\xi - \bar{\mu})(\xi - \bar{\mu})^T - \bar{\Sigma}$. Compared with the semi-definite constraint which specifies the property of the centered covariance of random variable, it may be more convenient to give a lower and upper bound for each component of the centered covariance. In some cases, the covariance between two random variables is precisely known, and the box constraint also allows different σ value to be set for each component of the covariance matrix.

In practice we may use samples to construct an estimate of the true mean and covariance. Let $\xi^1, \ldots, \xi^N$ be an independent and identically distributed sample of ξ and

$$\mu^N := \frac{1}{N}\sum_{s=1}^{N}\xi^s, \quad \Sigma^N := \frac{1}{N}\sum_{s=1}^{N}(\xi^s - \mu^N)(\xi^s - \mu^N)^T.$$

Then we construct a sample based ambiguity set as follows:

$$\mathcal{P}^N := \left\{ P \in \mathscr{P} : \begin{array}{ll} -\epsilon \le \mathbb{E}_P[\psi_i^N(\xi)] \le \epsilon, & i = 1, \ldots, p \\ -\sigma \le \mathbb{E}_P[\psi_j^N(\xi)] \le \sigma, & j = p+1, \ldots, q \end{array} \right\}. \tag{2.4.3}$$

where p and q are defined as above, $\psi_i^N(\xi) = \xi_i - \mu_i^N$ and $\psi_j^N(\xi)$, $j = p+1, \ldots, q$ are the elements of the upper triangular of matrix $(\xi - \mu^N)(\xi - \mu^N)^T - \Sigma^N$.

The next work is to address some theoretical questions: (i) does $\mathcal{P}^N$ converge to $\mathcal{P}^*$ as the sample size increases? (ii) is the true probability distribution of ξ lies in $\mathcal{P}^N$? (iii) does the optimal value and the optimal solutions obtained on the basis of $\mathcal{P}^N$ converge to their true counterpart?

We first address question (i). Let us recall the definition of total variation metric based on [10] for forthcoming discussion.

Definition 2.4.1 *(Total variation metric) Let $P, Q \in \mathscr{P}$ and $\mathscr{M}$ denote the set of measurable functions defined in the probability space $(\Xi, \mathcal{B})$. The total variation metric between P and Q is defined as*

$$d_{TV}(P, Q) := \sup_{g \in \mathscr{M}} (\mathbb{E}_P[g(\xi)] - \mathbb{E}_Q[g(\xi)]),$$

where $\mathscr{M} := \{g : \mathbb{R}^k \to \mathbb{R} \mid g \text{ is } \mathcal{B} \text{ measurable}, \sup_{\xi \in \Xi} |g(\xi)| \le 1\}$.

Using the total variation metric, we can define the distance from a point to a set, deviation form one set to another and Hausdorff distance between two sets in the space of $\mathscr{P}$ as follows:

$$d_{TV}(Q, \mathcal{P}^*) := \inf_{P \in \mathcal{P}^*} d_{TV}(Q, P),$$

$$\mathbb{D}_{TV}(\mathcal{P}^N, \mathcal{P}^*) := \sup_{Q \in \mathcal{P}^N} d_{TV}(Q, \mathcal{P}^*),$$

$$\mathbb{H}_{TV}(\mathcal{P}^N, \mathcal{P}^*) := \max\{\mathbb{D}_{TV}(\mathcal{P}^N, \mathcal{P}^*), \mathbb{D}_{TV}(\mathcal{P}^*, \mathcal{P}^N)\}.$$

Proposition 2.4.1 *Suppose that Ξ is a compact set. Then $\mathbb{H}_{TV}(\mathcal{P}^N, \mathcal{P}^*) \to 0$ as $N \to +\infty$.*

Proof. When Ξ is a compact set, both $\mathcal{P}^N$ and $\mathcal{P}^*$ are compact sets under the total variation metric. For any given $Q \in \mathcal{P}^N$, it follows by [80, Lemma 4.1] that there exists a positive constant C^* such that

$$\begin{aligned}
\mathbb{D}_{TV}(Q, \mathcal{P}^*) &\leq C^*(\|(\mathbb{E}_Q[\Psi_I(\xi)] - \epsilon e)_+\| + \|(-\mathbb{E}_Q[\Psi_I(\xi)] - \epsilon e)_+\|) \\
&\quad + C^*(\|(\mathbb{E}_Q[\Psi_J(\xi)] - \sigma e)_+\| + \|(-\mathbb{E}_Q[\Psi_J(\xi)] - \sigma e)_+\|) \\
&\leq C^*(\|(\mathbb{E}_Q[\Psi_I(\xi) - \Psi_I^N(\xi)])_+\| + \|(\mathbb{E}_Q[\Psi_I^N(\xi) - \Psi_I(\xi)])_+\|) \\
&\quad + C^*(\|(\mathbb{E}_Q[\Psi_J(\xi) - \Psi_J^N(\xi)])_+\| + \|(\mathbb{E}_Q[\Psi_J^N(\xi) - \Psi_J(\xi)])_+\|),
\end{aligned} \tag{2.4.4}$$

where

$$\Psi_I := \begin{bmatrix} \psi_1 \\ \vdots \\ \psi_p \end{bmatrix}, \Psi_J := \begin{bmatrix} \psi_{p+1} \\ \vdots \\ \psi_q \end{bmatrix}, \Psi_I^N := \begin{bmatrix} \psi_1^N \\ \vdots \\ \psi_p^N \end{bmatrix}, \Psi_J^N := \begin{bmatrix} \psi_{p+1}^N \\ \vdots \\ \psi_q^N \end{bmatrix},$$

for a vector a, $(a)_+ = \max\{a, 0\}$ with the maximum being taken componentwise. The first inequality of (2.4.4) means that a probability measure $Q \in \mathscr{P}$ deviating form $\mathcal{P}^*$ under the total variation metric is linearly bounded by the residual of the system of equalities and inequalities defining $\mathcal{P}^*$. The second inequality follows from the facts that $Q \in \mathcal{P}^N$ and $(a)_+ \leq (a-b)_+$ if $b \leq 0$. Likewise, for any given $Q \in \mathcal{P}^*$,

$$\begin{aligned}
\mathbb{D}_{TV}(Q, \mathcal{P}^N) &\leq C^*(\|(\mathbb{E}_Q[\Psi_I^N(\xi)] - \epsilon e)_+\| + \|(-\mathbb{E}_Q[\Psi_I^N(\xi)] - \epsilon e)_+\|) \\
&\quad + C^*(\|(\mathbb{E}_Q[\Psi_J^N(\xi)] - \sigma e)_+\| + \|(-\mathbb{E}_Q[\Psi_J^N(\xi)] - \sigma e)_+\|) \\
&\leq C^*(\|(\mathbb{E}_Q[\Psi_I^N(\xi) - \Psi_I(\xi)])_+\| + \|(\mathbb{E}_Q[\Psi_I(\xi) - \Psi_I^N(\xi)])_+\|) \\
&\quad + C^*(\|(\mathbb{E}_Q[\Psi_J^N(\xi) - \Psi_J(\xi)])_+\| + \|(\mathbb{E}_Q[\Psi_J(\xi) - \Psi_J^N(\xi)])_+\|).
\end{aligned}$$

Since $\mu^N \to \bar{\mu}$ and $\Sigma^N \to \bar{\Sigma}$, it is easy to see that Ψ_I^N and Ψ_J^N converge to Ψ_I and Ψ_J uniformly over Ξ as $N \to \infty$. Thus both $\mathbb{D}_{TV}(Q, \mathcal{P}^*)$ and $\mathbb{D}_{TV}(Q, \mathcal{P}^N)$ converge to zero as $N \to \infty$, which implies $\mathbb{H}_{TV}(\mathcal{P}^N, \mathcal{P}^*) \to 0$. □

To address question (ii), we give the following lemma based on [78].

Lemma 2.4.1 *If ξ is essentially bounded by a positive number β, then for any given positive small number δ,*

(i) with probability at least $1 - \delta$ over the choice of the samples of ξ,

$$\|\mathbb{E}[\xi] - \mu^N\|_\infty \leq \frac{\beta}{\sqrt{N}}\left(2 + \sqrt{2 \ln \frac{1}{\delta}}\right);$$

(ii) with probability at least $1 - 2\delta$ over the choice of the samples of ξ,

$$\|\mathbb{E}[(\xi - \mu^N)(\xi - \mu^N)^T] - \Sigma^N\|_\infty \leq \frac{3\beta^2}{\sqrt{N}}\left(2 + \sqrt{2 \ln \frac{1}{\delta}}\right).$$

Proof. Part (i). It follows straightforwardly from [78, Theorem 3].

Part (ii). Let

$$S^N := \frac{1}{N}\sum_{i=1}^{N} \xi_i \xi_i^T \text{ and } S := \mathbb{E}[\xi\xi^T].$$

Then

$$\begin{aligned}
& \|\mathbb{E}[(\xi-\mu^N)(\xi-\mu^N)^T] - \Sigma^N\|_\infty \\
= & \|\mathbb{E}[(\xi-\mu^N)(\xi-\mu^N)^T] - \tfrac{1}{N}\textstyle\sum_{i=1}^N(\xi_i-\mu^N)(\xi_i-\mu^N)^T\|_\infty \\
= & \|S - 2\mathbb{E}[\xi](\mu^N)^T + \mu^N(\mu^N)^T - S^N + \mu^N(\mu^N)^T\|_\infty \\
\le & \|S - S^N\|_\infty + 2\|\mathbb{E}[\xi](\mu^N)^T - \mu^N(\mu)^T\|_\infty \\
\le & \|S - S^N\|_F + 2\|\mathbb{E}[\xi] - \mu^N\| \cdot \|(\mu^N)^T\|.
\end{aligned}$$

Note that for any two events A and B, Bonferroni's inequality states that

$$P(AB) \ge P(A) + P(B) - 1.$$

Thus, it follows by [78, Corollary 5] and Part (i) that with probability at least $1-2\delta$

$$\begin{aligned}
& \|\mathbb{E}[(\xi-\mu^N)(\xi-\mu^N)^T] - \Sigma^N\|_\infty \\
\le & \tfrac{\beta^2}{\sqrt{N}}\left(2+\sqrt{2\ln\tfrac{1}{\delta}}\right) + \tfrac{2\beta^2}{\sqrt{N}}\left(2+\sqrt{2\ln\tfrac{1}{\delta}}\right) \\
= & \tfrac{3\beta^2}{\sqrt{N}}\left(2+\sqrt{2\ln\tfrac{1}{\delta}}\right),
\end{aligned}$$

where the inequality indicates that with high probability the covariance matrix of the randomly generated sample gives a good estimation of $\mathbb{E}[(\xi-\mu^N)(\xi-\mu^N)^T]$ in a way that does not depend on the dimension of the feature space. □

Theorem 2.4.1 *Suppose that ξ is essentially bounded by a positive number β and the parameters ϵ and σ are chosen as follows:*

$$\epsilon_N := \frac{\beta}{\sqrt{N}}\left(2+\sqrt{2\ln\frac{1}{\delta}}\right),\ \sigma_N := \frac{3\beta^2}{\sqrt{N}}\left(2+\sqrt{2\ln\frac{1}{\delta}}\right).$$

Then with probability at least $1-3\delta$ over the choice of the samples of ξ, the true distribution of ξ lies in the ambiguity set $\mathcal{P}^N$.

Proof. It follows from Bonferroni's inequality and Lemma 2.4.1 directly. □

Now we address question (iii). By the ambiguity set $\mathcal{P}^N$, we can derive the dual formulation of problem (2.2.13) coupled by the entropic approximation as follows:

$$\begin{array}{ll}
\inf\limits_{x\in X,\gamma\in\mathbb{R},\lambda\in\mathbb{R}^p\times\mathbb{R}_+^{q-p}} & -\gamma \\
\text{s.t.} & e_\alpha(-T^N(x,\gamma,\lambda,\xi)) \le 0,
\end{array} \quad (2.4.5)$$

where $T^N(x,\gamma,\lambda,\xi) := H(x,\xi,\gamma) - \sum_{i=1}^q \lambda_i \psi_i^N(\xi)$.

Let $\mathcal{F}$, S and ϑ denote the feasible set, the set of optimal solutions and the optimal value of problem (2.3.4) respectively. Likewise, we define $\mathcal{F}(\alpha)$, $S(\alpha)$ and $\vartheta(\alpha)$ for problem (2.3.7) and $\mathcal{F}^N(\alpha)$, $S^N(\alpha)$ and $\vartheta^N(\alpha)$ for problem (2.4.5). Let $\mathcal{F}^s(\alpha)$ denote the set of strictly feasible solutions [2] of problem (2.3.7).

Theorem 2.4.2 *Assume: (a) A is any compact set such that $A \cap S(\alpha) \neq \emptyset$ and $A \cap S^N(\alpha) \neq \emptyset$; (b) $\mathrm{cl}\mathcal{F}^s(\alpha) \cap \mathrm{S}(\alpha) \neq \emptyset$. Then*

(i) $\limsup_{N\to+\infty} S^N(\alpha) \cap A \subseteq S(\alpha) \cap A$ *and* $\lim_{N\to+\infty} \vartheta^N(\alpha) = \vartheta(\alpha)$*;*

(ii) if Assumption 2.3.2 holds, there exist $\hat{N}$ and $\hat{\alpha}$ sufficiently large such that for any $N \geq \hat{N}$ and $\gamma \geq \hat{\gamma}$,

$$\mathbb{D}(\mathcal{F}^N(\alpha), \mathcal{F}) \leq C \sup_{\mathfrak{m}\in\mathfrak{M}} \Delta_\alpha^N(\mathfrak{m})$$

and

$$\vartheta - \vartheta^N(\alpha) \leq C \sup_{\mathfrak{m}\in\mathfrak{M}} \Delta_\alpha^N(\mathfrak{m}),$$

where C is a positive constant and

$$\Delta_\alpha^N(\mathfrak{m}) := \left(\sup_{\xi\in\Xi} T(\mathfrak{m},\xi) - e_\alpha(-T^N(\mathfrak{m},\xi)) \right).$$

Proof. Part (i). Under condition (a), since $\Psi_I^N(\cdot)$ and $\Psi_J^N(\cdot)$ converge to $\Psi_I(\cdot)$ and $\Psi_J(\cdot)$ uniformly on Ξ, it is easy to show that $\limsup_{N\to+\infty} \mathcal{F}^N(\alpha) \subseteq \mathcal{F}(\alpha)$. We assume $(x^N,\gamma^N,\lambda^N) \in S^N(\alpha)$, $(x^N,\gamma^N,\lambda^N) \to (x^*,\gamma^*,\lambda^*) \in \mathcal{F}(\alpha)$ as $N \to +\infty$, which implies

$$\lim_{N\to+\infty} \vartheta^N(\alpha) = \lim_{N\to+\infty} -\gamma^N = -\gamma^* \geq \vartheta(\alpha). \tag{2.4.6}$$

Under condition (b), there exists $(\bar{x},\bar{\gamma},\bar{\lambda}) \in S(\alpha)$ and a sequence $\{(\hat{x}^s,\hat{\gamma}^s,\hat{\lambda}^s)\} \subset \mathcal{F}^s(\alpha)$, such that $(\hat{x}^s,\hat{\gamma}^s,\hat{\lambda}^s) \to (\bar{x},\bar{\gamma},\bar{\lambda})$, which implies that for any small positive number $\varepsilon > 0$, there exists a point $(\hat{x}^s,\hat{\gamma}^s,\hat{\lambda}^s) \in \mathcal{F}^s(\alpha)$ such that

$$-\hat{\gamma}^s - \vartheta(\alpha) = -\hat{\gamma}^s + \bar{\gamma} \leq \varepsilon.$$

Moreover, since $(\hat{x}^s,\hat{\gamma}^s,\hat{\lambda}^s) \in \mathcal{F}^s(\alpha)$ and $\Psi_I^N(\cdot)$ and $\Psi_J^N(\cdot)$ converge to $\Psi_I(\cdot)$ and $\Psi_J(\cdot)$ uniformly on Ξ as $N \to \infty$, then $(\hat{x}^s,\hat{\gamma}^s,\hat{\lambda}^s) \in \mathcal{F}^N(\alpha)$ for N sufficiently large. It implies that

$$\vartheta(\alpha) \geq -\hat{\gamma}^s - \varepsilon \geq -\gamma^N - \varepsilon = \vartheta^N(\alpha) - \varepsilon.$$

By deriving N to infinity, we arrive at $\vartheta(\alpha) \geq -\gamma^* - \varepsilon$, which implies $\vartheta(\alpha) \geq -\gamma^*$ in that ε can be chosen arbitrarily small. Together with (2.4.6), we have $\vartheta(\alpha) = -\gamma^*$,

[2] Strictly feasible solution is an interior point in $\mathfrak{M}$ such that the constraint of (2.3.7) holds strictly.

which implies $(x^*, \gamma^*, \lambda^*) \in S(\alpha)$, and it also shows $\lim_{N\to+\infty} \vartheta^N(\alpha) = \vartheta(\alpha)$.

Part (ii). By Part (i) and Theorem 2.3.1, there exist $\hat{N}$ and $\hat{\alpha}$ sufficiently large such that for any $N \geq \hat{N}$ and $\alpha \geq \hat{\alpha}$,

$$\mathbb{D}(\mathcal{F}^N(\alpha), \mathcal{F}) \leq \delta,$$

where δ is a positive constant. Since $\mathfrak{m} \in \mathcal{F}^N(\alpha)$ is equivalent to $e_\alpha(-T^N(\mathfrak{m}, \xi)) \leq 0$, then for any $\mathfrak{m} \in \mathcal{F}^N(\alpha)$, it follows by (2.3.9) that

$$\begin{aligned} d(\mathfrak{m}, \mathcal{F}) &\leq C\left(\sup_{\xi\in\Xi} T(\mathfrak{m}, \xi)\right)_+ - C\left(e_\alpha(-T^N(\mathfrak{m}, \xi))\right)_+ \\ &\leq C\left(\sup_{\xi\in\Xi} T(\mathfrak{m}, \xi) - e_\alpha(-T^N(\mathfrak{m}, \xi))\right)_+ \\ &= C\left(\sup_{\xi\in\Xi} T(\mathfrak{m}, \xi) - e_\alpha(-T^N(\mathfrak{m}, \xi))\right) \\ &\leq C \sup_{\mathfrak{m}\in\mathfrak{M}} \left(\sup_{\xi\in\Xi} T(\mathfrak{m}, \xi) - e_\alpha(-T^N(\mathfrak{m}, \xi))\right), \end{aligned}$$

where the second inequality follows from the fact $(a)_+ - (b)_+ \leq (a-b)_+$, the equality follows from the fact $\sup_{\xi\in\Xi} T(\mathfrak{m}, \xi) \geq e_\alpha(-T^N(\mathfrak{m}, \xi))$ for any $\mathfrak{m} \in \mathfrak{M}$. It shows

$$\mathbb{D}(\mathcal{F}^N(\alpha), \mathcal{F}) \leq C \sup_{\mathfrak{m}\in\mathfrak{M}} \left(\sup_{\xi\in\Xi} T(\mathfrak{m}, \xi) - e_\alpha(-T^N(\mathfrak{m}, \xi))\right). \tag{2.4.7}$$

Let $\mathfrak{m}^*$ and $\mathfrak{m}^N_\alpha$ be an optimal solution of problem (2.3.4) and problem (2.4.5) respectively. Let γ^* and γ^N_α be the corresponding second component of $\mathfrak{m}^*$ and $\mathfrak{m}^N_\alpha$ respectively. Then $\vartheta = -\gamma^*$ and $\vartheta^N(\alpha) = -\gamma^N_\alpha$. By (2.4.7), there exists $\bar{\mathfrak{m}} \in \mathcal{F}$ such that

$$\|\mathfrak{m}^N_\alpha - \bar{\mathfrak{m}}\| \leq C \sup_{\mathfrak{m}\in\mathfrak{M}} \Delta^N_\alpha(\mathfrak{m}).$$

Let $\bar{\gamma}$ be the corresponding second component of $\bar{\mathfrak{m}}$. Then we have

$$\vartheta \leq -\bar{\gamma} \leq -\gamma^N_\alpha + \left|\gamma^N_\alpha - \bar{\gamma}\right| \leq \vartheta^N(\alpha) + C \sup_{\mathfrak{m}\in\mathfrak{M}} \Delta^N_\alpha(\mathfrak{m}).$$

The conclusion follows. □

We are guaranteed that $\mathcal{F} \subset \mathcal{F}(\alpha), \forall\alpha > 0$ by definition and $\limsup_{N\to+\infty} \mathcal{F}^N(\alpha) \subseteq \mathcal{F}(\alpha)$ but we are not guaranteed that $\mathcal{F} \subseteq \mathcal{F}^N(\alpha)$ and hence we are unable to assess $\mathbb{D}(\mathcal{F}, \mathcal{F}^N(\alpha))$. Note that $\Delta^N_\alpha(\mathfrak{m}) \to 0$ as $N, \alpha \to +\infty$. This can be observed through the inequality below

$$\begin{aligned} \Delta^N_\alpha(\mathfrak{m}) \leq\ & \left(\sup_{\xi\in\Xi} T(\mathfrak{m}, \xi) - e_\alpha(-T(\mathfrak{m}, \xi))\right) \\ & + \left(e_\alpha(-T(\mathfrak{m}, \xi)) - e_\alpha(-T^N(\mathfrak{m}, \xi))\right)_+, \end{aligned}$$

where the first term goes to zero by Lemma 2.3.1 and the second term goes to zero as

$N \to +\infty$ by [73, Proposition 7].

2.5 Iterative Scheme

In this section, we propose an algorithm to solve problem (2.3.7). Let us first recall problem (2.3.7)

$$\begin{array}{ll} \inf\limits_{x\in X, \gamma\in\mathbb{R}, \lambda\in\Lambda} & -\gamma \\ \text{s.t.} & e_\alpha(-T(x,\gamma,\lambda,\xi)) \leq 0, \end{array}$$

where $\Lambda := \mathbb{R}^p \times \mathbb{R}_+^{q-p}$. The following algorithm presents an iterative scheme for solving this problem.

Algorithm 2.5.1 *(Iterative scheme)*

Step 1. Given x_0, λ_0, set $k = 0$.

Step 2. For given x_k, λ_k, solve γ_k as a solution to the following equation:

$$e_\alpha(-T(x_k,\gamma,\lambda_k,\xi)) = 0. \tag{2.5.1}$$

Step 3. For given γ_k, solve

$$\begin{array}{ll} \min & e_\alpha(-T(x,\gamma_k,\lambda,\xi)) \\ \text{s.t.} & x \in X \\ & \lambda \in \Lambda, \end{array} \tag{2.5.2}$$

and denote the optimal value and the optimal solution by $\mathcal{V}(\gamma_k)$ and (x_{k+1}, λ_{k+1}) respectively.

Step 4. If $\mathcal{V}(\gamma_k) = 0$, stop. Return $-\gamma_k$ as the optimal value of problem (2.3.7) and $(x_{k+1}, \gamma_k, \lambda_{k+1})$ as the optimal solution. Otherwise go to Step 2.

Note that the idea of the algorithm is rooted in the well known Dinkelbach method for quasi-convex fractional programming since our problem is reformulated from a fractional programming problem. The procedure is more complex here because we are handling $e_\alpha(-T(x,\gamma,\lambda,\xi))$ rather than $T(x,\gamma,\lambda,\xi)$. The bisection method can be applied in Step 2, where variables x and λ are fixed and we can find the root of equation (2.5.1) since the function is monotonically increasing in γ. Although our algorithm is analogous to the standard Dinkelbach method in Step 2 and Step 3, we solve a nonlinear equation in Step 2 to obtain γ_k and the objective function of (2.5.2) is nonlinear in γ or in (x,λ). Compared with [50, Algorithm 1] which updates the optimal value by a fixed increment and then solve a linear program, our iterative scheme updates the optimal value automatically at each iteration by solving a nonlinear equation.

Proposition 2.5.1 *Let γ_k be generated by Algorithm 2.5.1 and $\mathcal{V}(\gamma)$ be defined as the optimal value of (2.5.2). Then*

(i) $\mathcal{V}(\gamma)$ is continuous in γ;

(ii) equation (2.5.1) has a unique solution;

(iii) for each k, $-\gamma_{k+1} < -\gamma_k$;

(iv) $-\gamma_k$ is the optimal value of problem (2.3.7) if and only if $\mathcal{V}(\gamma_k) = 0$.

Proof. Part (i). Since X and Λ are assumed to be bounded, it is easy to verify that $e_\alpha(-T(x, \gamma, \lambda, \xi))$ is uniformly continuous in (x, λ). Then it follows by [70, Theorem 1.17] that $\mathcal{V}(\gamma)$ is continuous in γ.

Part (ii). By Proposition 2.2.3 and (2.3.3), for fixed x, λ, α, $e_\alpha(-T(x, \gamma, \lambda, \xi))$ is strictly increasing in γ. Thus equation (2.5.1) has a unique solution.

Part (iii) and (iv). By the definition of γ_k and $\mathcal{V}(\gamma_k)$, $\mathcal{V}(\gamma_k) \leq 0$. $-\gamma_k$ is the optimal value of problem (2.3.7) if and only if $\mathcal{V}(\gamma_k) \geq 0$. To see this, assume for the sake of a contradiction that $\mathcal{V}(\gamma_k) < 0$, that is

$$e_\alpha(-T(x_{k+1}, \gamma_k, \lambda_{k+1}, \xi)) < 0.$$

Then we can find a positive number δ such that

$$e_\alpha(-T(x_{k+1}, \gamma_k + \delta, \lambda_{k+1}, \xi)) = 0.$$

Thus $-\gamma_{k+1} = -\gamma_k - \delta < -\gamma_k$ which contradicts the assumption that $-\gamma_k$ is the optimal value. If $\mathcal{V}(\gamma_k) = 0$, then

$$\min_{x \in X, \lambda \in \Lambda} e_\alpha(-T(x, \gamma_k, \lambda, \xi)) = 0.$$

Since $e_\alpha(-T(x, \cdot, \lambda, \xi))$ is strictly increasing in γ, we have

$$e_\alpha(-T(x, \gamma_k + \delta, \lambda, \xi)) > 0, \ \forall \delta > 0, x \in X, \lambda \in \Lambda,$$

which means $\gamma_k + \delta$ is not feasible to problem (2.3.7). This shows $-\gamma_k$ is the optimal value and $(x_{k+1}, \gamma_k, \lambda_{k+1})$ is the optimal solution of problem (2.3.7). □

Theorem 2.5.1 *Let $\{-\gamma_k\}$ be a sequence generated by Algorithm 2.5.1. Under Assumption 2.2.1, the sequence is monotonically decreasing and it converges to the optimal value of problem (2.3.7).*

Proof. The Monotonicity follows from Proposition 2.5.1. Next, we show the convergence. Let us consider the case (a) that the Algorithm 2.5.1 terminates after k iterations, that is, $\mathcal{V}(\gamma_k) = 0$. By Proposition 2.5.1, γ_k is the optimal value.

Now we consider the case (b) that $\{-\gamma_k\}$ is an infinite sequence. Under Assumption 2.2.1, the sequence is lower bounded. Therefore, there exists a positive number γ^* such that $-\gamma_k \downarrow -\gamma^*$, which means γ^* is the upper bound of the sequence $\{\gamma_k\}$. It suffices to show that $\mathcal{V}(\gamma^*) \geq 0$.

Assume for the sake of a contradiction that $\mathcal{V}(\gamma^*) < 0$. Let (x^*, λ^*) denote the corresponding optimal solution to problem (2.5.2) for the given γ^*. Then there exists a positive constant c_0 such that

$$e_\alpha(-T(x^*, \gamma^*, \lambda^*, \xi)) \leq -c_0.$$

Since $e_\alpha(-T(x^*, \gamma, \lambda^*, \xi))$ is monotonically increasing w.r.t. γ, and $\gamma_k < \gamma^*$, then

$$e_\alpha(-T(x^*, \gamma_k, \lambda^*, \xi)) < -c_0, \ \forall k.$$

Let (x_{k+1}, λ_{k+1}) denote the corresponding optimal solution to problem (2.5.2) for the given γ_k. Then we have

$$e_\alpha(-T(x_{k+1}, \gamma_k, \lambda_{k+1}, \xi)) \leq e_\alpha(-T(x^*, \gamma_k, \lambda^*, \xi)) < -c_0.$$

Since $\gamma_k \to \gamma^*$ and $e_\alpha(\cdot)$ is continuous, there exists a sufficiently large $\hat{k}$ such that

$$e_\alpha(-T(x_{\hat{k}+1}, \gamma^*, \lambda_{\hat{k}+1}, \xi)) \leq -c_0/2.$$

At step 2 of the Algorithm 2.5.1 , $\gamma_{\hat{k}+1}$ is chosen to satisfy

$$e_\alpha(-T(x_{\hat{k}+1}, \gamma_{\hat{k}+1}, \lambda_{\hat{k}+1}, \xi)) = 0.$$

This shows $\gamma_{\hat{k}+1} > \gamma^*$, which contradicts the fact that γ^* is the upper bound of the sequence $\{\gamma_k\}$. The proof is complete. □

2.6 Numerical Tests

In this section, we carry out some numerical tests to evaluate the performance of the DRURR optimization model. We apply our model to a portfolio selection problem where limited funds are allocated among investment assets to maximize wealth.

2.6.1 Problem Description

Let us start by describing the problem setup and notations. Suppose that there are $n = 10$ assets and we let x_j be the proportion of the total fund invested in asset j. Since we do not consider short positions, the set of feasible fund allocations is a convex set

$$X := \left\{ x \in \mathbb{R}^n : \sum_{j=1}^{n} x_j = 1, x \geq 0 \right\}.$$

We let ξ_j denote the random return of asset j. To simplify the discussions, we ignore the transaction fee, then we let the random outcome

$$f(x, \xi) := \sum_{j=1}^{n} \xi_j x_j.$$

denote the return of the portfolio.

In this numerical study, we consider three models. The first one is the stochastic programming (SP) model where the problem (2.2.3) is solved by approximating the true probability distribution P with empirical data. The second model is to minimize the Conditional Value at Risk (CVaR):

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

where $\alpha \in (0, 1)$ is the level of confidence, in applications one typically sets $\alpha = 0.95$ or 0.99, see [68]. The third one is our DRURR optimization model:

$$\sup_{x \in X} \inf_{P \in \mathcal{P}} \frac{\mathbb{E}_P[u(f(x, \xi)) - u(Y(\xi))]}{\mathbb{E}_P[(u(Y(\xi)) - u(f(x, \xi)))_+]}.$$

To implement the DRURR optimization model and the iterative scheme, we use the ambiguity set defined in (2.4.3).

2.6.2 Data

We now describe the data set to be used in our experiments. We randomly select 10 assets in stock markets and use the data over a time horizon of 6 years with a total of 1500 records of historical asset returns (obtained from https://www.google.com/finance with adjustment for stock splitting).

For the experiments, we define the utility function $u(y) = -\exp(-\beta y)$ with $\beta = 0.05$. We let $Y(\xi)$ be the benchmark return calculated based on an equally weighted fund allocation strategy. We let the value of parameter δ in Theorem 2.4.1 be fixed at 0.01

which means with probability at least 97% the true distribution lies in the ambiguity set. We set the precision $\epsilon = 0.01$ and figure out the entropic approximation parameter $\alpha = 700$ (see details in Appendix B).

2.6.3 Experiments

We now describe the details of our experiments and present the results.

Experiment I: Comparison of the three models

In the first experiment, we compare our DRURR optimization model with the SP model and the CVaR minimization model. We first randomly select 100 samples from the total data set to calculate the optimal fund allocations for the three models. Then we randomly select another 100 samples as the test data set to figure out the lowest return and average return for each model. We present the results in Figure 2.1 after 100 simulations.

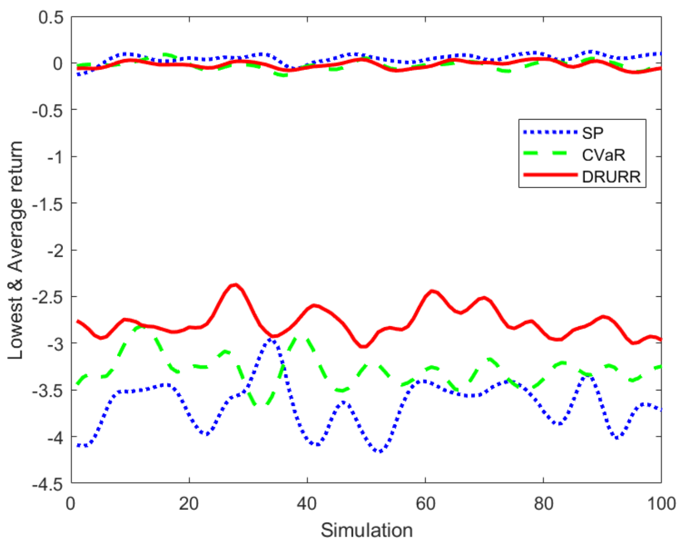
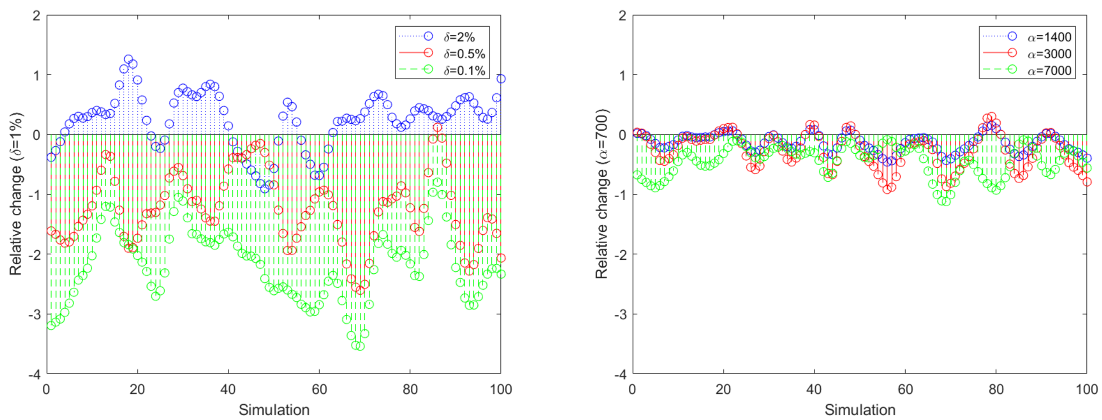


Figure 2.1: Lowest returns and average returns for three models: SP, CVaR and DRURR.

From Figure 2.1, we can see that the lowest returns under the optimal fund allocation plans of the DRURR model are higher than that of other two models. Compared with the lowest returns, the average returns of the three models show little change. It indicates that the DRURR model has lower risk than other two models and it is more suitable for the risk-averse decision makers.

Experiment II: Sensitivity test of the DRURR model

In the second experiment, we test the sensitivity of the DRURR model w.r.t. the change of parameters including α and δ. The change of parameter δ will determine the value of ϵ_N and σ_N [3] which in turn determine the size of the ambiguity set. We first randomly select 100 samples from the total data set to calculate the optimal fund allocations for the DRURR model based on different parameters. Then we randomly select another 100 samples as the test data set to figure out the total return for each parameter setting. We present the relative changes of the total returns w.r.t. different parameter settings in Figure 2.2 (a) and (b) after 100 simulations.



(a) **(b)**

Figure 2.2: **(a)** Relative change of total return w.r.t. parameter δ. **(b)** Relative change of total return w.r.t. parameter α.

From Figure 2.2 (a) and (b), we can observe that our DRURR model is relative stable w.r.t. the change of parameter α, while it is more sensitive to the variation of parameter δ. The results show that the size of ambiguity set has a more important influence on the return of the DRURR model than that of parameter α.

Experiment III: Runtime w.r.t the number of assets

In the third experiment, we test the computation time of the DRURR optimization model based on different numbers of assets. We implement Algorithm 2.5.1 on MATLAB R2019b installed in a generic laptop with Intel Core i5 processor, 4GM RAM, on a 64-bit Windows 7 operating system. We use the built-in optimization solver fmincon to solve minimization problem (2.5.2). We record the average computation time based on 100 simulations. The results in Table 2.1 and Figure 2.3 show that the computation time is approximately doubled for each increase of 5 in the number of assets.

[3] $\epsilon_N := \frac{\beta}{\sqrt{N}}\left(2+\sqrt{2\ln\frac{1}{\delta}}\right)$, $\sigma_N := \frac{3\beta^2}{\sqrt{N}}\left(2+\sqrt{2\ln\frac{1}{\delta}}\right)$, where N is the number of samples of ξ and ξ is essentially bounded by a positive number β.

Number of assets	5	10	15	20	25	30
Time (second)	36.47	67.18	132.36	271.63	553.62	1092.37

Table 2.1: Computation time w.r.t. the number of assets

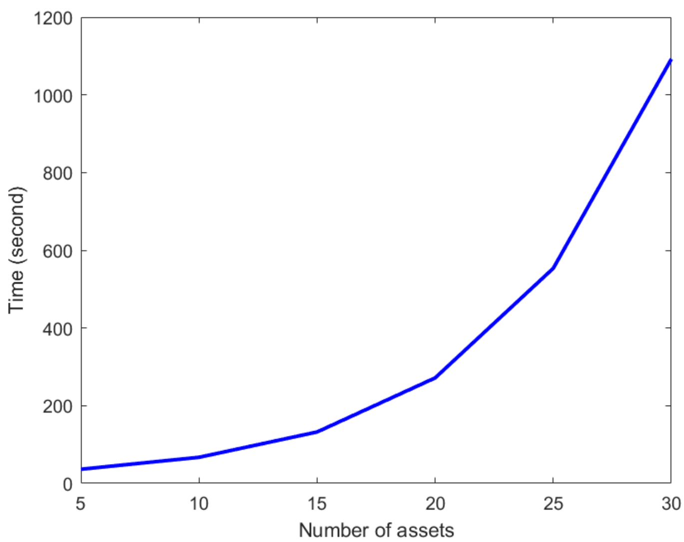


Figure 2.3: Computation time w.r.t. the number of assets

Chapter 3

Preference Robust Multivariate Utility-Based Shortfall Risk Optimization

The DRURR optimization model investigated in Chapter 2 is under the assumption that a decision maker's true utility function is known. In many practical applications, however, the true utility function is unavailable because there may not be enough information to specify it or a group of decision makers have difficulty agreeing on which utility function to use. In this chapter, we will concentrate on this issue in the context of the utility-based shortfall risk measure (SR) which has received increasing attention over the past few years. The SR model is capable of capturing the large potential loss and it is more sensitive to extreme events than CVaR. We consider the situation where the true loss function (represents a decision maker's risk preference) in SR is not available and use a robust scheme to mitigate the risk from the ambiguity. In addition, most studies of SR only deal with the univariate case, where profits/losses are expressed by a random variable. In natural applications, however, many financial positions possess multi-attributes, e.g., an insurance company typically has several business lines each of which has a distinct attribute. This motivates us to consider a multi-attribute SR model.

3.1 Introduction

3.1.1 Literature Review

Measuring or quantifying risk is important to understand the potential features of risk that an institution has. A proper assessment of the downside risk of financial positions is significant in the process of decision making. As discussed in [36], a good risk measure should be sensitive to excessive losses and encourage diversification. Over the past few decades since the pioneering work of Markowitz [58] on mean-variance risk measure, a great deal of effort has gone into achieving suitable methods of measuring risk.

One of the well known risk measures is value-at-risk (VaR), which has been widely used as an industry standard since the 1990s [39, 65]. However, there is a growing dissatisfaction about this measure, which is accused of neither encouraging diversification nor taking into account the size of extremely large losses [68, 69]. These shortcomings motivated an axiomatic analysis of risk measures with desirable properties. The axiomatic foundation study of risk measures was pioneered by Artzner et al. [9], which introduced the notion of coherent risk measure. In the work of Artzner et al. [9], a risk measure is called coherent if it satisfies four basic properties: translation invariance, monotonicity, subadditivity, positive homogeneity.

After the introduction of coherent risk measures, many variations and extensions of them have been proposed and studied in the literature, see e.g., [3, 27, 54, 63]. Conditional value-at-risk (CVaR), as one of the coherent risk measures, studied by [2, 8, 69] has gained popularity in recent years. CVaR has emerged as a better alternative to VaR because it measures the sizes of the potential losses beyond the threshold amount indicated by the VaR. Although CVaR is used in a wide range of applications such as supply chain [19], network design [11], healthcare [20], etc., it still suffers from severe deficiencies. One important drawback of CVaR is that it only captures a limited spectrum of risk attitudes and may not accurately represent the risk aversion of some decision makers. There are also many other interesting coherent risk measures, such as spectral risk measures [1, 4].

An important extension of the concept of coherent risk measure is the notion of convex risk measure, studied by Föllmer and Schied [32] and Frittelli and Gianin [34], who generalized coherent risk measures to the convex case by replacing the two properties of subadditivity and positive homogeneity with the property of convexity. A special category of convex risk measures, called utility-based shortfall risk measure (SR) proposed by Föllmer and Schied [32], has received increasing attentions over the past few years. For a given loss function (as a risk attitude) and a threshold value (as a prespecified risk level), the SR of a financial position is the minimal capital added to the position such that the new position's risk level is below the prespecified risk level [47].

The use of SR turns out to be very flexible because both loss function and threshold value can be tailored to the specific needs of any financial institution or regulating authority. Besides, SR has many appealing properties, including monotonicity, translation invariance and law invariance, and it moreover satisfies convexity if and only if the loss function is convex [32]. The work of Giesecke et al. [36] shows that SR is capable of capturing the large potential loss and it is more sensitive to extreme events than CVaR. The study on the estimation of SR is pioneered by Dunkel and Weber [29], who treated the estimation of SR as a stochastic root finding problem and applied the stochastic approximation algorithms combined with importance sampling techniques to solve the problem. While Hu and Zhang [47] treated the SR as the optimal value of a stochastic optimization problem and implemented the sample average approximation method to solve the problem.

Although sensitivity analysis and computational aspects of SR have been studied in the literature, an important challenge remains that it is unclear how to choose a loss function that faithfully represents a decision makers true risk attitude. In the work of Armbruster and Delage [6] and Hu and Mehrotra [46], the idea is to formulate a set of plausible risk preference relations and seek decisions that are optimal with respect to worst-case expected utility. Haskell et al. [42] extended this preference robust expected utility framework to cases where there is also ambiguity about the underlying probability distribution. Delage and Li [25] introduced the notion of preference robust risk measure: $\varrho_{\mathcal{R}}(X) := \sup_{\rho\in\mathcal{R}}(X)$, where $\mathcal{R}$ is an ambiguity set of all convex/coherent/law invariant risk measures. However, the authors did not provide a reformulation that could exploit the fact that the measure is known to be a utility-based shortfall risk measure.

3.1.2 Contribution

In this chapter, we consider the case that the true loss function in SR is not available since there is no enough information to specify it or decision makers have difficulty agreeing on which loss function to use. To mitigate the risk from the ambiguity, we exploit the idea of Armbruster and Delage [6] to construct a set of loss functions from empirical data or subjective judgements and compute the SR through the worst loss function from the set.

Compared with Föllmer and Schied [32]'s SR model which uses a fixed targeted utility loss λ to determine the acceptance set, Our SR model's acceptance set is determined by the expected value of utility benchmark loss $\mathbb{E}[l(-Y)]$ that captures not only the performance of benchmark position under the same random environment but also the decision maker's preference which is particularly important when there is an ambiguity about the loss function.

In addition, most studies of SR only deal with the univariate case, where profits/losses are expressed by a random variable, however, in many natural applications it is necessary to assess the risk of a vector whose components represent different gains or liabilities from various financial risk components. Thus in this work, we apply SR in a multivariate framework that models are dependence of several financial risk components, that is, we contribute to a systemic extension of SR based on multivariate loss functions.

Moreover, many tractable reformulations of SR optimization problems are established under the hypothesis of a discrete outcome space. We explore a tractable numerical scheme for the case when the underlying probability distribution is continuous. We also apply our proposed model to a multi-criteria resource allocation problem in homeland security and conduct some numerical tests to examine the performance of the model.

3.1.3 Structure

The remainder of this chapter is organised as follows: In section 3.2, we give a brief introduction to the multivariate utility-based shortfall risk measure (MSR) and propose a preference robust MSR model. In section 3.3, we introduce some properties of MSR and preference robust MSR. In section 3.4, we apply preference robust MSR to an optimization problem and discuss discrete approximation of the problem when the probability distribution is continuous. A tractable formulation is developed for the approximation problem in section 3.5, and some numerical studies are carried out in section 3.6.

3.2 MSR and Preference Robust MSR

Consider a financial position with multi-attributes represented by a random vector $X : \Omega \to \mathbb{R}^n$ for some integer $n \geq 1$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $(\Omega, \mathcal{F})$ is a measurable space with σ-algebra $\mathcal{F}$ and $\mathbb{P}$ is a probability measure. We let $\mathscr{L}^0 := \mathscr{L}^0(\Omega, \mathcal{F}, \mathbb{P})$ be the space of all measurable mappings $X : \Omega \to \mathbb{R}^n$, and $\mathscr{L}^\infty := \mathscr{L}^\infty(\Omega, \mathcal{F}, \mathbb{P}) \subset \mathscr{L}^0$ is the space of essentially bounded measurable functions. We generally treat $\mathscr{L}^0$ as a space of multi-attribute prospects with $n \geq 2$, although the case $n = 1$ is also covered.

Föllmer and Schied [32] first introduce utility-based shortfall risk measure (SR) over a single attribute prospect space. Let $l : \mathbb{R} \to \mathbb{R}$ be an increasing convex loss function and $Z : \Omega \to \mathbb{R}$ a financial position. Given a targeted utility loss λ, they define the set of acceptable positions as

$$\mathcal{S} := \{Z \in \mathscr{L}^\infty : \mathbb{E}[l(-Z)] \leq \lambda\}$$

and the utility-based shortfall risk measure (SR) as

$$\mathrm{SR}(Z) := \inf\{t \in \mathbb{R} : Z + t \in \mathcal{S}\}.$$

In practical applications, many financial positions possess multi-attributes [31], i.e., an insurance company typically has several business lines each of which has a distinct attribute. Moreover, there is often a benchmark position which provides a bottom line for the performance of a financial position. This motivates us to consider multi-attribuite utlity shortfall risk measure model with targeted utility loss determined by a benchmark position. We start by defining the acceptance set

$$\mathcal{A} := \{X \in \mathscr{L}^0 : \mathbb{E}[l(-X)] \leq \mathbb{E}[l(-Y)]\},$$

where $l : \mathbb{R}^n \to \mathbb{R}$ is an increasing loss function and $Y \in \mathscr{L}^0$ is a benchmark. Note that by increasing, we mean that $l(x) \leq l(y)$ for any $x, y \in \mathbb{R}^n$ with $x \leq y$. In comparison with the acceptance set which is determined by a fixed targeted utility loss in Föllmer and Schied's SR model, our acceptance set here is determined by $\mathbb{E}[l(-Y)]$ which captures not only performance of the benchmark position under the same random environment but also the investor's preference which is particularly important when there is an ambiguity of the loss function.

With the acceptable set $\mathcal{A}$, we can define the *multivariate utility-based shortfall risk measure (MSR)* as

$$\begin{aligned}\mathrm{MSR}_{l,d}(X) :=& \ \inf\{t \in \mathbb{R} : X + td \in \mathcal{A}\} \\ =& \ \inf\{t \in \mathbb{R} : \mathbb{E}[l(-X - td) - l(-Y)] \leq 0\},\end{aligned} \tag{3.2.1}$$

where $d \in \mathcal{D}$ is a preset weighting vector and

$$\mathcal{D} = \left\{ d \in \mathbb{R}^n : \sum_{i=1}^n d_i = 1, d_i > 0 \right\}.$$

The MSR may be interpreted as the smallest amount of cash t (when $t > 0$) to be injected to the financial position X or the largest amount of cash t (when $t < 0$) can be taken out so that the new position falls into the acceptance set $\mathcal{A}$. A slightly more general form of MSR is considered by Armenti et al. [7]

$$\rho(X) = \inf \left\{ \sum_{i=1}^n m_i : \mathbb{E}[l(X + m)] \leq 0 \right\}. \tag{3.2.2}$$

Note that by setting $m_i = td_i$ and using multivariate loss function $l(-X-td)-l(-Y)$ in (3.2.2), we get formulation (3.2.1). However, the two formulations are intrinsically different for two reasons: (a) In (3.2.2), the optimal solution may have one component being positive and another being negative, while in (3.2.1) the signs of the components at optimum are consistent; (b) In (3.2.1), d is fixed which may be interpreted as a preset weighting vector and the allocation is proportionate across the components, while in formulation (3.2.2), the allocation $m \in \mathbb{R}^n$ does not have to be proportionate and the overall aim is to minimize the aggregated liquidity cost $\sum_{i=1}^n m_i$. Although model (3.2.2) is more general, (3.2.1) does have some advantages: (i) The optimization problem has one variable t as opposed to n-variables in (3.2.2), this will significantly simplify (PRMSR-Opti) in the follow-up discussions; (ii) It allows one to choose optimal weighting vector in (3.2.1) although we do not explore in that direction in this thesis, this kind of research is interesting in asset allocation problem [13].

From the definition of MSR, it is easy to observe that choosing an appropriate loss function l is important in applying MSR. However, in practice, there could be considerable ambiguity in the choice of the loss function. For example, the decision maker does not gather enough information to uniquely specify the loss function, or a group of decision makers have difficulty agreeing which loss function to use. In these circumstances, there is no obvious choice for the true loss function. To overcome the risk arising from ambiguity about the decision maker's risk preference, we will construct a set of loss functions with available partial information or subjective judgements and define a preference robust MSR through the worst loss function from the set.

Definition 3.2.1 *Let L be a set of the loss functions $l : \mathbb{R}^n \to \mathbb{R}$ and $d \in \mathcal{D}$. The* preference robust multivariate shortfall risk measure *is defined as*

$$\text{(PRMSR)} \quad \text{MSR}_{L,d}(X) := \inf \left\{ t \in \mathbb{R} : \sup_{l \in L} \mathbb{E}[l(-X-td) - l(-Y)] \le 0 \right\}, \tag{3.2.3}$$

where the preference robust constraint means that the expected disutility of losses $\mathbb{E}[l(-X-td)]$ falls below $\mathbb{E}[l(-Y)]$ for every risk attitude defined as $l \in L$.

This worst-case approach is justified in robust optimization, and the above definition is an application of the philosophy of robust optimization to preference ambiguity. For a practical interpretation, which is especially relevant in the context of group decision making, the robust loss function means accommodating the preferences of the least favored member of L.

In practice, a risk measure is often associated with some decision making problems. Thus we consider an optimization problem with PRMSR as an objective function. Specifically, let $c(z, \xi(\omega))$ be a financial position associated with decision matrix $z \in$

$Z \subset \mathbb{R}^n$ and random matrix $\xi(\omega)$. We consider

$$\text{(PRMSR-Opti)} \quad \min_{z \in Z} \text{MSR}_{L,d}^{P}(c(z,\xi)), \tag{3.2.4}$$

where $\xi : (\Omega, \mathcal{F}, \mathbb{P}) \to \Xi \subset \mathbb{R}^s$, $P := \mathbb{P} \circ \xi^{-1}$ is the probability measure on $\mathbb{R}^s$ induced by ξ and $c(z,\xi) : Z \times \Xi \to \mathbb{R}^n$ is a continuous function. In this work, we focus on the case that P is continuously distributed and propose a sample average approximation scheme

$$\min_{z \in Z} \text{MSR}_{L,d}^{P_N}(c(z,\xi)), \tag{3.2.5}$$

where P_N is a discrete approximation of the true probability distribution P. We establish conditions under which (3.2.5) converges to (3.2.4) in terms of the optimal value and optimal solutions as the sample size increases, and we develop a tractable formulation for the optimization problem (3.2.5).

3.3 Properties of preference robust MSR

In this section, we will investigate some properties of MSR and its preference robust counterpart.

3.3.1 Convexity of MSR and PRMSR

We start by specifying a set of loss functions which will be used throughout this chapter to define our risk measures. Let $\mathcal{L}$ be the set of all convex increasing functions $l : \mathbb{R}^n \to \mathbb{R}$ that are strictly increasing along direction d [1] for each $d \in \mathcal{D}$.

Proposition 3.3.1 *Let $x \in \mathbb{R}^n$, $l \in \mathcal{L}$ and $d \in \mathcal{D}$, let $h(t) := l(x - td)$. Then the following hold.*

(i) h is a strictly decreasing convex function, and h is continuous on $\mathbb{R}$;

(ii) $h(t) \to +\infty$ as $t \to -\infty$.

Definition 3.3.1 (Multivariate convex risk measure) *A real-valued function $\rho : \mathscr{L}^0 \to \mathbb{R}$ is said to be a convex risk measure if it satisfies the following properties:*

(a) Convexity: $\rho(\alpha X + (1-\alpha)Y) \leq \alpha\rho(X) + (1-\alpha)\rho(Y)$ for all $X, Y \in \mathscr{L}^0$ and $\alpha \in [0,1]$;

[1] Increasing along direction $d \in \mathcal{D}$ means that $l(x + t_1 d) \leq l(x + t_2 d)$ for any $x \in \mathbb{R}^n$, $t_1, t_2 \in \mathbb{R}$ with $t_1 \leq t_2$.

(*b*) *Monotonicity:* $\rho(X) \le \rho(Y)$ *for* $X, Y \in \mathscr{L}^0$ *with* $X \ge Y$*;*

(*c*) *Translation invariance:* $\rho(X + md) = \rho(X) - m$ *for* $X \in \mathscr{L}^0$, $m \in \mathbb{R}$ *and* $d \in \mathcal{D}$.

Note that the inequality $X \ge Y$ is understood to mean $X(\omega) \ge Y(\omega)$ for all $\omega \in \Omega$, and we compare vectors on $\mathbb{R}^n$ in the component-wise order, i.e., $X(\omega) \ge Y(\omega)$ means $X_i(\omega) \ge Y_i(\omega)$ for all $i = 1, \ldots, n$. Here convexity means that diversification does not increase the risk, i.e., the risk of a diversified position $\alpha X + (1 - \alpha)Y$ is less or equal to the weighted average of the individual risks. Monotonicity means if financial position X is better than Y in every scenario then risk of position X is less than risk of position Y. Translation invariance means if financial position X can generate cash m deterministically then the risk will be reduced by m.

Proposition 3.3.2 *Let* $\rho(\cdot)$ *be a multivariate utility-based shortfall risk measure defined as in (3.2.1). Then* $\rho(\cdot)$ *is a convex risk measure for each associated* $l \in \mathcal{L}$.

Proof. By Definition 3.3.1, it suffices to verify the three properties specified there. Let us start with convexity and consider any $X_i \in \mathscr{L}^0$, $i = 1, 2$ such that

$$\begin{array}{rl} \rho(X_i) = & \inf\ t \\ \text{s.t.} & \mathbb{E}[l(-X_i - td)] \le \mathbb{E}[l(-Y)]. \end{array} \tag{3.3.1}$$

Let $t_1 = \rho(X_1)$, $t_2 = \rho(X_2)$ and $t_\alpha = \alpha t_1 + (1 - \alpha)t_2$, for any fixed $\alpha \in [0, 1]$. Then

$$\begin{aligned} & \mathbb{E}[l(-(\alpha X_1 + (1 - \alpha)X_2) - t_\alpha d)] \\ = \ & \mathbb{E}[l(\alpha(-X_1 - t_1 d) + (1 - \alpha)(-X_2 - t_2 d))] \\ \le \ & \alpha\mathbb{E}[l(-X_1 - t_1 d)] + (1 - \alpha)\mathbb{E}[l(-X_2 - t_2 d)] \\ \le \ & \alpha\mathbb{E}[l(-Y)] + (1 - \alpha)\mathbb{E}[l(-Y)] \\ = \ & \mathbb{E}[l(-Y)], \end{aligned}$$

which means that t_α is a feasible solution to the program (3.2.1) with $X = \alpha X_1 + (1 - \alpha)X_2$. Thus

$$\rho(\alpha X_1 + (1 - \alpha)X_2) \ \le \ t_\alpha \ = \ \alpha t_1 + (1 - \alpha)t_2 \ = \ \alpha\rho(X_1) + (1 - \alpha)\rho(X_2).$$

To show the monotonicity, let $X_1 \ge X_2$. Let $\mathcal{F}_i$ denote the feasible set of (3.3.1). Then $\mathcal{F}_2 \subseteq \mathcal{F}_1$ since

$$l(-X_1 - td) \le l(-X_2 - td),$$

which means that $\rho(X_1) \le \rho(X_2)$. The translation invariance follows by

$$\begin{aligned} \rho(X + md) &= \inf\{t : \mathbb{E}[l(-X - (t + m)d) - l(-Y)] \le 0\} \\ &= \inf\{t' - m : \mathbb{E}[l(-X - (t')d) - l(-Y)] \le 0\} \\ &= \rho(X) - m. \end{aligned}$$

Hence $\rho(\cdot)$ is a convex risk measure. □

Proposition 3.3.3 *Let $\rho(\cdot)$ be a multivariate utility-based shortfall risk measure with a loss function $l \in \mathcal{L}$. Assume that $\rho(X)$ is finite-valued. Then $t' := \rho(X)$ is the unique solution of the equation*

$$\mathbb{E}[l(-X - td) - l(-Y)] = 0. \tag{3.3.2}$$

Proof. Let

$$H(t) := \mathbb{E}[l(-X - td) - l(-Y)],$$

and t_0 be such that

$$H(t_0) = \mathbb{E}[l(-X - t_0 d) - l(-Y)] = 0.$$

The existence of t_0 is guaranteed by the fact: (a) $\rho(X)$ is finite which implies that the feasible set $\{t \in \mathbb{R} : \mathbb{E}[l(-X - td) - l(-Y)] \leq 0\}$ is nonempty and hence there exists a $\bar{t}$ such that

$$H(\bar{t}) = \mathbb{E}[l(-X - \bar{t}d) - l(-Y)] \leq 0;$$

(b) $H(t)$ is continuous and $H(t) \to +\infty$ as $t \to -\infty$ based on Proposition 3.3.1.

From Proposition 3.3.1, we know that for each $\omega \in \Omega$, the function $h_\omega(t) = l(-X(\omega) - td) - l(-Y(\omega))$ is a strictly decreasing convex function and satisfies $h_\omega(t) \to +\infty$ as $t \to -\infty$. Therefore, $H(\cdot)$ is a non-increasing continuous convex function. We first show that $H(\cdot)$ is strictly decreasing near t_0. Since $h_\omega(t)$ is strictly decreasing, for any sufficiently small positive number ϵ,

$$\begin{aligned} H(t_0 - \epsilon) &= \mathbb{E}[l(-X(\omega) - t_0 d + \epsilon d) - l(-Y)] \\ &= \int_\Omega h_\omega(t_0 - \epsilon) P(d\omega) \\ &> \int_\Omega h_\omega(t_0) P(d\omega) \\ &= \mathbb{E}[l(-X - t_0 d) - l(-Y)] = H(t_0). \end{aligned}$$

Likewise, we can show $H(t_0 + \epsilon) < H(t_0)$. This indicates that $H(\cdot)$ is strictly decreasing near t_0.

We now return to show $t' := \rho(X)$ is the unique solution of equation (3.3.2). It is easy to see that t' is a feasible solution to (3.2.1), which implies $H(t') \leq 0$. The strict inequality does not hold because otherwise t' could be further reduced and hence would not be optimal. This shows t' satisfies equation (3.3.2). Since we have shown that $H(\cdot)$ is non-increasing continuous function and strictly decreasing near t_0, hence t' is the unique solution to the equation. □

Theorem 3.3.1 *Let $L \subseteq \mathcal{L}$. Then* *(i)* $\mathrm{MSR}_{L,d}(X) = \sup_{l\in L} \mathrm{MSR}_{l,d}(X)$, *that is,*

$$\begin{aligned} & \inf\left\{t\in\mathbb{R} : \sup_{l\in L}\mathbb{E}[l(-X-td)-l(-Y)]\le 0\right\} \\ = & \sup_{l\in L}\inf\{t\in\mathbb{R} : \mathbb{E}[l(-X-td)-l(-Y)]\le 0\}; \end{aligned} \tag{3.3.3}$$

(ii) $\mathrm{MSR}_{L,d}(X)$ *is a convex risk measure.*

Proof. Part (i). Let $t^* := \mathrm{MSR}_{L,d}(X)$ and $\hat{t} := \sup_{l\in L}\mathrm{MSR}_{l,d}(X)$. We first consider the case (a) that there exists $t_0 \in \mathbb{R}$ such that

$$\sup_{l\in L}\mathbb{E}[l(-X-t_0d)-l(-Y)]\le 0. \tag{3.3.4}$$

Under this condition, the feasible set of $\mathrm{MSR}_{L,d}(X)$ is non-empty and hence $t^* \neq +\infty$. In what follows, we show that $t^* \neq -\infty$.

Since $l(\cdot)$ is convex, by the Jensen's inequality,

$$\mathbb{E}[l(-X-td)] \ge l(-\mathbb{E}[X]-td),$$

which enables us to deduce

$$\sup_{l\in L}\mathbb{E}[l(-X-td)-l(-Y)] \ge \sup_{l\in L} l(-\mathbb{E}[X]-td)-\mathbb{E}[l(-Y)]. \tag{3.3.5}$$

Based on Proposition 3.3.1, we have

$$\lim_{t\to-\infty}\sup_{l\in L} l(-\mathbb{E}[X]-td)-\mathbb{E}[l(-Y)] = +\infty,$$

and through (3.3.5), we have

$$\lim_{t\to-\infty}\sup_{l\in L}\mathbb{E}[l(-X-td)-l(-Y)] = +\infty. \tag{3.3.6}$$

Thus

$$\left\{t\in\mathbb{R} : \sup_{l\in L}\mathbb{E}[l(-X-td)-l(-Y)]\le 0\right\}$$

is a bounded set and hence $t^* \neq -\infty$. This shows t^* is finite.

Now we return to show $t^* = \hat{t}$. Since t^* is the smallest number such that

$$\sup_{l\in L}\mathbb{E}[l(-X-t^*d)-l(-Y)]\le 0,$$

then for any $l \in L$,

$$\mathbb{E}[l(-X-t^*d)-l(-Y)]\le 0,$$

which means that for each $l \in L$, t^* is a feasible solution for the inner minimization problem of $\sup_{l \in L} \mathrm{MSR}_{l,d}(X)$. If we use t_l to denote the optimal value of the inner minimization problem of $\sup_{l \in L} \mathrm{MSR}_{l,d}(X)$, then our discussion shows $t^* \geq t_l$ for all $l \in L$ and hence

$$t^* \geq \sup_{l \in L} t_l = \hat{t}. \tag{3.3.7}$$

Conversely, for each $l \in L$,

$$\mathbb{E}[l(-X - \hat{t}d) - l(-Y)] \leq \mathbb{E}[l(-X - t_l d) - l(-Y)] \leq 0.$$

Thus

$$\sup_{l \in L} \mathbb{E}[l(-X - \hat{t}d) - l(-Y)] \leq 0.$$

This shows $\hat{t}$ is a feasible solution of $\mathrm{MSR}_{L,d}(X)$ and hence $\hat{t} \geq t^*$.

We now consider the case (b) that condition (3.3.4) fails to hold, that is, the feasible set of $\mathrm{MSR}_{L,d}(X)$ is empty. In that case, the optimal value $t^* = +\infty$. Moreover, the emptiness of the feasible set means that for any $t \in \mathbb{R}$, there exist a constant $\delta_t > 0$ and $l_t \in L$ such that

$$\mathbb{E}[l_t(-X - td) - l_t(-Y)] > \delta_t. \tag{3.3.8}$$

In what follows, we show that (3.3.8) implies $\hat{t}$ is infinite. Indeed, if $\hat{t}$ is finite, then for all $l \in L$,

$$\mathbb{E}[l(-X - \hat{t}d) - l(-Y)] \leq 0,$$

which contradicts (3.3.8). Hence, without condition (3.3.4), both t^* and $\hat{t}$ must be infinite.

Part (ii). It follows from Part (i) in that the supremum operation preserves convexity, monotonicity and translation invariance. □

Theorem 3.3.1 Part (i) states that the robust shortfall risk measure is equal to the worst case shortfall risk measure associated with the loss function l chosen from the ambiguity set $\mathcal{L}$. It is important because the right hand side has a clear business insight whereas the left hand side will be effectively used to develop numerical procedures for evaluating the robust risk measure.

3.3.2 Domain of MSR and PRMSR

The risk measures are real-valued functions defined on the space of random variables. The domain of a risk measure specifies the set of random variables whose risks are finite valued. In practice, each random variable represents a prospect of gain/loss and the domain of a risk measure represents the scope of prospects where the risk measure can be sensibly used (not going to infinity which does not make a sense in practice). Specifically, let $\rho(\cdot)$ be a multivariate utility-based shortfall risk measure associated with $l \in \mathcal{L}$. It might be interesting to discuss the domain of $\rho(\cdot)$ over $\mathscr{L}^0$ where ρ is finite valued. To this end, we exploit the notion of Orlicz space.

Let $\Psi : \mathbb{R}^n_+ \to [0, \infty]$ be a Young function [2] such that Ψ is continuous, increasing convex with

$$0 = \Psi(0) = \lim_{x \downarrow 0} \Psi(x) \quad \text{and} \quad \lim_{x \uparrow \infty} \Psi(x) = \infty.$$

The Orlicz space associated with Ψ is

$$\mathscr{L}^\Psi := \mathscr{L}^\Psi(\Omega, \mathcal{F}, \mathbb{P}) = \{X \in \mathscr{L}^0 : \mathbb{E}[\Psi(c|X|)] < \infty \text{ for some } c > 0\},$$

where we write $|x|$ for the absolute value of vector x, that is, the i-th component of $|x|$ is $|x_i|$. The Orlicz heart is defined as

$$\mathscr{H}^\Psi := \mathscr{H}^\Psi(\Omega, \mathcal{F}, \mathbb{P}) = \{X \in \mathscr{L}^0 : \mathbb{E}[\Psi(c|X|)] < \infty \text{ for all } c > 0\}.$$

Note that when Ψ takes the value of ∞, $\mathscr{H}^\Psi = \{0\}$ and $\mathscr{L}^\Psi = \mathscr{L}^\infty$. For this reason, we mainly focus on the case when Ψ is finite. Then we have the following relationship of the spaces which are of our interest:

$$\mathscr{L}^\infty \subset \mathscr{H}^\Psi \subset \mathscr{L}^\Psi \subset \mathscr{L}^0.$$

We define a Young function

$$\Psi_l(x) := l(x) - l(0) \text{ for } x \in \mathbb{R}^n_+.$$

It is easy to see that

$$\begin{aligned} l(-X - td) &\leq \tfrac{1}{2} l(-2X) + \tfrac{1}{2} l(-2td) \\ &\leq \tfrac{1}{2} l(2|X|) - \tfrac{1}{2} l(0) + \tfrac{1}{2} l(2|td|) - \tfrac{1}{2} l(0) + l(0) \\ &= \tfrac{1}{2} \Psi_l(2|X|) + \tfrac{1}{2} \Psi_l(2|td|) + l(0), \end{aligned}$$

which implies $\mathbb{E}[l(-X - td)] < \infty$ for $t \in \mathbb{R}$ and $X \in \mathscr{H}^{\Psi_l}$ when $l(0) < \infty$. Hence $\rho(X) \neq -\infty$ for each $X \in \mathscr{H}^{\Psi_l}$. If in addition, $\mathcal{T}(X) := \{t \in \mathbb{R} : \mathbb{E}[l(-X - td) - l(-Y)] \leq 0\} \neq \emptyset$ for $X \in \mathscr{H}^{\Psi_l}$, then $\rho(X) \neq +\infty$. Thus the domain of $\rho(\cdot)$ is

$$\mathscr{H}^{\Psi_l} \cap \{X \in \mathscr{L}^0 : \mathcal{T}(X) \neq \emptyset\},$$

where ρ is finite valued.

Now we turn to the robust case, that is, PRMSR $\varrho(\cdot)$. Let $L \subset \mathcal{L}$ be the associated ambiguity set. Suppose that $\sup_{l \in L} l(0) < \infty$. Then following a similar analysis to the

[2] In literature, Young function is defined over $\mathbb{R}_+$. Here we extend the notion to multivariate case.

above, we can show that $\rho(X) \neq -\infty$ for each $X \in \bigcap_{l\in L} \mathscr{H}^{\Psi_l}$ and the domain of $\varrho(\cdot)$ is

$$\bigcap_{l\in L} \mathscr{H}^{\Psi_l} \bigcap \{X \in \mathscr{L}^0 : \mathscr{T}(X) \neq \emptyset\},$$

where

$$\mathscr{T}(X) := \left\{ t \in \mathbb{R} : \sup_{l\in L} \mathbb{E}[l(-X - td) - l(-Y)] \leq 0 \right\} \neq \emptyset.$$

3.4 Approximation of (PRMSR-Opti)

In this section, we discuss discrete approximation of (PRMSR-Opti) when the underlying probability distribution P is continuous. By the definition of the preference robust MSR, we rewrite problem (PRMSR-Opti) as

$$\begin{aligned} \min_{z\in Z, t\in\mathbb{R}} \quad & t \\ \text{s.t.} \quad & \sup_{l\in L} \mathbb{E}_P[l(-c(z,\xi) - td) - l(-Y(\xi))] \leq 0, \end{aligned} \tag{3.4.1}$$

and its approximate problem takes the form:

$$\begin{aligned} \min_{z\in Z, t\in\mathbb{R}} \quad & t \\ \text{s.t.} \quad & \sup_{l\in L} \mathbb{E}_{P_N}[l(-c(z,\xi) - td) - l(-Y(\xi))] \leq 0. \end{aligned} \tag{3.4.2}$$

To ease the exposition, we let

$$v(z,t) := \sup_{l\in L} \mathbb{E}_P[l(-c(z,\xi) - td) - l(-Y(\xi))] \tag{3.4.3}$$

and

$$v_N(z,t) := \sup_{l\in L} \mathbb{E}_{P_N}[l(-c(z,\xi) - td) - l(-Y(\xi))]. \tag{3.4.4}$$

Consequently, we rewrite (3.4.1) and (3.4.2) as

$$\text{(PRMSR-Opti)} \quad \begin{aligned} \min_{z\in Z, t\in\mathbb{R}} \quad & t \\ \text{s.t.} \quad & v(z,t) \leq 0, \end{aligned} \tag{3.4.5}$$

and

$$\text{(Opti-N)} \quad \begin{aligned} \min_{z\in Z, t\in\mathbb{R}} \quad & t \\ \text{s.t.} \quad & v_N(z,t) \leq 0, \end{aligned} \tag{3.4.6}$$

Let $\mathcal{F}$, $\mathcal{S}$ and ϑ denote the feasible set, the set of optimal solutions and the optimal value of problem (PRMSR-Opti) respectively. Likewise, we define $\mathcal{F}_N$, $\mathcal{S}_N$ and ϑ_N for problem (Opti-N). Throughout this section, we make the following assumption.

Assumption 3.4.1 *Assume that:*

(a) $L \subseteq \mathcal{L}$ *(as defined at the beginning of Section 3.3);*

(b) Z *is a compact set;*

(c) $c(\cdot,\cdot)$ *is a continuous function on* $Z \times \Xi$*;*

(d) Problem (PRMSR-Opti) satisfies Slater condition, that is, there exist a positive constant number θ *and* $z_0 \in Z$, $t_0 \in \mathbb{R}$ *such that*

$$\sup_{l\in L} \mathbb{E}_P[l(-c(z_0,\xi) - t_0 d) - l(-Y)] \leq -\theta. \tag{3.4.7}$$

Under Assumption 3.4.1 the optimal value ϑ is finite. To see this, we note that

$$\sup_{l\in L} \mathbb{E}_P[l(-X - t_0 d) - l(-Y)] \leq \sup_{l\in L} \mathbb{E}_P[l(-c(z_0,\xi) - t_0 d) - l(-Y)] \leq -\theta,$$

where $X := \max_{z\in Z} c(z,\xi)$. Following a similar proof to (3.3.6), we can show

$$\lim_{t\to-\infty} \sup_{l\in L} \mathbb{E}_P[l(-X - td) - l(-Y)] = +\infty,$$

which implies the t component of the feasible set $\mathcal{F}$ must have a lower bound and hence the optimal value $\vartheta > -\infty$. On the other hand, condition (3.4.7) ensures $\vartheta \leq t_0$.

3.4.1 Sample Average Approximation

We start by looking into sample average approximation (SAA) scheme. Let

$$P_N(\cdot) := \frac{1}{N}\sum_{k=1}^{N} \mathbb{1}_{\xi^k}(\cdot), \tag{3.4.8}$$

where $\xi^1, \ldots, \xi^N$ are iid samples of ξ and

$$\mathbb{1}_{\xi^k}(\xi(\omega)) := \begin{cases} 1, & \text{if } \xi(\omega) = \xi^k, \\ 0, & \text{if } \xi(\omega) \neq \xi^k. \end{cases}$$

Instead of deriving uniform approximation specifically for $v_N(z,t)$, we establish a general convergence result which may be of interest beyond this section.

Let $f : Z \times \Xi \to \mathbb{R}^n$ be a continuous function and $\mathcal{G}$ be a set of continuous functions $g : \mathbb{R}^n \to \mathbb{R}$. Let

$$w(z) := \sup_{g\in\mathcal{G}} \mathbb{E}_P[g(f(z,\xi))]$$

and

$$w_N(z) := \sup_{g\in\mathcal{G}} \mathbb{E}_{P_N}[g(f(z,\xi))].$$

In what follows, we show uniform convergence of $w_N(z)$ to $w(z)$ under some conditions. Note that in the literature of stochastic programming, there are some recent studies on uniform convergence of SAA of a random function, see e.g., [75, 85]. Here we focus on a slightly different situation where it considers uniform exponential convergence of the maximum of a class of SAA random functions as opposed to a single SAA function in the literature.

In order to derive uniform exponential convergence of $w_N(z)$ to $w(z)$, we need to impose some conditions on $\mathcal{G}$ and f.

Assumption 3.4.2 *Let Ξ be the support set of ξ.*

(*a*) *For any $\varepsilon > 0$, there exists a compact set $\Xi_\varepsilon \subset \Xi$ such that*

$$\sup_{N,z\in Z,g\in\mathcal{G}} \mathbb{E}_{P_N}[|g(f(z,\xi))\mathbb{1}_{\Xi\setminus\Xi_\varepsilon}(\xi)|] \le \varepsilon \tag{3.4.9}$$

with probability 1.

(*b*) *For any $m > 0$, there exist positive constants κ_m, B_m and $x_0 \in [-me, me]$ such that*

$$\sup_{g\in\mathcal{G}} |g(x_1) - g(x_2)| \le \kappa_m \|x_1 - x_2\|, \forall x_1, x_2 \in [-me, me] \tag{3.4.10}$$

and $\sup_{g\in\mathcal{G}_m} |g(x_0)| \le B_m$, where

$$\mathcal{G}_m := \{g(\cdot) \in \mathcal{G} : g(\cdot) \text{ is defined over interval } [-me, me]\}. \tag{3.4.11}$$

(*c*) *There exists a measurable function $\phi(\xi) : \Xi \to \mathbb{R}_+$ and a constant $\gamma > 0$ such that*

$$\|f(z,\xi) - f(z',\xi)\| \le \phi(\xi)\|z - z'\|^\gamma, \ \forall z, z' \in Z, \ \xi \in \Xi.$$

Assumption 3.4.2 (a) is a well known uniform integrability condition for all $g \in \mathcal{G}$ [15, Chapter 3]. Condition (b) requires the class of functions in $\mathcal{G}$ to be equi-continuous over $[-me, me]$ for any $m > 0$. Condition (c) means that $f(\cdot, \xi)$ is Hölder continuous in z [75, Theorem 5.2].

Under Assumption 3.4.2, the set $\mathcal{G}_m$ is bounded by

$$\sup_{g\in\mathcal{G}_m} \|g\|_\infty \le B_m + 2\kappa_m m. \tag{3.4.12}$$

By Ascoli-Arzela Theorem [18], the Lipschitz continuity (3.4.10) and the uniform boundedness (3.4.12) guarantee that $\mathcal{G}_m$ is relatively compact. The relative compactness ensures existence of ε-net of $\mathcal{G}_m$, that is, for any $\varepsilon > 0$, there exists a set of finite number of functions $\{g_1, \ldots, g_K\} \subset \mathcal{G}_m$ such that

$$\mathcal{G}_m = \bigcup_{k=1}^{K} (\mathcal{G}_m)_k$$

where $(\mathcal{G}_m)_k := \{g \in \mathcal{G}_m : \|g - g_k\|_\infty \leq \varepsilon\}$ for $k = 1, \ldots, K$.

Lemma 3.4.1 *Let Assumption 3.4.2 hold. When N is sufficiently large, then for any $\delta > 0$ and $\varepsilon > 0$, there exist positive constants $C(\varepsilon, \delta)$ and $\beta(\varepsilon, \delta)$ such that*

$$\text{Prob}\left(\sup_{z \in Z} |w_N(z) - w(z)| \geq \delta\right) \leq C(\varepsilon, \delta) e^{-N\beta(\varepsilon,\delta)} \tag{3.4.13}$$

Proof. Under Assumption 3.4.2 (a), let

$$m_\varepsilon := \sup_{z \in Z, \xi \in \Xi_\varepsilon} \|f(z, \xi)\|_\infty$$

and $\mathcal{G}_{m_\varepsilon}$ be defined as in (3.4.11). Then $\mathcal{G}_{m_\varepsilon}$ is relatively compact. Let

$$\sup_{g \in \mathcal{G}_{m_\varepsilon}} \sup_{x \in [-m_\varepsilon e, m_\varepsilon e]} |g(x)| \leq \lambda.$$

Then

$$\Xi_\varepsilon \subset \{\xi \in \Xi : |g(f(z, \xi))| < \lambda\}, \forall z \in Z, g \in \mathcal{G}$$

and hence

$$\{\xi \in \Xi : |g(f(z, \xi))| \geq \lambda\} \subset \Xi \backslash \Xi_\varepsilon, \forall z \in Z, g \in \mathcal{G}.$$

Under condition (3.4.9), we have

$$\begin{aligned} & \sup_{N, z \in Z, g \in \mathcal{G}} \int_{\{\xi \in \Xi : |g(f(z,\xi))| \geq \lambda\}} |g(f(z, \xi))| P_N(d\xi) \\ \leq \quad & \sup_{N, z \in Z, g \in \mathcal{G}} \int_{\Xi \backslash \Xi_\varepsilon} |g(f(z, \xi))| P_N(d\xi) \leq \varepsilon. \end{aligned}$$

Then it follows by [40, Lemma 2.1] that for any $z \in Z$, $g \in \mathcal{G}$,

$$\mathbb{E}_P[g(f(z, \xi)) \mathbb{1}_{\Xi \backslash \Xi_\varepsilon}(\xi)] = \lim_{N \to \infty} \mathbb{E}_{P_N}[g(f(z, \xi)) \mathbb{1}_{\Xi \backslash \Xi_\varepsilon}(\xi)],$$

which implies

$$\sup_{z \in Z, g \in \mathcal{G}} \mathbb{E}_P[|g(f(z, \xi)) \mathbb{1}_{\Xi \backslash \Xi_\varepsilon}(\xi)|] \leq \varepsilon.$$

By the definition of $w_N(z)$ and $w(z)$, we have

$$\begin{aligned}
& w_N(z) - w(z) \\
= \; & \sup_{g\in\mathcal{G}} \mathbb{E}_{P_N}[g(f(z,\xi))] - \sup_{g\in\mathcal{G}} \mathbb{E}_P[g(f(z,\xi))] \\
\le \; & \sup_{g\in\mathcal{G}} \mathbb{E}_{P_N}[g(f(z,\xi))\mathbb{1}_{\Xi_\varepsilon}(\xi)] - \sup_{g\in\mathcal{G}} \mathbb{E}_P[g(f(z,\xi))\mathbb{1}_{\Xi_\varepsilon}(\xi)] + 2\varepsilon \\
= \; & \sup_{g\in\mathcal{G}_{m_\varepsilon}} \mathbb{E}_{P_N}[g(f(z,\xi))\mathbb{1}_{\Xi_\varepsilon}(\xi)] - \sup_{g\in\mathcal{G}_{m_\varepsilon}} \mathbb{E}_P[g(f(z,\xi))\mathbb{1}_{\Xi_\varepsilon}(\xi)] + 2\varepsilon \\
= \; & \sup_{k\in K} \sup_{g\in(\mathcal{G}_{m_\varepsilon})_k} \mathbb{E}_{P_N}[g(f(z,\xi))\mathbb{1}_{\Xi_\varepsilon}(\xi)] - \sup_{k\in K} \sup_{g\in(\mathcal{G}_{m_\varepsilon})_k} \mathbb{E}_P[g(f(z,\xi))\mathbb{1}_{\Xi_\varepsilon}(\xi)] + 2\varepsilon \\
= \; & \sup_{k\in K} \sup_{g\in(\mathcal{G}_{m_\varepsilon})_k} \mathbb{E}_{P_N}[g(f(z,\xi))\mathbb{1}_{\Xi_\varepsilon}(\xi) - g_k(f(z,\xi))\mathbb{1}_{\Xi_\varepsilon}(\xi) + g_k(f(z,\xi))\mathbb{1}_{\Xi_\varepsilon}(\xi)] \\
& - \sup_{k\in K} \sup_{g\in(\mathcal{G}_{m_\varepsilon})_k} \mathbb{E}_P[g(f(z,\xi))\mathbb{1}_{\Xi_\varepsilon}(\xi) - g_k(f(z,\xi))\mathbb{1}_{\Xi_\varepsilon}(\xi) + g_k(f(z,\xi))\mathbb{1}_{\Xi_\varepsilon}(\xi)] + 2\varepsilon \\
\le \; & \sup_{k\in K}(\varepsilon + \mathbb{E}_{P_N}[g_k(f(z,\xi))\mathbb{1}_{\Xi_\varepsilon}(\xi)]) - \sup_{k\in K}(-\varepsilon + \mathbb{E}_P[g_k(f(z,\xi))\mathbb{1}_{\Xi_\varepsilon}(\xi)]) + 2\varepsilon \\
= \; & \sup_{k\in K} \mathbb{E}_{P_N}[g_k(f(z,\xi))\mathbb{1}_{\Xi_\varepsilon}(\xi)] - \sup_{k\in K} \mathbb{E}_P[g_k(f(z,\xi))\mathbb{1}_{\Xi_\varepsilon}(\xi)] + 4\varepsilon.
\end{aligned}$$

Likewise, we can also obtain

$$w_N(z) - w(z) \ge \sup_{k\in K} \mathbb{E}_{P_N}[g_k(f(z,\xi))\mathbb{1}_{\Xi_\varepsilon}(\xi)] - \sup_{k\in K} \mathbb{E}_P[g_k(f(z,\xi))\mathbb{1}_{\Xi_\varepsilon}(\xi)] - 4\varepsilon.$$

Combining the above two inequalities, we get

$$|w_N(z) - w(z)| \le \sup_{k\in K} |\mathbb{E}_{P_N}[g_k(f(z,\xi))\mathbb{1}_{\Xi_\varepsilon}(\xi)] - \mathbb{E}_P[g_k(f(z,\xi))\mathbb{1}_{\Xi_\varepsilon}(\xi)]| + 4\varepsilon.$$

Under Assumption 3.4.2 (b) and (c), for any $g \in \mathcal{G}_{m_\varepsilon}$,

$$\begin{aligned}
& |g(f(z,\xi))\mathbb{1}_{m_\varepsilon}(\xi) - g(f(z',\xi))\mathbb{1}_{m_\varepsilon}(\xi)| \\
\le \; & \mathbb{1}_{m_\varepsilon}(\xi)\kappa_m \|f(z,\xi) - f(z',\xi)\| \\
\le \; & \mathbb{1}_{m_\varepsilon}(\xi)\kappa_m \phi(\xi)\|z - z'\|^\gamma, \forall \xi \in \Xi.
\end{aligned}$$

Then by [75, Theorem 5.1], we have

$$\begin{aligned}
& \text{Prob}\left(\sup_{z\in Z} |w_N(z) - w(z)| \ge \delta\right) \\
\le \; & \text{Prob}\left(\sup_{z\in Z}\sup_{k\in K} |\mathbb{E}_{P_N}[g_k(f(z,\xi))\mathbb{1}_{m_\varepsilon}(\xi)] - \mathbb{E}_P[g_k(f(z,\xi))\mathbb{1}_{m_\varepsilon}(\xi)]| \ge \delta - 4\varepsilon\right) \\
= \; & \text{Prob}\left(\sup_{k\in K}\sup_{z\in Z} \mathbb{E}_{P_N}[g_k(f(z,\xi))\mathbb{1}_{m_\varepsilon}(\xi)] - \mathbb{E}_P[g_k(f(z,\xi))\mathbb{1}_{m_\varepsilon}(\xi)]| \ge \delta - 4\varepsilon\right) \\
\le \; & \sum_{k\in K} \text{Prob}\left(\sup_{z\in Z} |\mathbb{E}_{P_N}[g_k(f(z,\xi))\mathbb{1}_{m_\varepsilon}(\xi)] - \mathbb{E}_P[g_k(f(z,\xi))\mathbb{1}_{m_\varepsilon}(\xi)]| \ge \delta - 4\varepsilon\right) \\
\le \; & \sum_{k\in K} C(\varepsilon,\delta,g_k) e^{-N\beta(\varepsilon,\delta,g_k)},
\end{aligned}$$

which implies that (3.4.13) holds. □

Note that the work of Haskell et al. [44] proposes a similar discretization scheme to approximate integral stochastic dominance constraints where they derive uniform exponential convergence under the condition that the utility functions are Lipschitz continuous and defined on a compact set. Here we use the uniform integrability condition to relax the compactness condition so that the convergence result can be applied to the problem (PRMSR-Opti) where the loss functions are defined on $\mathbb{R}^n$ rather than a compact set.

3.4.2 Convergence of the Optimal Values and Optimal Solutions

We now return to discuss convergence of (Opti-N) to (PRMSR-Opti) in terms of the optimal values and optimal solutions. Let $v_N(z,t)$ and $v(z,t)$ be defined as in (3.4.4) and (3.4.3). We start by deriving uniform convergence of $v_N(z,t)$ to $v(z,t)$ using Lemma 3.4.1. To this end, we need to match the conditions required by the lemma.

(C1) Let T be a compact set in $\mathbb{R}$. For any $\epsilon > 0$, there exists a compact set $\Xi_\epsilon \subset \Xi$ such that

$$\sup_{N,z\in Z,t\in T,l\in L} \mathbb{E}_{P_N}[|l(-c(z,\xi)-td)\mathbb{1}_{\Xi\setminus\Xi_\epsilon}(\xi)|] \le \epsilon$$

and

$$\sup_{N,l\in L} \mathbb{E}_{P_N}[|l(-Y(\xi))\mathbb{1}_{\Xi\setminus\Xi_\epsilon}(\xi)|] \le \epsilon.$$

(C2) For any $m > 0$, there exist positive constants κ_m, B_m and $x_0 \in [-me, me]$ such that

$$\sup_{l\in L} |l(x_1) - l(x_2)| \le \kappa_m \|x_1 - x_2\|, \forall x_1, x_2 \in [-me, me]$$

and $\sup_{l\in L_m} |l(x_0)| \le B_m$, where

$$L_m := \{l(\cdot) \in L : l(\cdot)\ is\ defined\ over\ interval\ [-me, me]\}.$$

(C3) There exists a measurable function $r(\xi) : \Xi \to \mathbb{R}_+$ and a constant $\gamma > 0$ such that

$$\|c(z,\xi) - c(z',\xi)\| \le \phi(\xi)\|z - z'\|^\gamma,\ \forall z, z' \in Z,\ \xi \in \Xi.$$

Theorem 3.4.1 *Let conditions (C1)-(C3) hold. Let T be a compact set in $\mathbb{R}$. Then for any $\delta > 0$ and $\epsilon > 0$, there exist positive constants $N(\epsilon,\delta)$, $C(\epsilon,\delta)$ and $\beta(\epsilon,\delta)$ such that*

$$\text{Prob}\left(\sup_{z\in Z,t\in T} |v_N(z,t) - v(z,t)| \ge \delta\right) \le C(\epsilon,\delta)e^{-N\beta(\epsilon,\delta)} \tag{3.4.14}$$

for $N \ge N(\epsilon,\delta)$.

Proof. It is analogous to that of Lemma 3.4.1. □

The following theorem states convergence of (Opti-N) to (PRMSR-Opti) in terms of the optimal values and optimal solutions as the sample size N increases.

Theorem 3.4.2 *Let Assumption 3.4.1 and conditions (C1)-(C3) hold. Suppose that for almost every $\xi \in \Xi$, $c(\cdot, \xi)$ is a concave function. Then*

(i) For any $\delta \leq \theta$,

$$\text{Prob}(|\vartheta_N - \vartheta| \geq \delta) \leq C(\epsilon, \varepsilon)e^{-N\beta(\epsilon,\varepsilon)} \tag{3.4.15}$$

for $N \geq N(\epsilon, \varepsilon)$, where $N(\epsilon, \varepsilon)$, $C(\epsilon, \varepsilon)$ and $\beta(\epsilon, \varepsilon)$ are defined as in Theorem 3.4.1 and ε is some positive constant depending on δ, and θ is given in (3.4.7).

(ii) Let $\{z_N, t_N\}$ be a sequence of optimal solution obtained from solving problem (Opti-N). Then with probability 1, a cluster point of the sequence is an optimal solution of problem (PRMSR-Opti).

Proof. Part (i). Let $t^* = \vartheta$. Following the discussions immediately after Assumption 3.4.1, we know that t^* is finite and $t^* \leq t_0$. Let δ be given as in Theorem 3.4.1 with $\delta \leq \theta$ and η be any fixed positive constant such that $\eta \geq \delta$. Then there exists a constant $c_\eta > 0$ such that

$$\inf_{z \in Z} v(z, t^* - c_\eta) \geq \eta. \tag{3.4.16}$$

To see the existence, notice that

$$\begin{aligned}
& \inf\nolimits_{z \in Z} v(z, t^* - c_\eta) \\
= \;& \inf_{z \in Z} \sup_{l \in L} \mathbb{E}_P[l(-c(z, \xi) - (t^* - c_\eta)d)] - \mathbb{E}_P[l(-Y(\xi))] \\
\geq \;& \inf_{z \in Z} \mathbb{E}_P[l(-c(z, \xi) - (t^* - c_\eta)d)] - \mathbb{E}_P[l(-Y(\xi))] \; (for\ any\ l \in L) \\
\geq \;& \inf_{z \in Z} l(\mathbb{E}_P[-c(z, \xi)] - (t^* - c_\eta)d) - \mathbb{E}_P[l(-Y(\xi))] \; (by\ convexity\ of\ l) \\
= \;& l\left(\inf_{z \in Z} \mathbb{E}_P[-c(z, \xi)] - (t^* - c_\eta)d\right) - \mathbb{E}_P[l(-Y(\xi))] \; (by\ monotonicity\ of\ l).
\end{aligned}$$

Since Z is a compact set and $\mathbb{E}_P[-c(z, \xi)]$ is continuous, then $\inf_{z \in Z} \mathbb{E}_P[-c(z, \xi)]$ is bounded. Moreover, since $\lim_{t \to +\infty} l(t) = +\infty$, the last term goes beyond η for a sufficiently large c_η and hence (3.4.16) holds.

Let T in Theorem 3.4.1 be chosen such that $[t^* - c_\eta, t_0] \subset T$. Then by Theorem 3.4.1

$$\begin{aligned}
& \inf_{z \in Z} v_N(z, t^* - c_\eta) \\
= \;& \inf_{z \in Z} v(z, t^* - c_\eta) + \inf_{z \in Z} v_N(z, t^* - c_\eta) - \inf_{z \in Z} v(z, t^* - c_\eta) \\
\geq \;& \eta - \sup_{z \in Z} |v_N(z, t^* - c_\eta) - v(z, t^* - c_\eta)| \\
> \;& \eta - \delta/2
\end{aligned}$$

with probability at least $1 - C(\epsilon, \delta/2)e^{-N\beta(\epsilon,\delta/2)}$ for $N \geq N(\epsilon, \delta/2)$. Let $(z_N, t_N) \in S_N$ be the optimal solution of (Opti-N). The inequality above shows

$$v_N(z_N, t^* - c_\eta) \geq \inf_{z \in Z} v_N(z, t^* - c_\eta) > \eta - \delta/2, \tag{3.4.17}$$

which implies $t_N > t^* - c_\eta$.

On the other hand, it follows by (3.4.7) and Theorem 3.4.1,

$$\begin{aligned}
& \sup_{l \in L} \mathbb{E}_{P_N}[l(-c(z_0, \xi) - t_0 d) - l(-Y)] \\
= \;& \sup_{l \in L} \mathbb{E}_P[l(-c(z_0, \xi) - t_0 d) - l(-Y)] + \sup_{l \in L} \mathbb{E}_{P_N}[l(-c(z_0, \xi) - t_0 d) - l(-Y)] \\
& - \sup_{l \in L} \mathbb{E}_P[l(-c(z_0, \xi) - t_0 d) - l(-Y)] \\
\leq \;& -\theta + v_N(z_0, t_0) - v(z_0, t_0) \\
\leq \;& -\theta + \sup_{z \in Z, t \in T} |v_N(z, t) - v(z, t)| \\
< \;& -\theta + \delta/2 \;\leq\; -\delta/2
\end{aligned} \tag{3.4.18}$$

with probability at least $1 - C(\epsilon, \delta/2)e^{-N\beta(\epsilon,\delta/2)}$ for $N \geq N(\epsilon, \delta/2)$. The inequality (3.4.18) implies (z_0, t_0) is a feasible solution to (Opti-N) and hence $t_N \leq t_0$. Summarizing the discussions above, we have

$$t_N \in [t^* - c_\eta, t_0] \tag{3.4.19}$$

with probability at least $1 - C(\epsilon, \delta/2)e^{-N\beta(\epsilon,\delta/2)}$.

Let us consider the systems of inequalities

$$v(z, t) \leq 0, \;\; (z, t) \in Z \times T$$

and

$$v_N(z, t) \leq 0, \;\; (z, t) \in Z \times T.$$

The set of solutions to the systems of inequalities are equal to $\mathcal{F} \cap Z \times T$ and $\mathcal{F}_N \cap Z \times T$ respectively. Since $l(-c(z, \xi) - td)$ is convex in (z, t), both $v(z, t)$ and $v_N(z, t)$ are convex functions. By the Slater condition (3.4.7) and Robinson's error bound theorem for convex systems [67], for any $(z, t) \in Z \times T$,

$$d((z, t), \mathcal{F} \cap Z \times T) \leq \frac{\Delta}{\theta} \max\{v(z, t), 0\},$$

where Δ denotes the diameter of $\mathcal{F} \cap Z \times T$ and we write $d(a, A)$ for the distance from a point a to a set A. Likewise, by the Slater condition (3.4.18), for any $(z, t) \in Z \times T$,

$$d((z, t), \mathcal{F}_N \cap Z \times T) \leq \frac{2\Delta}{\delta} \max\{v_N(z, t), 0\}.$$

with probability at least $1-C(\epsilon,\delta/2)e^{-N\beta(\epsilon,\delta/2)}$ for $N \geq N(\epsilon,\delta/2)$. Combining the two error bounds, we obtain (see Appendix C)

$$\mathbb{H}(\mathcal{F}_N \cap Z \times T, \mathcal{F} \cap Z \times T) \leq \frac{2\Delta}{\delta} \sup_{z\in Z, t\in T} |v_N(z,t) - v(z,t)|, \tag{3.4.20}$$

where $\mathbb{H}$ denotes the Hausdorff distance. Thus

$$|\vartheta_N - \vartheta| = |t_N - t^*| \leq \mathbb{H}(\mathcal{F}_N \cap Z \times T, \mathcal{F} \cap Z \times T) \leq \frac{2\Delta}{\delta} \sup_{z\in Z, t\in T} |v_N(z,t) - v(z,t)|. \tag{3.4.21}$$

Let $\varepsilon := \min(\frac{\delta^2}{2\Delta}, \frac{\delta}{2})$. We deduce from (3.4.14) and (3.4.21)

$$\begin{aligned} \text{Prob}(|\vartheta_N - \vartheta| \geq \delta) &\leq \text{Prob}\left(\sup_{z\in Z, t\in T} |v_N(z,t) - v(z,t)| \geq \varepsilon\right) \\ &\leq C(\epsilon,\varepsilon)e^{-N\beta(\epsilon,\varepsilon)} \end{aligned}$$

for $N \geq N(\epsilon,\varepsilon)$.

Part (ii). The exponential rate of convergence (3.4.15) implies $t_N \to t^*$ almost surely. Moreover, since $v_N(z_N, t_N) \leq 0$ and v_N uniformly converges to v over $Z \times T$, then $v(\hat{z}, t^*) \leq 0$ for every cluster point $\hat{z}$ of $\{z_N\}$. □

3.5 Tractable Formulation of (Opti-N)

In this section, we develop a tractable formulation for the approximated optimization problem (Opti-N). Let us start by specifying the ambiguity set of multivariate loss functions L. We require the multivariate loss functions to satisfy certain properties. Below is a list of them.

(C1) $L_1 \subseteq \mathcal{L}$;

(C2) L_2 is a set of uniformly Lipschitz continuous functions with modulus κ, that is,

$$\sup_{l\in L_2} |l(x_1) - l(x_2)| \leq \kappa\|x_1 - x_2\|, \forall x_1, x_2 \in \mathbb{R}^n.$$

Condition (C2) means that the multivariate loss functions will not change rapidly at any point. Moreover, we suppose that the information on the decision makers preference can be elicited through pairwise comparison of questionnaires or lotteries and add some additional conditions as follows:

(C3) $L_3 := \{l : \mathbb{E}_P[l(-G_0)] - \mathbb{E}_P[l(-B_0)] = -1, \mathbb{E}_P[l(-O)] = 0\}$;

(C4) $L_4 := \{l : \mathbb{E}_P[l(-G_k)] \leq \mathbb{E}_P[l(-B_k)], k = 1, \ldots, K\}$.

Here $G_0, \ldots, G_K$, $B_0, \ldots, B_K$ are given random vectors representing lotteries, and O is a zero vector with dimension n. Condition (C3) is used to normalise the multivariate loss functions which specifies that the utility difference between $-G_0$ and $-B_0$ is 1. Condition (C4) means admissible multivariate loss functions must satisfy the pairwise preference elicitation conditions, that is, L_4 denotes the set of utilities that prefer $-B_k$ to $-G_k$ for all k. Conditions (C3) and (C4) are also used in [6] to define the set of decision maker's utility functions.

We consider L has the specified structure that satisfies Conditions (C1)-(C4), that is,

$$L = L_1 \cap L_2 \cap L_3 \cap L_4. \tag{3.5.1}$$

Following [70, Theorem 9.13], l is locally Lipschitz continuous at $\bar{x}$ with $l(\bar{x})$ finite if and only if the convex subdifferential mapping $\partial l : x \to \partial l(x)$ is locally bounded at $\bar{x}$. Furthermore, the Lipschitz modulus $\bar{\kappa}$ of l at $\bar{x}$ is equal to

$$\bar{\kappa} = \max_{a \in \partial l(\bar{x})} \|a\|.$$

Hence, Condition (C2) can be ensured by

$$\max_{a \in \partial l(x)} \|a\| \leq \kappa, \forall x \in \mathbb{R}^n, l \in L. \tag{3.5.2}$$

We define

$$\Theta := \bigcup_{k=0}^{K} (\text{supp}(-G_k) \cup \text{supp}(-B_k)) \cup \text{supp}(-O) \cup \text{supp}(-Y)$$

to be the joint support of all random vectors. For convenience, let θ_j denote the j-th entry of Θ, and the size of the support is denoted by $J := |\Theta|$.

As in Delage et al. [24], we will use support function approach to reformulate the robust constraint in (Opti-N) as a linear programming problem. The main challenge that we need to tackle here is that the loss functions are multivariate. To this effect, let us recall some preliminary definitions associated with support functions.

Definition 3.5.1 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. A function g is said to be majorized by f if*

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

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

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

and we denote the set of subgradients of f at x by $\partial f(x)$ which is known as subdifferential of f at x.

When f is convex and subdifferentiable at x, the linear function

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

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

Theorem 3.5.1 *Let $f : \mathbb{R}^n \to \mathbb{R}$. The following assertions hold:*

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

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

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

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

$$\hat{f}(x) = \max_{a,b} \{ \langle a, x \rangle + b : \langle a, \theta \rangle + b \leq v_\theta, \ \forall \theta \in \Theta \} \tag{3.5.3}$$

is convex.

These results are well known, see [17] for instance. Part (i) states that a convex function can be recovered by taking the supremum of its support functions, and Part (ii) gives conditions for constructing the "highest" convex function that contains a set of values $\{v_\theta\}_{\theta \in \Theta}$ over a finite set Θ. Note that in (3.5.3), point x is not necessarily in set Θ. The next theorem states that any positive combination of the function values at a set of points may be calculated by solving a single linear programming problem. Moreover if there is a set of convex functions each of which has some specified values over Θ, then the worst value of the positive combination may be obtained by solving a single linear programming problem with v_θ being treated as a vector of variables.

Theorem 3.5.2 *Let L be a set of proper convex functions from $\mathbb{R}^n$ to $\mathbb{R}$ and $\Theta \subset \mathbb{R}^n$ be a discrete set of points. For each $l \in L$, let $v = \{v_1, \cdots, v_J\} = \{l(\theta) : \theta \in \Theta\}$ be the set of values of l over Θ, where $J = |\Theta|$ denotes the cardinality of Θ. Then for given*

$t_k \in \mathbb{R}^n$, *positive number* α_k, $k = 1, \cdots, K$ *and* β_j, $j = 1, \cdots, J$,

$$\sum_{k=1}^{K} \alpha_k l(t_k) - \sum_{j=1}^{J} \beta_j l(\theta_j) = \sup_{(A,b)\in\mathcal{F}(v)} \left\{ \sum_{k=1}^{K} \alpha_k (a_k^T t_k + b_k) - \sum_{j=1}^{J} \beta_j v_j \right\} \tag{3.5.4}$$

and

$$\sup_{l\in L} \left\{ \sum_{k=1}^{K} \alpha_k l(t_k) - \sum_{j=1}^{J} \beta_j l(\theta_j) \right\} = \sup_{v,(A,b)\in\mathcal{F}(v)} \left\{ \sum_{k=1}^{K} \alpha_k (a_k^T t_k + b_k) - \sum_{j=1}^{J} \beta_j v_j \right\}, \tag{3.5.5}$$

where we write (A, b) *for* $\{(a_j, b_j)\}_{j=1}^J$ *and*

$$\mathcal{F}(v) := \{(a_j, b_j), j = 1, \cdots, J : a_j^T \theta_i + b_j \le v_i, \text{ for } i = 1, \cdots, J\}.$$

Proof. Observe that (3.5.5) follows from (3.5.4) by taking supremum with respect to l over L at the left hand side of the equation and v (function values of l over Θ) on the right had side, it suffices to show (3.5.4).

For the fixed $t_k \in \mathbb{R}^n$, $k = 1, \cdots, K$, it follows by Theorem 3.5.1 (ii) that

$$l(t_k) = \sup \left\{ a_k^T t_k + b_k : (A, b) \in \mathcal{F}(v) \right\},$$

Since (a_i, b_i) and (a_j, b_j) are two independent pairs of variables for $i \neq j$ and their feasible ranges are not affected each other, then the maximization problem at the right hand side of equation (3.5.4) is decomposeable (v is fixed there). This shows (3.5.4). □

Let us recall problem (Opti-N):

$$\begin{aligned} \min_{z\in Z, t\in\mathbb{R}} \quad & t \\ \text{s.t.} \quad & v_N(z, t) \le 0, \end{aligned}$$

where $v_N(z, t) := \sup_{l\in L} \mathbb{E}_{P_N}[l(-c(z, \xi) - td) - l(-Y(\xi))]$. With the specified structure of L (defined as in (3.5.1)), we first give a tractable formulation for $v_N(z, t)$.

Proposition 3.5.1 *Assume that* ξ *is discretely distributed with* $p_i = P(\xi = \xi_i)$ *for*

$i = 1, \dots, N$. Then $v_N(z,t)$ can be reformulated as

$$
\begin{aligned}
\max_{a_i\in\mathbb{R}^n, b_i\in\mathbb{R}, v_j\in\mathbb{R}, s_j\in\mathbb{R}^n} \quad & \sum_{i=1}^{N} p_i(\langle a_i, -c(z,\xi_i) - td\rangle + b_i) - \sum_{j=1}^{J} P(-Y=\theta_j)v_j \\
\text{s.t.} \quad & \sum_{j=1}^{J} P(-G_0=\theta_j)v_j - \sum_{j=1}^{J} P(-B_0=\theta_j)v_j = -1, \\
& \sum_{j=1}^{J} P(-O=\theta_j)v_j = 0, \\
& \sum_{j=1}^{J} P(-G_k=\theta_j)v_j \le \sum_{j=1}^{J} P(-B_k=\theta_j)v_j, \quad k=1,\dots,K, \\
& \langle a_i, \theta_j\rangle + b_i \le v_j, \quad j=1,\dots,J; i=1,\dots,N, \\
& \langle s_j, \theta_q - \theta_j\rangle + v_j \le v_q, \quad j=1,\dots,J; q=1,\dots,J, \\
& 0 \le a_i \le \kappa e, \quad i=1,\dots,N, \\
& 0 \le s_j \le \kappa e, \quad j=1,\dots,J,
\end{aligned}
\tag{3.5.6}
$$

where $\theta_j \in \Theta$, v_j is the value of l on θ_j, and $J = (2K+3)N+1$.

Proof. Let $L(v) := \{l : l(\theta_j) = v_j, j = 1, \dots, J\}$. Then

$$
\begin{aligned}
v_N(z,t) &= \left\{ \begin{array}{ll} \sup\limits_{v} \sup\limits_{l\in L(v)\cap L} & \mathbb{E}_{P_N}[l(-c(z,\xi)-td) - l(-Y(\xi))] \\ \quad \text{s.t.} & L(v)\cap L \neq \emptyset, \end{array} \right. \\
&= \left\{ \begin{array}{ll} \sup\limits_{v} \sup\limits_{l\in L(v)\cap L_1\cap L_2} & \mathbb{E}_{P_N}[l(-c(z,\xi)-td) - l(-Y(\xi))] \\ \quad \text{s.t.} & L(v)\cap L_1\cap L_2 \neq \emptyset, L(v)\subset L_3, L(v)\subset L_4, \end{array} \right.
\end{aligned}
$$

where the first equality is due to the fact that $\bigcup_v L(v)\cap L = L$ and the second equality holds since $L(v)$ is either a subset of L_3 or is disjoint from it and the same is true with respect to L_4. Let us characterize the feasible set of the last program in the sequel.

Based on Theorem 3.5.1 and (3.5.2), constraint $L(v)\cap L_1\cap L_2 \neq \emptyset$ can be represented as

$$
\begin{aligned}
& \langle s_j, \theta_q - \theta_j\rangle + v_j \le v_q, \quad j=1,\dots,J; q=1,\dots,J, \\
& \|s_j\|_\infty \le \|s_j\| \le \kappa, \quad j=1,\dots,J, \\
& s_j \ge 0, \quad j=1,\dots,J.
\end{aligned}
$$

Here the reason why we adopt the infinity norm rather than Euclidean norm is that the resulting constraint with the former is linear which is easy to handle, otherwise the resulting problem (3.5.6) will be a quadratic optimization problem. Constraint

$L(v) \subset L_3$ can be characterized by

$$\begin{array}{l}\sum_j (P(-G_0 = \theta_j)v_j - P(-B_0 = \theta_j)v_j) = -1,\\ \sum_j P(-O = \theta_j)v_j = 0,\end{array}$$

and $L(v) \subset L_4$ by

$$\sum_j P(-G_k = \theta_j)v_j \le \sum_j P(-B_k = \theta_j)v_j, \quad k = 1, \ldots, K.$$

All of the three sets of the constraints form the feasible set of v, denoted by V.

Next we look at the objective function of problem (3.5.6). Note that for $l \in L(v)$,

$$\begin{aligned}\mathbb{E}_{P_N}[l(-Y(\xi))] &= \sum_{i=1}^{N} p_i l(-Y(\xi_i))\\ &= \sum_{j=1}^{J} P(-Y = \theta_j)v_j,\end{aligned}$$

where the second equality follows from the fact that for $j = 1, \ldots, J$, $i = 1, \ldots, N$,

$$P(-Y = \theta_j) = \begin{cases} p_i, & -Y(\xi_i) = \theta_j,\\ 0, & otherwise,\end{cases}$$

and $l(-Y(\xi_i)) = v_j$ if $-Y(\xi_i) = \theta_j$. Then for fixed $v \in V$, it follows by Theorem 3.5.2 that

$$\begin{aligned}&\mathbb{E}_{P_N}[l(-c(z,\xi) - td) - l(-Y(\xi))]\\ =\;& \max_{(A,b)\in\mathcal{F}(v)} \sum_{i=1}^{N} p_i(\langle a_i, -c(z,\xi_i) - td\rangle + b_i) - \sum_{j=1}^{J} P(-Y = \theta_j)v_j,\end{aligned} \tag{3.5.8}$$

where

$$\mathcal{F}(v) := \{(a_j, b_j), j = 1, \cdots, J : \langle a, \theta_j\rangle + b \le v_j, j = 1, \ldots, J, \|a\|_\infty \le \kappa, a \ge 0\}.$$

By taking supremum with respect to l over L (where L satisfies $L(v) \cap L_1 \cap L_2 \neq \emptyset, L(v) \subset L_3, L(v) \subset L_4$) at the left hand side of (3.5.8) and v over V at the right hand side of (3.5.8), we obtain (3.5.6). □

Proposition 3.5.1 indicates that for each fixed z, the value of $v_N(z,t)$ can be computed by solving a finite dimensional linear program of reasonable size with $(n+1)(N+J)$ variables and $J^2 + NJ + (N+J)n + K + 1$ constraints (not counting the non-negativity constraints). However, if we merge program (3.5.6) directly into the constraint of (Opti-N), then the latter becomes an optimization problem with semi-infinite constraints (indexed by parameters (a_i, b_i, v_j, s_j)) which is undesirable. This motivates us to consider the Lagrangian dual of (3.5.6) which is a minimization problem and can be better integrated into (Opti-N). To this end, we define the Lagrangian of program (3.5.6)

$$
\begin{aligned}
& L(a,b,v,s;\gamma,\tau,\lambda,\mu,\eta,\alpha,\beta) \\
= \; & \sum_{i=1}^{N} p_i(\langle a_i, -c(z,\xi_i)) - td\rangle + b_i) - \sum_{j=1}^{J} P(-Y=\theta_j)v_j + \gamma\left(\sum_{j=1}^{J} P(-O=\theta_j)v_j\right) \\
& +\tau\left(\sum_{j=1}^{J} P(-G_0=\theta_j)v_j - \sum_{j=1}^{J} P(-B_0=\theta_j)v_j + 1\right) + \sum_{j=1}^{J}\alpha_j(\kappa e - s_j) \\
& +\sum_{k=1}^{K}\lambda_k\left(\sum_{j=1}^{J} P(-B_k=\theta_j)v_j - \sum_{j=1}^{J} P(-G_k=\theta_j)v_j\right) + \sum_{i=1}^{N}\beta_i(\kappa e - a_i) \\
& +\sum_{j=1}^{J}\sum_{j'=1}^{J}\mu_{jj'}(v_{j'} - v_j + s_j(\theta_j - \theta_{j'})) + \sum_{i=1}^{N}\sum_{j=1}^{J}\eta_{ij}(v_j - b_i - a_i^{\mathrm{T}}\theta_j) \\
= \; & \sum_{i=1}^{N} a_i^{\mathrm{T}}\left(p_i(-c(z,\xi_i) - td) - \sum_{j=1}^{J}\eta_{ij}\theta_j - \beta_i\right) + \sum_{i=1}^{N} b_i\left(p_i - \sum_{j=1}^{J}\eta_{ij}\right) \\
& +\sum_{j=1}^{J} v_j[-P(-Y=\theta_j) + \gamma P(-O=\theta_j) + \tau(P(-G_0=\theta_j) - P(-B_0=\theta_j)) \\
& +\sum_{k=1}^{K}\lambda_k(P(-B_k=\theta_j) - P(-G_k=\theta_j)) + \sum_{j'=1}^{J}\mu_{j'j} - \sum_{j'=1}^{J}\mu_{jj'} + \sum_{i=1}^{N}\eta_{ij}] \\
& +\sum_{j=1}^{J} s_j\left(\sum_{j'=1}^{J}\mu_{jj'}(\theta_j - \theta_{j'}) - \alpha_j\right) + \tau + \kappa\langle e, \sum_{j=1}^{J}\alpha_j + \sum_{i=1}^{N}\beta_i\rangle.
\end{aligned}
$$

Since strong duality holds to problem (3.5.6), then the optimal value of (3.5.6) is equal to the optimal value of its Lagrange dual

$$
\begin{aligned}
\min_{\gamma,\tau,\lambda,\mu,\alpha,\eta,\beta} \quad & \tau + \kappa\langle e, \sum_{j=1}^{J} \alpha_j + \sum_{i=1}^{N} \beta_i \rangle \\
\text{s.t.} \quad & \sum_{j'=1}^{J} \mu_{j'j} - \sum_{j'=1}^{J} \mu_{jj'} + \sum_{i=1}^{N} \eta_{ij} \\
& + \sum_{k=1}^{K} \lambda_k (P(-B_k = \theta_j) - P(-G_k = \theta_j)) - P(-Y = \theta_j) \\
& + \gamma P(-O = \theta_j) + \tau(P(-G_0 = \theta_j) - P(-B_0 = \theta_j)) = 0, \quad j = 1, \ldots, J, \\
& p_i(c(z, \xi_i) + td) + \sum_{j=1}^{J} \eta_{ij}\theta_j + \beta_i \geq 0, \quad i = 1, \ldots, N, \\
& \sum_{j=1}^{J} \eta_{ij} = p_i, \quad i = 1, \ldots, N, \\
& \sum_{j'=1}^{J} \mu_{jj'}(\theta_j - \theta_{j'}) - \alpha_j \leq 0, \quad j = 1, \ldots, J, \\
& \lambda_k \geq 0, \quad k = 1, \ldots, K, \\
& \mu_{jq} \geq 0, \quad j = 1, \ldots, J; q = 1, \ldots, J, \\
& \eta_{ij} \geq 0, \quad i = 1, \ldots, N; j = 1 \ldots, J, \\
& \alpha_j \geq 0, \quad j = 1, \ldots, J, \\
& \beta_i \geq 0, \quad i = 1, \ldots, N.
\end{aligned}
\tag{3.5.9}
$$

By merging (3.5.9) into (Opti-N), we obtain

$$
\begin{aligned}
\min_{z \in Z, t \in \mathbb{R}} \quad & t \\
\text{s.t.} \quad & \min_{(\gamma,\tau,\lambda,\mu,\alpha,\eta,\beta) \in \mathcal{F}(z)} \tau + \kappa\langle e, \sum_{j=1}^{J} \alpha_j + \sum_{i=1}^{N} \beta_i \rangle \leq 0,
\end{aligned}
\tag{3.5.10}
$$

where $\mathcal{F}(z)$ denotes the feasible set of program (3.5.9). It is easy to show that problem (3.5.10) is equivalent to

$$
\begin{aligned}
\min_{z \in Z, t \in \mathbb{R}, \gamma,\tau,\lambda,\mu,\alpha,\eta,\beta} \quad & t \\
\text{s.t.} \quad & \tau + \kappa\langle e, \sum_{j=1}^{J} \alpha_j + \sum_{i=1}^{N} \beta_i \rangle \leq 0, \\
& (\gamma, \tau, \lambda, \mu, \alpha, \eta, \beta) \in \mathcal{F}(z).
\end{aligned}
\tag{3.5.11}
$$

This enables us to give tractable formulation of (Opti-N) as stated in the following theorem.

Theorem 3.5.3 *Let $p_i = P(\xi = \xi_i)$ for $i = 1, \ldots, N$. If problem (Opti-N) is feasible, then it equals to the following problem:*

$$
\begin{aligned}
\min_{z,t,\gamma,\tau,\lambda,\mu,\alpha,\eta,\beta} \quad & t \\
\text{s.t.} \quad & \tau + \kappa\langle e, \sum_{j=1}^{J} \alpha_j + \sum_{i=1}^{N} \beta_i \rangle \leq 0, \\
& p_i(c(z, \xi_i) + td) + \sum_{j=1}^{J} \eta_{ij}\theta_j + \beta_i \geq 0, \quad i = 1, \ldots, N, \\
& -P(-Y = \theta_j) + \gamma P(-O = \theta_j) + \tau(P(-G_0 = \theta_j) - P(-B_0 = \theta_j)) \\
& + \sum_{k=1}^{K} \lambda_k (P(-B_k = \theta_j) - P(-G_k = \theta_j)) \\
& + \sum_{j'=1}^{J} \mu_{j'j} - \sum_{j'=1}^{J} \mu_{jj'} + \sum_{i=1}^{N} \eta_{ij} = 0, \quad j = 1, \ldots, J, \\
& \sum_{j'=1}^{J} \mu_{jj'}(\theta_j - \theta_{j'}) - \alpha_j \leq 0, \quad j = 1, \ldots, J, \\
& \sum_{j=1}^{J} \eta_{ij} = p_i, \quad i = 1, \ldots, N, \\
& \lambda_k \geq 0, \quad k = 1, \ldots, K, \\
& \mu_{jq} \geq 0, \quad j = 1, \ldots, J; q = 1, \ldots, J, \\
& \eta_{ij} \geq 0, \quad i = 1, \ldots, N; j = 1 \ldots, J, \\
& \alpha_j \geq 0, \quad j = 1, \ldots, J, \\
& \beta_i \geq 0, \quad i = 1, \ldots, N, \\
& z \in Z.
\end{aligned}
\tag{3.5.12}
$$

Note that problem (3.5.12) is a convex programming problem and it is linear when $c(z, \xi_i)$ is linear in z. Theorem 3.5.3 establishes a tractable reformulation of problem (Opti-N) and we will use it in the later numerical tests.

3.6 Numerical Tests

In this section, we apply our model to a multi-criteria resource allocation problem in homeland security where we aim to allocate limited budget among major cities in the United States to protect against terrorist threats. This application is based on the case study investigated in Hu et al. [45] and Haskell et al. [43]. We use this example to examine the performance of our proposed model.

3.6.1 Problem Setup

Let us start by describing the problem setup and notations. As mentioned by Hu et al. [45], there are four criteria giving a comprehensive measure of the health of a city, thus in this case study, we believe that there are $n = 4$ criteria for estimating the effect of a terrorist attack. In order to make the background of the case study clearer, we extract the definition of these criteria from Haskell et al. [43]: *property loss* measures the impact on economic structures and individual property; *fatalities* measures the loss of human lives; *air departures* measures the loss for the citys airport; *bridge traffic* measures the loss for the citys bridge traffic. Based on Hu et al. [45], there are three possible underlying loss scenarios corresponding to different levels of terrorist attacks: reduced loss, standard loss and increased loss, which means the size of sample space is $|\Omega| = 3$.

Suppose that we have a fund A to be allocated among $m = 10$ cities. Let z_{ij} denote the fund to be allocated to city i corresponding to criterion j and $z \in Z \subset \mathbb{R}^{m\times n}$, where Z represents the set of feasible fund allocations:

$$Z := \left\{ z \in \mathbb{R}^{m\times n} : \sum_{i=1}^{m}\sum_{j=1}^{n} z_{ij} \leq A \right\}.$$

The random loss at city i with respect to criterion j for scenario ω is denoted by $\xi_{ij}(\omega)$. Thus a random matrix $\xi \in \Xi \subset \mathbb{R}^{m\times n}$ captures the loss for all criteria in all cities. To simplify the discussions, we assume that the benefits per monetary unit to be allocated to city i corresponding to criterion j are the same.

Given a particular budget allocation $z \in Z$, similar to Willis et. al. [84], the cost of budget misallocation with respect to criterion j is measured by the function:

$$c_j(z)(\omega) := \sum_{i=1}^{m} (\xi_{ij}(\omega) - z_{ij})_+, \quad \omega \in \Omega \text{ and } j = 1, \dots, n. \tag{3.6.1}$$

The quantity $c_j(z)(\omega)$ can be viewed as the sum of the shortfall with respect to criterion j over all cities. We combine the shortfall for each criterion into the vector-valued mapping

$$c(z)(\omega) = (c_1(z)(\omega), \dots, c_n(z)(\omega)), \quad \omega \in \Omega. \tag{3.6.2}$$

The negative value of the shortfall, i.e. $-c(z)(\omega)$ can be viewed as the “reward”.

In this numerical study, we consider three models. The first one is the risk-neutral and minimizes the expected shortfall:

$$\min_{z\in Z} \mathbb{E}[w^T c(z)(\omega)], \tag{3.6.3}$$

where $w \in \mathbb{R}^n$ with $w \geq 0$ and $\|w\| = 1$. The second model is to minimize the aggregated CVaR:

$$\min_{z\in Z} \sum_{j=1}^{n} \mathrm{CVaR}_{\alpha_j}(c_j(z)(\omega)) = \min_{z\in Z} \sum_{j=1}^{n} \inf_{\eta_j \in \mathbb{R}} \left\{ \eta_j + \frac{1}{1-\alpha_j} \mathbb{E}[(c_j(z)(\omega) - \eta_j)_+] \right\}, \quad (3.6.4)$$

where $\alpha_j \in (0, 1)$ is the level of confidence for criterion j. The third one is our PRMSR model:

$$\min_{z\in Z} \mathrm{MSR}^{P}_{L,d}(c(z)(\omega)), \quad (3.6.5)$$

where L is an ambiguity set of multivariate loss functions l and d is a given weighting vector in $\mathcal{D}$.

3.6.2 Data

We now describe the data set used in our experiments. We assume that the probability distribution over three loss scenarios are equal and use the same data set considered by Hu et al. [45] and Haskell et al. [43]. We give the data set for completeness of the paper in Appendix E.

In our framework, the ambiguity set L is defined as in (3.5.1). To construct L, we assume that the decision maker has a true multivariate loss function denoted by $l_{\text{true}} : \mathbb{R}^n \to \mathbb{R}$, which is unknown. However, the decision maker has her/his choice function denoted by $\rho : \mathscr{L}^0(\mathbb{R}^n) \to \mathbb{R}$ when s/he is asked to choose the preferred prospect from each pair of prospects (the preferred prospect is denoted by G_k and the other is denoted by B_k, $k = 0, \ldots, M$, the first pair G_0 and B_0 is used in the normalization property). We record G_k and B_k in an "elicited comparison data set" and an example is presented in Appendix F.

For simplification, in this case study we take a continuous, increasing and convex function $u : \mathbb{R} \to \mathbb{R}$ and choose weights $w \in \mathbb{R}^n$ with $w \geq 0$ and $\|w\| = 1$ to define the choice function

$$\rho(X) := \mathbb{E}[u(\langle w, X\rangle)], \quad X \in \mathscr{L}^0(\mathbb{R}^n). \quad (3.6.6)$$

Here we use the exponential function $u(y) = \exp(\beta y)$ with $\beta = 0.01$. Moreover, we set the fund to be $A = \$500$ million and assume the fund allocated to each city i corresponding to criterion j not less than \$1 million and no more than \$30 million. We let $Y(\xi)$ be the loss calculated based on an equally weighted fund allocation strategy. We take the modulus of Lipschitz continuity to be $\kappa = 1/12$. We let all elements in vector w and vector d to be $1/n$. The computational details of these three models are given in Appendix D.

3.6.3 Experiments

In this subsection, we describe the details of our experiments and present the results.

Experiment I: Sensitivity test of three models to data perturbation

In the first experiment, we compare the sensitivity of the three models (3.6.3), (3.6.4) and (3.6.5) to data perturbation. The elicited comparison data set used for this experiment is presented in Table F.1 in Appendix F, it contains five pairs of prospects for pairwise comparison and one pair for normalization. The optimal budget allocations are presented in Table 3.1. From the Table 3.1 we can see that more fund are allocated to the Bridge traffic compared with other three criteria and New York, Chicago and San Francisco are the three cities that are allocated the most fund.

To make the comparison, we consider a perturbed data set of the terrorism losses. Specifically, we assume that the perturbed data set is generated from the original data set by randomly changing ±5% of the data in Table E.1 for each city, each criterion and each scenario, since the distribution of the terrorism losses is fully described in Table E.1 and has three different levels of terrorist attacks: reduced loss, standard loss and increased loss. For example, the element $\widetilde{\xi}_{ij}(\omega_k)$ for city i, criterion j and scenario ω_k in the perturbed data set is given by

$$\widetilde{\xi}_{ij}(\omega_k) = \xi_{ij}(\omega_k) \times (1 + r \times s), \tag{3.6.7}$$

where $r = 5\%$ is a perturbation ratio and s is from the uniform distribution on $[-1, 1]$.

In this experiment, we randomly generate 100 perturbed data sets. By solving models (3.6.3), (3.6.4) and (3.6.5), we obtain the optimal budget allocations for each perturbed data set. We then calculate the average optimal budget allocation for each model under 100 simulations. To analyze the impact of the perturbations of original data set to the optimal budget allocations for the three models, we consider the relative change of the average optimal budget allocation to the original optimal budget allocation for each city and each criterion. The results are presented in Figure 3.1. We can see that the proposed PRMSR model is more sensitive to the perturbed data set compared with risk-neutral model and aggregated CVaR model, which confirms that the PRMSR model captures changes in data more effectively from shortfall risk perspective.

Experiment II: Effect of the perturbation range on optimal value

In the second experiment, we examine the impact of the data perturbation on the optimal value of the PRMSR model. Specifically, we generate perturbed data sets

Cities	Property losses ($ million)			Fatalities ($ million)			Air departures ($ million)			Bridge traffic ($ million)		
	RN	CVaR	PRMSR	RN	CVaR	PRMSR	RN	CVaR	PRMSR	RN	CVaR	PRMSR
New York	25.87	23.77	25.68	25.88	23.39	27.79	4.61	3.93	12.18	25.83	22.70	15.74
Chicago	25.66	21.87	25.09	24.33	19.06	26.85	9.59	6.45	20.48	25.62	21.26	13.79
San Francisco	24.42	20.26	23.18	9.83	12.94	16.54	3.63	3.29	10.00	25.57	20.84	13.18
Washington	14.68	18.63	17.26	10.25	14.87	18.02	3.26	3.05	8.91	25.52	20.54	12.77
Los Angeles	10.44	18.40	14.73	3.82	11.17	10.37	5.96	4.69	14.85	25.64	21.41	14.01
Philadelphia	4.42	12.72	8.48	2.64	5.74	5.63	2.61	2.59	7.05	25.18	19.17	11.22
Boston	4.53	11.99	7.89	4.02	7.33	7.89	2.28	2.35	6.00	25.86	22.89	15.97
Houston	3.41	7.43	5.04	2.98	5.35	5.66	4.01	3.56	10.83	25.61	21.17	13.65
Newark	1.00	5.97	3.42	1.00	4.08	2.98	2.48	2.49	6.62	25.79	22.44	15.41
Seattle	2.15	5.01	3.29	1.66	2.87	2.80	2.60	2.59	7.02	25.35	19.73	11.77

Table 3.1: Optimal budget allocations: Risk-Neutral(RN), CVaR and PRMSR.

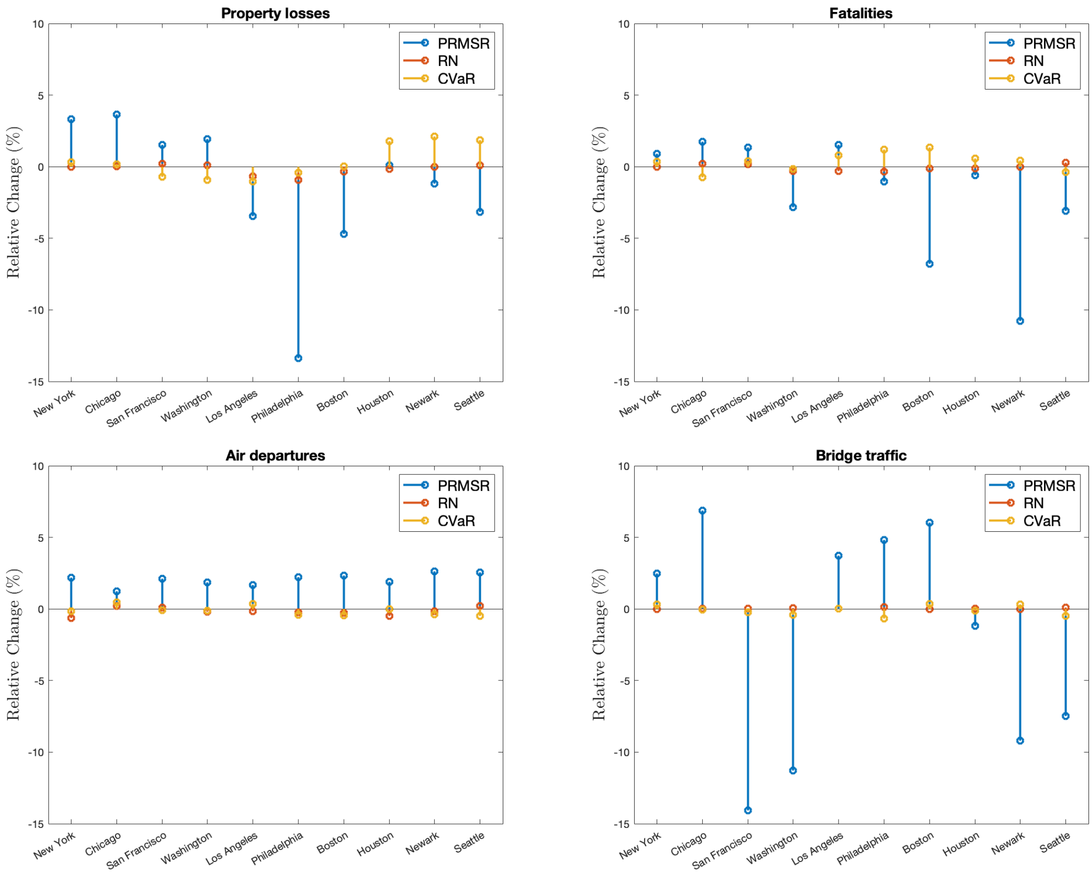
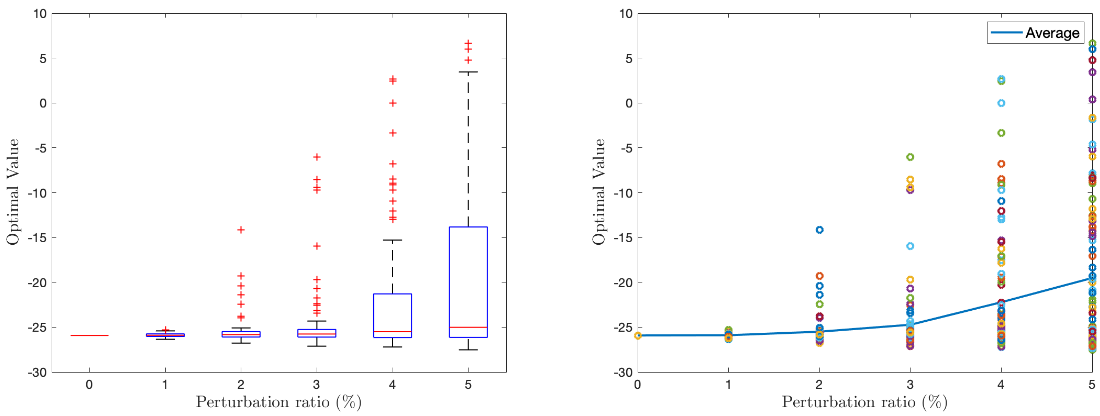


Figure 3.1: Comparative analysis of impact of data perturbation: average relative change out of 100 simulations for the optimal budget allocations to each criterion and each city from different models: PRMSR, Risk-Neutral (RN) and CVaR.

by following the same procedures as in Experiment I with the perturbation ratio $r \in \{1\%, 2\%, 3\%, 4\%, 5\%\}$. We randomly generate 100 perturbed data sets for each r and solve the model (3.6.5) to obtain the optimal value for each perturbed data set. Figure 3.2 depicts boxplot and dot plot of the optimal values as r increases. We can see that the change of the optimal value is not drastic.

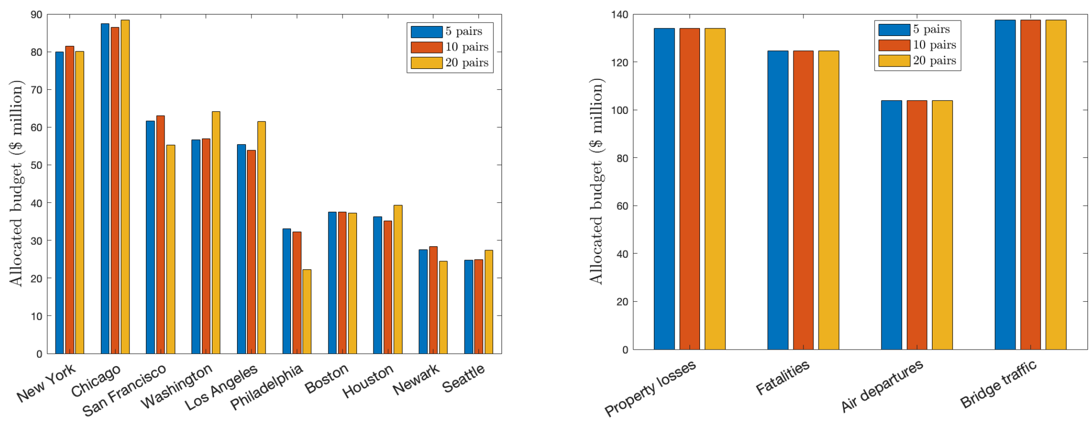


(a) **(b)**

Figure 3.2: **(a)** Boxplot of the optimal values of the PRMSR model w.r.t. the perturbation ratio (%). **(b)** Dot plot of the optimal values of the PRMSR model w.r.t. the perturbation ratio (%).

Experiment III: Effect of the number of pairs on optimal budget allocation

In the third experiment, we consider the impact of the number of pairs for eliciting the ambiguity set L on the optimal budget allocation. We select K pairs of prospects, $K \in \{5, 10, 20\}$, to form the elicited comparison data set, and then solve the problem (3.6.5) to compare their effects on budget allocations for cities and criteria. The results are shown in Figure 3.3.



(a) **(b)**

Figure 3.3: **(a)** Optimal budget allocation to cities w.r.t. the number of pairs. **(b)** Optimal budget allocation to criteria w.r.t. the number of pairs.

From Figure 3.3, we observe that an increase in the number of pairs has minor effect on the optimal budget allocations for cities and criteria. This phenomenon confirms that in practical decision making, even with less extra preference information, our model still can effectively capture the worst decision maker's risk preference.

Experiment IV: effect of d on optimal budget allocation

In the fourth experiment, we consider the effect of the choice of the weighting vector d on the optimal budget allocation. To this end, we consider three weighting vectors: $d_1 = (1/4, 1/4, 1/4, 1/4)$ equally treating each criterion, $d_2 = (1/8, 1/8, 3/8, 3/8)$ emphasizing the air departures and bridge traffic, and $d_3 = (3/8, 3/8, 1/8, 1/8)$ giving more consideration to the property losses and fatalities as Hu et al. [45] did for reflecting the importance of each criterion. The influences of d on optimal budget allocations to cities are shown in Figure 3.4(a) and the influences of d on optimal budget allocations to criteria are shown in Figure 3.4(b).

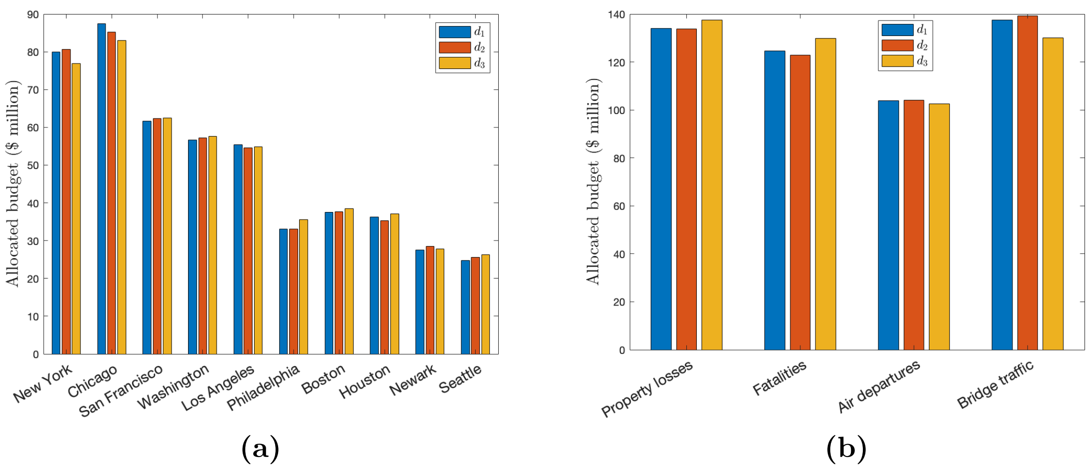


Figure 3.4: **(a)** Optimal budget allocation to cities w.r.t. weighting vector d. **(b)** Optimal budget allocation to criteria w.r.t. weighting vector d.

From Figure 3.4(a) we can see that no matter which weighting vector used, more budget is allocated to New York and Chicago whereas less budget is allocated to Philadelphia, Newark and Seattle. We observe form Figure 3.4(b) that the effect of d on optimal budget allocation to Air departures is small compared with other three criteria.

Experiment V: Runtime w.r.t the number of pairs

In the fifth experiment, we test the solution time of problem (3.6.5) based on the size of the elicited comparison data set. The experiments are performed on a generic laptop with Intel Core i5 processor, 4GM RAM, on a 64-bit Windows 7 operating system. Specifically, we solve problem (D4) by the SDPT3 4.0 solver in `CVX` on Matlab R2019b. The average computation time w.r.t. the number of pairs under 100 simulations is shown in Table 3.2 and Figure 3.5. We can observe that the solution time grows gradually as the size of the elicited comparison data set increases because the proposed tractable formulation to problem (3.6.5) is a finite dimensional linear program of reasonable size.

Number of pairs	2	4	6	8	10	12	14	16	18	20
Time (second)	3.62	6.91	8.34	10.97	11.73	13.84	16.75	17.82	19.78	20.83

Table 3.2: Computation time w.r.t. the number of pairs

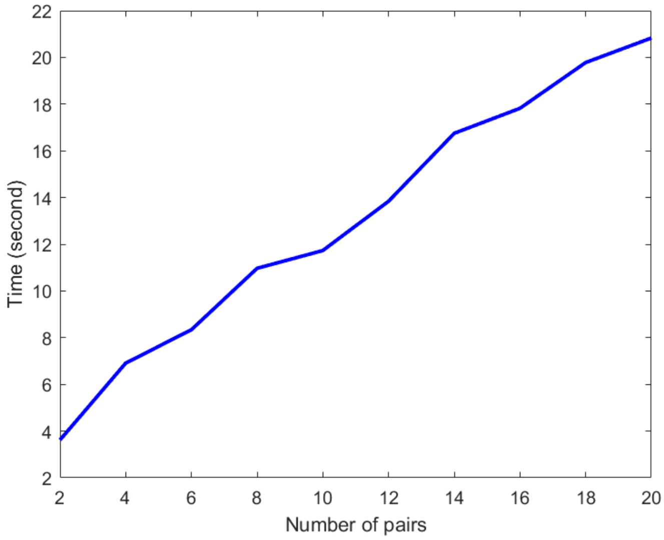


Figure 3.5: Computation time w.r.t. the number of pairs

Chapter 4

Statistical Robustness in Preference Robust Optimization Models

Preference robust optimization has recently received increasing attention in decision analysis. It deals with the ambiguity of decision maker's utility preference or risk attitude. In preference robust optimization models, e.g., the PRMSR optimization model in Chapter 3, the true probability distribution of random variable is often assumed to be either known or can be recovered via empirical data which do not contain any noise. In practice, however, empirical data may contain some noise, and it is unclear whether a statistical estimator such as the optimal value of a preference robust optimization model based on those data is still reliable. Thus in this chapter, we will investigate this issue in the context of the PRMSR optimization model. We aims to derive moderate conditions and identify appropriate metrics under which the optimal value of PRMSR optimization model obtained based on the perceived data is close to that based on real data.

4.1 Introduction

4.1.1 Literature Review

Consider the following expected utility maximization problem:

$$\max_{x \in X} \mathbb{E}_P[u(f(x, \xi))],$$

where $u : \mathbb{R} \to \mathbb{R}$ is an increasing utility function, $x \in X \subset \mathbb{R}^n$ is a decision vector, $\xi : \Omega \to \mathbb{R}^s$ is a vector of random variables defined over probability space $(\Omega, \mathcal{F}, \mathbb{P})$, $P := \mathbb{P} \circ \xi^{-1}$ is the probability measure on $\mathbb{R}^s$ induced by ξ and $f : \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}$ is a continuous function.

By the well-known expected utility theory [83], there exists a utility function $u(\cdot)$ such that the decision maker prefers prospect A to prospect B if and only if $\mathbb{E}[u(A)] \geq \mathbb{E}[u(B)]$. A convex utility function means the decision maker is risk taking whereas a concave utility function means risk averse and an affine utility function means risk neutral. Our focus here is on the situation where the true utility function is unknown, but it is possible to construct an ambiguity set of utility functions such that the true utility function lies in the ambiguity set with high probability. Thus we consider the following maximin preference robust optimization model:

$$\mathcal{V}(P) := \max_{x \in X} \inf_{u \in \mathcal{U}} \mathbb{E}_P[u(f(x, \xi))],$$

where $\mathcal{U}$ denotes the ambiguity set of utility functions.

This kind of preference robust optimization model is first considered by Armbruster and Delage [6] and it is rooted in stochastic dominance [28]. The structure of the above model is largely determined by the ambiguity set $\mathcal{U}$ of the utility functions. The work of Armbruster and Delage [6] considers an ambiguity set of utility functions which meets some criteria such as risk averse, S-shaped and prudent. Instead of trying to identify a single utility function the criteria, they develop a minimax preference robust optimization model where the optimal decision is based on the worst utility function from the ambiguity set and demonstrate how the minimax optimization problem can be reformulated as a finite dimensional linear programming problem. Delage and Li [25] extend the research to risk management problem where the objective is the Fölmer and Schied's convex risk measure [32] and the uncertainty arises from decision maker's risk attitude. Haskell et al. [42] take it further by taking into account of ambiguity about both decision maker's utiltiy and probability distribution of underlying exogenous uncertainty.

Hu and Mehortra [46] tackle the issue in a different manner. First, they propose a moment type framework for constructing the ambiguity set of decision maker's preference which cover a wide range of approaches such as pairwise comparison, certainty equivalent and stochastic dominance. Second, they consider a probabilistic representation of the class of increasing convex utility functions by confining them to a compact interval and scaling them to being bounded by 1. Third, they consider lower and upper bound of the true unknown utility function, and propose a piecewise linear approximation of the functions in the bounds for deriving tractable reformulation of the preference robust optimization model.

In all these preference robust optimization models, the true probability distribution is assumed to be either known or can be recovered via empirical data which do not contain any noise. However, it is unclear whether a statistical estimator such as the optimal value of a preference robust optimization model based on the data with noise is still reliable. Specifically, let Q_N denote the the empirical distribution based on the data with noise and P_N the data with the noise detached, we ask whether $\mathcal{V}(Q_N)$ is close to $\mathcal{V}(P_N)$ under some metric when N is sufficiently large. This issue is also known as statistical robustness in the literature of statistics [48]. The concept of statistical robustness can be traced back to the work of Hampel [41], and it has been popularized over the years particularly with monographs [48, 49].

A research by Cont et al. [22] defines the notion of qualitative robustness [1] of a risk estimator and uses it to examine the robustness of various risk estimators derived from empirical data. They demonstrate that the historical estimator of any spectral risk measure is not robust. The work of Krätschmer et al. [53] proposes and analyses a refined notion of qualitative robustness that applies to tail-dependent law-invariant convex risk measures on Orlicz spaces.

4.1.2 Contribution

Our research in this chapter concerns the issue that the optimal value obtained from solving the PRMSR optimization model with perceived data may perform differently from it's true counterpart. We fill up this significant gap by providing a comprehensive study of the statistical robustness of the PRMSR optimization model. Specifically, we derive moderate sufficient conditions under which the optimal value changes continuously against small variation of the probability distribution which paves the way for the analysis of statistical robustness. Compared with the standard stability analysis in stochastic programming [71], our result shows interactions between the tail behaviours of probability distribution and loss function through some specified topology under which the continuity is established.

[1] Throughout this chapter, we use terminology statistical robustness to avoid confusion with other notions of robustness.

Moreover, we identify appropriate metrics under which the statistical estimator of the optimal value is uniformly asymptotically consistent. This result is associated with convergence analysis of sample average approximation [74], while the uniformity requirement makes the analysis much more challenging. This result gives a theoretical basis for discrete approximation in the PRMSR optimization model where true probability distribution is continuous, and it is important in developing tractable numerical schemes for the PRMSR optimization model.

We demonstrate statistical robustness of the optimal value of the PRMSR optimization model obtained based on the data generated by some distributions close to the true distribution. Although the analysis follows from the framework of qualitative robustness in [52, 53], the PRMSR optimization model includes minimax operations w.r.t. z, t and l which needs more complex mathematical treatment.

4.1.3 Structure

The remainder of this chapter is structured as follows: In section 4.2, we recall the PRMSR optimization model and make a preliminary analysis of the statistical robustness. In section 4.3, we recall some basic concepts and results about ψ-weak topology and show continuity of $\vartheta(\cdot)$ (the optimal value of PRMSR optimization model) near P. In section 4.4, we introduce the Uniform Glivenko-Cantelli property and derive uniform consistency of $\vartheta(P_N)$ to $\vartheta(P)$. In section 4.5, we establish the statistical robustness of the estimator of $\vartheta(\cdot)$.

4.2 Preliminary Description

Consider a financial position $c(z, \xi(\omega))$ associated with decision variable $z \in Z \subset \mathbb{R}^n$ and random variable $\xi(\omega)$. Let L be a set of increasing and not constant loss functions $l : \mathbb{R}^n \to \mathbb{R}$. We first recall the preference robust multivariate utility-based shortfall risk measure (PRMSR) optimization model defined as in Chapter 3:

$$\text{(PRMSR-Opti)} \quad \begin{aligned} \min_{z \in Z, t \in \mathbb{R}} \quad & t \\ \text{s.t.} \quad & \sup_{l \in L} \mathbb{E}_P[l(-c(z, \xi) - td) - l(-Y(\xi))] \leq 0, \end{aligned} \tag{4.2.1}$$

where $\xi : (\Omega, \mathcal{F}, \mathbb{P}) \to \Xi \subset \mathbb{R}^s$, $P := \mathbb{P} \circ \xi^{-1}$ is the probability measure on $\mathbb{R}^s$ induced by ξ and $c(z, \xi) : Z \times \Xi \to \mathbb{R}^n$ is a continuous function, Y is a benchmark, $d \in \mathcal{D}$ is a preset weighting vector and

$$\mathcal{D} = \left\{ d \in \mathbb{R}^n : \sum_{i=1}^{n} d_i = 1, d_i > 0 \right\}.$$

In the PRMSR optimization model, we assume the true probability distribution P can be recovered from empirical data. In practice, however, the perceived empirical data may contain some noise, and it is differ from the real data generated by P. The discrepancy may have an impact on the quality of the optimal value obtained from solving the PRMSR optimization model with perceived empirical data. Thus it may be interesting to examine insensitivity of the optimal value of the PRMSR optimization model to the deviation of empirical distributions from the true.

For simplicity of exposition, we assume that the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is atomless. The assumption is common in the literature of statistical robustness, see [53]. The assumption excludes all discrete distributions and we will focus on the case where P follows a continuous distribution in the following discussion. Throughout this chapter, we use $\mathscr{P}(\mathbb{R}^s)$ to denote the set of all probability measures on $\mathbb{R}^s$.

Let $P, Q \in \mathscr{P}(\mathbb{R}^s)$. Let $\xi^1, \ldots, \xi^N$ and $\tilde{\xi}^1, \ldots, \tilde{\xi}^N$ be iid samples generated by P and Q respectively. We write empirical distribution

$$P_N(\cdot) := \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\xi^i}(\cdot) \tag{4.2.2}$$

if the samples are generated by P and

$$Q_N(\cdot) := \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\tilde{\xi}^i}(\cdot) \tag{4.2.3}$$

if the samples are generated by Q, where $\mathbb{1}_{\tilde{\xi}}(\cdot)$ denotes the Dirac measure at $\tilde{\xi}$. We refer to P as the true probability distribution and Q its perturbation. In practice, samples are often obtained from empirical data which contain some noise. This means the samples are generated by Q rather than P. In this setup, samples of P are not obtainable and $P_N(\cdot)$ is defined only for theoretical analysis.

Let $\vartheta(\cdot)$ denote the optimal value of (PRMSR-Opti). It is obvious that $\vartheta(Q_N)$ is a statistical estimator of $\vartheta(Q)$ rather than the true robust optimal value $\vartheta(P)$, so we are interested in how close $\vartheta(Q_N)$ is to $\vartheta(P_N)$ (a statistical estimator of $\vartheta(P)$) under some metric. If $\vartheta(Q_N)$ is close to $\vartheta(P_N)$, then it is safe to use $\vartheta(Q_N)$ as an estimate of $\vartheta(P)$. Note that there is a distinction between traditional stability analysis in stochastic programming and statistical robustness. The former concentrate on the convergence of $\vartheta(P_N)$ to $\vartheta(P)$ or continuity of $\vartheta(\cdot)$ near P. The latter requires not only the continuity of $\vartheta(\cdot)$ near P but also convergence of $\vartheta(Q_N)$ to $\vartheta(Q)$ uniformly for all Q near P.

4.3 Continuity of $\vartheta(P)$

A key step to establish statistical robustness of the estimator of $\vartheta(\cdot)$ is to show continuity of $\vartheta(P')$ when P' is near the true probability distribution P under some metric.

4.3.1 ψ-Weak Topology

Let us recall some basic concepts of ψ-weak topology which are needed in the forthcoming discussions. The materials are mainly extracted from [21], and references therein for a more comprehensive discussion on the subject.

Definition 4.3.1 *Let $\psi : \mathbb{R}^s \to [0, \infty)$ be a continuous function and*

$$\mathcal{M}^\psi := \left\{ P' \in \mathscr{P}(\mathbb{R}^s) : \int_{\mathbb{R}^s} \psi(\xi) P'(d\xi) < \infty \right\}.$$

Note that $\mathcal{M}^\psi$ defines a subset of probability measures in $\mathscr{P}(\mathbb{R}^s)$ which satisfies the generalized moment condition of ψ.

Definition 4.3.2 (ψ-weak topology) *Let $\psi : \mathbb{R}^s \to [0, \infty)$ be a gauge function, that is, ψ is continuous and there exists a compact set such that $\psi \geq 1$ holds outside the compact set. Let H^ψ be the linear space of all continuous functions $h : \mathbb{R}^s \to \mathbb{R}$ for which there exists a positive constant c such that*

$$|h(\xi)| \leq c(\psi(\xi) + 1), \quad \forall \xi \in \mathbb{R}^s.$$

The ψ-weak topology, denoted by τ_ψ, is the coarsest topology on $\mathcal{M}^\psi$ for which the mapping $\mathcal{H} : \mathcal{M}^\psi \to \mathbb{R}$ defined by

$$\mathcal{H}(P') := \int_{\mathbb{R}^s} h(\xi) P'(d\xi), h \in H^\psi$$

is continuous. A sequence $\{P_k\} \subset \mathcal{M}^\psi$ is said to convergence ψ-weakly to $P \in \mathcal{M}^\psi$ written $P_k \xrightarrow{\psi} P$ if it converges w.r.t. τ_ψ.

By Krätschmer et al. [52], ψ-weak convergence implies weak convergence, denoted by $P_k \xrightarrow{w} P$, and

$$\int_{\mathbb{R}^s} \psi(\xi) P_k(d\xi) \to \int_{\mathbb{R}^s} \psi(\xi) P(d\xi)$$

as $k \to \infty$.

Definition 4.3.3 (Uniform integrating set) *Let $\psi : \mathbb{R}^s \to [0, \infty)$ be a gauge function. A set $\mathcal{M} \subset \mathcal{M}^\psi$ is said to be locally uniformly ψ-integrating if for any $P' \in \mathcal{M}$ there exists some open neighborhood $\mathcal{N}$ of P' w.r.t. the topology of weak convergence such that*

$$\lim_{\gamma\to\infty} \sup_{P''\in\mathcal{N}\cap\mathcal{M}} \int_{\{\xi\in\mathbb{R}^s:\psi(\xi)\geq\gamma\}} \psi(\xi)P''(d\xi) = 0.$$

It is well known that the relative ψ-weak topology on $\mathcal{M} \subset \mathcal{M}^\psi$ coincides with the relative weak topology of weak convergence on $\mathcal{M}$ if and only if $\mathcal{M}$ is locally uniformly ψ-integrating [87, Lemma 3.4]. In the forthcoming discussions, we will use this equivalence.

4.3.2 Continuity of $\vartheta(P)$

Let us start by deriving a gauge function which majorizes $l(-c(z,\xi)-td)$ and $l(-Y(\xi))$. To ease the exposition, let

$$v_P(z,t) := \sup_{l\in L} \mathbb{E}_P[l(-c(z,\xi) - td) - l(-Y(\xi))], \tag{4.3.1}$$

and rewrite (4.2.1) as

$$\text{(PRMSR-Opti)} \quad \begin{array}{ll} \min\limits_{z\in Z,t\in\mathbb{R}} & t \\ \text{s.t.} & v_P(z,t) \leq 0 \end{array} \tag{4.3.2}$$

Let $\mathcal{F}_P$ denote the feasible set of problem (PRMSR-Opti). Throughout this section, we let Assumption 3.4.1 hold and making the following assumptions.

Assumption 4.3.1 (Growth condition) *Let $c(\cdot,\cdot)$ and $Y(\cdot)$ be defined as in the problem (PRMSR-Opti). Let $t \in T \subset \mathbb{R}$ and T be a compact set. There is an exponent $\beta > 0$ and some locally bounded function $g : \mathbb{R}^n \to \mathbb{R}$ such that*

$$\|-c(z,\xi) - td\| \leq g(z)(\|\xi\|^\beta + 1), \quad \forall (z,\xi) \in Z \times \Xi, t \in T, d \in \mathcal{D} \tag{4.3.3}$$

and

$$\|-Y(\xi)\| \leq g(z)(\|\xi\|^\beta + 1), \quad \forall \xi \in \Xi. \tag{4.3.4}$$

Note that $g(\cdot)$ is said to be a locally bounded function if the convergence of $\{z^N\}$ implies the boundedness of $\{g(z^N)\}$.

Let $\mathcal{C} := \sup_{z \in Z} g(z)$. Since Z is assumed to be a compact set, then $\mathcal{C} < \infty$. We assume without loss of generality that $\mathcal{C} \geq 1$. Let

$$\psi(\xi) := \max\left(\sup_{l \in L} l(\mathcal{C}(\|\xi\|^\beta + 1)), \sup_{l \in L} -l(-\mathcal{C}(\|\xi\|^\beta + 1))\right). \tag{4.3.5}$$

Then the growth condition and monotonic increasing property of $l(\cdot)$ imply for all $(z, \xi) \in Z \times \Xi, t \in T, d \in \mathcal{D}$ and $l \in L$,

$$\begin{array}{rclclll} l(-c(z,\xi) - td) & \leq & l(\mathcal{C}(\|\xi\|^\beta + 1)) & \leq & \psi(\xi) & \text{for } c(z,\xi) + td \leq 0, \\ -l(-c(z,\xi) - td) & \leq & -l(-\mathcal{C}(\|\xi\|^\beta + 1)) & \leq & \psi(\xi) & \text{for } c(z,\xi) + td \geq 0, \end{array}$$

which means

$$|l(-c(z,\xi) - td)| \leq \psi(\xi), \quad \forall (z,\xi) \in Z \times \Xi, t \in T, d \in \mathcal{D}, l \in L. \tag{4.3.6}$$

Likewise, we have

$$|l(-Y(\xi))| \leq \psi(\xi), \quad \forall \xi \in \Xi, l \in L. \tag{4.3.7}$$

Assumption 4.3.2 *Let $\psi(\cdot)$ be defined as in (4.3.5) and $P \in \mathcal{M}^\psi$. Let $\{P_k\} \subset \mathcal{M}^\psi$ be a sequence of probability measures such that $P_k \xrightarrow{\psi} P$. There exists a sequence of monotonically increasing numbers $\{r_N\}$ where $r_N \to \infty$ such that for any fixed $r \in \{r_N\}$, $P(\|\xi\| = r) = 0$,*

$$\lim_{k \to \infty} \sup_{z,t,d} \sup_{l \in L} \left| \int_{\{\xi \in \mathbb{R}^s : \|\xi\| \leq r\}} l(-c(z,\xi) - td) P_k(d\xi) - \int_{\{\xi \in \mathbb{R}^s : \|\xi\| \leq r\}} l(-c(z,\xi) - td) P(d\xi) \right| = 0 \tag{4.3.8}$$

and

$$\lim_{k \to \infty} \sup_{l \in L} \left| \int_{\{\xi \in \mathbb{R}^s : \|\xi\| \leq r\}} l(-Y(\xi)) P_k(d\xi) - \int_{\{\xi \in \mathbb{R}^s : \|\xi\| \leq r\}} l(-Y(\xi)) P(d\xi) \right| = 0 \tag{4.3.9}$$

Based on [66, Theorem 3.2] we can establish equalities (4.3.8) and (4.3.9) under conditions: (a) $l(\cdot)$ is uniformly Lipschitz continuous with a bounded modulus for all $l \in L$ and (b) Assumption 4.3.1 holds. Condition $P(\|\xi\| = r) = 0$ is fulfilled when P is non-atomic. Note that the setting of (4.3.8) and (4.3.9) does not fit to the framework of [66, Theorem 3.2] exactly as the range of the integral in our case is bounded whereas the range of the integral in [66, Theorem 3.2] is over the whole space. However, under condition (a) we are able to show the class of loss functions l are equicontinuous and together with (b), we can show that (4.3.8) and (4.3.9) hold by a similar proof to [66, Theorem 3.2] based on the equicontinuity condition and condition (b).

We let $P_k \xrightarrow{\psi} P$ and consider the problem

$$\begin{aligned} \min_{z\in Z, t\in\mathbb{R}} \quad & t \\ \text{s.t.} \quad & \sup_{l\in L} \mathbb{E}_{P_k}[l(-c(z,\xi) - td) - l(-Y(\xi))] \le 0. \end{aligned} \tag{4.3.10}$$

To ease the exposition, let

$$v_{P_k}(z,t) := \sup_{l\in L} \mathbb{E}_{P_k}[l(-c(z,\xi) - td) - l(-Y(\xi))] \tag{4.3.11}$$

and rewrite (4.3.10) as

$$(\text{Opti-P}_\text{k}) \quad \begin{aligned} \min_{z\in Z, t\in\mathbb{R}} \quad & t \\ \text{s.t.} \quad & v_{P_k}(z,t) \le 0. \end{aligned} \tag{4.3.12}$$

We use $\mathcal{F}_{P_k}$ and $\vartheta(P_k)$ to denote the feasible set and the optimal value of problem (Opti-P_k) respectively.

Theorem 4.3.1 (Continuity) *Let $v_P(\cdot,\cdot)$ and $\psi(\cdot)$ be defined as in (4.3.1) and (4.3.5). Suppose that $c(\cdot,\xi)$ is a concave function. Then under Assumptions 3.4.1, 4.3.1 and 4.3.2,*

(i)

$$\lim_{P' \xrightarrow{\psi} P} v_{P'}(z,t) = v_P(z,t) \tag{4.3.13}$$

(ii)

$$\lim_{P' \xrightarrow{\psi} P} \vartheta(P') = \vartheta(P). \tag{4.3.14}$$

Proof. It follows by [49, Theorem 2.15] that $\mathcal{M}^\psi$ is a Polish space. Thus it suffices to show (4.3.13) and (4.3.14) for any sequence $\{P_k\} \subset \mathcal{M}^\psi$ converging to $P \in \mathcal{M}^\psi$ under ψ-weak topology.

Part (i). Since $P_k \xrightarrow{\psi} P$, then $P_k \xrightarrow{w} P$ and

$$\lim_{k\to\infty} \int_{\mathbb{R}^s} \psi(\xi) P_k(d\xi) = \int_{\mathbb{R}^s} \psi(\xi) P(d\xi).$$

Under the growth condition, we have via (4.3.6)

$$\begin{aligned}&\lim_{r\to\infty}\sup_{z,t,d}\sup_{l\in L}\int_{\{\xi\in\mathbb{R}^s:\|\xi\|>r\}}|l(-c(z,\xi)-td)|P_k(d\xi)\\ \leq\ &\lim_{r\to\infty}\int_{\{\xi\in\mathbb{R}^s:\|\xi\|>r\}}\psi(\xi)P_k(d\xi)=0.\end{aligned}\tag{4.3.15}$$

The last equality is due to the fact that $\{P_k\}\subset\mathcal{M}^\psi$ and [52, Lemma 3.4], that is, ψ-weak convergence implies weak convergence and uniform integrating property. Note that since $\psi(\cdot)$ is defined as in (4.3.5), we have $\psi(\xi)\to\infty$ as $\|\xi\|\to\infty$. Similarly, since $P\in\mathcal{M}^\psi$, we have

$$\begin{aligned}&\lim_{r\to\infty}\sup_{z,t,d}\sup_{l\in L}\int_{\{\xi\in\mathbb{R}^s:\|\xi\|>r\}}|l(-c(z,\xi)-td)|P(d\xi)\\ \leq\ &\lim_{r\to\infty}\int_{\{\xi\in\mathbb{R}^s:\|\xi\|>r\}}\psi(\xi)P(d\xi)=0.\end{aligned}\tag{4.3.16}$$

On the other hand, under Assumption 4.3.2, we can choose those r such that $P(\|\xi\|=r)=0$ and

$$\begin{aligned}\lim_{k\to\infty}\sup_{z,t,d}\sup_{l\in L}\ &\left|\int_{\{\xi\in\mathbb{R}^s:\|\xi\|\leq r\}}l(-c(z,\xi)-td)P_k(d\xi)\right.\\ &\left.-\int_{\{\xi\in\mathbb{R}^s:\|\xi\|\leq r\}}l(-c(z,\xi)-td)P(d\xi)\right|=0.\end{aligned}\tag{4.3.17}$$

Combing (4.3.15), (4.3.16) and (4.3.17), we obtain

$$\begin{aligned}\lim_{k\to\infty}\sup_{z,t,d}\sup_{l\in L}\ &\left|\int_{\mathbb{R}^s}l(-c(z,\xi)-td)P_k(d\xi)\right.\\ &\left.-\int_{\mathbb{R}^s}l(-c(z,\xi)-td)P(d\xi)\right|=0.\end{aligned}\tag{4.3.18}$$

Likewise, we have

$$\begin{aligned}\lim_{k\to\infty}\sup_{l\in L}\ &\left|\int_{\mathbb{R}^s}l(-Y(\xi))P_k(d\xi)\right.\\ &\left.-\int_{\mathbb{R}^s}l(-Y(\xi))P(d\xi)\right|=0.\end{aligned}\tag{4.3.19}$$

Combing (4.3.18) and (4.3.19), we obtain

$$\begin{aligned}\lim_{k\to\infty}\sup_{z,t,d}\ &\left|\sup_{l\in L}\mathbb{E}_{P_k}[l(-c(z,\xi)-td)-l(-Y(\xi))]\right.\\ &\left.-\sup_{l\in L}\mathbb{E}_{P}[l(-c(z,\xi)-td)-l(-Y(\xi))]\right|=0,\end{aligned}\tag{4.3.20}$$

which implies (4.3.13).

Part (ii). The proof is similar to that of Theorem 3.4.2 Part (i), we include a proof in Appendix G for self-containedness. □

Note that the stability result (4.3.14) requires perturbation of P' from P under topology of ψ-weak convergence. Since ψ captures the growth of $c(z,\xi)$ and l, it means that if ψ is steep w.r.t. growth of ξ either due to the significance of loss or due to the sharp increase of the loss function l at the tail would make the optimal value function $\vartheta(\cdot)$ less stable.

4.4 Uniform Consistency

In this section, we show the uniform consistency of $v_{P_N}(z,t)$ to $v_P(z,t)$, that is, the convergence of $v_{P_N}(z,t)$ to $v_P(z,t)$ for all P in a subset of $\mathscr{P}(\mathbb{R}^s)$, and its uniformity w.r.t. z, the latter leads to uniform consistency of $\vartheta(P_N)$ to $\vartheta(P)$. Here

$$v_{P_N}(z,t) := \sup_{l\in L} \mathbb{E}_{P_N}[l(-c(z,\xi)-td) - l(-Y(\xi))] \tag{4.4.1}$$

and the approximate problem of (PRMSR-Opti) takes the form

$$\text{(Opti-P}_\text{N}\text{)} \quad \begin{array}{ll} \min\limits_{z\in Z, t\in\mathbb{R}} & t \\ \text{s.t.} & v_{P_N}(z,t) \le 0. \end{array} \tag{4.4.2}$$

The convergence result is important because in the literature of preference robust optimization models, tractable formulations rely on P_N (discrete approximation of P) when P is continuously distributed.

It follows by [52, 53] that the ψ-weak topology on $\mathcal{M}^\psi$ is generated by the metric $\mathbf{d}_\psi : \mathcal{M}^\psi \times \mathcal{M}^\psi \to \mathbb{R}$ defined by

$$\mathbf{d}_\psi(P',P'') := \mathbf{d}_{\text{Prok}}(P',P'') + \left| \int_{\mathbb{R}^s} \psi(\xi)P'(d\xi) - \int_{\mathbb{R}^s} \psi(\xi)P''(d\xi) \right|, \tag{4.4.3}$$

for $P', P'' \in \mathcal{M}^\psi$, where $\mathbf{d}_{\text{Prok}} : \mathscr{P}(\mathbb{R}^s) \times \mathscr{P}(\mathbb{R}^s) \to \mathbb{R}_+$ is the Prokhorov metric

$$\mathbf{d}_{\text{Prok}}(P',P'') := \inf\{\epsilon > 0 : P'(A) \le P''(A^\epsilon) + \epsilon, \forall A \in \mathcal{B}(\mathbb{R}^s)\}, \tag{4.4.4}$$

where $A^\epsilon := A + B_\epsilon(0)$ denotes the Minkowski sum of A and the open ball centred at 0 (w.r.t. Euclidean norm) and $\mathcal{B}(\mathbb{R}^s)$ denotes the Borel σ-algebra on $\mathbb{R}^s$.

Let $(\mathbb{R}^s)^{\otimes N}$ denote the Cartesian product $\mathbb{R}^s \times \ldots \times \mathbb{R}^s$ and $\mathcal{B}(\mathbb{R}^s)^{\otimes N}$ its Borel σ-algebra. Let $P^{\otimes N}$ denote the probability measure on the measurable space $((\mathbb{R}^s)^{\otimes N}, \mathcal{B}(\mathbb{R}^s)^{\otimes N})$ with marginal P on each $(\mathbb{R}^s, \mathcal{B}(\mathbb{R}^s))$ and $Q^{\otimes N}$ with marginal Q. We recall the definition of Uniform Glivenko-Cantelli (UGC) property based on [52] for forthcoming discussion.

Definition 4.4.1 (UGC property) *Let ψ be a gauge function and $\mathbf{d}_\psi$ be defined as in (4.4.3). Let $\mathcal{M}$ be a subset of $\mathcal{M}^\psi$. The metric space $(\mathcal{M}, \mathbf{d}_\psi)$ has UGC property if for any $\epsilon > 0$ and $\delta > 0$, there exists $N_0 \in \mathbb{N}$ such that for all $P \in \mathcal{M}$ and $N \geq N_0$*

$$P^{\otimes N}[(\xi^1, \ldots, \xi^N) : \mathbf{d}_\psi(P, P_N) \geq \delta] \leq \epsilon. \tag{4.4.5}$$

The UGC property states that for every empirical probability measure P_N generated by $P \in \mathcal{M}$, P_N is close to P under the metric $\mathbf{d}_\psi$ when N is sufficiently large.

Theorem 4.4.1 (Uniform consistency) *Let ψ be defined as in (4.3.5) and*

$$\mathcal{M}_\kappa^{\psi^p} := \left\{ P \in \mathscr{P}(\mathbb{R}^s) : \int_{\mathbb{R}^s} \psi(\xi)^p P(d\xi) \leq \kappa \right\} \tag{4.4.6}$$

for some fixed $p > 1$ and $\kappa > 0$. Let $\mathcal{M} \subset \mathcal{M}_\kappa^{\psi^p}$ be a compact set. If Assumptions 4.3.1 and 4.3.2 hold, then for any $\epsilon > 0$ and $\delta > 0$, there exists $N_0 \in \mathbb{N}$ such that for all $P \in \mathcal{M}$ and $N \geq N_0$

(i)

$$P^{\otimes N}\left[\left\{ \xi \in (\mathbb{R}^s)^N : \sup_{z,t,d} |v_{P_N}(z,t) - v_P(z,t)| \geq \delta \right\}\right] \leq \epsilon \tag{4.4.7}$$

(ii)

$$P^{\otimes N}\left[\left\{ \xi \in (\mathbb{R}^s)^N : |\vartheta(P_N) - \vartheta(P)| \geq \delta \right\}\right] \leq \epsilon. \tag{4.4.8}$$

Proof. Part (i). By [52, Corollary 3.5], $(\mathcal{M}_\kappa^{\psi^p}, \mathbf{d}_{\psi^p})$ has the UGC property which means that for any $\epsilon > 0$ and $\delta > 0$, there exists $N_0 \in \mathbb{N}$ such that

$$P^{\otimes N}[\{\xi \in (\mathbb{R}^s)^N : \mathbf{d}_{\psi^p}(P, P_N) \geq \delta\}] \leq \epsilon, \forall P \in \mathcal{M}_\kappa^{\psi^p} \tag{4.4.9}$$

for all $N \geq N_0$. For each fixed P as such, we may set r large enough so that $P(\|\xi\| = r) = 0$ and

$$\sup_{z,t,d} \sup_{l \in L} \left| \int_{\{\xi \in \mathbb{R}^s : \|\xi\| > r\}} l(-c(z,\xi) - td) P(d\xi) \right| \leq \int_{\{\xi \in \mathbb{R}^s : \|\xi\| > r\}} \psi(\xi) P(d\xi) \leq \frac{\delta}{8}. \tag{4.4.10}$$

On the other hand, by (4.4.9) and [21, Lemma 2.61],

$$\begin{aligned} & P^{\otimes N}\left[\sup_{z,t,d}\sup_{l\in L}\left|\int_{\{\xi\in\mathbb{R}^s:\|\xi\|>r\}} l(-c(z,\xi)-td)P_N(d\xi)\right| \geq \frac{\delta}{8}\right] \\ \leq\; & P^{\otimes N}\left[\int_{\{\xi\in\mathbb{R}^s:\|\xi\|>r\}} \psi(\xi)P_N(d\xi) \geq \frac{\delta}{8}\right] \leq \frac{\epsilon}{4} \end{aligned} \tag{4.4.11}$$

for all $N \geq N_0$ and N_0 sufficiently large. Moreover, by Assumption 4.3.2 and (4.4.9),

$$\begin{aligned} P^{\otimes N}\Bigg[\sup_{z,t,d}\sup_{l\in L}\Bigg|&\int_{\{\xi\in\mathbb{R}^s:\|\xi\|\leq r\}} l(-c(z,\xi)-td)P_N(d\xi) \\ &-\int_{\{\xi\in\mathbb{R}^s:\|\xi\|\leq r\}} l(-c(z,\xi)-td)P(d\xi)\Bigg| \geq \frac{\delta}{4}\Bigg] \leq \frac{\epsilon}{4} \end{aligned} \tag{4.4.12}$$

for all $N \geq N_0$ and N_0 sufficiently large. Combining (4.4.10), (4.4.11) and (4.4.12), we obtain

$$\begin{aligned} P^{\otimes N}\Bigg[\sup_{z,t,d}\sup_{l\in L}\Bigg|&\int_{\mathbb{R}^s} l(-c(z,\xi)-td)P_N(d\xi) \\ &-\int_{\mathbb{R}^s} l(-c(z,\xi)-td)P(d\xi)\Bigg| \geq \frac{\delta}{2}\Bigg] \leq \frac{\epsilon}{2} \end{aligned} \tag{4.4.13}$$

for fixed $P \in \mathcal{M}$.

Next we show the uniformity of (4.4.13) w.r.t. P. Assume for the sake of a contradiction that there exist some positive numbers ϵ_0 and δ_0 such that for any $v \in \mathbb{N}$, there exist $v' > v$, $P_{v'} \in \mathcal{M}$ and some $N_{v'} > v$ such that

$$\begin{aligned} P_{v'}^{\otimes N_{v'}}\Bigg[\sup_{z,t,d}\sup_{l\in L}\Bigg|&\int_{\mathbb{R}^s} l(-c(z,\xi)-td)P_{N_{v'}}(d\xi) \\ &-\int_{\mathbb{R}^s} l(-c(z,\xi)-td)P_{v'}(d\xi)\Bigg| \geq \delta_0\Bigg] > \epsilon_0. \end{aligned} \tag{4.4.14}$$

Let v increase. Then we obtain a sequence of $\{P_{v'}\}$ which satisfies (4.4.14). Since $\mathcal{M}$ is a compact set under the ψ-weak topology, then $\{P_{v'}\}$ has a converging subsequence. Assume without loss of generality that $P_{v'} \xrightarrow{\psi} P_* \in \mathcal{M}$. It follows by [40, Lemma 2.1] that the convergence implies

$$\int_{\mathbb{R}^s} l(-c(z,\xi)-td)P_v(d\xi) \to \int_{\mathbb{R}^s} l(-c(z,\xi)-td)P_*(d\xi).$$

On the other hand, since $\{P_{N_v}\}$ converges ψ-weakly to P_* as v increase, we can obtain

$$\begin{aligned} P_*^{\otimes N_v}\Bigg[\sup_{z,t,d}\sup_{l\in L}\Bigg|&\int_{\mathbb{R}^s} l(-c(z,\xi)-td)P_{N_v}(d\xi) \\ &-\int_{\mathbb{R}^s} l(-c(z,\xi)-td)P_*(d\xi)\Bigg| \geq \frac{\delta_0}{2}\Bigg] < \frac{\epsilon_0}{2}, \end{aligned} \tag{4.4.15}$$

which leads to a contradiction as desired. Thus the inequality (4.4.13) holds for all $P \in \mathcal{M}$.

Likewise, we have

$$P^{\otimes N}\left[\sup_{l\in L}\left|\int_{\mathbb{R}^s} l(-Y(\xi))P_N(d\xi) - \int_{\mathbb{R}^s} l(-Y(\xi))P(d\xi)\right| \geq \delta\right] \leq \epsilon, \forall P \in \mathcal{M} \quad (4.4.16)$$

for all $N \geq N_0$ and N_0 sufficiently large. Hence the inequality (4.4.7) holds.

Part (ii). It is analogous to the proof of Theorem 4.3.1. □

We make a few comments on the conditions and results of Theorem 4.4.1. First, the set of probability measures $\mathcal{M}_\kappa^{\psi_p}$ is determined by the nature of function ψ which is in turn dependent on the growth of $c(z,\xi)$ and the loss function l. In fact only the tail properties of l and c affect the set. Roughly speaking, the heavier the tails, the smaller the set $\mathcal{M}_\kappa^{\psi_p}$ will be. Second, in comparison with the convergence results in Theorem 3.4.2, the convergence result (4.4.8) focuses on uniform convergence of $\vartheta(P_N)$ to $\vartheta(P)$ for all P in $\mathcal{M}$ rather than the rate of convergence for a fixed P. Third, when Ξ is a compact set, $\mathcal{M}_\kappa^{\psi_p} = \mathscr{P}(\Xi)$ for κ sufficiently large.

4.5 Statistical Robustness

In this section, we use the UGC property to derive statistical robustness of the estimator of $\vartheta(\cdot)$. Let us start by recalling the definition of statistical robustness based on Krätschmer et al. [53].

Definition 4.5.1 (Statistical robustness) *Let $\mathfrak{M} \subset \mathscr{P}(\mathbb{R}^s)$ be a set of probability measures and $\mathbf{d}_\psi$ be defined as in (4.4.3) for some gauge function $\psi : \mathbb{R}^s \to \mathbb{R}$. A statistical estimator $\mathcal{V}(\cdot) : \mathfrak{M} \to \mathbb{R}$ is said to be robust on $\mathfrak{M}$ with respect to $\mathbf{d}_\psi$ and $\mathbf{d}_{\text{Prok}}$ if for all $P \in \mathfrak{M}$ and $\epsilon > 0$, there exist $\delta > 0$ and $N_0 \in \mathbb{N}$ such that*

$$Q \in \mathfrak{M}, \mathbf{d}_\psi(P,Q) \leq \delta \Longrightarrow \mathbf{d}_{\text{Prok}}\left(P^{\otimes N} \circ \mathcal{V}(P_N)^{-1}, Q^{\otimes N} \circ \mathcal{V}(Q_N)^{-1}\right) \leq \epsilon, \forall N \geq N_0,$$

where $\mathcal{V}(P_N)$ and $\mathcal{V}(Q_N)$ map from $(\mathbb{R}^s)^{\otimes N}$ to $\mathbb{R}$ and provide an estimator for $\mathcal{V}(P)$ and $\mathcal{V}(Q)$ respectively.

The definition is based on Hampel's classical concept of qualitative robustness [41] of an estimator requires that a small change in the law of the data results in only small changes in the law of the estimator.

Theorem 4.5.1 (Statistical robustness) *Let ψ and $\mathcal{M}_\kappa^{\psi^p}$ be defined as in (4.3.5) and (4.4.6). Let $\mathcal{M} \subset \mathcal{M}_\kappa^{\psi^p}$ be a compact set with $p > 1$. Assume in addition that Assumptions 4.3.1 and 4.3.2 hold. Then for any $P \in \mathcal{M}$ and $\epsilon > 0$, there exist positive numbers δ and $N_0 \in \mathbb{N}$ such that*

$$Q \in \mathcal{M}, \mathbf{d}_{\psi^p}(P,Q) \le \delta \Longrightarrow \mathbf{d}_{\text{Prok}}\left(P^{\otimes N} \circ \vartheta(P_N)^{-1}, Q^{\otimes N} \circ \vartheta(Q_N)^{-1}\right) \le \epsilon \tag{4.5.1}$$

for all $N \ge N_0$.

Proof. By the triangle inequality, we have

$$\begin{aligned} & \mathbf{d}_{\text{Prok}}\left(P^{\otimes N} \circ \vartheta(P_N)^{-1}, Q^{\otimes N} \circ \vartheta(Q_N)^{-1}\right) \\ \le\ & \mathbf{d}_{\text{Prok}}\left(P^{\otimes N} \circ \vartheta(P_N)^{-1}, \mathbb{1}_{\vartheta(P)}\right) + \mathbf{d}_{\text{Prok}}\left(\mathbb{1}_{\vartheta(P)}, Q^{\otimes N} \circ \vartheta(Q_N)^{-1}\right), \end{aligned}$$

where $\mathbb{1}_x$ denote the Dirac measure at $x \in \mathbb{R}$. It suffices to show existence of appropriate positive numbers δ and $N_0 \in \mathbb{N}$ such that for $Q \in \mathcal{M}$ with $\mathbf{d}_{\psi^p}(P,Q) \le \delta$, we have

$$\mathbf{d}_{\text{Prok}}\left(P^{\otimes N} \circ \vartheta(P_N)^{-1}, \mathbb{1}_{\vartheta(P)}\right) \le \frac{\epsilon}{2} \tag{4.5.2}$$

and

$$\mathbf{d}_{\text{Prok}}\left(\mathbb{1}_{\vartheta(P)}, Q^{\otimes N} \circ \vartheta(Q_N)^{-1}\right) \le \frac{\epsilon}{2} \tag{4.5.3}$$

for all $N \ge N_0$. By Strassen's theorem [49, Theorem 2.13], (4.5.2) and (4.5.3) are implied respectively by

$$P^{\otimes N}\left[\left\{\xi \in (\mathbb{R}^s)^N : |\vartheta(P_N) - \vartheta(P)| \le \frac{\epsilon}{2}\right\}\right] \ge 1 - \frac{\epsilon}{2} \tag{4.5.4}$$

and

$$Q^{\otimes N}\left[\left\{\xi \in (\mathbb{R}^s)^N : |\vartheta(P) - \vartheta(Q_N)| \le \frac{\epsilon}{2}\right\}\right] \ge 1 - \frac{\epsilon}{2}. \tag{4.5.5}$$

Note that (4.5.4) follows from Theorem 4.4.1. Thus we are left to prove (4.5.5). Note that

$$|\vartheta(P) - \vartheta(Q_N)| \le |\vartheta(P) - \vartheta(Q)| + |\vartheta(Q) - \vartheta(Q_N)|.$$

It follows by Theorem 4.3.1 that we can choose δ sufficiently small such that when $\mathbf{d}_{\psi^p}(P,Q) \le \delta$, we have

$$|\vartheta(P) - \vartheta(Q)| \le \frac{\epsilon}{4}. \tag{4.5.6}$$

Analogous to (4.5.4), we have from Theorem 4.4.1

$$Q^{\otimes N}\left[\left\{\xi \in (\mathbb{R}^s)^N : |\vartheta(Q) - \vartheta(Q_N)| \le \frac{\epsilon}{4}\right\}\right] \ge 1 - \frac{\epsilon}{2} \tag{4.5.7}$$

for all $N \ge N_0$. Combining (4.5.6) and (4.5.7), we have (4.5.5) holds. □

Theorem 4.5.1 provides a theoretical guarantee that if the perceived data is generated by some probability distribution Q which is close to the true distribution P and Q

satisfies moment condition (4.4.6), then the optimal value obtained with the perceived data is close to the one with real data.

As we discussed after Theorem 4.4.1, the set of probability measures $\mathcal{M}_{\kappa}^{\psi^p}$ is determined by the tail properties of $c(z, \xi)$ and loss function l. The heavier the tails, the smaller the set $\mathcal{M}_{\kappa}^{\psi^p}$ will be, which means less plausible probability distributions may be considered for the perturbation in the left hand side of (4.5.1) and consequently less chances for the statistical estimator of the optimal value to be robust (satisfying the inequality at the right hand side of (4.5.1)).

Chapter 5

Summary and Future Directions

Optimal decision making problems naturally arise in many fields of engineering and management science. Decision makers often face underlying exogenous and endogenous uncertainties when taking decisions. Such uncertainties are outside the deciders' control and knowledge at the time of the decisions and will affect the desired outcome. In this thesis, we focus on this issue and propose some efficient models to deal with the uncertainties. The following sections summarize this thesis and identify some promising directions for future research.

5.1 Summary

Chapter 2. In Chapter 2, we introduce a utility-based reward-risk ratio (URR) optimization model and consider a situation where the true probability distribution of the underlying random variables is unknown. To mitigate the risk arising from ambiguity of the true probability distribution, we propose a distributionally robust utility-based reward-risk ratio (DRURR) optimization model where the ambiguity set of probability distribution is constructed through prior moment information.

We reformulate the DRURR optimization model as a mathematical program with robust inequality constraints and further transform it into a nonlinear semi-infinite programming problem through the Lagrange dualization. We then apply the entropic approximation scheme to deal with the semi-infinite constraints and show the stability of the approximated optimization problem, and consequently propose a numerical scheme to solve it.

We also investigate a specific case that the ambiguity set is determined by the mean and covariance, and use box constraints for each component of the two quantities. We demonstrate that the true probability distribution lies in the ambiguity set with a high likelihood, and the optimal value and the optimal solutions obtained on the basis of the ambiguity set converge to their true counterpart. We apply the DRURR optimization model to a portfolio selection problem, and the numerical test results show some promising performance of our model and numerical scheme.

Chapter 3. In Chapter 3, we introduce a multivariate utility-based shortfall risk measure (MSR) and consider a situation where a decision makers true loss function is unknown but it is possible to elicit a set of plausible loss functions with partial information, and consequently propose a preference robust multivariate utility-based shortfall risk measure (PRMSR). We demonstrate that MSR and PRMSR are convex risk measures and define the domains of MSR and PRMSR.

We apply the PRMSR to an optimal decision making problem. Considering the case that the underlying probability distribution is continuous, we propose a sample average approximation scheme and show that it converges to the true problem in terms of the optimal value and the optimal solutions as the sample size increases. A tractable formulation is developed for the approximated optimization problem when the ambiguity set of loss functions is defined with some specified characteristics. Some numerical studies are also given to examine the efficiency of the proposed robust model.

Chapter 4. In Chapter 4, we consider a situation where the true probability distribution P is approximated by empirical data and the data may contain some noise which means the data are not exactly generated by P, rather they are generated by a perturbed distribution Q from P. Under these circumstances, we examine the quality of the optimal value $\vartheta(\cdot)$ obtained from solving the PRMSR optimization model with the empirical data.

We derive moderate sufficient conditions under which $\vartheta(\cdot)$ changes continuously against small variation of the true probability distribution P and identify appropriate metrics under which $\vartheta(P_N)$ uniformly converges $\vartheta(P)$. We establish statistical robustness of estimator of $\vartheta(\cdot)$, that is, $\vartheta(Q_N)$ is close to $\vartheta(P_N)$ under the Prokhorov metric when N is sufficiently large as long as Q is close to P.

5.2 Future Directions

There are several research directions which can be taken to extend the works in this thesis:

- In Chapter 2, we consider the situation where there is an ambiguity about the underlying probability distribution in the URR optimization model but assume the utility function in the definition of URR is known. We may take it further by taking into account of ambiguity about both decision makers risk preference and underlying probability distribution in the URR optimization model.

- In Chapter 3, we assume that the loss function in the definition of MSR is convex. In the future work, we may consider the case that the loss function has different shapes, such as quasi-convex or S-shaped. In addition, we may take it further by considering the situation where the direction d in the definition of MSR is also ambiguous since in this work we assume that the direction d is known.

- In Chapter 3, we demonstrate that when $L \subseteq \mathcal{L}$ then PRMSR is equal to the worst case MSR associated with the loss function l chosen from the ambiguity set L. In the future work, we may investigate whether the worst of a set of MSR with some specified characteristics such as positive homogeneity can be represented as PRMSR by defining an appropriate ambiguity set of loss functions.

- In Chapter 4, we derive moderate sufficient conditions under which the optimal value of (PRMSR-Opti) is statistically robust against perturbation of the exogenous uncertainty data. In the future work, we may investigate statistical robustness of the estimators of the optimal solutions for (PRMSR-Opti) as well.

Appendix A

Reformulation as A Semi-Infinite Programming Problem

Let us recall the problem (2.2.13)

$$\begin{array}{ll} \inf_{x\in X,\gamma\in\mathbb{R}} & -\gamma \\ \text{s.t.} & \sup_{P\in\mathcal{P}} \mathbb{E}_P[H(x,\xi,\gamma)] \le 0, \end{array}$$

and the ambiguity set $\mathcal{P}$ constructed in (2.3.1)

$$\mathcal{P} := \left\{ P \in \mathscr{P} : \begin{array}{ll} \mathbb{E}_P[\psi_i(\xi)] = 0, & i = 1,\dots,p \\ \mathbb{E}_P[\psi_i(\xi)] \le 0, & i = p+1,\dots,q \end{array} \right\}.$$

Consider the constraint of probelm (2.2.13) with the ambiguity set $\mathcal{P}$,

$$\begin{array}{lll} \sup_{P\in\mathscr{P}} & \mathbb{E}_P[H(x,\xi,\gamma)] & \\ \text{s.t.} & \mathbb{E}_P[\psi_i(\xi)] = 0, & i = 1,\dots,p \\ & \mathbb{E}_P[\psi_i(\xi)] \le 0, & i = p+1,\dots,q \\ & \mathbb{E}_P[\mathbb{1}_\Xi(\xi)] = 1, & \end{array}$$

where $\mathbb{1}_\Xi(\xi) := \begin{cases} 1, & \xi \in \Xi, \\ 0, & \xi \notin \Xi. \end{cases}$

The Lagrangian function can be defined as

$$L(P,\lambda,\mu) = \mathbb{E}_P[H(x,\xi,\gamma) - \Sigma_{i=1}^q \lambda_i \psi_i(\xi) - \mu \mathbb{1}_\Xi(\xi)] + \mu,$$

and the Lagrangian dual problem is

$$\begin{array}{lll} \inf\limits_{\mu,\lambda} & \mu & \\ \text{s.t.} & H(x,\xi,\gamma) - \Sigma_{i=1}^{q}\lambda_i\psi_i(\xi) \leq \mu \mathbb{1}_{\Xi}(\xi), & \forall \xi \in \Xi, \\ & \lambda_i \geq 0, & i = p+1,\ldots,q. \end{array} \tag{A1}$$

Problem (A1) can be reformulated as

$$\begin{array}{lll} \sup\limits_{\xi\in\Xi} & H(x,\xi,\gamma) - \Sigma_{i=1}^{q}\lambda_i\psi_i(\xi) & \\ \text{s.t.} & \lambda_i \geq 0, & i = p+1,\ldots,q. \end{array}$$

Then we can reformulate problem (2.2.13) as follows:

$$\begin{array}{ll} \inf\limits_{x\in X,\gamma\in\mathbb{R},\lambda\in\mathbb{R}^p\times\mathbb{R}_+^{q-p}} & -\gamma \\ \text{s.t.} & \sup_{\xi\in\Xi} H(x,\xi,\gamma) - \sum_{i=1}^{q}\lambda_i\psi_i(\xi) \leq 0. \end{array}$$

Appendix B

Entropic Approximation Parameter α

In this work, we use the samples of the random variables to compute sample average approximation (SAA) of the expected value in $e_\alpha(\cdot)$. In this case, the value of parameter α may be estimated by the sample size and the tolerance. For example, suppose there are 10 samples "$Z_1, \ldots, Z_{10}$" in decreasing order, then

$$e_\alpha(-Z) - Z_1 = \frac{1}{\alpha}\ln(1 + e^{\alpha(Z_2 - Z_1)} + \ldots + e^{\alpha(Z_{10}-Z_1)}) - \frac{1}{\alpha}\ln 10 \leq 0.$$

Thus

$$|e_\alpha(-Z) - Z_1| = Z_1 - e_\alpha(-Z) = \frac{1}{\alpha}\ln 10 - \frac{1}{\alpha}\ln(1 + e^{\alpha(Z_2 - Z_1)} + \ldots + e^{\alpha(Z_{10}-Z_1)}) \leq \frac{1}{\alpha}\ln 10.$$

If the precision is ϵ, then we may set $\alpha \geq \frac{\ln 10}{\epsilon}$.

In the numerical tests, we set $\epsilon = 0.01$ and use SAA method with 1000 samples to calculate the expectation in $e_\alpha(Z) := \frac{1}{\alpha}\ln \mathbb{E}[e^{-\alpha Z}]$. $\alpha = 700$ ensures $\frac{1}{700}\ln(1000) < 0.01$.

Appendix C

Proof of Inequality (3.4.20)

Under Robinson's error bound theorem, we have

$$d((z,t), \mathcal{F} \cap Z \times T) \leq \frac{\Delta}{\theta} \max\{v(z,t), 0\}, \ \forall (z,t) \in Z \times T \tag{C1}$$

and

$$d((z,t), \mathcal{F}_N \cap Z \times T) \leq \frac{2\Delta}{\delta} \max\{v_N(z,t), 0\}, \ \forall (z,t) \in Z \times T, \tag{C2}$$

where δ is a positive number and $\delta < \theta$.

Let $(z', t') \in \mathcal{F}_N \cap Z \times T$, then

$$\begin{array}{rcl} d((z',t'), \mathcal{F} \cap Z \times T) & \leq & \frac{\Delta}{\theta} \max\{v(z',t'), 0\} - \frac{\Delta}{\theta} \max\{v_N(z',t'), 0\} \\ & \leq & \frac{\Delta}{\theta} |v(z',t') - v_N(z',t')| \\ & \leq & \frac{\Delta}{\theta} \sup_{z \in Z, t \in T} |v(z,t) - v_N(z,t)|. \end{array}$$

The first inequality follows from (C1) and the fact that $v_N(z', t') \leq 0$. Thus

$$\mathbb{D}(\mathcal{F}_N \cap Z \times T, \mathcal{F} \cap Z \times T) \leq \frac{\Delta}{\theta} \sup_{z \in Z, t \in T} |v(z,t) - v_N(z,t)|. \tag{C3}$$

Conversely, let $(z', t') \in \mathcal{F} \cap Z \times T$. Then

$$\begin{array}{rcl} d((z',t'), \mathcal{F}_N \cap Z \times T) & \leq & \frac{2\Delta}{\delta} \max\{v_N(z',t'), 0\} - \frac{2\Delta}{\delta} \max\{v(z',t'), 0\} \\ & \leq & \frac{2\Delta}{\delta} |v_N(z',t') - v(z',t')| \\ & \leq & \frac{2\Delta}{\delta} \sup_{z \in Z, t \in T} |v_N(z,t) - v(z,t)|. \end{array}$$

Thus

$$\mathbb{D}(\mathcal{F} \cap Z \times T, \mathcal{F}_N \cap Z \times T) \leq \frac{2\Delta}{\delta} \sup_{z \in Z, t \in T} |v_N(z,t) - v(z,t)|. \tag{C4}$$

Combining (C3) and (C4), we obtain

$$\mathbb{H}(\mathcal{F}_N \cap Z \times T, \mathcal{F} \cap Z \times T) \leq \frac{2\Delta}{\delta} \sup_{z \in Z, t \in T} |v_N(z,t) - v(z,t)|.$$

Appendix D

Computational Details

The first model (3.6.3): Let $\tau_{ij}^k := (\xi_{ij}(\omega_k) - z_{ij})_+$ for $i = 1, \ldots, m;\ j = 1, \ldots, n;\ k = 1, \ldots, N$. Then τ_{ij}^k can be written as

$$\tau_{ij}^k \geq 0, \text{ and } \tau_{ij}^k \geq \xi_{ij}(\omega_k) - z_{ij}.$$

Consequently (3.6.3) can be rewritten as

$$\begin{array}{ll}\min\limits_{z,\tau} & \sum\limits_{k=1}^{N} p_k \sum\limits_{j=1}^{n} w_j \sum\limits_{i=1}^{m} \tau_{ij}^k \\ \text{s.t.} & \tau_{ij}^k \geq 0, \quad i = 1, \ldots, m;\ j = 1, \ldots, n;\ k = 1, \ldots, N, \\ & \tau_{ij}^k \geq \xi_{ij}(\omega_k) - z_{ij}, \quad i = 1, \ldots, m;\ j = 1, \ldots, n;\ k = 1, \ldots, N, \\ & z \in Z. \end{array} \tag{D1}$$

Problem (D1) is a LP and can be easily solved by using e.g., `fmincon` or `CVX` in Matlab.

The second model (3.6.4): Let $\mu_j = \mathrm{CVaR}_{\alpha_j}(c_j(z)(\omega))$ for $j = 1, \ldots, n$. Then there exists a $\eta_j^* \in \mathbb{R}$ such that

$$\mu_j \geq \eta_j^* + \frac{1}{1-\alpha_j}\mathbb{E}[(c_j(z)(\omega) - \eta_j^*)_+], \quad j = 1, \ldots, n.$$

Let $\lambda_j^k = (c_j(z)(\omega_k) - \eta_j)_+$ for $j = 1, \ldots, n,\ k = 1, \ldots, N$. Then (3.6.4) can be represented as

$$
\begin{aligned}
\min_{z,\tau,\lambda,\mu,\eta} \quad & \sum_{j=1}^{n} \mu_j \\
\text{s.t.} \quad & \mu_j \geq \eta_j + \frac{1}{1-\alpha_j} \sum_{k=1}^{N} p_k \lambda_j^k, \quad j = 1, \ldots, n, \\
& \lambda_j^k \geq \sum_{i=1}^{m} \tau_{ij}^k - \eta_j, \quad j = 1, \ldots, n; k = 1, \ldots, N, \\
& \lambda_j^k \geq 0, \quad j = 1, \ldots, n; k = 1, \ldots, N, \\
& \tau_{ij}^k \geq \xi_{ij}(\omega_k) - z_{ij}, \quad i = 1, \ldots, m; j = 1, \ldots, n; k = 1, \ldots, N, \\
& \tau_{ij}^k \geq 0, \quad i = 1, \ldots, m; j = 1, \ldots, n; k = 1, \ldots, N, \\
& z \in Z.
\end{aligned}
\tag{D2}
$$

Problem (D2) is a LP and can be easily solved by using e.g., `fmincon` or `CVX` in Matlab.

The third model (3.6.5): We using the proposed tractable reformulation (3.5.12) to solve (3.6.5). We assign values to components in Θ as follows:

$$
\theta_j = \begin{cases} G_k(\omega_i), & \text{for } j = 2kN + i, \\ B_k(\omega_i), & \text{for } j = (2k+1)N + i, \\ Y(\omega_i), & \text{for } j = 2(K+1)N + i, \\ 0, & \text{for } j = (2K+3)N + 1. \end{cases} \quad \text{with } i = 1, \ldots, N; k = 0, \ldots, K. \tag{D3}
$$

To make the numerical formulation clear, we introduce some notations for our experiments. The index of cities is $i = 1, \ldots, m = 10$; the index of criteria is $j = 1, \ldots, n = 4$; the index of scenarios is $k = 1, \ldots, N = 3$; the index of θ is $l = 1, \ldots, J = 40$; the index of elicited pairs is $q = 1, \ldots, K = 5$. There the numerical formulation for our experiments of (3.5.12) is

$$
\begin{aligned}
\min_{z,t,\gamma,\tau,\lambda,\mu,\eta,\alpha,\beta,\nu} \quad & t \\
\text{s.t.} \quad & \tau + \langle \kappa e, \sum_{k=1}^{N} \alpha_k + \sum_{l=1}^{J} \beta_l \rangle \leq 0, \\
& \sum_{k=1}^{N} \mu_{kl} + \sum_{l'=1}^{J} \eta_{l'l} - \sum_{l'=1}^{J} \eta_{ll'} + \sum_{q=1}^{K} \lambda_q (P(-B_q = \theta_l) \\
& -P(-Y = \theta_l) + \gamma P(-O = \theta_l) + \tau P(-G_0 = \theta_l) \\
& -\tau P(-B_0 = \theta_l) - P(-G_q = \theta_l)) = 0, \quad l = 1, \ldots, J, \\
& -p_k (\sum_{i=1}^{m} \nu_{ij}^k - t d_j) + \sum_{l=1}^{J} \mu_{kl} \theta_{lj} + \alpha_{kj} \geq 0, \quad k = 1, \ldots, N; j = 1, \ldots, n, \\
& \sum_{l'=1}^{J} \eta_{ll'} (\theta_{l'} - \theta_l) + \beta_l \geq 0, \quad l = 1, \ldots, J, \\
& \sum_{l=1}^{J} \mu_{kl} = p_k, \quad k = 1, \ldots, N, \\
& \nu_{ij}^k \geq \xi_{ij}(\omega_k) - z_{ij}, \quad i = 1, \ldots, m; j = 1, \ldots, n; k = 1, \ldots, N, \\
& \nu_{ij}^k \geq 0, \quad i = 1, \ldots, m; j = 1, \ldots, n; k = 1, \ldots, N, \\
& \lambda_q \geq 0, \quad q = 1, \ldots, K; \\
& \mu_{kl} \geq 0, \quad k = 1, \ldots, N; l = 1, \ldots, J, \\
& \eta_{ll'} \geq 0, \quad l = 1, \ldots, J; l' = 1, \ldots, J, \\
& \alpha_k \geq 0, \quad k = 1, \ldots, N, \\
& \beta_l \geq 0, \quad l = 1, \ldots, J, \\
& z \in Z.
\end{aligned}
\tag{D4}
$$

Problem (D4) is a LP and can be easily solved by using e.g., `fmincon` or `CVX` in Matlab.

Appendix E

Terrorism Losses Data Set

Table E.1 presents the raw data obtained from Hu et al. [45] and Haskell et al. [43] which is used as the loss scenarios.

Cities	Property losses ($ million)			Fatalities ($ million)		
	Reduced	Standard	Increased	Reduced	Standard	Increased
New York	265	413	550	221	304	401
Chicago	77	115	150	38	54	73
San Francisco	38	57	81	16	24	36
Washington	21	36	59	16	29	48
Los Angeles	16	34	58	7	17	31
Philadelphia	8	21	28	5	9	13
Boston	8.3	18	26	8	12	17
Houston	6.7	11	15	6	9	12
Newark	0.8	7.3	12	0.1	4	9
Seattle	4	6.7	10	3	4	6

Cities	Air departures ($ million)			Bridge traffic ($ million)		
	Reduced	Standard	Increased	Reduced	Standard	Increased
New York	10.71	11.78	12.96	162.41	178.65	196.51
Chicago	18.13	19.94	21.94	86.81	95.50	105.04
San Francisco	9.56	8.69	10.51	75.62	83.18	91.50
Washington	7.83	8.61	9.47	69.43	76.38	84.01
Los Angeles	13.08	14.39	15.82	91.50	100.65	110.71
Philadelphia	6.19	6.81	7.49	52.34	57.57	63.33
Boston	5.28	5.80	6.38	182.18	200.40	220.44
Houston	9.52	10.47	11.52	84.04	92.44	101.68
Newark	5.82	6.40	7.04	141.09	155.19	170.71
Seattle	6.16	6.78	7.46	57.37	63.50	69.85

Table E.1: The random loss $\xi_{ij}(\omega)$ for $i = 1, \ldots, 10$, $j = 1, \ldots, 4$ and $|\Omega| = 3$.

Appendix F

Elicited Comparison Data Set

In this case study, we define the prospects over three scenarios where each scenario has an equal one-third probability of being realized. Since the elicited comparison data set is designed as a questionnaire to ask the decision maker to obtain his/her preference, the precise values of the comparison data set have slightly effect on the elicited preference for decision makers. Therefore, we assume that for each prospect, we generate the loss for attribute j for each scenario from the uniform distribution on

$$\left[r_1 \times \min_{i=1,\ldots,10;\omega\in\Omega}\{\xi_{ij}(\omega)\},\ r_2 \times \max_{i=1,\ldots,10;\omega\in\Omega}\{\xi_{ij}(\omega)\}\right], \tag{D1}$$

where $r_1 < 1 < r_2$. To generate the elicited comparison data set in our experiments, we set $r_1 = 0.5$ and $r_2 = 2$ and then the ranges to generate the realization for each criterion are $[0.4, 1100]$, $[0.05, 802]$, $[2.64, 43.88]$ and $[26.17, 440.88]$. Specifically, for each criterion, we firstly generate three random realizations from (D1) and then assign these values to different scenarios according to the risk levels, i.e., increased loss level has the largest value, and so on. In our setting, the elicited comparison data set is given in Table F.1.

	Criteria	Losses ($ million) (G_k)			Losses ($ million) (B_k)		
		Reduced	Standard	Increased	Reduced	Standard	Increased
Pair 0	Property losses	83.93	264.21	442.23	371.76	406.42	990.10
	Fatalities	98.95	147.53	192.48	89.23	312.60	625.77
	Air departures	4.69	19.85	39.87	6.62	12.61	19.30
	Bridge traffic	229.07	229.74	417.98	80.90	416.85	422.69
Pair 1	Property losses	286.28	394.39	1017.51	562.94	618.95	865.63
	Fatalities	26.21	72.53	597.01	29.93	344.68	782.59
	Air departures	6.18	28.42	31.47	13.51	24.55	39.28
	Bridge traffic	239.75	307.93	393.77	78.20	338.13	431.26
Pair 2	Property losses	102.00	753.40	959.89	106.62	930.53	1037.25
	Fatalities	307.40	418.94	729.53	204.56	709.34	729.06
	Air departures	12.32	23.08	37.27	8.92	12.16	29.77
	Bridge traffic	105.33	350.24	369.16	32.17	161.07	203.93
Pair 3	Property losses	423.91	714.37	715.52	12.87	576,24	1000.37
	Fatalities	171.80	455.23	717.40	319.64	501.28	717.44
	Air departures	14.22	19.64	32.16	14.68	25.86	28.30
	Bridge traffic	200.24	247.02	354.63	82.03	213.00	263.50
Pair 4	Property losses	306.31	469.75	923.88	633.42	695.26	839.23
	Fatalities	196.42	244.26	514.03	3.14	188.61	706.23
	Air departures	14.58	23.24	39.84	8.23	17.58	34.86
	Bridge traffic	47.66	361.51	399.88	252.12	399.03	410.07
Pair 5	Property losses	441.12	488.47	609.86	297.50	694.93	916.88
	Fatalities	126.05	313.45	642.67	662.16	708.72	758.19
	Air departures	7.96	31.37	41.65	8.82	30.56	34.96
	Bridge traffic	37.84	94.76	170.59	65.06	121.01	367.23

Table F.1: Elicited comparison data set example

Appendix G

Proof of Theorem 4.3.1 Part (ii)

As we discussed after Assumption 3.4.1 that $\vartheta(P)$ is finite and $\vartheta(P) \leq t_0$. Let α be any fixed positive constant. Then there exists a constant $c_\alpha > 0$ such that

$$\inf_{z \in Z} v_P(z, \vartheta(P) - c_\alpha) \geq \alpha. \tag{G1}$$

Let T be chosen as $[\vartheta(P) - c_\alpha, t_0] \subset T$. Since $P_k \xrightarrow{\psi} P$, then by (4.3.13), there exists k_0 sufficiently large such that for all $k \geq k_0$,

$$\begin{aligned}
& \inf_{z \in Z} v_{P_k}(z, \vartheta(P) - c_\alpha) \\
\geq \;& \alpha - \sup_{z \in Z} |v_{P_k}(z, \vartheta(P) - c_\alpha) - v_P(z, \vartheta(P) - c_\alpha)| \\
\geq \;& \alpha - \varepsilon > 0.
\end{aligned}$$

Let $(z_{P_k}, \vartheta(P_k))$ be the optimal solution of problem (Opti-P$_\text{k}$). The inequality above shows

$$v_{P_k}(z_{P_k}, \vartheta(P) - c_\alpha) \geq \inf_{z \in Z} v_{P_k}(z, \vartheta(P) - c_\alpha) > 0, \quad \forall k \geq k_0, \tag{G2}$$

which implies $\vartheta(P_k) > \vartheta(P) - c_\alpha$.

On the other hand, it follows by (3.4.7) and (4.3.13),

$$\begin{aligned}
& \sup_{l \in L} \mathbb{E}_{P_k}[l(-c(z_0, \xi) - t_0 d) - l(-Y)] \\
\leq \;& -\theta + v_{P_k}(z_0, t_0) - v_P(z_0, t_0) \\
\leq \;& -\theta + \sup_{z \in Z, t \in T} |v_{P_k}(z, t) - v_P(z, t)| \\
\leq \;& -\theta + \varepsilon < 0
\end{aligned} \tag{G3}$$

for all $k \geq k_0$. The inequality (G3) implies (z_0, t_0) is a feasible solution to problem

(Opti-P_k) and hence $\vartheta(P_k) \le t_0$. Summarizing the discussions above, we have

$$\vartheta(P_k) \in [\vartheta(P) - c_\alpha, t_0] \tag{G4}$$

for all $k \ge k_0$.

Consider the systems of inequalities

$$v_P(z,t) \le 0, \quad (z,t) \in Z \times T$$

and

$$v_{P_k}(z,t) \le 0, \quad (z,t) \in Z \times T.$$

The set of solutions to the systems of inequalities are equal to $\mathcal{F}_P \cap Z \times T$ and $\mathcal{F}_{P_k} \cap Z \times T$ respectively. By Robinson's error bound theorem for convex systems [67], for any $(z,t) \in Z \times T$,

$$d((z,t), \mathcal{F}_P \cap Z \times T) \le \frac{\Delta}{\theta} \max\{v_P(z,t), 0\},$$

where Δ denotes the diameter of $\mathcal{F} \cap Z \times T$, θ is the parameter in the Slater condition (3.4.7). Likewise, for any $(z,t) \in Z \times T$, we have

$$d((z,t), \mathcal{F}_{P_k} \cap Z \times T) \le \frac{\Delta}{\delta} \max\{v_{P_k}(z,t), 0\},$$

for all $k \ge k_0$, where δ is a positive number and $\delta < \theta$. Combining the two error bounds, we obtain

$$\mathbb{H}(\mathcal{F}_{P_k} \cap Z \times T, \mathcal{F}_P \cap Z \times T) \le \frac{\Delta}{\delta} \sup_{z \in Z, t \in T} |v_{P_k}(z,t) - v_P(z,t)|.$$

Thus

$$|\vartheta(P_k) - \vartheta(P)| \le \mathbb{H}(\mathcal{F}_{P_k} \cap Z \times T, \mathcal{F}_P \cap Z \times T) \le \frac{\Delta}{\delta} \sup_{z \in Z, t \in T} |v_{P_k}(z,t) - v_P(z,t)|. \tag{G5}$$

We deduce from (4.3.13) and (G5)

$$\lim_{k \to \infty} |\vartheta(P_k) - \vartheta(P)| = 0$$

which implies (4.3.14).

Bibliography

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

[2] C. Acerbi and D. Tasche. Expected shortfall: a natural coherent alternative to value at risk. *Economic notes*, 31(2):379–388, 2002.

[3] C. Acerbi and D. Tasche. On the coherence of expected shortfall. *Journal of Banking and Finance*, 26(7):1487–1503, 2002.

[4] A. Adam, M. Houkari, and J.-P. Laurent. Spectral risk measures and portfolio selection. *Journal of Banking and Finance*, 32(9):1870–1882, 2008.

[5] E. J. Anderson. *Business Risk Management: Models and Analysis*. John Wiley & Sons, 2013.

[6] B. Armbruster and E. Delage. Decision making under uncertainty when preference information is incomplete. *Management science*, 61(1):111–128, 2015.

[7] Y. Armenti, S. Crépey, S. Drapeau, and A. Papapantoleon. Multivariate shortfall risk allocation and systemic risk. *SIAM Journal on Financial Mathematics*, 9(1):90–126, 2018.

[8] P. Artzner, F. Delbaen, J. Eber, and D. Heath. Thinking coherently. *Risk*, pages 68–71, 1997.

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

[10] K. B. Athreya and S. N. Lahiri. *Measure theory and probability theory*. Springer Science & Business Media, 2006.

[11] R. Babazadeh, J. Razmi, and M. S. Pishvaee. Robust facility location in supply chain network design under risk. *International Journal of Mechanic Systems Engineering*, 1(1):35–47, 2011.

[12] A. Ben-Tal, T. Margalit, and A. Nemirovski. Robust modeling of multi-stage portfolio problems. In *High performance optimization*, pages 303–328. Springer, 2000.

[13] F. Biagini, J.-P. Fouque, M. Frittelli, and T. Meyer-Brandis. A unified approach to systemic risk measures via acceptance sets. *Mathematical Finance*, 29(1):329–367, 2019.

[14] A. Biglova, S. Ortobelli, S. T. Rachev, and S. Stoyanov. Different approaches to risk estimation in portfolio theory. *The Journal of Portfolio Management*, 31(1):103–112, 2004.

[15] P. Billingsley. Convergence of probability measures. 1999.

[16] F. Black and R. Litterman. Global portfolio optimization. *Financial analysts journal*, 48(5):28–43, 1992.

[17] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[18] R. F. Brown. A topological introduction to nonlinear analysis. 2004.

[19] M. C. Carneiro, G. P. Ribas, and S. Hamacher. Risk management in the oil supply chain: a cvar approach. *Industrial and Engineering Chemistry Research*, 49(7):3286–3294, 2010.

[20] T. C. Chan, H. Mahmoudzadeh, and T. G. Purdie. A robust-cvar optimization approach with application to breast cancer therapy. *European Journal of Operational Research*, 238(3):876–885, 2014.

[21] M. Claus. Advancing stability analysis of mean-risk stochastic programs: Bilevel and two-stage models. 2012.

[22] R. Cont, R. Deguest, and G. Scandolo. Robustness and sensitivity analysis of risk measurement procedures. *Quantitative finance*, 10(6):593–606, 2010.

[23] E. De Giorgi. Reward–risk portfolio selection and stochastic dominance. *Journal of Banking & Finance*, 29(4):895–926, 2005.

[24] E. Delage, S. Guo, and H. Xu. *Shortfall Risk Models when Information of Loss Function is Incomplete*. GERAD HEC Montréal, 2018.

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

[26] E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations research*, 58(3):595–612, 2010.

[27] F. Delbaen. Coherent risk measures on general probability spaces. In *Advances in finance and stochastics*, pages 1–37. Springer, 2002.

[28] D. Dentcheva and A. Ruszczyński. Robust stochastic dominance and its application to risk-averse optimization. *Mathematical Programming*, 123(1):85, 2010.

[29] J. Dunkel and S. Weber. Stochastic root finding and efficient estimation of convex risk measures. *Operations Research*, 58(5):1505–1521, 2010.

[30] J. Dupačová. Uncertainties in minimax stochastic programs. *Optimization*, 60(10-11):1235–1250, 2011.

[31] J. Fliege and R. Werner. Robust multiobjective optimization & applications in portfolio optimization. *European Journal of Operational Research*, 234(2):422–433, 2014.

[32] H. Föllmer and A. Schied. Convex measures of risk and trading constraints. *Finance and stochastics*, 6(4):429–447, 2002.

[33] H. Föllmer and A. Schied. *Stochastic finance: an introduction in discrete time*. Walter de Gruyter, 2011.

[34] M. Frittelli and E. R. Gianin. Putting order in risk measures. *Journal of Banking and Finance*, 26(7):1473–1486, 2002.

[35] L. E. Ghaoui, M. Oks, and F. Oustry. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations research*, 51(4):543–556, 2003.

[36] K. Giesecke, T. Schmidt, and S. Weber. Measuring the risk of large losses. *Journal of Investment Management*, 6(4), 2008.

[37] D. Goldfarb and G. Iyengar. Robust portfolio selection problems. *Mathematics of operations research*, 28(1):1–38, 2003.

[38] B. L. Gorissen. Robust fractional programming. *Journal of Optimization Theory and Applications*, 166(2):508–528, 2015.

[39] T. Guldimann. The story of riskmetrics. *Risk*, 13(1):56–58, 2000.

[40] S. Guo, H. Xu, and L. Zhang. Convergence analysis for mathematical programs with distributionally robust chance constraint. *SIAM Journal on optimization*, 27(2):784–816, 2017.

[41] F. R. Hampel. A general qualitative definition of robustness. *The Annals of Mathematical Statistics*, pages 1887–1896, 1971.

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

[43] W. B. Haskell, W. Huang, and H. Xu. Preference elicitation and robust optimization with multi-attribute quasi-concave choice functions. *Available at https://arxiv.org/pdf/1805.06632.pdf*, 2018.

[44] W. B. Haskell, J. G. Shanthikumar, and Z. M. Shen. Aspects of optimization with stochastic dominance. *Annals of Operations Research*, 253(1):247–273, 2017.

[45] J. Hu, T. Homem-de Mello, and S. Mehrotra. Risk-adjusted budget allocation models with application in homeland security. *IIE Transactions*, 43(12):819–839, 2011.

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

[47] Z. Hu and Z. Dali. Convex risk measures: Efficient computations via monte carlo. 2016.

[48] P. J. Huber. Robust statistics. *Wiley Series in Probability and Mathematical Statistics, New York: Wiley*, 1981.

[49] P. J. Huber and E. M. Ronchetti. *Robust statistics*. A John Wiley & Sons, INC., Publication, 2009.

[50] M. Kapsos, N. Christofides, and B. Rustem. Worst-case robust omega ratio. *European Journal of Operational Research*, 234(2):499–507, 2014.

[51] H. Konno, H. Shirakawa, and H. Yamazaki. A mean-absolute deviation-skewness portfolio optimization model. *Annals of Operations Research*, 45(1):205–220, 1993.

[52] V. Krätschmer, A. Schied, and H. Zähle. Qualitative and infinitesimal robustness of tail-dependent statistical functionals. *Journal of Multivariate Analysis*, 103(1):35–47, 2012.

[53] V. Krätschmer, A. Schied, and H. Zähle. Comparative and qualitative robustness for law-invariant risk measures. *Finance and Stochastics*, 18(2):271–295, 2014.

[54] S. Kusuoka. On law invariant coherent risk measures. In *Advances in mathematical economics*, pages 83–95. Springer, 2001.

[55] Y. Liu and H. Xu. Entropic approximation for mathematical programs with robust equilibrium constraints. *SIAM Journal on Optimization*, 24(3):933–958, 2014.

[56] M. S. Lobo and S. Boyd. The worst-case risk of a portfolio. *Unpublished manuscript. Available from http://faculty. fuqua. duke. edu/%7Emlobo/bio/researchfiles/rsk-bnd. pdf*, 2000.

[57] Z. Lu. A new cone programming approach for robust portfolio selection. *Optim. Methods Soft*, 26(1):89–104, 2006.

[58] H. Markowitz. Portfolio selection. *The journal of finance*, 7(1):77–91, 1952.

[59] L. Martellini and B. Urošević. Static mean-variance analysis with uncertain time horizon. *Management Science*, 52(6):955–964, 2006.

[60] J. M. Mulvey and H. G. Erkan. Applying cvar for decentralized risk management of financial companies. *Journal of Banking & Finance*, 30(2):627–644, 2006.

[61] K. Natarajan, D. Pachamanova, and M. Sim. Incorporating asymmetric distributional information in robust value-at-risk optimization. *Management Science*, 54(3):573–585, 2008.

[62] J.-S. Pang. Error bounds in mathematical programming. *Mathematical Programming*, 79(1-3):299–332, 1997.

[63] G. C. Pflug. Some remarks on the value-at-risk and the conditional value-at-risk. In *Probabilistic constrained optimization*, pages 272–281. Springer, 2000.

[64] I. Popescu. Robust mean-covariance solutions for stochastic optimization. *Operations Research*, 55(1):98–112, 2007.

[65] M. Pritsker. Evaluating value at risk methodologies: accuracy versus computational time. *Journal of Financial Services Research*, 12(2-3):201–242, 1997.

[66] R. R. Rao. Relations between weak and uniform convergence of measures with applications. *The Annals of Mathematical Statistics*, pages 659–680, 1962.

[67] S. M. Robinson. An application of error bounds for convex programming in a linear space. *SIAM Journal on Control*, 13(2):271–273, 1975.

[68] R. T. Rockafellar and S. Uryasev. Conditional value-at-risk for general loss distributions. *Journal of banking and finance*, 26(7):1443–1471, 2002.

[69] R. T. Rockafellar, S. Uryasev, et al. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 2000.

[70] R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.

[71] W. Römisch. Stability of stochastic programming problems. *Handbooks in operations research and management science*, 10:483–554, 2003.

[72] A. Shapiro. On duality theory of conic linear problems. In *Semi-infinite programming*, pages 135–165. Springer, 2001.

[73] A. Shapiro. Monte carlo sampling methods. *Handbooks in operations research and management science*, 10:353–425, 2003.

[74] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on stochastic programming: modeling and theory*. SIAM, 2009.

[75] A. Shapiro and H. Xu. Stochastic mathematical programs with equilibrium constraints, modelling and sample average approximation. *Optimization*, 57(3):395–418, 2008.

[76] W. F. Sharpe. Mutual fund performance. *The Journal of business*, 39(1):119–138, 1966.

[77] W. F. Sharpe. The sharpe ratio. *Journal of portfolio management*, 21(1):49–58, 1994.

[78] J. Shawe-Taylor and N. Cristianini. Estimating the moments of a random vector with applications. *URL. Available from documents.irevues.inist.fr/bitstream/handle/2042/13775/A478.pdf*, 2003.

[79] S. V. Stoyanov, S. T. Rachev, and F. J. Fabozzi. Optimal financial portfolios. *Applied Mathematical Finance*, 14(5):401–436, 2007.

[80] H. Sun and H. Xu. Convergence analysis for distributionally robust optimization and equilibrium problems. *Mathematics of Operations Research*, 41(2):377–401, 2015.

[81] X. Tong and F. Wu. Robust reward–risk ratio optimization with application in allocation of generation asset. *Optimization*, 63(11):1761–1779, 2014.

[82] R. H. Tütüncü and M. Koenig. Robust asset allocation. *Annals of Operations Research*, 132(1-4):157–187, 2004.

[83] J. Von Neumann and O. Morgenstern. *Theory of games and economic behavior (commemorative edition)*. Princeton university press, 2007.

[84] H. H. Willis, A. R. Morral, T. K. Kelly, and J. J. Medby. *Estimating terrorism risk*. Rand Corporation, 2006.

[85] H. Xu. Uniform exponential convergence of sample average random functions under general sampling with applications in stochastic programming. *Journal of Mathematical Analysis and Applications*, 368(2):692–710, 2010.

[86] H. Xu, Y. Liu, and H. Sun. Distributionally robust optimization with matrix moment constraints: Lagrange duality and cutting plane methods. *Mathematical Programming*, 169(2):489–529, 2018.

[87] H. Zähle. A definition of qualitative robustness for general point estimators, and examples. *Journal of Multivariate Analysis*, 143:12–31, 2016.

[88] S. Zhu and M. Fukushima. Worst-case conditional value-at-risk with application to robust portfolio management. *Operations research*, 57(5):1155–1168, 2009.

[89] S. Zhu, D. Li, and S. Wang. Robust portfolio selection under downside risk measures. *Quantitative Finance*, 9(7):869–885, 2009.