OPTIMIZATION FOR DECISION MAKERS WITH AMBIGUOUS PREFERENCES

YANG ZHIYUE

NATIONAL UNIVERSITY OF SINGAPORE

2017

OPTIMIZATION FOR DECISION MAKERS WITH AMBIGUOUS PREFERENCES

YANG ZHIYUE

(*B.Eng. Shanghai Jiao Tong University*)

A THESIS SUBMITTED
FOR THE DEGREE OF PHD OF ENGINEERING
DEPARTMENT OF INDUSTRIAL AND SYSTEMS ENGINEERING
NATIONAL UNIVERSITY OF SINGAPORE

2017

SUPERVISORS:
Assistant Professor Tan Chin Hon, Main Supervisor
Assistant Professor William Benjamin Haskell, Co-Supervisor

DECLARATION

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

YANG ZHIYUE

12th January 2017

Acknowledgment

The four-year period PhD study in ISE department was an impressive experience. Firstly, I would like to extend my sincere gratitude to my main supervisor, Assistant Prof. Tan Chin Hon. I am thankful that Prof. Tan could accept me as his student after my previous supervisor left NUS. I have benefited intellectually from his invaluable guidance and sharp insights on academic research and emotionally from his great and selfless care. His talent and enthusiasm in work have a deep influence on my attitude towards challenges and obstacles as an individual. I want also express my sincere gratitude to my co-supervisor, Assistant Prof. William B. Haskell. His wisdom and expertise in optimization have enlightened a lot in my research directions. Sharing his mathematical insights and research experiences opens my way to a better understanding of optimization.

I shall also deeply thank the ISE department and National University of Singapore for providing me with the prestigious Graduate Research Scholarship. This great opportunity offers me support so that I can devote to my research. This opportunity also provides me with a platform so that I could meet my excellent colleagues from whom I can always learn, including my seniors, my labmates as well as PhD candidates under our common supervisors.

Finally, I am immensely indebted to my parents and friends for their unconditional love and support. Thanks to them, I can achieve this.

Contents

Abstract

The fundamental goal of this thesis is to build optimization frameworks for decision makers with unknown preferences under different conditions. Specifically, we firstly propose a new stochastic dominance relationship in which utility functions are weighted against a reference utility for risk-averse decision makers. The necessary and sufficient conditions are provided. We then formulate our proposed weighted almost stochastic dominance in our optimization framework by convex function interpolation and subgradient characterization. We will resort to linear programming and its duality as our technique.

We next extend the concept of almost stochastic dominance to random variables with normal and log-normal probability distributions, and apply the results to mean-variance analysis and Maximum Geometric Mean (MGM) strategy. We show how we could calculate the amount of dominance by which a normally or log-normally distributed reward dominates another by almost stochastic dominance and determine the set of utility functions such that one prospect dominates the other.

We then provide a more general optimization framework that considers the following four factors: multivariate prospects, preference uncertainty, computational tractability, and target-oriented measure. Two approaches, subgradient characterization and acceptance set approach, are considered.

List of Tables

List of Figures

List of Symbols

X, Y	Upper-case letters denote common vectors
F_X, F_Y	Cumulative distribution function of X and Y
$\mathbb{R}^n$	Set of all real valued vectors with n entries
Ω	Sample space
ω	Outcomes in sample space
$\mathcal{B}$	Borel σ-algebra
P	Probability measure
$\mathcal{L}^n$	Linear space of n-dimension random vectors
$\mathcal{C}$	Set of continuous functions
$\mathcal{C}^1$	Set of continuously differentiable functions
$\mathcal{Z}$	Decision set
m	Weight function
supp Y	Support of Y
e_i	Unit vector with i^{th} component 1
U	Set of all differentiable nondecreasing utility functions
U_{icv}	Set of all differentiable nondecreasing concave functions
$\mathcal{R}_{icx}$	Set of all differentiable nondecreasing convex risk measures
Θ	Discretization of the support
$\mathcal{A}$	Acceptance set
W_k, Y_k	Random vectors that are used for elicited comparisons
$X \succeq_U Y$	$\mathbb{E}[u(X)] \geq \mathbb{E}[u(Y)]$ for all $u \in U$

List of Abbreviations

SD	Stochastic Dominance
ASD	Almost Stochastic Dominance
WASD	Weighted Almost Stochastic Dominance
WRASD	Weighted Risk-averse Almost Stochastic Dominance
LP	Linear Programming
MGM	Maximum Geometric Mean

Chapter 1

Introduction

1.1 Background

Comparing random variables, or more precisely comparing distributions, has been fundamental and of great interest as it is prevalent in real life. For instance, we may encounter situations to choose between investments, or to compare Monte Carlo simulation results. Some of the early works on this topic include Mann and Whitney (1947), who gave a test of whether a random variable is stochastically larger than another; and Lehmann (1955), who made a comparison of several definitions of ordered sets of distributions. Their discussion laid the foundation of the subsequent approaches for comparing random variables.

One common way of ranking random variables is the notion of stochastic dominance (SD). SD is a form of stochastic ordering. The term is used in decision analysis to refer to situations where one prospect (a probability distribution over possible outcomes) can be considered stochastically superior over another based on preferences regarding outcomes (Levy, 1992). A preference might be a simple ranking of outcomes from most to least favored, or it might also employ a value measure such as different kinds of utility functions. As more assumptions such as risk aversion are needed, higher order SD is taken into account.

The concept of SD was first introduced in a test of whether one of two random variables is stochastically larger than the other. However, traditional SD rules may not reveal dominance between two options even though there is an obvious preference for one of these options because the corresponding utility sets contain "extreme" utility functions that do not correspond to decision makers observed in practice. One example would be that most "reasonable" investors would prefer a higher proportion of stocks as the investment horizon increases (Leshno and Levy, 2002). An "extreme" utility function could be in the form of not assigning a relatively high marginal utility to very low values or a relatively low marginal utility to large values. These relationships are thus quite restrictive and leads to challenges for use in practice. In optimization problems, they may even make the problem infeasible.

This leads to various paradoxes in decision making and the introduction of almost SD (ASD) by Leshno and Levy (2002). ASD resolves this difficulty by giving relaxations of traditional stochastic dominance relationships. Several notions of ASD that are more flexible than conventional SD have been proposed. The idea is to choose a strict subset of the corresponding utility set of conventional SD that gives rise to a weaker stochastic dominance relation but that only includes realistic utility functions. ASD is an ordering that reveals the preferences among most decision makers so that the extreme cases of decision makers, which are theoretically possible but rarely observed in practice, could be left out.

Even though ASD possesses several advantages over traditional SD, it still does not reveal the preferences observed in the St. Petersburg paradox. Tan (2015) generalizes the conditions of ASD by introducing a weight function to address this issue. The proposed the necessary and sufficient conditions for WASD imply the unanimous preference by all individuals with non-decreasing utility functions whose marginal utilities are bounded by a reference marginal utility.

It can be noted that along the development of SD, its definition has been tailored to be more practical. Meanwhile, we note that individuals have been observed to be risk-averse under multiple settings (Brenner, 2015; Rieger et al., 2015). People may wish to check for unanimous preference across all non-extreme and non-decreasing concave utility functions, rather than across all non-extreme and non-decreasing utility functions, when screening a set of feasible acts. Therefore, the prevalence of risk-aversion among decision makers makes considering concavity restrictions in WASD an interesting topic.

The concept of SD can also be applied between specific common probability distributions. The normal distribution is common and well-known. In particular, the central limit theorem states that the arithmetic mean of a large enough number of independently and identically distributed random variables will be approximately normally distributed, regardless of the underlying distribution of random variables. This attractive property makes the normal distribution important and useful. For instance, the connection between SD and the normal distribution gives insights in the mean-variance efficient frontier analysis. The log-normal distribution is popular in describing natural phenomena as well. In particular, many natural growth processes are in the form of the accumulation of multiple small percentage changes, which become additive on a log scale. For example, in the world of finance, the log-normal property of stock prices is among the assumptions of the famous Black–Scholes model. Therefore, linking SD with these specific distributions has practical advantages.

Since SD plays an important role in the decision analysis literature, it is highly considered in the stochastic optimization models regarding decision making and this connection has already been made. For example, we could employ a parametric representation of the set of utility functions used to define the dominance constraint (Dentcheva and Ruszczyński, 2003; Hu et al., 2012; Haskell and Jain, 2013). SD has appeared in the dominance constraints of convex optimization problems. Multi-variate SD constraints are developed along a similar fashion as well. Under this parametric representation, increasing concave utility functions show up in the dual problems as the Lagrange multipliers of SD constraints. From another perspective, representations of SD constraints based on linear and integer programming are developed in which the entire family of increasing concave functions is represented (Luedtke, 2008; Armbruster and Luedtke, 2015).

Rather than comparing random variables, one might be interested in evaluating each random variable independently. This could be done via risk measures which can take into account the risk preferences of the decision maker. We propose that the following four considerations be taken into account for a general risk measure optimization framework:

Multivariate prospects. A general risk-aware optimization framework should be able to handle multiple criteria. Many key problems in stochastic optimization have multiple criteria. For instance, a portfolio in the financial market whose assets cannot be aggregated may be deemed a multivariate case. An effective risk management paradigm must be able to handle multivariate random prospects.

Preference uncertainty. Risk-averse preference is of great interest due to its prevalence across decision makers. However, risk-seeking preferences (or a mixture of them) have also been observed. Even for risk-averse decision makers, the extent of their aversion towards risk also differs. In particular, it is difficult for a decision maker to precisely express his risk preferences. Hence, a practical risk management scheme must be robust against uncertainty in risk preferences.

Computational tractability. This is a major consideration in the related literature on robust optimization as many problems in practice involve a large number of variables and constraints. Therefore a risk-aware framework should be computationally tractable. Ideally, it should be evaluated with convex optimization techniques.

Target-oriented measure. In real life, a decision maker may evaluate the fitness of his decision with respect to a target or goal. Under this context, risk may be interpreted as failure to meet the desired targets or goals. A comprehensive optimization framework should be able to take targets into account.

Next, we discuss the main research objectives.

1.2 Research objectives

While there is a rich literature on stochastic dominance relationships as well as optimization models based on it, there remain areas of improvements. Specifically, we address the following questions:

1. Is there a SD relationship for which the decision maker is known to be risk-averse and whose utility is approximately known? We aim to reveal the unanimous preference between prospects by all risk-averse decision makers whose utility functions are unknown but not deviating too much from a reference utility (e.g. constant relative risk aversion utility). As mentioned, risk aversion is often observed in practice. Although the utility of the decision maker is unknown, it could often be approximated via a series of lottery comparisons. Surprisingly, SD for such settings is rarely discussed.

2. Can we develop an optimization framework that introduces a set of constraints for this new proposed SD relationship? This model could provide a decision making scheme for all risk-averse decision makers whose utility functions are unknown but remain in some cer-

tain range. Specifically, under reasonable assumptions, this framework is computationally tractable by employing convex optimization techniques with a non-parametric representation of dominance constraints.

3. How can we apply the concept of ASD to specific probability distributions, e.g. the log-normal distribution? Understanding the conditions for ASD between specific distributions allows us to study the preferences of individuals in settings where these distributions are valid. For example, there is both theoretic and empirical evidence that, under a sufficiently long horizon, investment returns are approximately log-normal distributed.

4. Based on the optimization framework of SD relationship, can we generalize it to one that considers the four considerations mentioned in the previous section? With this model, the value of the target-oriented risk measure given multivariate prospects by decision makers who have convex risk measures, which has multiple applications in practice, could be determined.

As the decision making process is becoming more sophisticated, a broad range of requirements should be taken into account. The purpose of this thesis is to propose a new SD condition which can guide decision making as well as improve existing optimization formulations. We then apply the concept of ASD to specific probability distributions. Lastly, we provide a unifying framework that satisfies all four of our considerations just outlined.

We combine the work on different perspectives of optimization and SD. Previously the two different bodies of literature were, for the most part, restricted to different communities. However, we feel that there exists a natural link between them and a combination of these two fields is promising. Although the study here is theoretically oriented, these models indeed provide some significant insights in real life, as we will illustrate in this thesis. For instance, we are interested in a portfolio optimization problem where we aim to maximize the expected return such that the portfolio stochastically dominates a given benchmark.

After the work in this dissertation, the following contributions can be summarized.

1. We have proposed a new SD relationship in which utility functions are weighted against a reference utility for risk-averse decision makers and an optimization framework that introduces a set of constraints of the above mentioned SD relationship. In particular, new SD relationship addresses the gap of SD for risk-averse decision makers. The optimization framework here could address the limitation in conventional SD optimization approaches when the preference information is incomplete. In this way, we could avoid to make decision makers to evaluating a series of pairwise comparisons between lotteries mentioned in Armbruster and Delage (2015). We use a utility range to describe preference that does not deviate too much from a popular pattern.

2. We have extended the concept of ASD to log-normal probability distributions and applied the results to mean-variance analysis and MGM strategy. We note that though MGM portfolio could maximize expected return, but does not maximize expected utility across all nondecreasing utility functions, even in the long run. Levy (2016) attempted to address this concern but under some conditions. By applying the concept of ASD to log-normal probability distributions, we could relax some of these conditions to make it more reasonable.

3. We provided a more general optimization framework that considers the four considerations we have mentioned. This framework deals

with the multivariate case and target-based measure that are not considered in Armbruster and Delage (2015). It also addresses the computational tractability that is not focused in Brown and Sim (2009) and Brown et al. (2012). Specifically, this is done by subgradient characterization and acceptance set approach.

The remainder of this thesis is organized as follows. Chapter 2 presents the notation and some critical concepts that are used in this dissertation. Chapter 3 provides a comprehensive literature review of related previous work. Chapter 4 proposes a new stochastic dominance relationship and the corresponding optimization model. Chapter 5 discusses the application of the concept of ASD to some well-known probability distributions, namely the log-normal distribution. Chapter 6 proposes a unifying optimization framework with good properties. Chapter 7 concludes the whole thesis and highlights possible topics for future research.

Chapter 2

Preliminary

In this chapter, we introduce basic notation and assumptions, as well as some critical concepts and definitions that are used in this dissertation. We list them here to make the dissertation self-contained and convenient to check.

2.1 Notation and assumptions

Let Ω denote a sample space with elements $\omega \in \Omega$, $\mathcal{B}$ a Borel σ-algebra defined on Ω, and P a probability measure defined on $(\Omega, \mathcal{B})$. Hence we have introduced a probability space $(\Omega, \mathcal{B}, P)$.

Next, we define $\mathcal{L}^n$ to be the linear space of essentially bounded $\mathcal{B}$-measurable mappings $X : \Omega \to \mathbb{R}^n$ for $n \geq 1$. When $n = 1$, it corresponds to the univariate random variable case where we write it as $\mathcal{L}$. When $n \geq 2$, we write $\mathcal{L}^n$ to denote the space of multivariate prospects. For any $X, Y \in \mathcal{L}^n$, we write $X \leq Y$ when $X(\omega) \leq Y(\omega)$ for all $\omega \in \Omega$.

Let $\mathcal{Z} \subset \mathbb{R}^m$ be a decision set determined by decision makers. Let $G : \mathcal{Z} \to \mathcal{L}^n$ be a random-variable-valued mapping with realizations denoted by $[G(z)](\omega)$ for all $\omega \in \Omega$. Therefore, the mapping $G(z)$ inherits the randomness in the underlying decision-making problem and is a random variable. We introduce a random variable $Y \in \mathcal{L}^n$ to be the benchmark of

$G(z)$.

For any compact set $\mathcal{X} \subset \mathbb{R}$, let $\mathcal{C}(\mathcal{X})$ be the space of continuous functions on $\mathcal{X}$ in the supremum norm,

$$\|u\|_{\mathcal{C}(\mathcal{X})} \triangleq \sup_{x \in \mathcal{X}} |u(x)|.$$

In addition, we let $\mathcal{C}^1(\mathcal{X})$ denote the set of continuously differentiable functions in $\mathcal{C}(\mathcal{X})$ and $\mathcal{C}^2(\mathcal{X})$ denote the set of continuously second-order differentiable functions in $\mathcal{C}(\mathcal{X})$.

Let $f \in \mathcal{C}(\mathcal{Z})$ be a deterministic objective function on the decision set and $m : \mathcal{X} \to \mathbb{R}$ be a nonnegative weight function which bounds the marginal utilities of risk-averse decision makers.

Let H be a closed convex cone in $\mathbb{R}^n$ such that $\mathbb{R}^n_+ \subset H$ and $H \neq \mathbb{R}^n$ with an induced partial ordering by $X \succeq 0$ if and only if $X \in H$ almost surely.

We now make some assumptions in our decision-making optimization problem for technical convenience.

A1 $\mathcal{Z}$ is closed and convex.

A2 $G : \mathcal{Z} \to \mathcal{L}^n$ is convex.

A3 f is concave.

Therefore, $[G(z)](\omega) : \mathcal{Z} \to \mathbb{R}^n$ is convex in $z \in \mathcal{Z}$ for P-almost all $\omega \in \Omega$. This key convexity assumption ensures that convex optimization techniques could be used. We want to choose $z \in \mathcal{Z}$ so that $G(z)$ has a favorable distribution in some sense. To evaluate the distribution of $G(z)$, we use mappings $\rho : \mathcal{L}^n \to \mathbb{R}$ which we refer to as "risk functions" in line with Ruszczyński and Shapiro (2006).

We make the following key assumption on the underlying sample space:

A4 Ω is finite. Random variables on $(\Omega, \mathcal{B}, P)$ have bounded and finite support in the interval $\mathcal{X} \triangleq [x_{\min}, x_{\max}]$.

For later use, we introduce a discretization $\Theta \triangleq \{\theta_1, \ldots, \theta_K\} \subset \mathcal{X}$ of $\mathcal{X}$ where $\theta_1 < \theta_2 < \cdots < \theta_K$, $\theta_1 = x_{\min}$, and $\theta_K = x_{\max}$. Note that the length of the longest subinterval of Θ, $\max_{i=2,\ldots,K} (\theta_i - \theta_{i-1})$, is a measure of the granularity of Θ. We make an additional technical assumption regarding the support of the benchmark Y:

A5 $\operatorname{supp} Y$ is finite and $\operatorname{supp} Y \subset \Theta$.

We can meet this assumption by construction since the benchmark Y is a user input. The discretization Θ is also a user input, and it can be constructed a priori to include the support of Y.

We will make further assumptions on the weight function m.

A6 m is Lipschitz continuous with constant L.

This is a relatively strong form of function uniform continuity. Imagine a double cone with slopes of L and $-L$. When the vertex moves along the function curve, the curve will always remain entirely outside the cone.

A7 Cone H that defines the acceptance set $\mathcal{A}$ (containing $\mathcal{L}^n(H)$) possesses the substitutability property:

$$-e_i + \alpha e_1 \in H,\ e_i - \beta e_1 \in H,\ \alpha,\ \beta > 0,\ i = 2, ..., n,$$

where e_i denotes the unit vector with i^{th} component 1.

This means that any entry other than the one in the first position, can be substituted or compensated by some position in the first entry. More precisely, it states that the unitary prices of the assets for $i \geq 2$ in terms of the first asset must be bounded. We give an example of H in two-dimension Euclidean space as following.

Example 2.1.1. Let $H = \{(x_1, x_2) \in \mathbb{R}^2 : 5x_1 + x_2 \geq 0, \frac{1}{5}x_1 + x_2 \geq 0\}$. Then the linear space of essentially bounded H-valued $\mathcal{B}$-measurable random vectors $\mathcal{L}^n(H)$ can be used to define the acceptance set $\mathcal{A}$ in two-dimension case.

2.2 Definitions of concepts

In this section, we will start with stochastic dominance and its related concepts. This is the key idea in revealing the dominance relationship between prospects. Definitions as well as the corresponding sufficient and necessary conditions will be discussed. Different types of stochastic dominance are associated with different categories of utility functions of decision makers. We could apply these concepts to different decision makers accordingly. In addition, we will also introduce some other concepts that used in the optimization frameworks and the application of stochastic dominance, such as robust certainty equivalent, maximum geometric mean (MGM) portfolio, etc.

2.2.1 Stochastic dominance

Let F_X and F_Y denote the cumulative distribution function (*cdf*) of random variables X and Y, respectively. Let S_1 denote the set of outcomes where $F_X > F_Y$ (i.e., $S_1 = \{t : F_X(t) > F_Y(t)\}$).

We denote a collection of utility functions $U \subset \mathcal{C}(\mathcal{X})$ the set of all non-decreasing utility functions, and $U_{icv} \subset \mathcal{C}(\mathcal{X})$ the set of all non-decreasing convex utility functions.

The theory of SD was developed by Hadar and Russell (1969), Hanoch and Levy (1969), Rothschild and Stiglitz (1970). They provide the definition as well as the criteria rules and proof. We begin with the following definition of conventional SD:

Definition 2.2.1 (Stochastic dominance)**.** *For $X, Y \in \mathcal{L}$, X stochastically dominates Y, denoted by $X \succeq_U Y$, if and only if $\mathbb{E}[u(X)] \geq \mathbb{E}[u(Y)]$ for all u in U.*

Conventional SD is defined via the expected utility theory here. We know that the set of utility functions such that X is clearly preferred over Y is set of all non-decreasing utility functions. We can also conclude the dominance relationship from another perspective, which leads to the introduction of its sufficient and necessary condition.

It can be showed that X stochastically dominates Y if and only if $F_X(t) \leq F_Y(t) \, \forall t$, and there exists a t_0 such that the strict inequality $F_X(t_0) < F_Y(t_0)$ holds. In particular, it can be noted that the *cdf* curve of X cannot go above that of Y everywhere. The conditions from both the expected utility theory and the *cdf* relationship are equivalent in showing the dominance relationship between X and Y.

We then introduce the following defition of ASD:

Definition 2.2.2 (Almost stochastic dominance)**.** *For $X, Y \in \mathcal{L}$ and $0 < \varepsilon < 0.5$, X dominates Y with ε-ASD, denoted by $X \succeq_{U^*(\varepsilon)} Y$, if and only if $\mathbb{E}[u(X)] \geq \mathbb{E}[u(Y)]$ for all u in $U^*(\varepsilon)$, where:*

$$U^*(\varepsilon) = \left\{ u \in U \cap \mathcal{C}^1(\mathcal{X}) : u'(t) \leq \inf\{u'(t)\} \left[\frac{1}{\varepsilon} - 1\right] \forall t \right\}.$$

i.e. when the expected utility of X is greater than or equal to the expected utility of Y for all $u \in U^*(\varepsilon)$.

In particular, $U^*(\varepsilon)$ is the set of all differentiable nondecreasing utility functions whose marginal utility deviates by a maximum factor of $\left[\frac{1}{\varepsilon} - 1\right]$. For example, all utility functions whose marginal utility is no greater than 2 and no lesser than 0.5 are contained in $U^*(0.2)$.

Leshno and Levy (2002) showed that for $0 < \varepsilon < 0.5$, X dominates Y

with ε-ASD if and only if:

$$\varepsilon \geq \frac{\int_{S_1} [F_X(t) - F_Y(t)]dt}{\int_{-\infty}^{\infty} |F_X(t) - F_Y(t)|dt}. \tag{2.1}$$

In particular, note that the denominator in Equation (2.1) corresponds to the total area between F_X and F_Y while the numerator in Equation (2.1) corresponds to the area between F_X and F_Y where $F_X > F_Y$. Comparing to the conventional SD case where the *cdf* curve of X cannot go above that of Y everywhere, it is allowed to violate this rule if the violation is within some predetermined level. As shown in Figure 2.1, area A_1 is the violation area which is the numerator in Equation (2.1), and $A_1 + A_2$ is the total area which is the denominator. Therefore, X dominates Y with ε-ASD if and only if the ratio of the two areas is less than the predetermined level ε.

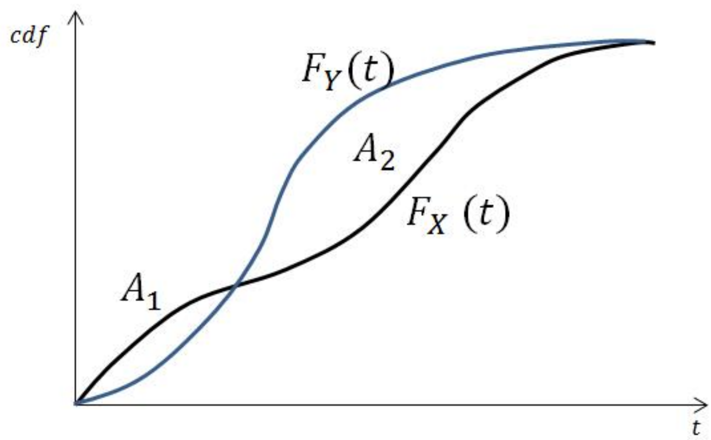

Fig. 2.1. Almost stochastic dominance.

Definition 2.2.3 (Weighted almost stochastic dominance)**.** *For $X, Y \in \mathcal{L}$ and $0 < \varepsilon < 0.5$, X dominates Y with (m, ε)-WASD, denoted by $X \succeq_{U^*(m,\varepsilon)}$*

Y, if and only if $\mathbb{E}[u(X)] \geq \mathbb{E}[u(Y)]$ *for all u in $U^*(m,\varepsilon)$, where:*

$$U^*(m,\varepsilon) = \Big\{ u \in U \cap \mathcal{C}^1(\mathcal{X}) : \\ \left[\frac{1}{\varepsilon} - 1\right]^{-0.5} m(t) \leq u'(t) \leq \left[\frac{1}{\varepsilon} - 1\right]^{0.5} m(t), \forall t \Big\}.$$

$U^*(m,\varepsilon)$ is the set of all differentiable nondecreasing utility functions whose marginal utility differs from $m(t)$ by a maximum factor of $\left[\frac{1}{\varepsilon} - 1\right]^{0.5}$. In particular, $m(t)$ denotes a nonnegative function that describes the marginal utility of some canonical utility function, which $U^*(m,\varepsilon)$ is constructed around.

Tan (2015) showed that for $0 < \varepsilon < 0.5$, X dominates Y with (m,ε)-WASD if and only if:

$$\varepsilon \geq \frac{\int_{S_1} m(t)[F_X(t) - F_Y(t)]dt}{\int_{-\infty}^{\infty} m(t)|F_X(t) - F_Y(t)|dt}. \tag{2.2}$$

Equation (2.2) is very similar to Equation (2.1) other than a weight function is required in the integral. Note that WASD relationship becomes ASD if the weight function is a constant. Thus ASD is a special case of WASD.

2.2.2 Other concepts

Armbruster and Delage (2015) proposes the concept of robust certainty equivalent which removes the need of a benchmark by which the desired prospect depending on our decisions should dominate.

The certainty equivalent is the amount for sure such that one would be indifferent between it and the random prospect:

Definition 2.2.4 (Robust Certainty Equivalent (RCE))**.** *The Robust Certainty Equivalent of a prospect X is:*

$$\mathbb{C}_u[X] \triangleq \sup\{s : u(s) \leq \mathbb{E}[u(X)]\}, \forall u \in \mathcal{U}.$$

This can be formulated as an RCE maximization problem which is tractable.

Maximum geometric mean (MGM) strategy has been proposed for investors with a sufficiently long horizon (see Kelly (1956), Breiman (1960), and Markowitz (1976)). It aims for maximal terminal wealth by investing in each period based on the logarithm of returns.

Definition 2.2.5 (Maximum Geometric Mean (MGM) Portfolio)**.** *For a long investment horizon with multi-period, MGM portfolio is a portfolio investing in each period aiming to maximize*

$$\mathbb{E}[log(1 + R_t)],$$

where R_t represents the one-period portfolio rate of return. R_t's are independently and identically distributed and have finite mean and variance.

In particular, it follows from the law of large numbers that a MGM portfolio will almost surely outperform other portfolios in the long run under mild conditions.

Chapter 3

Literature review

This chapter presents a survey of literature pertinent to studies on stochastic dominance (SD) as well as risk measures. Related optimization frameworks, especially classical linear programming, will also be discussed. In particular, we look into how the concept of SD and risk measures can be adopted within various optimization frameworks.

3.1 Stochastic dominance (SD)

In this section, we start with the development of stochastic dominance (SD). In particular, we review how the concept of SD has evolved. In addition to the review on the development of SD, previous studies that have attempted to develop SD-related optimization framework are also discussed.

3.1.1 Development of SD

As more complicated and practical needs are considered in decision making, the concept of SD has evolved along the way. The development of SD can be broadly broken down as follows: Conventional SD and Almost SD.

Conventional SD

In 1932, Karamata proved a theorem, i.e. Karamata's inequality, which is very similar to second-order SD (SSD). Then the concept of SD was introduced in mathematics by Mann and Whitney (1947) and Lehmann (1955). Mann and Whitney gave a test of whether a random variable is stochastically larger than another. A statistic depending on the relative ranks of the two random variables is proposed for testing the hypothesis that the continuous cumulative distribution functions of the two are equivalent. Lehmann made a comparison of several definitions of ordered sets of distributions. These definitions attempt to make precise the intuitive notion that large values of the parameter which labels the distributions go together with large values of the random variables themselves.

Originating from the majorization theory, the theory of SD and its many theoretical and empirical extensions in economics and finance were formally developed when four papers were independently published by Hadar and Russell (1969), Hanoch and Levy (1969), Rothschild and Stiglitz (1970) and Whitmore (1970). The first three papers provide the definition of the first-order SD (FSD) and second-order SD (SSD) as well as the criteria rules and proof. The fourth paper develops those for the third-order SD (TSD). The SD rules can easily be extended to higher orders, i.e. n^{th}-order SD. Since then, hundreds of papers have been written on the topic, as highlighted in a survey paper by Levy (1992). Levy pointed out the following four main areas of development:

1. further theoretical development;
2. application of SD rules to empirical data;
3. application of SD rules to other economic and financial issues;
4. application of SD rules in statistics.

These traditional SD rules have been developed to offer, in many cases, efficient criteria on decision making in both theory and application (Levy, 1992). However, these conventional rules sometimes have limitation in failing to reveal some obvious dominance between prospects due to some extreme utility functions in the case of even a very small violation of these rules. Such examples can be seen in Leshno and Levy (2002).

Almost SD

The theory of ASD developed by Leshno and Levy (2002) plays an important role in several fields, particularly in finance. They provide a new way of imposing restrictions on the first and second derivatives of utility functions so that the preferences that do not represent most decision makers are excluded.

Since Leshno and Levy's paper, other works have further drawn several important applications. For example, Levy (2009) employs ASD to make the case for "stocks for the long run". It is shown that ASD and the geometric-mean argument do not necessarily support long-run investment in equities. Specifically, bonds may be preferred to stocks over a short horizon, but stocks are preferred in the long run. Regarding investment strategies, Bali et al. (2009) use data from the United States to show that the ASD approach unambiguously supports the popular practice that suggests a higher stock-to-bond ratio for long investment horizons. Levy et al. (2010) construct several experiments to show that the ASD rule corresponds to sets of non-pathological preferences. Bali et al. (2011) further adopt the ASD rule to examine the practice of investing in stock market anomalies. They found that the ASD rule provides evidence for "the significance of size, short-term reversal, and momentum for short investment horizons and the significance of book-to-market and long-term reversal for longer term horizons" (pp. 18). Lizyayev and Ruszczyński (2012) comment on the difficulty of introducing ASD constraints within an optimization model and provide an alternative tractable formulation.

Tzeng et al. (2013) show that the almost second degree SD introduced by Leshno and Levy (2002) does not possess the property of expected-utility maximization. They modify the definition of the ASD to achieve this property. Nonetheless, Guo et al. (2013a) construct some examples to show that the ASD definition modified by Tzeng et al. (2013) does not possess any hierarchy property and establish necessary conditions for ASD criteria of various orders in Guo et al. (2013).

The advantages of ASD over SD and the mean-variance rule are as follows:

1. ASD is able to rank otherwise unrankable alternatives.

2. ASD can eliminate alternatives that are considered to be inferior by most investors.

3. ASD sheds light on the debate related to optimal portfolio composition and the planned investor horizon. It is possible to establish a functional relationship between the percentage of equity in the portfolio and the planned investors horizon. Namely, ASD may be employed by financial advisors in choosing portfolios for young versus old investors.

However, Tan (2015) points out that ASD fails to reveal the preference relationships observed in St. Petersburg paradox. As a result, the related concept of weighted almost stochastic dominance (WASD) has been proposed. Specifically, a reference function is introduced and the set of utility functions whose marginal utility differs from the reference marginal function by a maximum factor is considered.

Tsetlin et al. (2015) develop generalized almost stochastic dominance which combines different rules of almost stochastic dominance into one framework. This concept addresses the implementation issues and inconsistencies between integral conditions and their associated utility classes of different rules of ASD.

Risk preference among decision makers

Risk preference is the tendency to choose a risky or less risky position. Different decision makers may have different risk attitudes. There are three basic types of risk preferences: risk-averse, risk seeking and risk-neutral.

Risk-averse decision makers dislike risk. They will stay away from adding high-risk investments to portfolios and, in turn, will often forfeit higher return. Such decision makers are generally characterized by concave utility functions. Risk-seeking decision makers, in contrast, are generally characterized by convex utility functions. Theoretical works linking SD theory to the selection rules for risk-averse and risk-seeking decision makers under different restrictions on the utility functions has also been well investigated. Quirk and Saposnik (1962), Hanoch and Levy (1969) develop the theory of SD related to economics and obtained SD rules for risk-averse decision makers. To distinguish SD theory for risk-averse decision makers from that for risk-seeking decision makers, they call this ascending stochastic dominance because the cumulative distribution function is integrated in ascending order from the leftmost point of downside risk.

On the other hand, Hammond (1974), Li and Wong (1999), and Wong (2007) develop the stochastic dominance rules for risk-seeking decision makers, which they call descending stochastic dominance because the cumulative distribution function is integrated in descending order from the rightmost point of upside profit.

The application of ASD to risk-averse and risk seeking decision makers has been discussed. The theory of second-order ASD developed by Leshno and Levy (2002) can be considered as the case for risk-averse decision makers. ASD relationships for risk-seeking decision makers to the first three orders are developed by Guo et al. (2013b).

Among different risk preferences, risk-aversion is of great interest due to its prevalence across decision makers. Eckel and Grossman (2008) show that decision makers are generally risk-averse via abstract gamble experiments and contextual environment experiments. In particular, their study suggests that women generally have greater risk aversion then men. Moreover, these results from both kinds of experiments are consistent with field studies. Rieger et al. (2015) conduct an extensive international survey on risk preferences which reveals an attitude of risk aversion in gains and risk seeking in losses on average. Furthermore, they highlight that the degree of risk attitudes is affected by different factors such as economic conditions and cultural factors. Brenner (2015) proposes a subjective option valuation model to probe the risk preferences of U.S. executives. The author concludes that the observed behavior among 7000 U.S. executives is basically consistent with moderate relative risk aversion.

In conclusion, risk aversion is prevalent among most decision makers. Hence, it is interesting to consider risk aversion in weighted almost stochastic dominance relationships. Moreover, we expect that we could control the degree of risk aversion in this new SD relationship so that it better describes the behavior of decision makers.

3.1.2 Optimization framework of ASD

The connection between stochastic dominance relationships and optimization problems has been broadly discussed. One possible way is to form the stochastic dominance relationships in the constraints.

We start with the univariate case. Dentcheva and Ruszczyński (2003) introduce SD constraints into stochastic optimization problems and develop necessary and sufficient conditions of optimality and duality theory for their proposed optimization framework. They also show that Lagrange multipliers regarding dominance constraints are concave non-decreasing utility functions, which is consistent with risk-averse preference. Again Dentcheva and Ruszczyński (2004) introduce SD relationships with second order into stochastic optimization problems which leads to nonlinear constraints. In this case, they develop a new splitting approach and provide the corresponding optimality and duality theory.

The time horizon is another factor that could be of interest. For example, Dentcheva and Ruszczyński (2008) are interested in a finite horizon stochastic programming problem which improves the model by introducing a random reward sequence into the constraints. The dominance in the constraints is defined by discounting two processes: a family of discount sequences, and by applying a univariate order. Optimality conditions are also provided. Haskell and Jain (2013) formulate stochastic dominance constraints for infinite horizon discrete time Markov decision processes (MDPs). They use a linear programming model to obtain the optimal policy and compute the dual of this linear program to obtain average dynamic programming optimality equations that reflect the dominance constraints.

This work has also been extended to the multivariate case. Dentcheva and Ruszczyński (2009) extend the optimization model they proposed in Dentcheva and Ruszczyński (2003) to the multivariate case. They identify a suitable multivariate stochastic order and describe its generator in terms of utility functions. With the assumption of convexity, they reveal that the Lagrange multipliers regarding dominance constraints are elements of the generator.

Haskell et al. (2013) introduce stochastic order constraints defined in terms of parametrized families of increasing concave functions. They show that utility functions behave as the Lagrange multipliers of the corresponding constraints, and that the dual problem is a search over utility functions.

Homem-de Mello and Mehrotra (2009) introduce the cut generation problem. In particular, they discuss linear optimization problems with a newly introduced concept of multidimensional polyhedral linear second-order stochastic dominance constraints and propose a cutting-surface algorithm to deal with this problem.

Hu et al. (2012) study optimization problems with multivariate stochastic dominance constraints of second order. They apply the Sample Average Approximation (SAA) method to this problem which is a semi-infinite program and develop a finitely convergent method to find an ϵ-optimal solution of the SAA problem.

It can be noted that the above mentioned works implement a parametric representation of the utility functions used to reflect the dominance constraints. From a different perspective, a non-parametric representation of the utility functions can be employed. Luedtke (2008) suggests new integer and linear programming formulations for optimization under first- and second-order stochastic dominance constraints. They also present a specialized branching strategy and heuristics. Armbruster and Luedtke (2015) proposes a constraint using a new version of multivariate stochastic dominance which connects to expected utility maximization theory and is relatively tractable. The good thing here is that such a constraint can be formulated with linear constraints for second-order dominance relations and with mixed-integer constraints for first-order relations. Haskell et al. (2014) investigate the optimization problem with an infinite number of constraints indexed by a function space of non-decreasing concave utility functions. They focus on effective numerical methods of SAA formulation and Lagrangian duality theory.

In the above mentioned previous studies, one issue that can be noted is that the optimization frameworks generally require a benchmark in order to introduce the stochastic dominance constraints (i.e. the desired prospect depending on our decisions should stochastically dominate the given benchmark). However, the choice of a benchmark is not trivial to make and remains an interesting topic to explore. So far we do not have reasonable guidance in constructing a benchmark, in particular, for different risk preferences. The robust certainty equivalent firstly introduced in Armbruster and Delage (2015) is an alternative, which is a different concept from Ben-Tal and Teboulle (2007).

Another issue is that there is limited work in formulating almost stochastic dominance constraints, or more generally weighted almost stochastic dominance, in optimization problems. Therefore, it is interesting to formulate almost, as well as, weighted almost stochastic dominance constraints in our optimization framework.

3.2 Risk measures

In this section, we firstly review the previous work on risk measure with some good properties (e.g. coherent risk measure). Specifically, we are interested in scalar-valued and vector-valued risk measures. In addition to the review on risk measures, previous studies that aim to develop the optimization framework for general risk measures are also discussed. In particular, we focus on studies related to our four main considerations respectively.

3.2.1 Scalar-valued risk measures

We start with the review on scalar-valued risk measure functions. Risk measures in this category return a real scalar to reflect the degree of risk. Artzner et al. (1999) is a pioneering paper in coherent risk measure which discusses methods of measurement of market and non-market risks. In particular they determine a set of four desirable properties and refer to the measures satisfying these properties "coherent". Follmer and Schied (2002) propose the concept of a convex risk measure which is an extension of the idea in Artzner et al. (1999). They prove a corresponding extension of the representation theorem in terms of probability measures on the underlying space of scenarios. Ruszczyński and Shapiro (2006) introduce convex risk functions in optimization problems. Specifically they develop new representation theorems for risk models, and optimality and duality theory for problems with convex risk functions with convex analysis and optimization theory.

3.2.2 Vector-valued risk measures

When dealing with multivariate prospects it is sometimes more natural to use vector-valued risk functions. Risk functions are extended to the vector-valued setting in Jouini et al. (2004) and Burgert and Rschendorf (2006). Jouini et al. (2004) defines coherent risk measures as set-valued maps from $\mathcal{L}_n^\infty$ into $\mathbb{R}^d$ satisfying some axioms. They also discuss the aggregation issue as well as necessary and sufficient conditions of coherent aggregation. Burgert and Rschendorf (2006) introduce convex risk measures for portfolio vectors defined axiomatically. They further illustrate two natural classes of examples of risk measures for portfolio vectors which are easy to interpret and investigate their corresponding properties. Cascos and Molchanov (2007) develop risk functions that take values in abstract cones, which include the classical risk measures and set-valued risk measures and obtains a natural definition of vector-valued risk measures. They also demonstrate that the idea of depth-trimmed regions from multivariate statistics is closely associated with the definition of risk measures. Hamel and Heyde (2010) define set-valued convex measures of risk as well as the corresponding acceptance sets. They also show their dual representation theorems. Set-valued measures of risk are also provided based on primal and dual descriptions. Ararat et al. (2014) shed light on multi-asset financial market with frictions. The utility-based risk of a financial position with such conditions can be quantified by set-valued risk measures, and market frictions are modeled by convex random solvency regions representing proportional transaction costs or illiquidity effects. Molchanov and Cascos (2016) consider risky positions in multivariate portfolios and give a constructive approach for vector-valued risk functions where a set-valued portfolio is acceptable if it possesses a selection where all individual marginals are acceptable. The obtained risk measure possesses the good properties of coherency, law invariance and having values being upper convex closed sets. The dual representation was also provided.

3.3 Optimization of general risk measures

It is not always suitable to guarantee the reliability of a decision when we maximize expected performance in stochastic optimization. As a result, a number of risk measures, which generalize expected performance, have been proposed to give the decision maker flexibility in expressing his risk preferences. However, it is still challenging to do risk-aware optimization in practice from both a modeling and a computational perspective. In this section, previous studies pertaining to the four considerations (i.e. multivariate prospects, preference uncertainty, computational tractability, target-oriented measure) mentioned in the introduction will be reviewed.

3.3.1 Multivariate prospects

A general risk-aware optimization framework must be able to handle multiple criteria. Many key problems in stochastic optimization have multiple criteria. Thus, an effective risk management paradigm must be able to handle multivariate random prospects.

Gutjahr and Pichler (2013) survey many key stochastic multi-objective optimization problems and their solution techniques, and focuses on the analysis of decision-making problems that simultaneously consider multiple objectives and stochastically represented uncertainty. It is argued that many key problems are naturally multi-objective. In Liefooghe et al. (2007), a multi-objective formulation of the flow-shop scheduling problem subjected to a wide range of uncertainties is proposed. Several multi-objective methods that are able to handle any type of probability distribution are also discussed. Chen et al. (2010) consider three stochastic multi-objective models for designing transportation network under demand uncertainty, and show how to compute Pareto optimal solutions that explicitly optimize all objectives under demand uncertainty by simultaneously generating a family of optimal solutions.

It can be noted that the combination of multiple decision criteria and uncertainty is of great interest for its practical usefulness. Many financial positions, for instance, a portfolio whose assets cannot be aggregated, fall into the multivariate case.

3.3.2 Preference uncertainty

It is difficult for a decision maker to precisely express his risk preferences. In response, a practical risk management system must be robust against ambiguity in risk preferences. The case of expected utility maximization has received major attention.

In Armbruster and Delage (2015), ambiguity in risk preferences is considered in expected utility maximization. They propose finding a solution that is robust to a set of possible utility functions obtained by preference elicitation. In particular, they showed that the worst-case utility can be expressed as the maximum of a reasonable sized linear program. In Delage and Li (2015), the discussion is extended to ambiguity in risk preferences over general risk functions. They seek financial positions that perform best given the worst-case of the risk measure potentially perceived by the decision maker and show how this robust risk minimization problem can be solved numerically by formulating convex optimization problems in a tractable way.

The assumption of incomplete preference information is necessary. Although it has been highlighted that risk-aversion is prevalent in people's behavior, risk-seeking behavior or a mixture of the two has also been observed (Kahneman and Tversky, 1979). Even for risk-averse decision makers, the extent to which they are risk-averse also differs.

3.3.3 Computational tractability

Some research work has been done focusing on the computational tractability of robust optimization, which is a topic of major concern in the literature. Goh and Sim (2010) deals with a linear programming problem with uncertainties, which has expected values both in the objective and constraints. They obtain an approximate solution to the problem that is distributionally robust which is more flexible than using linear decision rules. In Wiesemann et al. (2014), distributionally robust optimization is studied where the true probability distribution lies in standardized ambiguity sets.

In particular, we are interested in risk-aware frameworks that can be evaluated with convex optimization techniques. For instance, the risk-aware formulations in Armbruster and Delage (2015) and Delage and Li (2015) can be cast as convex optimization problems on finite probability spaces. In Haskell et al. (2014), the technique in Armbruster and Delage (2015) is combined with sample average approximation and extended to the multivariate setting on general probability spaces.

3.3.4 Target-oriented measure

In real life, it is often the case to make a specific amount of monetary reward and to interpret risk in terms of loss to meet this target. Therefore, a decision maker often evaluates the fitness of his decision with respect to a target or goal. A comprehensive optimization framework should ideally be able to take such targets into account.

Brown and Sim (2009) firstly propose satisficing measures to quantify risk of financial positions based on the ability to achieve financial targets. Brown et al. (2012) continue this work and develop a general class of aspirational risk preferences by showing that these preferences share a representation in terms of targets. It can be noted that a target-based risk measure has the advantage of easy interpretation and is more natural to specify than risk tolerance parameters.

Based on the above mentioned four considerations, we are interested in the formulations that can be evaluated with convex optimization techniques. Basically, the approaches that deal with risk measure can be categorized into two types. One type is by subgradient characterization, as shown in Armbruster and Delage (2015), Delage and Li (2015) and Haskell et al. (2014). The other one is the acceptance approach. The previous related work are Jouini et al. (2004), Burgert and Rschendorf (2006), Hamel et al. (2011), and Molchanov and Cascos (2016).

The review above has revealed that many aspects are discussed in terms of the optimization framework. However, a unified framework of almost stochastic dominance that satisfies all four considerations is still not well studied. Furthermore, the algorithms with respect to the framework are also not yet deeply investigated.

Chapter 4

Weighted risk-averse almost stochastic dominance and its optimization

4.1 Introduction

As reviewed in Section 3.1, the theory of stochastic dominance has been investigated extensively and deeply. The introduction of almost and weighted almost stochastic dominance provides more flexibility in decisions rules as well as more reasonable restrictions on the set of utility functions considered. In particular, ASD requires that the marginal utilities do not deviate by a maximum factor while WASD requires that the marginal utilities do not deviate from that of a reference utility by a maximum factor. WASD reduces to ASD when the reference utility is linear. However, these concepts do not consider the risk-aversion factor which, as highlighted previously, is prevalent among decision makers. Therefore, it is natural and reasonable to focus on risk-averse decision makers in WASD.

The combination of stochastic dominance relationships and optimization methods has been broadly discussed in the previous chapter. One natural way is to form the stochastic dominance relationships in the constraints. However, there is limited work in formulating ASD nor WASD constraints in optimization problems.

In this chapter, we will first propose WASD relationship under risk-averse preferences. The necessary and sufficient conditions would be provided. Next, we will formulate our proposed WASD in the constraints in our optimization framework. We will resort to linear programming and its duality as our technique. Generalized stochastic dominance relationship introduced by Tsetlin et al. (2015) can be investigated in a similar way. As noted, these is no reasonable guidance in the construction of a benchmark in SD constraints. As a remedy, an optimization framework of robust certainty equivalent maximization problem will also be discussed.

In the following, the conditions for WASD for risk-averse decision makers are defined in Section 4.2. A WASD constrained convex optimization framework for risk averse decision makers will be obtained in Section 4.3. Section 4.3.3 formulates the WASD as a robust certainty equivalent maximization problem.

4.2 Weighted risk-averse almost stochastic dominance

4.2.1 Preliminary

There are a variety of SD relationships, each accounting for a different set of utility functions. We would like to provide a flexible optimization framework that can deal with these relationships by providing a particular category of stochastic dominance relationships on $\mathcal{L}$.

There is a difficulty with using conventional SD relationships because the corresponding utility sets contain "extreme" utility functions that do not correspond to decision makers observed in practice. One example would be that most "reasonable" investors would prefer a higher proportion of stocks as the investment horizon increases (Leshno and Levy, 2002). An "extreme" utility function could be in the form of not assigning a relatively high marginal utility to very low values or a relatively low marginal utility to large values. These relationships on $\mathcal{L}$ are thus quite restrictive and lead to challenges for use in practice. In optimization problems, they may even make the problem infeasible.

ASD resolves this difficulty by giving relaxations of traditional stochastic dominance relationships. Several notions of ASD that are more flexible than conventional SD have been proposed. The idea is to choose a strict subset of the corresponding utility set of conventional SD that gives rise to a weaker stochastic dominance relation but that only includes realistic utility functions.

Recently, Tan (2015) proposed the necessary and sufficient conditions for WASD, which implies the unanimous preference by all individuals with non-decreasing utility functions whose marginal utilities are bounded by a reference marginal utility. However, it can be noted that risk-aversion is prevalent among decision makers and therefore we suggest including concavity restrictions in WASD.

4.2.2 Necessary and sufficient conditions

Here, we develop a new almost stochastic dominance relationship based on the marginal utilities of the decision maker. There are multiple reasons why our proposed condition is of interest:

1. Individuals have been observed to be risk-averse under multiple settings (Brenner, 2015; Rieger et al., 2015). Therefore, the analyst may wish to check for unanimous preference across all non-extreme and non-decreasing concave utility functions, rather than across all non-

extreme and non-decreasing utility functions, when screening a set of feasible acts. In particular, the efficient set (i.e., set of non-dominated acts) obtained by WASD may contain acts that are only preferred by risk-seeking individuals and can be further eliminated if the decision maker is known to be risk-averse. Hence, our proposed condition can reduce the size of the efficient set, simplifying the decision process.

2. From the perspective of optimization, our proposed condition has advantages over ASD and second-degree stochastic dominance (SSD). Firstly, our condition leads to convex optimization problems because it is defined in terms of non-decreasing concave functions. ASD includes non-convex functions, and thus we may not obtain convex optimization problems under ASD. Secondly, our condition is more flexible than SSD and it gives the decision maker more modeling power. Specifically, even if the decision maker sets a reasonable benchmark for performance, there may not be any feasible decisions which dominate the benchmark in SSD. However, feasible optimization problems can be obtained for reasonable benchmarks under our proposed condition by adjusting the parameter ε, when necessary.

We present a necessary and sufficient condition for unanimous preference by all rational risk-averse decision makers whose utility does not deviate too much from a reference utility. We term our condition weighted risk-averse almost stochastic dominance (WRASD).

Definition 4.2.1. *Weighted risk-averse almost stochastic dominance. Given a constant* $0 < \varepsilon < 0.5$*, we define the family of utility functions*

$$U_w(m, \varepsilon) \triangleq \Bigg\{ u \in U_{icv} \cap \mathcal{C}^1(\mathcal{X}) : \left(\frac{1}{\varepsilon} - 1\right)^{-0.5} m(x) \leq u'(x) \leq \left(\frac{1}{\varepsilon} - 1\right)^{0.5} m(x), \, \forall x \in \mathcal{X} \Bigg\}$$

We call the resulting relationship "weighted risk-averse almost stochastic dominance" (WRASD) if $\mathbb{E}\left[u\left(X\right)\right] \geq \mathbb{E}\left[u\left(Y\right)\right]$ *for all* $u \in U_w\left(m,\, \varepsilon\right)$*.*

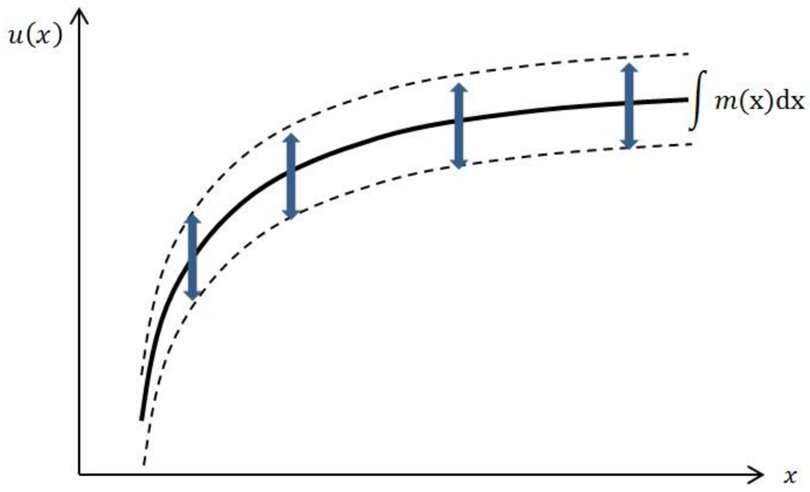

Fig. 4.1. Illustration of set $U_w\left(m,\, \varepsilon\right)$.

The set $U_w\left(m,\, \varepsilon\right)$ is the set of all differentiable nondecreasing concave utility functions whose marginal utility differs from $m(t)$ by a maximum factor of $\left[\frac{1}{\varepsilon} - 1\right]^{0.5}$, as shown in Figure 4.1. It provides greater flexibility in expressing risk preferences than U_{icv} through user control over the weight function m and the parameter ε. From the definition, we see that functions in $U_w\left(m,\, \varepsilon\right)$ are bounded from above and below:

$$\left(\frac{1}{\varepsilon} - 1\right)^{-0.5} \int_{x_{min}}^{x} m\left(\xi\right) d\xi \leq u\left(x\right) \leq \left(\frac{1}{\varepsilon} - 1\right)^{0.5} \int_{x_{min}}^{x} m\left(\xi\right) d\xi,\ \forall x \in \mathcal{X}.$$

Example 4.2.1. CRRA (Constant relative risk-averse) utility functions have the general form:

$$u(x) = \begin{cases} \frac{1}{1-\alpha} x^{1-\alpha} & if\ \alpha > 0,\ \alpha \neq 1, \\ \ln x & if\ \alpha = 1, \end{cases}$$

where α is the risk aversion parameter. In this case the weight function $m(x)$ has the form:

$$m(x) = x^{-\alpha},\ \alpha > 0.$$

In the following theorem, we give necessary and sufficient conditions for the preference of X over Y by WRASD to hold in terms of the cumulative distribution functions F_X and F_Y, respectively.

Theorem 4.2.1. *For $0 < \varepsilon < 0.5$, $X \succeq_{U_w(m,\varepsilon)} Y$ if and only if*

$$\begin{aligned} &\left[\left(\frac{1}{\varepsilon}-1\right)^{0.5} m\left(x_{min}\right)-\left(\frac{1}{\varepsilon}-1\right)^{-0.5} m\left(x_{max}\right)\right] \\ &\cdot \max_{x\in[x_{min},\,x_{max}]}\left\{\int_{x_{min}}^{x}\left[F_X(t)-F_Y(t)\right]dt\right\} \\ \leq &\left(\frac{1}{\varepsilon}-1\right)^{-0.5} m\left(x_{max}\right)\int_{x_{min}}^{x_{max}}\left[F_Y(t)-F_X(t)\right]dt. \end{aligned} \tag{4.1}$$

Proof. See Appendix A. □

Theorem 4.2.1 provides a condition that can be used to check for unanimous preference by all rational risk-averse decision makers whose utility does not deviate too much from a reference utility by a maximum factor $\left(\frac{1}{\varepsilon}-1\right)$. The proof of the theorem is based on the observation that the second derivative of a concave utility function is non-increasing. Hence, it is sufficient to ensure preference for all utility functions whose marginal utility at the smallest and largest element in $\mathcal{X}$, rather than across the whole outcome space, differs by a maximum factor $\left(\frac{1}{\varepsilon}-1\right)$.

It is noted that WRASD is based on the difference between the area under F_X and F_Y (i.e., $F_Y(t)-F_X(t)$) and is different from ASD, which is based on the area between F_X and F_Y (i.e., $||F_X - F_Y||$). In particular, the max term in the left hand side of Equation (4.1) is the largest violation area, which is A_1 as shown in Figure 4.2; the integral term in the right hand side of Equation (4.1) is the real area of intersection between the two *cdf*s, not the area in absolute value, which is $A_2 - A_1$ as shown in Figure 4.2.

Example 4.2.2. We provide an example to illustrate WRASD relationship. Suppose we have two discrete random variables X and Y and their probability mass functions:

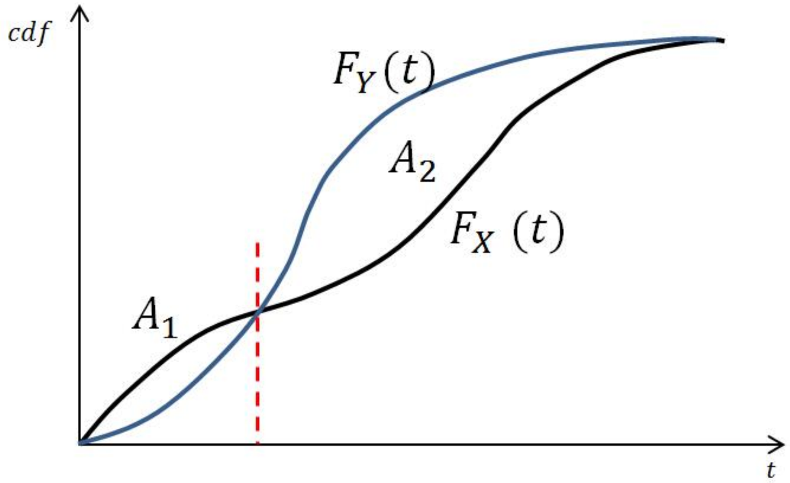

Fig. 4.2. Illustration of WRASD.

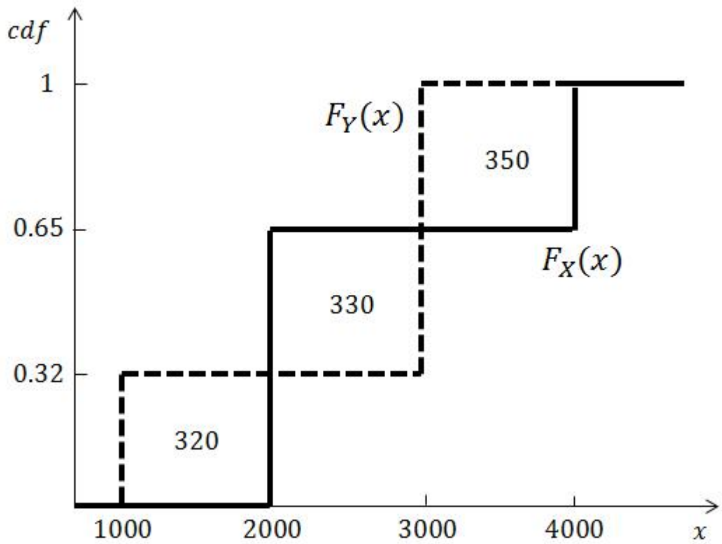

Fig. 4.3. Cumulative distribution functions of X and Y

Table 4.1
Probability mass functions of X and Y.

t	1000	2000	3000	4000
$Pr(X = t)$	0	0.65	0	0.35
$Pr(Y = t)$	0.32	0	0.68	0

The cumulative distribution functions of X and Y can be seen in Figure 4.3. In this case, $x_{min} = 1000$, $x_{max} = 4000$. According to the cumulative distribution functions of X and Y, we obtain

$$\max_{x\in[x_{min},x_{max}]} \left\{ \int_{x_{min}}^{x} [F_X(t) - F_Y(t)]\, dt \right\} = \int_{1000}^{3000} [F_X(t) - F_Y(t)]\, dt = 10,$$

and

$$\int_{x_{min}}^{x_{max}} [F_Y(t) - F_X(t)]\, dt = 340.$$

In addition, we suppose the utility function of the decision maker can be approximated by a logarithm function. Therefore, we have $m(x) = x^{-1}$. It follows that $m(x_{min}) = \frac{1}{1000}$ and $m(x_{max}) = \frac{1}{4000}$.

According to Theorem 4.2.1, we have

$$\left[\left(\frac{1}{\varepsilon} - 1\right)^{0.5} \cdot \frac{1}{1000} - \left(\frac{1}{\varepsilon} - 1\right)^{-0.5} \cdot \frac{1}{4000} \right] \cdot 10 \leq \left(\frac{1}{\varepsilon} - 1\right)^{-0.5} \cdot \frac{1}{4000} \cdot 340.$$

Therefore, $\varepsilon \geq \varepsilon^* = \frac{4}{39} = 0.103$. We conclude that the preference for X over Y can be guaranteed for all rational risk-averse decision makers whose utility functions belong to the set $U_w\left(\frac{1}{x}, 0.103\right)$.

4.3 Optimization with stochastic dominance

We believe that Armbruster and Delage (2015) addresses an important limitation in conventional stochastic optimization approaches and that their proposed approach is suitable when we have very limited information regarding the utility function of the decision maker (e.g., only know that the decision maker is risk averse) and it is possible to have the decision maker evaluate a series of pairwise comparisons between lotteries. A difficulty in applying the approach proposed by Armbruster and Delage (2015) is that decision makers may not be willing or may be biased to make a series of pairwise comparisons between lotteries in practice. In such a case, the analyst may wish to determine the optimal solution based on a CRRA utility function, which can be appropriate for describing the preferences of individuals across multiple settings (see Wakker (2008)). The appropriate γ value to use in the CRRA function can be based on earlier work, like the study by Brenner (2015) which provides comprehensive risk aversion parameters estimates based on option exercising data.

It is important to note that, in this work, we do not assume that the preference of the decision maker is perfectly described by a CRRA utility function. If that were so, the robustness of a solution with respect to deviation in the value of α can easily be studied by conventional sensitivity analysis. Rather, we merely assume that the utility function of the decision maker deviates marginally from a CRRA utility function with predefined risk aversion parameter. This assumption is more appropriate as it is unlikely that the preference of the decision maker is perfectly described by a CRRA utility function in practice.

In this section, we consider the case where the decision maker's utility function can be estimated reasonably well by some parametric function (e.g., CRRA utility function). We seek a solution that is robust to deviations from our estimated utility function. In particular, we consider a set of utility functions whose marginal utility deviates from a reference marginal utility by no greater than some predefined factor τ. We note that this set is similar to that which was proposed by Tan (2015) in defining WASD to explain the unanimous preference for the non-risky reward observed in the classical St. Petersburg paradox (Bernoulli, 1954).

Our result is particularly helpful in practice because constraints based on conventional stochastic dominance can be too strict, resulting in poor solutions. In the extreme case, the introduction of these stochastic dominance constraints can result in infeasible optimization problems (i.e., no feasible solution stochastically dominates the benchmark), as we will illustrate in a numerical example using a simple portfolio optimization problem. The introduction of WASD constraints, rather than constraints based on conventional stochastic dominance, provides the analyst with the flexibility to adjust the degree of dominance desired.

4.3.1 Problem description

In the stochastic optimization literature, it is often assumed that the utility function of the decision maker is available. However, the utility function of the decision maker is generally unknown in practice. One way to resolve this problem is to elicit the utility function by asking the decision maker to make a series of pairwise comparisons between lotteries.

As defined in Section 2.1, we have a random variable $G(z)$ determined by the elements in the decision set $\mathcal{Z}$. In this chapter, we consider $G(z)$ as reward. In particular, we will benchmark G against Y with respect to almost stochastic dominance. We first shed light on the set of all nondecreasing convex utility functions $U_{icv} \subset \mathcal{C}(\mathcal{X})$. Given a set $U \subset U_{icv}$, we obtain

$$\sup_{z \in \mathcal{Z}} \{f(z) : G(z) \succeq_U Y\}. \tag{4.2}$$

Problem (4.2) has finitely many variables and infinitely many constraints via the SD relationship of $G(z)$ and Y. As a semi-infinite programming problem, Problem (4.2) is generally difficult to solve because we cannot enumerate all of the constraints. The main difficulty is that the constraint index set U in Problem (4.2) is an infinite-dimensional space of functions. However, if U is suitably chosen then we can use linear programming duality to get a more tractable representation of the dominance constraint.

Since the SD relationship can be revealed by the expected utility given the corresponding utility set, we can deal with the SD relationship by using the shortfall function

$$\psi\left(G\left(z\right);\,U,\,Y\right)\triangleq\inf_{u\in U}\mathbb{E}\left[u\left(G\left(z\right)\right)-u\left(Y\right)\right]. \tag{4.3}$$

It can be noted that function $\psi\left(G\left(z\right);\,U,\,Y\right)$ is concave with respect to z. Each $z\to\mathbb{E}\left[u\left(G\left(z\right)\right)-u\left(Y\right)\right]$ is concave by Assumption **A2** and the fact that $u\in U_{icv}$. The infimum of concave functions is concave.

Hence, we can rewrite Problem (4.2) with a single functional constraint as

$$\sup_{z\in\mathcal{Z}}\left\{f\left(z\right):\psi\left(G\left(z\right);\,U,\,Y\right)\geq 0\right\}. \tag{4.4}$$

Problems (4.2) and (4.4) are equivalent, but Problem (4.4) is in a more computationally advantageous form as we will see. Immediately, we can see that Problem (4.4) is a convex optimization problem. The objective f is concave by Assumption **A3**, and the implicit constraints $z\in\mathcal{Z}$ are convex by Assumption **A1**.

For Problem (4.4), the general scheme is as follows with two stages:

1. Take the dual of the minimization in Problem (4.3) to obtain a minimization problem

2. Amalgamate the two maximization objectives to obtain a single minimization problem

When the minimization in Problem (4.3) can be written as the linear programming problem (abusing notation)

$$\min_{x}\left\{\left\langle g\left(G\left(z\right)\right),\,x\right\rangle :A\,x\geq b\right\}$$

with dual

$$\max_{p}\left\{\langle b,\, p\rangle : A^{*}p \leq g\left(G\left(z\right)\right)\right\}$$

then we can write Problem (4.4) as

$$\max_{z\in\mathcal{Z},\, p}\left\{f\left(z\right) : \langle b,\, p\rangle \geq 0; A^{*}p \leq g\left(G\left(z\right)\right)\right\},$$

which is a convex optimization problem.

In addition, Assumptions **A4** and **A5** hold here. Since, we introduce a discretization $\Theta \triangleq \{\theta_1, \ldots, \theta_K\} \subset \mathcal{X}$ of $\mathcal{X}$ where $\theta_1 < \theta_2 < \cdots < \theta_K$, $\theta_1 = x_{\min}$, and $\theta_K = x_{\max}$. In this case, we can produce Θ by dividing the interval $\mathcal{X}$ uniformly. Then the length of the uniform sub-intervals is a measure of the granularity of Θ. Since both the benchmark Y and Θ are user inputs, we could ensure that the discretization Θ includes the support of Y.

In order to solve Problem (4.4), we start with the subproblem (4.3). Basically we have two steps in solving Problem (4.3), as in Haskell et al. (2014):

1. Determine the values of u on the finite set Θ, $v_k = u\left(\theta_k\right)$ for $k = 1, \ldots, K$.

2. Interpolate among $\{v_k\}_{k=1}^{K}$ to compute the term

$$\mathbb{E}\left[u\left(G(z)\right)\right] = \sum_{\omega\in\Omega} P\left(\{\omega\}\right) u\left(G(z)\left(\omega\right)\right)$$

   by setting the values $\{u\left(G(z)\left(\omega\right)\right)\}_{\omega\in\Omega}$.

Since interpolation is required, we resort to piecewise linear function. Let:

- v_k be the value of u at θ_k for $k = 1, \ldots, K$ (without loss of generality we just take $v_0 = 0$);

- s_k be the slope of u on $(\theta_k,\, \theta_{k+1})$ for $k = 1, \ldots, K-1$.

Then a piecewise linear function $u \in \mathcal{C}(\mathcal{X})$ will be like

$$u(x) = s_k (x - \theta_k) + v_k,\; x \in [\theta_k,\, \theta_{k+1}],\; \forall k = 1, \ldots,\, K-1,$$

In this way, we can ensure the continuity of utility functions. This piecewise linear u therefore is completely determined by its values at its breakpoints $v = \{v_k\}_{k=1}^{K} \in \mathbb{R}^K$ and the subgradients at the breakpoints $s = \{s_k\}_{k=1}^{K-1} \in \mathbb{R}^{K-1}$.

The following result is a necessary condition for a piecewise linear u to lie in U_{icv}. With this result, the first step above mentioned is done. The values of u on the finite set Θ can be determined.

Lemma 4.3.1 (Haskell et al. (2016)). *(i) Let $u \in U_{icv}$, then there exist $v = \{v_k\}_{k=1}^{K} \in \mathbb{R}^K$ and $s = \{s_k\}_{k=1}^{K-1} \in \mathbb{R}^{K-1}$ such that*

$$v_{k+1} = s_k (\theta_{k+1} - \theta_k) + v_k,\; \forall k = 1, \ldots, K-1,$$

$$s_k \geq s_{k+1} \geq 0,\; \forall k = 1, \ldots, K-2.$$

(ii) Given $\{v_k\}_{k=1}^{K} \in \mathbb{R}^K$ and $\{s_k\}_{k=1}^{K-1} \in \mathbb{R}^{K-1}$, define

$$u^*(x) = \min_k \{s_k (\theta_{k+1} - \theta_k) + v_k\},$$

then $u^* \in U_{icv}$.

Proof. See Boyd and Vandenberghe (2004), Subsection 6.5.5. We see that this condition is simply the requirement that a subgradient exists for u at all $\theta \in \Theta$, and that the subgradients are decreasing. □

In Lemma 4.3.1, we set up the framework of a non-decreasing concave utility function, that is, the discrete values $\{v_k\}_{k=1}^{K}$ would lie on a non-decreasing concave function. In the first step, since the domain is

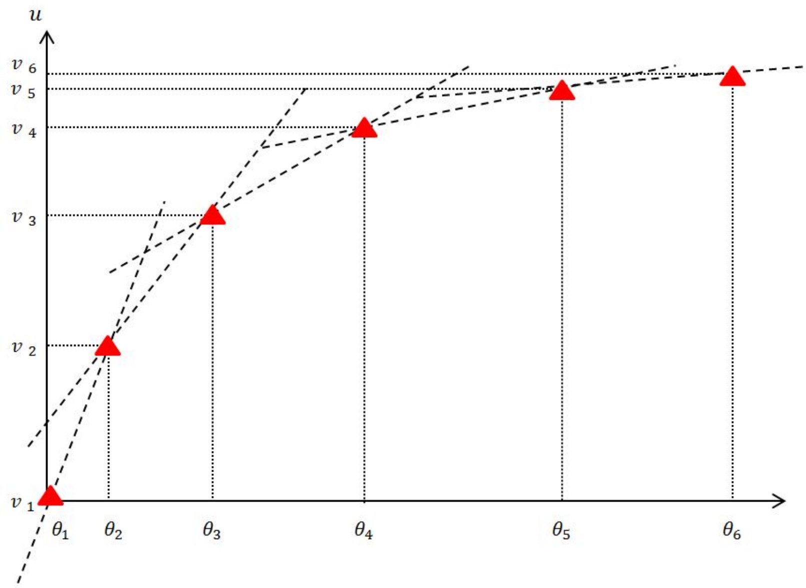

Fig. 4.4. Illustration of Lemma 4.3.1 (i).

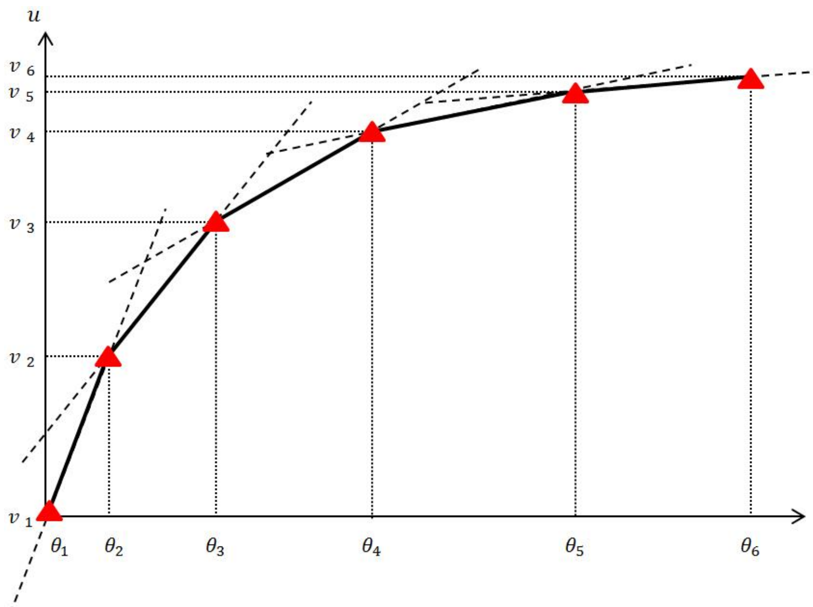

Fig. 4.5. Illustration of Lemma 4.3.1 (ii).

dicretized, there are many intervals. In each interval, there is a linear function determined by the values of two end points. Among the intervals, the slopes of linear functions are in the non-increasing order, as shown in Figure 4.4. In part (ii), we aim to find the lowest part of all linear functions in each interval via the minimization problem. Thus the resulting piecewise function is concave, as shown in Figure 4.5.

The next lemma describes a procedure for linearly interpolating between values $\{u(\theta)\}_{\theta\in\Theta}$ for any $u \in U_{icv}$. With this result, the second step above mentioned can be done.

Lemma 4.3.2 (Haskell et al. (2016)). *Let $u \in U_{icv}$, and define*

$$u^*(x) \triangleq \min_{a\geq 0,\, b\in\mathbb{R}} a\,x + b$$
$$s.t.\, a\,\theta + b \geq u(\theta), \forall\theta \in \Theta.$$

i) u^* is nondecreasing and concave.

ii) u^* is equal to $-\infty$ outside conv$\{\Theta\} \cup \mathbb{R}_+$.

iii) If $\hat{u}$ is another increasing concave function with $\hat{u}(\theta) \geq u(\theta)$ for all $\theta \in \Theta$, then $\hat{u} \geq u^*$.

Proof. i) The function u^* is increasing and convex as it is the supreme of increasing linear functions. It is also immediate that $u^*(\theta) = u(\theta)$ for all $\theta \in \Theta$.

ii) In the domain where it is outside conv $\{\Theta\} \cup \mathbb{R}_+$, since we minimize on a such that $a \geq 0$, we will have $a \to \infty$ which results in $u^* \to -\infty$.

iii) Moreover, $u^* \leq \hat{u}$ for any $\hat{u} \in U_{icv}$ with $\hat{u}(\theta) \geq u(\theta)$ for all $\theta \in \Theta$. We see that the hypograph of u^* is by definition the intersection of the hypographs of all $\hat{u} \in U_{icv}$ with $\hat{u}(\theta) \geq u(\theta)$ for all $\theta \in \Theta$. □

In Lemma 4.3.2, for any values determined by decision makers, we interpolate them among the framework we have set up in Lemma 4.3.1, that

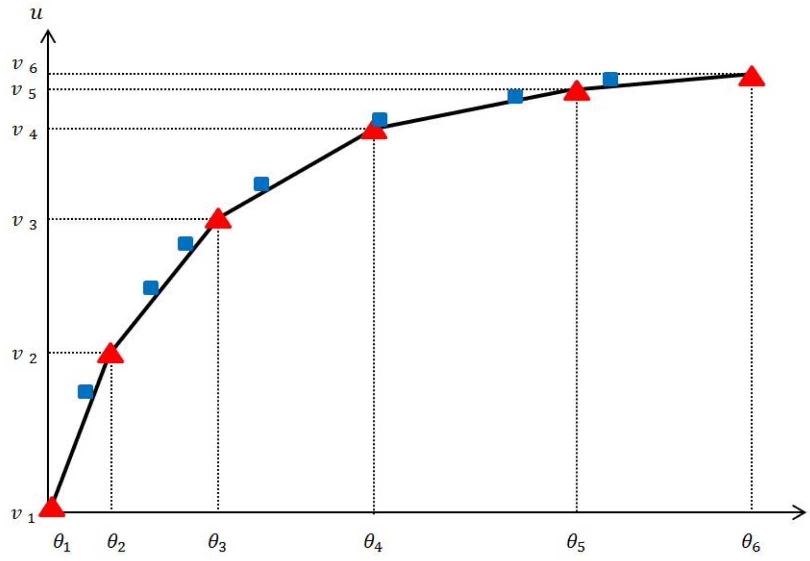

Fig. 4.6. Illustration of Lemma 4.3.2.

is, they will lie one the piecewise concave function, as shown in Figure 4.6.

Theorem 1 in Armbruster and Delage (2015) gives a similar proof as well. By observation, u^* has the following form according to Lemma 4.3.2:

$$u^*\left(x\right)=u\left(\theta_k\right)+\frac{u\left(\theta_{k+1}\right)-u\left(\theta_k\right)}{\theta_{k+1}-\theta_k}\left(x-\theta_k\right),\; x\in\left[\theta_k,\,\theta_{k+1}\right],\;\forall k=1,\ldots,\,K-1.$$

It follows that Lemma 4.3.2 can be used to linearly interpolate between the values of a piecewise linear u given by $\{v_k\}_{k=1}^K$.

To achieve this, we introduce some additional decision variables for the later development: $x\to a_\omega x+b_\omega$ represents a linear function corresponding to each scenario $\omega\in\Omega$. Let $a=\{a_\omega\}_{\omega\in\Omega}\in\mathbb{R}^{|\Omega|}$ and $b=\{b_\omega\}_{\omega\in\Omega}\in\mathbb{R}^{|\Omega|}$ be the parameters for the family of linear functions just defined. We use these functions to linearly interpolate among the values $\{v_k\}_{k=1}^K$ of a piecewise linear function.

So far, we have done the two steps so that the utility function linearly interpolated is non-decreasing and concave. For the sake of convenience, we define the following set:

For this result, let

$$\mathcal{V}_p \triangleq \Big\{(v,\, s,\, a,\, b) \in \mathbb{R}^K \times \mathbb{R}^{K-1} \times \mathbb{R}^{|\Omega|} \times \mathbb{R}^{|\Omega|} : \downarrow$$

$$v_{k+1} = s_k\left(\theta_{k+1} - \theta_k\right) + v_k,\ \forall k = 1, \ldots, K-1,$$

$$a_\omega \theta_k + b_\omega \geq v_k,\ \forall k = 1, \ldots, K,\ \forall \omega \in \Omega,$$

$$s_k \geq s_{k+1} \geq 0,\ \forall k = 1, \ldots, K-2,$$

$$a \geq 0, v_1 = 0\Big\}.$$

represent a set of constraints on the parameter values that determine a piecewise linear u. We can use it in our upcoming linear programming formulation because this set consists of only linear constraints on the variables.

4.3.2 Weighted risk-averse almost stochastic dominance

In the previous section, we have investigated how to interpolate nondecreasing and concave utility functions piecewise linearly. This corresponds to conventional SD relationships. However, this is not practical since the set of nondecreasing and concave utility functions contains "extreme" functions, which does not describe the preferences of most individuals in practice. WRASD, which can reveal unanimous preference by all rational risk-averse decision makers whose utility does not deviate too much from a reference utility by a maximum factor $\left(\frac{1}{\varepsilon} - 1\right)$, can be more appropriate. Therefore, the SD we proposed in Section 4.2 provides a good solution to this problem. We are specifically interested in

$$\sup_{z \in \mathcal{Z}} \left\{f\left(z\right) : \psi\left(G\left(z\right);\, U_w\left(m,\, \varepsilon\right),\, Y\right) \geq 0\right\}, \tag{4.5}$$

where the dominance constraints are generated by WRASD.

In this circumstance, the piecewise linear utility function should fall in some range determined by the weight function. This can be realized by imposing constraints on the subgradients at the breakpoints of the piecewise linear utility function. Since the utility function is interpolated piecewise linearly, it is reasonable to assume a piecewise linear reference utility. This leads to the piecewise constant weight function $m_0 : \mathcal{X} \to \mathbb{R}$.

$$m_0(x) = w_k, \forall x \in [\theta_k,\, \theta_{k+1}),\ \forall k = 1, \ldots, K-2,$$
$$m_0(x) = w_{K-1}, \forall x \in [\theta_{K-1},\, \theta_K].$$

This is the most basic possible form. We will see that by imposing constraints on the subgradients at the breakpoints of the piecewise linear utility function leads to linear programming.

In the next theorem, we show that $\psi(X; U_w(m_0, \varepsilon), Y)$ can be solved by a linear programming problem:

Theorem 4.3.1. *Suppose Ω is finite and $X, Y \in \mathcal{L}$, then $\psi(X; U_w(m_0, \varepsilon), Y)$ is equal to the optimal value of the following linear programming problem:*

$$\min_{v,\,s,\,a,\,b} \sum_{\omega\in\Omega} P(\{\omega\})(a_\omega X(\omega) + b_\omega) - \sum_{k=1}^{K} Pr\{Y = \theta_k\}\, v_k \tag{4.6}$$

$$s.t.\ (v,\, s,\, a,\, b) \in \mathcal{V}_p, \tag{4.7}$$

$$\left(\frac{1}{\varepsilon} - 1\right)^{-0.5} w_k \le s_k \le \left(\frac{1}{\varepsilon} - 1\right)^{0.5} w_k,\ \forall k = 1, \ldots, K-1. \tag{4.8}$$

Proof. The objective term $\sum_{\omega\in\Omega} P(\{\omega\})(a_\omega X(\omega) + b_\omega)$ and constraint (4.7) correspond to Lemma 4.3.1 and Lemma 4.3.2, and the objective term $\sum_{k=1}^{K} \Pr\{Y = \theta_k\}\, v_k$ corresponds to $\mathbb{E}[u(Y)]$. Finally, constraint (4.8) corresponds to the weight function and expresses the weights on the subgradients of $u \in U_w(m_0, \varepsilon)$ required by WRASD. □

The proceeding theorem discusses the dual to Problem (4.6) - (4.8)and shows how to solve Problem (4.5) directly by using linear programming duality. We introduce dual decision variables $\lambda = \{\lambda_k\}_{k=1}^{K-1} \in \mathbb{R}^{K-1}$, $\mu = \{\mu_k\}_{k=1}^{K-2} \in \mathbb{R}_+^{K-2}$, $\delta = \{\delta_{k\omega}\}_{k\in\mathbb{K},\omega\in\Omega} \in \mathbb{R}_+^{K|\Omega|}$, $\gamma = \{\gamma_k^l, \gamma_k^l\}_{k=1}^{K-1} \in \mathbb{R}_+^{2K-2}$.

Theorem 4.3.2. *(i) Problem (4.5) is equivalent to*

$$\max_{z\in\mathcal{Z},\,\lambda,\,\mu\geq 0,\,\delta\geq 0,\,\gamma\geq 0} f(z) \tag{4.9}$$

$$s.t. \sum_{k=1}^{K-1}\left(\frac{1}{\varepsilon}-1\right)^{-0.5} w_k\gamma_k^l - \sum_{k=1}^{K-1}\left(\frac{1}{\varepsilon}-1\right)^{0.5} w_k\gamma_k^u \geq 0, \tag{4.10}$$

$$-Pr\left(\{Y=\theta_k\}\right) + \lambda_{k-1} - \lambda_k + \sum_{\omega\in\Omega}\delta_{k\omega} = 0,\ \forall k = 1,\ldots,K, \tag{4.11}$$

$$-\lambda_k\left(\theta_{k+1}-\theta_k\right) + \mu_{k-1} - \mu_k - \gamma_k^l + \gamma_k^u \geq 0,\ \forall k = 1,\ldots,K-1, \tag{4.12}$$

$$P\left(\{\omega\}\right)\left[G\left(z\right)\right]\left(\omega\right) - \sum_{k=1}^{K}\theta_k\delta_{k\omega} \geq 0,\ \forall\omega\in\Omega, \tag{4.13}$$

$$P\left(\{\omega\}\right) - \sum_{k=1}^{K}\delta_{k\omega} = 0,\ \forall\omega\in\Omega. \tag{4.14}$$

(ii) Problem (4.9) - (4.14) is a convex optimization problem.

Proof. See Appendix B. □

Remember that we have just introduced the piecewise constant weight function, which is the most basic form. We now evaluate how well a piecewise constant weight function m_0 can approximate a general weight function m.

We have assumed that m is Lipschitz continuous with constant L, which is a relatively strong form of function uniform continuity. In order to compare a piecewise constant weight function m_0 with a general weight function m, we construct $m_0(x) = m(\theta_k)$ for $x \in [\theta_k, \theta_{k+1})$ for $k = 1,\ldots,K-1$. With this construction, we can also show that any utility function could be approximated by a piecewise linear one. The following proposition gives an error estimate.

Proposition 4.3.1. *For $\epsilon > 0$ such that $\max_{i=2,\ldots,K}(\theta_i - \theta_{i-1}) < \epsilon$, we have:*

(i) $\|m - m_0\|_{\mathcal{C}(\mathcal{X})} < \epsilon\, L$.

(ii) For any $u \in U_w(m, \varepsilon)$, there is a $\hat{u} \in U_w(m_0, \varepsilon)$ such that $\|u - \hat{u}\|_{\mathcal{C}(\mathcal{X})} < \left(\frac{1}{\varepsilon} - 1\right)^{0.5} \epsilon\, L\, (x_{\max} - x_{\min})$.

Proof. See Appendix C. □

The preceding result shows that we could approximate a general weight function with a piecewise constant one. Note that it is not always the case that $U_w(m, \varepsilon) \subset U_w(m_0, \varepsilon)$ in general, since $m_0 \geq m$ by our construction. However, since $m_0 \geq m$, for any $u \in U_w(m, \varepsilon)$ there exists a $\hat{u} \in U_w(m_0, \varepsilon)$ with $\hat{u} \geq u$.

4.3.3 Robust certainty equivalent maximization problem

So far we have discussed the optimization framework which introduces stochastic dominance relationships. All of them require a given benchmark by which the desired prospect depending on our decisions should dominate. Armbruster and Delage (2015) proposes the concept of robust certainty equivalent which removes the need of a benchmark.

We briefly comment on the robust certainty equivalent (RCE) for G. The certainty equivalent is the amount for sure such that one would be indifferent between it and the random prospect:

$$\mathbb{C}_u[G(z)] \triangleq \sup\{s : u(s) \leq \mathbb{E}[u(G(z))]\}, \forall u \in \mathcal{U}.$$

Since we do not know the exact utility functions but a prevalent class of utility functions, we adopt the worst-case utility function for convenience, which is tractable in the formulation. The RCE maximization problem looks like

$$\max_{z \in \mathcal{Z}} \inf_{u \in \mathcal{U}} \mathbb{C}_u \left[G\left(z \right) \right]. \tag{4.15}$$

Problem (4.15) avoids the difficult problem of constructing a benchmark as in stochastic dominance constraints.

A big advantage of the formulation of robust certainty equivalent is that it avoids the difficult problem of constructing a benchmark as in stochastic dominance constraints. The choice of a benchmark is not trivial to make and remains an interesting topic to explore. So far we do not have reasonable guidance in constructing a benchmark, in particular, for different risk preferences. Therefore, it is natural that we aim at the highest amount of guaranteed return. Unlike utilities that can be scaled arbitrarily, this measure has a meaningful set of units.

Now we discuss the shape of our formulation. A function $f : \mathbb{R}^n \to \mathbb{R}$ is called quasiconcave if its domain and all its super-level sets

$$S_\xi = \left\{ x \in \mathbf{dom}\, f \mid f(x) \geq \xi \right\}, \forall \xi \in \mathbb{R},$$

are convex. We have the following lemma on the objective of Problem (4.15).

Lemma 4.3.3. *$z \to \inf_{u \in \mathcal{U}} \mathbb{C}_u[G\left(z\right)]$ is quasiconcave.*

Proof. Since all $u \in \mathcal{U}$ are increasing and concave, the function $z \to u\left(\left[G\left(z\right)\right]\left(\omega\right)\right)$ is concave as the composition of the increasing concave function u with the concave function $\left[G\left(z\right)\right]\left(\omega\right)$ for all $\omega \in \Omega$. Then

$$\mathbb{E}\left[u\left(G\left(z\right)\right)\right] = \sum_{\omega \in \Omega} P\left(\{\omega\}\right) u\left(\left[G\left(z\right)\right]\left(\omega\right)\right)$$

is concave as the nonnegative sum of concave functions. We know that

$\mathbb{C}_u[G(z)] \geq t$ is equivalent to $\mathbb{E}[u(G(z))] \geq u(t)$ and therefore $\mathbb{C}_u[G(z)]$ is quasiconcave in z. Thus, the point-wise infimum $\inf_{u\in\mathcal{U}} \mathbb{C}_u[G(z)]$ is quasiconcave in z with $\inf_{u\in\mathcal{U}} \mathbb{C}_u[G(z)] \geq t$. $\square$

When $\mathcal{U} = \mathcal{U}_w(m_0, \varepsilon)$, the optimal value of Problem (4.15) under WRASD is greater than t if and only if the constraint

$$\psi(G(z); \mathcal{U}_w(m_0, \varepsilon), t) \geq 0$$

has a feasible solution in $z \in \mathcal{Z}$ because $\inf_{u\in\mathcal{U}_w(m_0,\varepsilon)} \mathbb{C}_u[G(z)] \geq t$ is equivalent to $\psi(G(z); \mathcal{U}_w(m_0, \varepsilon), t) \geq 0$. Therefore, we can solve Problem (4.15) with binary search via a series of feasibility problems like

$$\max_{z\in\mathcal{Z}} \{0 : \psi(G(z); U_w(m_0, \varepsilon), t) \geq 0\}. \tag{4.16}$$

It can be shown that the cut generation problem $\psi(X; U_w(m_0, \varepsilon), t)$ in Problem (4.16) can be given by the optimal value of the following linear programming problem:

$$\begin{aligned}
\min_{v,\, s,\, a,\, b} \; & \sum_{\omega\in\Omega} P(\{\omega\})(a_\omega X(\omega) + b_\omega) - t \\
\text{s.t. } & (v,\, s,\, a,\, b) \in \mathcal{V}_p, \\
& \left(\frac{1}{\varepsilon} - 1\right)^{-0.5} \beta_k \leq s_k \leq \left(\frac{1}{\varepsilon} - 1\right)^{0.5} \beta_k,\ \forall k = 1, \ldots, K-1.
\end{aligned}$$

The last constraints correspond to the weight function in the definition of $\mathcal{U}_w(m_0, \varepsilon)$.

In fact we aim to find out the value of t such that $\psi(X; U_w(m_0, \varepsilon), t) = 0$, which is the worst case. To solve Problem (4.16), we compute the dual of $\psi(G(z); \mathcal{U}_w(m_0, \varepsilon), t)$ (which is a linear programming problem) which then gives a system of convex inequality constraints. Then we can combine this dual problem with our original problem and we have the following theorem.

Theorem 4.3.3. *(i)Problem (4.16) is equivalent to*

$$\max_{z\in\mathcal{Z},\,\lambda,\,\mu\geq 0,\,\delta\geq 0,\,\gamma\geq 0} 0 \tag{4.17}$$

$$s.t. \sum_{k=1}^{K-1}\left(\frac{1}{\varepsilon}-1\right)^{-0.5}\beta_k\gamma_k^l-\sum_{k=1}^{K-1}\left(\frac{1}{\varepsilon}-1\right)^{0.5}\beta_k\gamma_k^u\geq 0, \tag{4.18}$$

$$-\mathbf{1}_t\left(\theta_k\right)+\lambda_{k-1}-\lambda_k+\sum_{\omega\in\Omega}\delta_{k\omega}=0,\ \forall k=1,\ldots,K, \tag{4.19}$$

$$-\lambda_k\left(\theta_{k+1}-\theta_k\right)+\mu_{k-1}-\mu_k-\gamma_k^l+\gamma_k^u\geq 0,\ \forall k=1,\ldots,K-1, \tag{4.20}$$

$$P\left(\{\omega\}\right)\left[G\left(z\right)\right]\left(\omega\right)-\sum_{k=1}^{K}\theta_k\delta_{k\omega}\geq 0,\ \forall\omega\in\Omega, \tag{4.21}$$

$$P\left(\{\omega\}\right)-\sum_{k=1}^{K}\delta_{k\omega}=0,\ \forall\omega\in\Omega, \tag{4.22}$$

(ii) Problem (4.17) - (4.22) is a convex optimization problem.

The proof is very similar to that of Theorem 4.3.2, which can be seen in Appendix C.

By Problem (4.17) - (4.22), we can conclude whether the RCE of prospect $G(z)$ depending on our decision is less than the given t or not. If there is a feasible solution, then the RCE of $G(z)$ is greater than or equal to the value of given t; if there is no solution, then the RCE of $G(z)$ is less than the value of given t.

In order to figure out the worst case of RCE, we require that the difference of the expected utility of the random prospect and the value of t be small enough. Thus we apply a binary search method. We summarize the binary search algorithm next.

Theorem 4.3.4. *Suppose that Problem (4.6) - (4.8) is feasible. Algorithm*

1 finds a robust certainty equivalent $\bar{t} = \frac{t_1+t_2}{2}$ of X such that $|\bar{t} - t^*| < \zeta$ in at most $O(\log(1/\zeta))$ computations of $\psi(X; \mathcal{U}_w(m_0, \varepsilon), t)$, where t^* is the true robust certainty equivalent of X.

input : A routine that solves model optimally and $\zeta > 0$.
output: $\frac{t_1+t_2}{2}$

Step1: Set $t_1 := \tau_1$ and $t_2 := \tau_2$;

Step2: If $t_2 - t_1 < \zeta$, stop. Output: $\frac{t_1+t_2}{2}$;

Step3: Let $t := \frac{t_1+t_2}{2}$. Solve Problem (4.16) by Problem (4.17) - (4.22);

Step4: If $\psi(G(z); \mathcal{U}_w(m_0, \varepsilon), t) \geq 0$ is feasible, update $t_1 := t$. Otherwise, update $t_2 := t$;

Step5: Go to Step 2.

Algorithm 1: Binary search

Proof. It can be noted that every loop in Algorithm 1 reduces the gap between t_1 and t_2 by half. We now show the correctness of the binary search. Note that $\psi(G(z); \mathcal{U}_w(m_0, \varepsilon), t)$ is non-increasing in t. If

$$\psi(G(z); \mathcal{U}_w(m_0, \varepsilon), t) \geq 0,$$

then $t \leq t^*$; otherwise, we have $t^* < t$; we always have $t_1 < t^* < t_2$. It follows that $\bar{t} = \frac{t_1+t_2}{2}$ is close enough to t^* after sufficiently many loops. □

4.4 Examples

4.4.1 Portfolio optimization problem with WRASD

A portfolio optimization problem can be represented as:

$$z \in Z \left\{ \mathbb{E}\left[\sum_{i=1}^{m} R_i z_i\right] : \mathbb{E}\left[u\left(\sum_{i=1}^{m} R_i z_i\right)\right] \geq \mathbb{E}[u(Y)], \forall u \in U_w(m, \varepsilon) \right\}.$$

We consider a simple portfolio consisting of two stocks. In addition, a bond fund will be used as the benchmark. Their returns are listed in Table

4.2 with $G(z) = R_1 z_1 + R_2 z_2$ and $\mathbb{E}[G(z)] = 0.113z_1 + 0.105z_2$.

Table 4.2
Returns of two stocks and a bond fund.

Rate of return: r	5%	6%	7%	9%	11%	12%
$Pr,\ (R_1 = r)$	0.1	0	0	0	0	0.9
$Pr,\ (R_2 = r)$	0	0.1	0	0	0.9	0
$Pr,\ (Y = r)$	0	0	0.4	0.6	0	0

We formulate the cut generation problem $\psi(G(z); U_w(m_0, \varepsilon), Y)$ in which $U_w(m_0, \varepsilon)$ is the set of all non-decreasing concave utility functions with marginal utilities bounded by a piecewise constant weight function m_0 for any given $0 < \varepsilon < 0.5$. In this example, we firstly assume m_0 is constant, which reduces to the almost first-order stochastic dominance (AFSD) proposed by Leshno and Levy (2002). Then we assume that the piecewise constant values of m_0 are obtained from the logarithmic utility function (i.e. $u(x) = \log x$) which is proposed in Tan (2015). According to Theorem 4.3.2, we solve the portfolio optimization problem as ε changes. The results are shown in Figure 4.7 and Figure 4.8, respectively.

As can be seen in Figure 4.7 and Figure 4.8, the vertical axis shows the optimal weights for both assets (z_1^*, z_2^*) and the horizontal axis denotes ε. In Figure 4.7, it can be noted that when $\varepsilon = 0.04$, we have $z_1^* = 0$ and $z_2^* = 1$ and when $\varepsilon = 0.057$, we have $z_1^* = 1$ and $z_2^* = 0$. This is consistent with the fact that R_2 dominates Y by AFSD with $\varepsilon \geqslant 0.04$, and R_1 dominates Y by AFSD with $\varepsilon \geqslant 0.057$. When $\varepsilon < 0.04$, there is no AFSD relationship and the problem is infeasible. In Figure 4.8, the solution pattern is similar to that in Figure 4.7 except that the corresponding ε value for feasible solutions becomes greater. This is because m_0 imposes stricter restrictions on the utility functions.

Therefore, neither R_1 nor R_2, nor the combination of R_1 and R_2 will dominate the benchmark Y by the conventional stochastic dominance.

Hence, our simple portfolio example is infeasible under conventional SD constraints. However, the introduction of WASD constraints allows us to obtain solutions that are reasonable in practice.

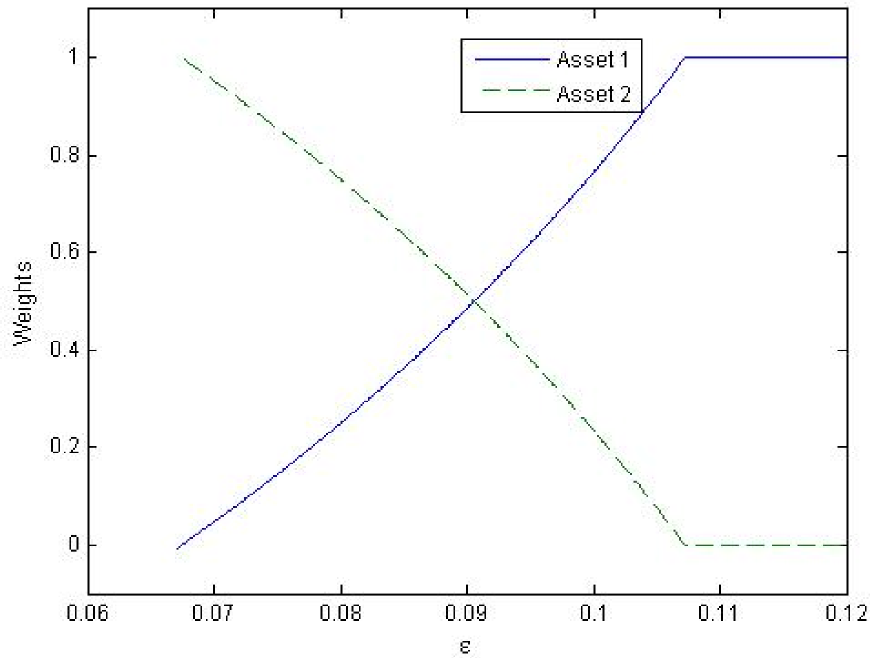

Fig. 4.7. m_0 is constant.

4.4.2 Portfolio optimization problem with RCE

Now we provide an illustration example of the robust certainty equivalent maximization problem. Suppose that there are assets $i = 1, \ldots, m$. The random return rate for asset i is R_i, and the return rate for asset i on $\omega \in \Omega$ is $R_i(\omega)$. Let $Z = \{z \in \mathbb{R}^m : \sum_{i=1}^m z_i = 1,\, z \geq 0\}$ be the set of feasible portfolio allocations. The overall portfolio return rate is

$$G(z) = \sum_{i=1}^{m} R_i z_i$$

We see that the mapping $G(z)$ is linear in the sense that $z \to \sum_{i=1}^m R_i(\omega) z_i$ is linear for all $\omega \in \Omega$. We aim to find the combination of these assets with the largest robust certainty equivalent. We consider the two stocks used in

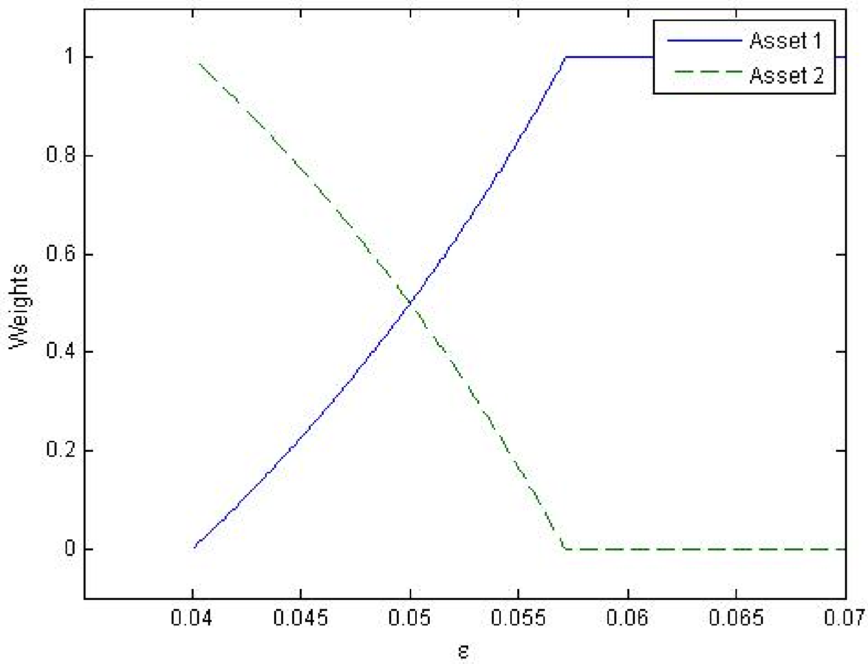

Fig. 4.8. m_0 is obtained from the logarithmic utility function.

the previous example (i.e. Table 4.2).

We formulate the robust certainty equivalent maximization problem $\max_{z\in Z} \inf_{u\in U_w(m_0,\,\varepsilon)} \mathbb{C}_u[G\,(z)]$, where $U_w\,(m_0,\,\varepsilon)$ is the set of all nondecreasing concave utility functions with marginal utilities bounded by a piecewise constant weight function m_0 for any given $0 < \varepsilon < 0.5$. Here we assume that the piecewise constant values of m_0 are obtained from the CRRA utility function (see Example 4.2.1) as proposed in Tan (2015). We solve the robust certainty equivalent maximization problem as ε changes and risk aversion parameter α changes. In this solution, we choose $\zeta = 10^{-5}$ and a discretization of $\mathcal{X}$ with evenly spaced sub-intervals of 0.2. Figure 4.9 and Figure 4.10 show the optimal RCE and the optimal weights of assets as ε changes by fixing $\alpha = 0.8$, respectively. Figure 4.11 shows the optimal RCE as α changes by fixing $\varepsilon = 0.1$. In this case where α is in the range of $[0.7, 0.9]$, we always have $z_1^* = 0$, $z_2^* = 1$.

It can be noted that RCE is a measure with a meaningful set of units, unlike utility measures that can be scaled arbitrarily. In particular, we note that RCE increases when allowable deviation from the reference utility decreases (i.e., increasing ε). In addition, RCE decreases when degree of relative risk aversion increases (i.e., increasing α).

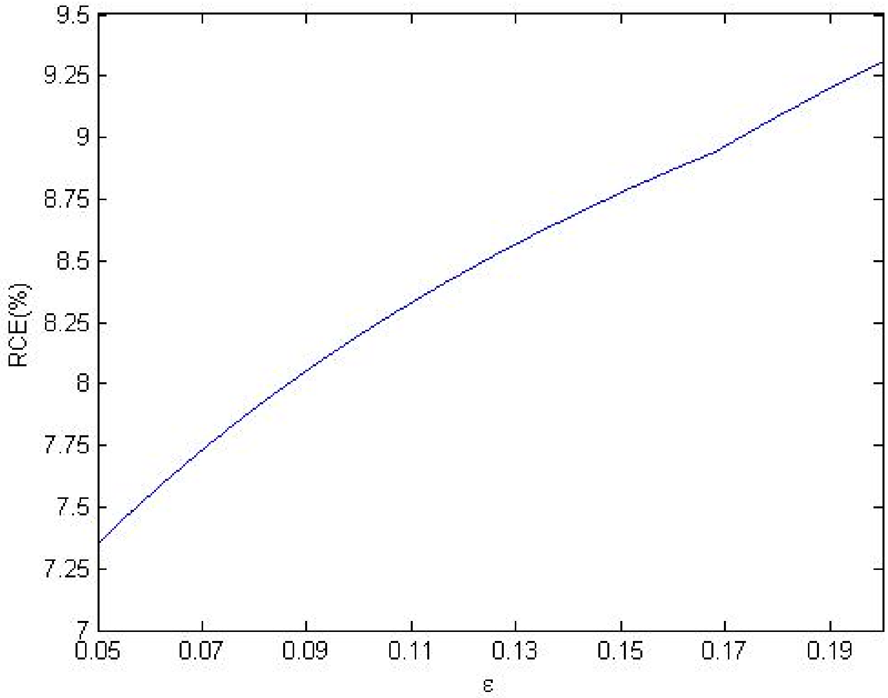

Fig. 4.9. Optimal RCE solutions as ε changes.

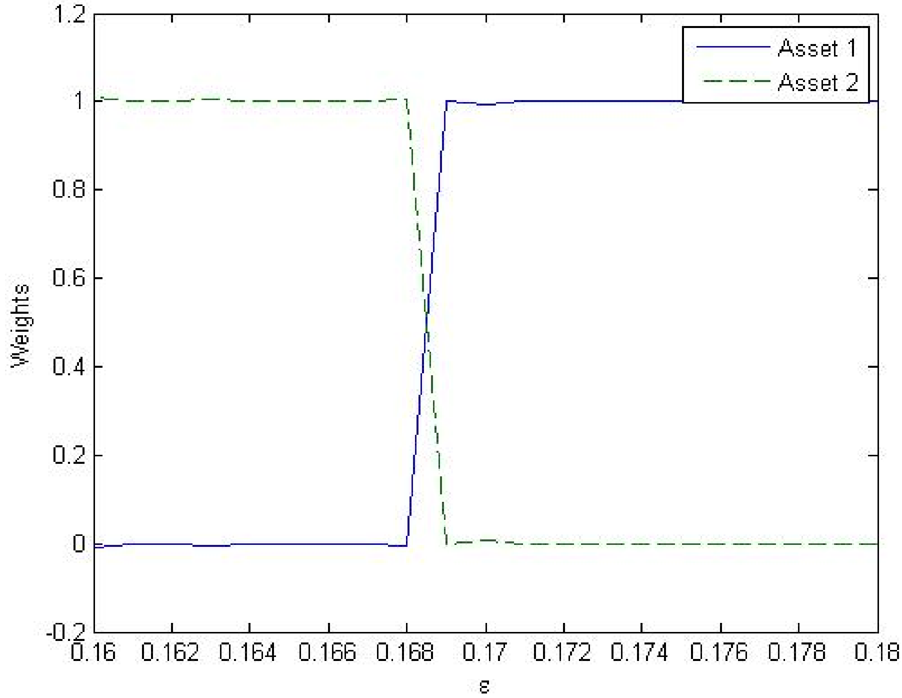

Fig. 4.10. Optimal weights as ε changes.

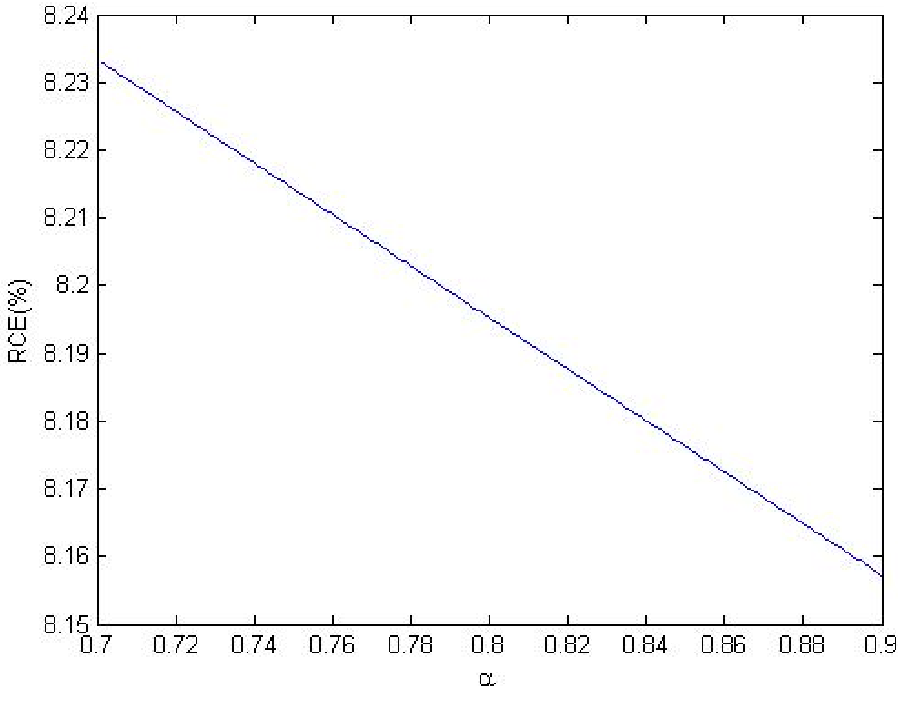

Fig. 4.11. Optimal RCE solutions as α changes.

Chapter 5

Dominance of the Maximum Geometric Mean Portfolio in the long run

5.1 Introduction

It has been proposed by many that an investor with a sufficiently long horizon should adopt a maximum geometric mean (MGM) strategy, which aims for maximal terminal wealth by investing in each period based on the logarithm of returns (see Kelly (1956), Breiman (1960), and Markowitz (1976)). In particular, it follows from the law of large numbers that a MGM portfolio will almost surely outperform other portfolios in the long run under mild conditions.

One of the main criticism of the "clear preference" for the MGM strategy was raised by Merton and Samuelson (1974) who noted that the MGM strategy does not maximize expected utility across all nondecreasing utility functions, even in the long run. Levy (2016) attempted to address this concern by showing that, in the long run, the expected utility of the MGM portfolio is no less than the expected utility of all other portfolios under certain conditions. Namely, when (i) terminal wealth of portfolios are log-normally distributed, (ii) geometric standard deviation of the MGM portfolio is no less than geometric standard deviation of the other portfolios and (iii) marginal utility is bounded.

For infinitely long investment horizons, it follows from the central limit theorem that terminal wealth is lognormally distributed under mild conditions. Furthermore, this assumption appears to be reasonable under a sufficiently long investment horizon in practice. Based on the annual rates of returns of various assets from 1926 to 2012, Levy (2016) observed that, across a 20-year horizon or longer, deviation between the log-normal distribution and empirical distributions based on actual returns appear negligible.

The requirement for higher geometric standard deviation is more problematic as the terminal wealth of the MGM portfolio will almost surely be greater than the terminal wealth of all other portfolios, and not just those with smaller geometric standard deviation, in the long run. However, the superiority of the MGM portfolio over portfolios with larger geometric standard deviation in the long run is not addressed by Levy (2016).

In addition, we believe that there is merit in relaxing the requirement for bounded marginal utility. Although a case for bounded marginal utility was presented in Levy (2016), we believe that there are cases where unbounded marginal utility is reasonable. For example, since an investor who has lost all capital is "out of the game", it is reasonable to assume $u'(w) \to \infty$ when $w \to 0$.

In this chapter, we refine the argument presented by Levy (2016) regarding the superiority of the MGM strategy in the long run by relaxing the requirements of higher geometric standard deviation and bounded marginal utility. In particular, we show that a higher geometric mean is necessary and sufficient for log-weighted almost stochastic dominance and the MGM strategy is preferred by all investors whose utility function deviates marginally from the logarithm utility function. Furthermore, the maximum allowable deviation increases in the investment horizon and is unbounded.

In the following, the problem will be described in Section 5.2. The main results by applying ASD relationship for log-normal distribution will be presented in Section 5.3. We will conclude this chapter in Section 5.4.

5.2 Problem description

Let X_t denote the portfolio return (end of period value) in period t. For an investment horizon of T periods, the terminal wealth of the portfolio is $W_X(T) = \Pi_{t=1}^{T} X_t$. If each X_t is independent and follows a log-normal distribution with parameters μ_X and σ_X^2, $W_X(T)$ also follows a log-normal distribution with parameters $T\mu_X$ and $T\sigma_X^2$. Consider a second portfolio with terminal wealth $W_Y(T)$ that is log-normally distributed with parameters $T\mu_Y$ and $T\sigma_Y^2$.

Assuming that μ_X, μ_Y, σ_X and σ_Y are finite, it follows from the law of large numbers that the terminal wealth of the portfolio with higher geometric mean will almost surely be greater (see Levy (2016) for details). Stated formally, if $\mu_X > \mu_Y$ then:

$$P[W_X(T) > W_Y(T)] \to 1, \text{as}\, T \to \infty.$$

Here, we note that the assumption on log-normally distributed returns is not particularly restrictive since it follows from the central limit theorem that the terminal wealth distribution of both portfolios approach the log-normal distribution in the long run, even if X_t and Y_t are not log-normally distributed.

Although the argument above appears compelling, Merton and Samuelson (1974) noted that the preferred investment strategy of an investor with iso-elastic utility is independent of the investment horizon and the MGM strategy does not maximize the expected utility of investors under some iso-elastic utility functions. Therefore, the MGM strategy does not dominate under conventional stochastic dominance rules in the long run and it is not immediately clear that the MGM strategy should be preferred.

Levy (2016) attempted to resolve this issue by proving that the expected utility of a log-normally distributed prospect with higher geometric mean and geometric standard deviation will be at least as large as the expected utility of another log-normally distributed prospect in the long run under bounded marginal utility. However, this does not explain why the MGM portfolio is also preferred over portfolios with higher geometric standard deviation and by investors with unbounded marginal utility. In the next section, we explain why the superiority of the MGM strategy in the long run is better explained via log-weighted almost stochastic dominance.

5.3 Main results

We begin by introducing the concept of weighted almost stochastic dominance (WASD) proposed by Tan (2015). Let U denote the set of all differentiable nondecreasing utility functions and $U^*(m,\varepsilon)$ denote the set of all differentiable nondecreasing utility functions whose marginal utility differs from $m(t)$ by a maximum factor of $\left[\frac{1}{\varepsilon}-1\right]^{0.5}$ for some constant $\varepsilon \in (0, 0.5]$:

$$U^*(m,\varepsilon) = \left\{ u \in U : \left[\frac{1}{\varepsilon}-1\right]^{-0.5} m(t) \leq u'(t) \leq \left[\frac{1}{\varepsilon}-1\right]^{0.5} m(t), \forall t \right\}.$$

We note that $m(t)$ is a nonnegative function that describes the marginal utility of some canonical utility function, which $U^*(m,\varepsilon)$ is constructed around. For example, if the decision maker's preference is described approximately by logarithm utility, we set $m(t) = \frac{d}{dt}\ln(t) = \frac{1}{t}$.

Tan (2015) showed that X dominates Y with (m, ε)-WASD if and only if $\mathbb{E}[u(X)] \geq \mathbb{E}[u(Y)]$ for all u in $U^*(m, \varepsilon)$. Since $U^*(m, \varepsilon)$ increases as ε decreases, a lower ε implies a higher maximum allowable deviation from the canonical utility function such that preference for X over Y is clear.

Consider two random variables X and Y that are log-normally distributed with parameters μ_X, σ_X^2, μ_Y and σ_Y^2.

Theorem 5.3.1. *Suppose $X \sim \ln N(\mu_X, \sigma_X^2)$, $Y \sim \ln N(\mu_Y, \sigma_Y^2)$, $\mu_X > \mu_Y$ and $\sigma_X \neq \sigma_Y$. X dominates Y with $\left(\frac{1}{t}, \varepsilon\right)$-WASD if and only if:*

$$\varepsilon \geq \frac{1}{2}\left[1 - \frac{\phi\sqrt{\pi}}{\phi\sqrt{\pi}\cdot erf(\frac{\phi}{\sqrt{2}}) + \sqrt{2}e^{-\frac{\phi^2}{2}}}\right],$$

where $\phi = \frac{\mu_X - \mu_Y}{|\sigma_X - \sigma_Y|}$ and $erf(\cdot)$ denotes the Gauss error function.

Theorem 5.3.1 highlights the set of utility functions such that X is clearly preferred over Y. For mathematical convenience, let

$$h(\phi) = \frac{1}{2}\left[1 - \frac{\phi\sqrt{\pi}}{\phi\sqrt{\pi}\cdot erf(\frac{\phi}{\sqrt{2}}) + \sqrt{2}e^{-\frac{\phi^2}{2}}}\right].$$

For any two portfolios with log-normal returns, one could easily determine the largest $U^*(\frac{1}{t}, h(\phi))$ such that preference for the portfolio with higher geometric mean can be guaranteed by computing $h(\phi)$. The corresponding values of $h(\phi)$ for different ϕ are illustrated in Figure 5.1.

The following two propositions highlight that $h(\phi)$ decreases in ϕ and approaches 0 as ϕ grows infinitely large.

Proposition 5.3.1. *$h(\phi)$ is strictly decreasing in ϕ.*

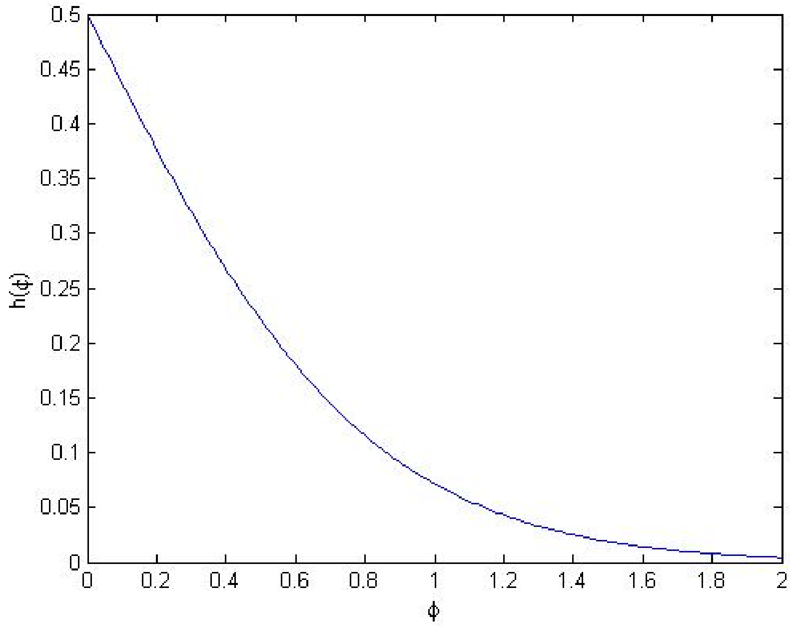

Fig. 5.1. The relationship between $h(\phi)$ and ϕ.

Proof of Proposition 5.3.1. We take the derivative of $h(\phi)$.

$$\begin{aligned} h'(\phi) &= -\frac{1}{2} \cdot \frac{\sqrt{\pi}\left(\phi\sqrt{\pi} \cdot erf\left(\frac{\phi}{\sqrt{2}}\right) + \sqrt{2}e^{-\frac{\phi^2}{2}}\right)}{\left(\phi\sqrt{\pi} \cdot erf\left(\frac{\phi}{\sqrt{2}}\right) + \sqrt{2}e^{-\frac{\phi^2}{2}}\right)^2} \\ &\quad -\frac{1}{2} \cdot \frac{-\phi\sqrt{\pi}\left(\sqrt{\pi} \cdot erf\left(\frac{\phi}{\sqrt{2}}\right) + \sqrt{2}\phi e^{-\frac{\phi^2}{2}} - \sqrt{2}\phi e^{-\frac{\phi^2}{2}}\right)}{\left(\phi\sqrt{\pi} \cdot erf\left(\frac{\phi}{\sqrt{2}}\right) + \sqrt{2}e^{-\frac{\phi^2}{2}}\right)^2} \\ &= -\frac{1}{2} \cdot \frac{\pi\phi \cdot erf\left(\frac{\phi}{\sqrt{2}}\right) + \sqrt{2\pi}e^{-\frac{\phi^2}{2}} - \pi\phi \cdot erf\left(\frac{\phi}{\sqrt{2}}\right)}{\left(\phi\sqrt{\pi} \cdot erf\left(\frac{\phi}{\sqrt{2}}\right) + \sqrt{2}e^{-\frac{\phi^2}{2}}\right)^2} \\ &= -\frac{1}{2} \cdot \frac{\sqrt{2\pi}e^{-\frac{\phi^2}{2}}}{\left(\phi\sqrt{\pi} \cdot erf\left(\frac{\phi}{\sqrt{2}}\right) + \sqrt{2}e^{-\frac{\phi^2}{2}}\right)^2} \\ &< 0. \end{aligned}$$

□

Proposition 5.3.2. $\lim_{\phi\to\infty} h(\phi) = 0$.

Proof of Proposition 5.3.2. Follows from the observation that

$$\lim_{\phi\to\infty} erf(\frac{\phi}{\sqrt{2}}) = 1$$

and

$$\lim_{\phi\to\infty} \phi^{-1}e^{-\frac{\phi^2}{2}} = 0.$$

□

Having obtained the results above, we are now ready to study the superiority of the MGM portfolio in the long run from the perspective of WASD. Let $W_X(T)$ denote the terminal wealth of the MGM portfolio at period T where $W_X(T)$ is log-normally distributed with parameters $T\mu_X$ and $T\sigma_X^2$. In addition, let $W_Y(T)$ denote the terminal wealth of another portfolio at period T where $W_Y(T)$ is log-normally distributed with parameters $T\mu_Y$ and $T\sigma_Y^2$. Since the MGM portfolio has maximal geometric mean, $\mu_X > \mu_Y$.

If $\sigma_X = \sigma_Y$, preference for $W_X(T)$ over $W_Y(T)$ is clear as the former dominates the latter with first-degree stochastic dominance (Levy, 1973). If $\sigma_X \neq \sigma_Y$, it follows from Theorem 5.3.1 that $W_X(T)$ dominates $W_Y(T)$ with $\left(\frac{1}{t}, h(\phi)\right)$-WASD, where:

$$\phi = \frac{T\mu_X - T\mu_Y}{\left|\sqrt{T}\sigma_X - \sqrt{T}\sigma_Y\right|} = \frac{\sqrt{T}(\mu_X - \mu_Y)}{|\sigma_X - \sigma_Y|}.$$

Since ϕ increases with T, it follows from Proposition 5.3.1 that the maximum allowable tolerance from logarithm utility, such that there is clear preference for $W_X(T)$ over $W_Y(T)$, increases with T. Therefore, preference for the portfolio with higher geometric mean becomes clearer as the investor's investment horizon increases.

Furthermore, since $\phi \to \infty$ as $T \to \infty$, it follows from Proposition 5.3.2 that the maximum allowable tolerance is unbounded. Hence, the maximum allowable tolerance from logarithm utility becomes infinitely large as the investment horizon grows infinitely long. From the perspective of log-weighted almost stochastic dominance, a higher geometric mean is necessary and sufficient for clear preference across an infinitely long investment horizon, which is consistent with the observation that a portfolio with higher geometric mean will almost surely have higher terminal wealth in the long run.

5.4 Discussion

In this chapter, we address the gap between the preference for the MGM strategy in the long run from the perspectives of the law of large numbers and stochastic dominance. In particular, the former states that the MGM strategy is almost surely to be better in the long run but preference for the MGM strategy is unclear under conventional stochastic dominance rules. Here, we explain why the clear preference for the MGM strategy in the long run can be explained via log-weighted almost stochastic dominance.

Besides adding to the theoretic debate regarding the superiority of the MGM strategy in the long run, this work also adds to the stochastic dominance literature by providing additional support for the use of log-weighted almost stochastic dominance to explain clear preferences between risky prospects in practice for a wide range of decision makers. In particular, log-weighted almost stochastic dominance can reveal the clear preference for the MGM strategy in the long run but, as noted by Merton and Samuelson (1974), WASD based on other CRRA utility functions may not.

Finally, this work presents an alternative to the mean-variance framework proposed by Markowitz (1952), which is often criticized for assumptions on normality and quadratic utility. In our work, we highlight that a geometric-mean-geometric-standard-deviation framework is suitable for comparing between investments with log-normal returns, which is reasonable for investors with a sufficiently long investment horizon. Here, we do not assume that the utility of the decision maker follows any particular form but only assume that it can be approximated by logarithm utility. One key insight is that the maximum allowable deviation from logarithm utility can be expressed as a function of the geometric mean and geometric standard deviations of investments under consideration (see, Theorem 5.3.1). In particular, the greater the difference in geometric mean and the smaller the difference in geometric standard deviation, the clearer the preference for the investment with higher geometric mean.

Chapter 6

Optimization frameworks for scalar-valued risk measures

6.1 Introduction

As reviewed in Section 3.3, we point out four aspects that our optimization framework should take into account: multivariate prospects, preference uncertainty, computational tractability, target-oriented measure. Specifically, previous work that is pertinent to each area has been reviewed respectively. It can be noted that risk-aware optimization is a highly developed field. However, it is not easy to do risk-aware optimization in practice for models that simultaneously address the above mentioned four considerations.

One of the difficulties lies in determining the risk preferences of the decision maker. Here, we will offer intuitive frameworks for doing risk-aware optimization that are easy to apply. All the prospects we consider in this chapter are multivariate. Specifically, we aim to construct a set of appropriate risk measures that well characterizes possible decision-maker risk preferences, and then construct a robust stochastic optimization problem using this risk measure set. The set of risk measures should be carefully chosen so that we can appeal to the special structure of convex interpolation problems to make our robust formulations computationally tractable.

In this chapter, we focus on scalar-valued risk measures (although the inputs are multivariate). In particular, we look into two approaches: subgradient characterization and acceptance set approach. In the subgradient characterization approach, the set of risk measures is constructed by convex function interpolation as well as other conditions such as elicited comparison information. Then we aim to minimize the worst-case of the constructed risk measures. In the acceptance set approach, we aim to figure out the acceptance set which is determined by the elicited comparisons of given prospects.

In the following, the preliminary will be provided in Section 6.2. The two approaches that deal with the scalar-valued risk measures will be discussed in Section 6.3.

6.2 Preliminary

We define any mapping $\rho : \mathcal{L}^n \to \mathbb{R}^d$ as a risk measure where $d \leq n$. When $d = 1$, ρ is a scalar-valued risk measure. When $d \geq 2$, ρ is a vector-valued risk measure.

We define the following category of risk measures:

$$\mathcal{R}_{icx} \triangleq \{\rho : \rho \text{ is nondecreasing, convex, and } \rho(0) = 0\},$$

defined to be the set of all monotonic and convex risk functions such that the risk of the zero portfolio is zero. Therefore, the risk measures in $\mathcal{R}_{icx}$ possess the following two key properties:

- Convexity: $\rho(\lambda X + (1-\lambda) Y) \leq \lambda \rho(X) + (1-\lambda) \rho(Y)$ for all $X, Y \in \mathcal{L}^n$ and $\lambda \in [0, 1]$.
- Monotonicity: If $X, Y \in \mathcal{L}^n$ and $X \leq Y$, then $\rho(X) \leq \rho(Y)$.

Now, we present several other conditions to be used in the construction of our uncertainty sets:

- Positive homogeneity: If $\alpha \geq 0$ and $X \in \mathcal{L}^n$, then $\rho(\alpha X) = \alpha \rho(X)$. Positive homogeneity is well studied for univariate risk functions, see Ruszczyński and Shapiro (2006). Notice that positive homogeneity implies that $\rho(0) = 0$.

- Translation equivariance: If $\alpha \in \mathbb{R}$ and $X \in \mathcal{L}^n$, then $\rho(X + \alpha e_1) = \rho(X) + \alpha$. We define

$$\mathcal{R}_{eq} \triangleq \{\rho : \rho(X + \alpha e_1) = \rho(X) + \alpha\}.$$

  Translation equivariance is also well studied for univariate risk functions, where we modify the definition for multivariate risk functions.

- Normalization: For fixed $W_0, Y_0 \in \mathcal{L}^n$, $\rho(W_0) - \rho(Y_0) = 1$ (e.g. we can take $W_0 = 1$ and $Y_0 = 0$). We define

$$\mathcal{R}_{nor} \triangleq \{\rho : \rho(W_0) - \rho(Y_0) = 1\}.$$

  We include normalization to avoid risk measures which take arbitrarily large values, which incurs difficulties for our upcoming optimization problems. In implementation, we often take $W_0 = 1$ and $Y_0 = 0$ to require that $\rho(1) = 1$.

- Elicited preference: Given a finite collection of pairs of random vectors $W_k, Y_k \in \mathcal{L}^n$ indexed by $\mathbb{K}$, $\rho(W_k) \leq \rho(Y_k)$ for all $k \in \mathbb{K}$. We define

$$\mathcal{R}_{elc}(\Sigma) \triangleq \{\rho : \rho(W_k) \leq \rho(Y_k), \forall k \in \mathbb{K}\} \text{ with } \Sigma \triangleq \{(W_k, Y_k)\}_{k=1}^{K}.$$

  Preference elicitation is explored for expected utility and univariate risk functions in Armbruster and Luedtke (2015) and Delage and Li (2015). Its extension to the multivariate setting is quite natural.

6.3 Problem formulations

Suppose we have a random variable $G(z)$ defined in Section 2.1. In this chapter, we consider $G(z)$ as loss. Given a risk function $\rho : \mathcal{L}^n \to \mathbb{R}$, $\rho(G(z))$ is a measure of the fitness of $G(z)$, where lower values of $\rho(G(z))$ are preferred. The corresponding risk-aware optimization problem is

$$\min_{z \in \mathcal{Z}} \rho(G(z)) . \tag{6.1}$$

Problem (6.1) is the main object of attention in this paper. Next, we give conditions for Problem (6.1) to be a convex optimization problem.

Proposition 6.3.1. *Suppose $\rho \in \mathcal{R}_{icx}$, and that assumptions A1 and A2 hold. Then Problem (6.1) is a convex optimization problem.*

Proof. We need only to show that $z \to \rho(G(z))$ is convex since $\mathcal{Z}$ is a convex set by assumption. Choose $z_1, z_2 \in \mathcal{Z}$ and $\lambda \in [0, 1]$ and compute

$$\begin{aligned} \rho(G(\lambda z_1 + (1-\lambda) z_2)) &\leq \rho(\lambda G(z_1) + (1-\lambda) G(z_2)) \\ &\leq \lambda \rho(G(z_1)) + (1-\lambda) \rho(G(z_2)), \end{aligned}$$

where the first inequality follows by convexity of G and monotonicity of ρ, and the second inequality follows by convexity of ρ. □

Next we give several specific examples of risk functions to better motivate Problem (6.1).

Example 6.3.1. For a utility function $u : \mathbb{R}^n \to \mathbb{R}$, the expected utility $\rho(\cdot) = \mathbb{E}[u(\cdot)]$ is a risk measure. This is closely related to the theory of stochastic dominance that we have discussed in Chapter 4.

Example 6.3.2. For the univariate case, the conditional value-at-risk with the form of $\phi_\alpha(X) = \inf_{\eta\in\mathbb{R}} \left\{\eta + (1-\alpha)^{-1}\mathbb{E}\left[(X-\eta)_+\right]\right\}$ is a risk measure on $\mathcal{L}$.

Example 6.3.3. For $n \geq 1$, we let $\mathcal{F}$ be a collection of convex functions $g : \mathbb{R}^n \to \mathbb{R}$. Then $\rho : \mathcal{L}^n \to \mathbb{R}$ defined by $\rho(X) = \sup_{f\in\mathcal{F}} \mathbb{E}[g(X)]$ is a convex function on $\mathcal{L}^n$ since the supremum of convex functions is convex.

We emphasize that Problem (6.1) requires exact specification of the decision maker's risk preferences via the risk function ρ. However, this information is difficult to elicit in practice. In this subsection we propose a family of uncertainty sets for the decision maker's risk preferences. Then, we propose a robust optimization problem over this uncertainty set and show how to reformulate it as a tractable convex optimization problem.

We are interested in the risk measures in $\mathcal{R}_{icx}$. It is reasonable to restrict to monotonic risk functions since greater loss should be associated with greater risk. The requirement of convexity is also reasonable for our setting since diversification should not increase risk.

We will use the notation $\mathcal{R} \subset \mathcal{R}_{icx}$ to denote an uncertainty set of risk functions. We will consider more restrictions on $\mathcal{R}$ later. The robust optimization problem can be represented as:

$$\min_{z\in\mathcal{Z}} \sup_{\rho\in\mathcal{R}} \rho(G(z)). \tag{6.2}$$

which minimizes the worst-case risk over the uncertainty set $\mathcal{R}$. Problem (6.2) can be interpreted as being robust against uncertainty in the decision maker's risk preferences. This type of robust formulation was proposed for the univariate case in Delage and Li (2015), while now we extend this framework to the multivariate case.

Next we establish convexity of Problem (6.2).

Proposition 6.3.2. *Suppose assumptions A1 and A2 hold, and $\mathcal{R} \subset \mathcal{R}_{icx}$,* *then Problem (6.2) is a convex optimization problem.*

Proof. For any $\rho \in \mathcal{R}$, each $z \to \rho(G(z))$ is convex since

$$\begin{aligned}\rho(G(\lambda z_1 + (1-\lambda) z_2)) &\leq \rho(\lambda G(z_1) + (1-\lambda) G(z_2)) \\ &\leq \lambda \rho(G(z_1)) + (1-\lambda) \rho(G(z_2)),\end{aligned}$$

for all $z_1, z_2 \in \mathcal{Z}$ and $\lambda \in [0, 1]$, where the first inequality follows from convexity of G and monotonicity of ρ, and the second inequality follows by convexity of ρ. Then, we see that $z \to \sup_{\rho \in \mathcal{R}} \rho(G(z))$ is convex since the supremum of convex functions is convex. □

6.3.1 Subgradient characterization

Next, we discuss tractable reformulations of the robust optimization Problem (6.2) for various uncertainty sets $\mathcal{R}$. The key to the following development is found in Boyd and Vandenberghe (2004), Subsection 6.5.5 (for example) which shows that interpolation problems with convex functions can be solved with convex optimization. This observation allows us to solve Problem (6.2) with convex optimization techniques.

Again, Assumptions **A4** and **A5** also hold here. They vastly simplify our development. For the benchmark, we can meet Assumption **A5** by construction since the benchmark Y is user input. The discretization Θ is also user input, and it can be constructed a priori to include the support of Y.

Since Ω is finite, we identify a random variables $X \in \mathcal{L}^n$ with a vector $X(\omega) \in \mathbb{R}^{n\,|\Omega|}$ where we list the realizations of X component-wise. We define

$$\Theta = \{X\} \cup \left\{ \bigcup_{k \in \mathbb{K}} \{W_k, Y_k\} \right\} \cup \{0\}$$

to be the union of the supports of random vector X, $\bigcup_{k \in \mathbb{K}} \{W_k, Y_k\}$, and the origin. Since Ω is finite, the set Θ is finite as well.

For the robust optimization problem, the general scheme is as follows with two stages:

1. Take the dual of the inner maximization in Problem (6.2) to obtain a minimization problem
2. Amalgamate the two minimization objectives to obtain a single minimization problem

When the inner maximization in Problem (6.2) can be written as the linear programming problem

$$\max \{\langle f(G(z)), x\rangle : A\, x \geq b\}$$

with dual

$$\min \{\langle b, p\rangle : A^{*}p \leq f(G(z))\}$$

then we can write Problem (6.2) as

$$\min_{z\in\mathcal{Z}} \min_{p} \{\langle b, p\rangle : A^{*}p \leq f(G(z))\} \equiv \min_{z\in\mathcal{Z},\, p} \{\langle b, p\rangle : A^{*}p \leq f(G(z))\},$$

which is a convex optimization problem.

Likewise, we could use the interpolation techniques in Chapter 4 to cope with the inner maximization in Problem (6.2). However, we make two changes here: the functions that are interpolated here are convex, and the prospects we are interested in are multivariate. The main two steps for the linear interpolation technique here are as the following:

1. Determine the values of $\rho(\theta)$ on the finite set Θ for $k = 1, \ldots, K$.
2. Interpolate to obtain the worst-case risk over the uncertainty set.

We introduce the following decision variables:

- s_θ be the subgradient of ρ at θ for $\theta \in \Theta$.

- t_θ be the intercept of ρ at θ for $\theta \in \Theta$.

The following lemma is a necessary condition for a piecewise linear ρ to lie in $\mathcal{R}_{icx}$. With this result, the first step above mentioned is done. The values of ρ on the finite set Θ can be determined.

Lemma 6.3.1. *(i) Let $\rho \in \mathcal{R}_{icx}$, then there exist $\{s_\theta\}_{\theta\in\Theta} \subset \mathbb{R}_+^{n\,|\Omega|}$ and $\{t_\theta\}_{\theta\in\Theta} \subset \mathbb{R}$ such that*

$$\rho\left(\theta\right) = \langle s_\theta,\, \theta\rangle + t_\theta,\, \forall \theta \in \Theta,$$

$$\rho\left(\theta'\right) \geq \langle s_\theta,\, \theta'\rangle + t_\theta,\, \forall \theta,\, \theta' \in \Theta,\, \theta \neq \theta'.$$

(ii) Given $\{s_\theta\}_{\theta\in\Theta} \subset \mathbb{R}_+^{n\,|\Omega|}$ and $\{t_\theta\}_{\theta\in\Theta} \subset \mathbb{R}$, define

$$\rho^*\left(x\right) = \max_{\theta\in\Theta}\left\{\langle s_\theta,\, x\rangle + t_\theta\right\},$$

then $\rho^* \in \mathcal{R}_{icx}$.

Proof. See Boyd and Vandenberghe (2004), Subsection 6.5.5. It can be noted that the subgradients for all $\rho \in \mathcal{R}_{icx}$ are finite on conv $\{\Theta\}$. Specifically, at all $x \in$ conv $\{\Theta\}$ there are $s_x \in \mathbb{R}_+^{n\,|\Omega|}$ and $t_x \in \mathbb{R}$ such that $\rho\left(x\right) = \langle s_x,\, x\rangle + t_x$ and $\rho\left(y\right) \geq \langle s_x,\, y\rangle + t_x,\, \forall y$. □

Given a set of values $\{\rho\left(\theta\right)\}_{\theta\in\Theta}$ where $\rho \in \mathcal{R}_{icx}$, the next lemma constructs the largest increasing convex function ρ^* with $\rho^*\left(\theta\right) \leq \rho\left(\theta\right)$ for all $\theta \in \Theta$.

Lemma 6.3.2. *Let $\rho \in \mathcal{R}_{icx}$, and define*

$$\rho^*\left(x\right) \triangleq \max_{a\geq 0,\, b}\langle a,\, x\rangle + b$$
$$s.t.\ \langle a,\, \theta\rangle + b \leq \rho\left(\theta\right), \forall \theta \in \Theta.$$

i) ρ^* *is increasing and convex.*

ii) ρ^* *is equal to* ∞ *outside conv*$\{\Theta\} \cup \mathbb{R}_-^{n\,|\Omega|}$.

iii) If $\bar{\rho}$ *is another increasing convex function with* $\bar{\rho}(\theta) \leq \rho(\theta)$ *for all* $\theta \in \Theta$*, then* $\bar{\rho} \leq \rho^*$.

Proof. i) The function ρ^* is increasing and convex as it is the supremum of increasing linear functions. It is also immediate that $\rho^*(\theta) = \rho(\theta)$ for all $\theta \in \Theta$.

ii) In the domain where it is outside conv $\{\Theta\} \cup \mathbb{R}_-^{n\,|\Omega|}$, since we maximize on a such that $a \geq 0$, we will have $a \to \infty$ which results in $\rho^* \to \infty$.

iii) Moreover, $\rho^* \geq \bar{\rho}$ for any $\bar{\rho} \in \mathcal{R}_{icx}$ with $\bar{\rho}(\theta) \leq \rho(\theta)$ for all $\theta \in \Theta$. We see that the epigraph of ρ^* is by definition the intersection of the epigraphs of all $\bar{\rho} \in \mathcal{R}_{icx}$ with $\bar{\rho}(\theta) \leq \rho(\theta)$ for all $\theta \in \Theta$. □

We now proceed to discuss the specifics of the construction of the uncertainty set $\mathcal{R}$. We begin by identifying some desirable properties of risk functions that should belong to this set. We already discussed the monotonicity and convexity in the previous section, since these conditions are necessary for Problem (6.1) to be convex. Now, we present several other conditions to be used in the construction of our uncertainty sets:

We are interested in the set of risk measures $\mathcal{R}_1 = \mathcal{R}_{icx} \cap \mathcal{R}_{nor} \cap \mathcal{R}_{elc}$. We introduce decision variables $s = \left(s_\theta \in \mathbb{R}^{n\,|\Omega|}\right)_{\theta \in \Theta}$, $t = (t_\theta \in \mathbb{R})_{\theta \in \Theta}$, $a \in \mathbb{R}^{n\,|\Omega|}$, and $b \in \mathbb{R}$.

Theorem 6.3.1. *Fix* $X \in \mathcal{L}^n$*, then* $\sup_{\rho \in \mathcal{R}_1} \rho(X)$ *is equal to the optimal* *value of the following linear programming problem:*

$$\sup_{a,\, b,\, s,\, t} \langle a,\, X\rangle + b \tag{6.3}$$

$$s.t.\ \langle s_\theta,\, \theta'\rangle + t_\theta \leq \langle s_{\theta'},\, \theta'\rangle + t_{\theta'},\ \forall\theta,\, \theta' \in \Theta,\, \theta \neq \theta', \tag{6.4}$$

$$\langle a,\, \theta\rangle + b \leq \langle s_\theta,\, \theta\rangle + t_\theta,\ \forall\theta \in \Theta, \tag{6.5}$$

$$\langle s_\theta,\, \theta\rangle + t_\theta \leq \langle s_{\theta'},\, \theta'\rangle + t_{\theta'},\ \forall\theta \in \{W_k\}\,,\, \theta' \in \{Y_k\}\,, \tag{6.6}$$

$$\langle s_\theta,\, \theta\rangle + t_\theta = 0,\ \theta = 0, \tag{6.7}$$

$$\langle s_\theta,\, \theta\rangle + t_\theta - \langle s_{\theta'},\, \theta'\rangle - t_{\theta'} = 1,\ \theta = W_0,\, \theta' = Y_0, \tag{6.8}$$

$$s_\theta \geq 0,\ \forall\theta \in \Theta, \tag{6.9}$$

$$a \geq 0. \tag{6.10}$$

If Problem (6.3) - (6.10) has an optimal solution $\left(\hat{a},\, \hat{b},\, \hat{s},\, \hat{t}\right)$*, then*

$$\hat{\rho}\,(X) = \langle \hat{a},\, X\rangle + \hat{b}$$

is the value of the worst-case risk measure of X*.*

Proof. Constraints (6.4) and (6.8) ensure that $\rho\,(\theta)$ takes the values of a risk measure in $\mathcal{R}_{icx}$ evaluated on Θ by Lemma 6.3.1. Constraint (6.5) ensures that we find out the worst-case risk measure in $\mathcal{R}_{icx}$ evaluated on Θ by Lemma 6.3.2. Constraint (6.6) requires that $\rho\,(\theta)$ takes the values of a risk measure in $\mathcal{R}_{elc}$. Constraint (6.7) require the risk measure takes the values of a risk measure in $\mathcal{R}_{nor}$. □

The preceding proposition allows us to compute the worst-case risk by solving a linear programming problem. Next we aim to optimize over the decision set and turn attention back to Problem (6.2). We introduce decision variables $\alpha = (\alpha_{\theta,\theta'})_{\theta,\theta'\in\Theta,\,\theta\neq\theta'} \in \mathbb{R}^{|\Theta|(|\Theta|-1)}$, $\beta = (\beta_\theta)_{\theta\in\Theta} \in \mathbb{R}^{|\Theta|}$, $\gamma = (\gamma_{\theta_k})_{k\in\mathbb{K}} \in \mathbb{R}^{|\mathbb{K}|}$, $\delta,\, \varepsilon \in \mathbb{R}$ and z. Then Problem (6.2) is equivalent to a convex optimization problem by the following theorem.

Theorem 6.3.2. *Problem (6.2) is equivalent to*

$$\min_{\alpha,\beta,\gamma,\delta,\varepsilon,z} \quad \varepsilon$$

$$s.t. \sum_{\theta' \neq \theta} \alpha_{\theta,\theta'}\theta' - \sum_{\theta} \alpha_{\theta,\theta'}\theta' - \beta_\theta\theta + \sum_k \gamma_{\theta_k}\theta\mathbf{1}\left\{\theta_k \in \bigcup W_k\right\} - \sum_k \gamma_{\theta_k}\theta\mathbf{1}\left\{\theta_k \in \bigcup Y_k\right\} + \delta\theta\mathbf{1}\left\{\theta = 0\right\} + \varepsilon\theta\mathbf{1}\left\{\theta = W_0\right\} - \varepsilon\theta\mathbf{1}\left\{\theta = Y_0\right\}\rangle \geq 0,$$

$$\forall\theta,\ \theta' \in \Theta,\ \theta \neq \theta',$$

$$\sum_{\theta' \neq \theta} \alpha_{\theta,\theta'} - \sum_{\theta} \alpha_{\theta,\theta'} - \beta_\theta + \sum_k \gamma_{\theta_k}\mathbf{1}\left\{\theta_k \in \bigcup W_k\right\} - \sum_k \gamma_{\theta_k}\mathbf{1}\left\{\theta_k \in \bigcup Y_k\right\} + \delta\mathbf{1}\left\{\theta = 0\right\} + \varepsilon\mathbf{1}\left\{\theta = W_0\right\} - \varepsilon\mathbf{1}\left\{\theta = Y_0\right\} = 0, \forall\theta \in \Theta,$$

$$\sum_{\theta} \beta_\theta\theta - G(z) \geq 0, \forall\theta \in \Theta,$$

$$\sum_{\theta} \beta_\theta - 1 = 0,$$

$$\alpha,\ \beta,\ \gamma \geq 0.$$

The proof is very similar to that of Theorem 4.3.2, which can be seen in Appendix B.

6.3.2 Acceptance set approach

An alternative way of defining risk measures is provided by the notion of acceptance set, that is, the set of random variables $X \in \mathcal{L}^n$ which are considered risk free by decision makers. In this subsection, we begin by defining the acceptance set for scalar-valued risk measures with multi-variate prospects. Then, we propose a tractable convex optimization problem showing how to solve Problem (6.1) by the acceptance set method.

In Jouini et al. (2004), they proved a representation result for the risk sets which have monotone, homogeneous, translation invariant and substitutability properties. We start with the definition of acceptance set.

Definition 6.3.1. *An acceptance set for scalar-valued risk measures with multi-variate prospects is a closed convex cone $\mathcal{A}$ of $\mathcal{L}^n$, containing $\mathcal{L}^n(H)$, and such that $\mathbb{R} \times \{0\}^{n-1} \not\subset \mathcal{A}$. Here, $\mathcal{L}^n(H)$ denotes the linear space of essentially bounded H-valued $\mathcal{B}$-measurable random vectors.*

Intuitively, the elements in the set $\mathcal{A}$ are considered acceptable because no additional capital is required in order to be risk free.

In Burgert and Rschendorf (2006), given an acceptance set $\mathcal{A}$ we can define a corresponding risk measure

$$\rho_{\mathcal{A}}(X) = \inf \{m \in \mathbb{R} : X + m\, e_1 \in \mathcal{A}\},$$

which is convex when $\mathcal{A}$ is convex. Here $X + m\, e_1$ denotes that all the realizations of the first entry in X are added by m. In this case, the set of all convex risk measures can be denoted as $\mathcal{R}_{icx} \cap \mathcal{R}_{eq}$. We will similarly use the notation $\mathcal{R} \subset \mathcal{R}_{icx} \cap \mathcal{R}_{eq}$ to denote an uncertainty set of risk functions.

In this case, we think that investors are able to aggregate their portfolios to the first position. In order for a risky random vector X to be acceptable in terms of risk, the decision maker determines that some deterministic capital m should be added to the first position. We say that this deterministic capital m cancels the risk induced by X if $X + me_1$ is acceptable by decision makers in the sense of the risk measure. The risk measure of X will be the lowest amount of such deterministic m.

Using this representation, we can adapt the worst-case acceptance set approach from Delage and Li (2015). In particular, the worst-case risk measure is of the form

$$\varrho_{\mathcal{R}}(X) = \sup_{\mathcal{A}:\rho_{\mathcal{A}} \in \mathcal{R}} \inf \{m \in \mathbb{R} : X + m\, e_1 \in \mathcal{A}\}.$$

This worst-case risk measure obtains the greatest value of deterministic capital that would be required to make X risk free. Since we are interested in convex risk measures, the good thing here is that the worst-case risk measure can be equivalently considered as looking for a worst-case acceptance set $\mathcal{A}$ such that $\rho_{\mathcal{A}} = \varrho_{\mathcal{R}}$. It follows that we are going to construct an acceptance set $\mathcal{A}$ effectively. It can be noted that the larger the acceptance set $\mathcal{A}$ is, the smaller the value of m will be. Therefore, the worst-case acceptance set $\mathcal{A}$ should be the smallest convex set that covers all the acceptable points.

Meanwhile, we still consider the elicited comparison information. Given a finite collection of pairs of random vectors $W_k,\, Y_k \in \mathcal{L}^n$ indexed by $\mathbb{K}$, we define a set of the risk measures revealing these elicited comparisons: $\mathcal{R}_{elc}(\varSigma) = \{\rho(W_k) \leq \rho(Y_k)$ for all $k \in \mathbb{K}\}$ with $\varSigma \triangleq \{(W_k,\, Y_k)\}_{k=1}^{K}$. In particular, we start with a special case of risk measures $\mathcal{R}_2 = \mathcal{R}_{icx} \cap \mathcal{R}_{eq} \cap \mathcal{R}_{elc}(\varSigma_0)$ with $\varSigma_0 \triangleq \{(W_k,\, 0)\}_{k=1}^{K}$. In addition, we list the realizations of X component-wisely. Therefore, we consider $X \in \mathcal{L}$ with a vector $(X(\omega)) \in \mathbb{R}^{n\,|\Omega|}$. We show in this case how $\sup_{\mathcal{A}:\, \rho_{\mathcal{A}} \in \mathcal{R}} \rho_{\mathcal{A}}(X)$ can be solved by linear programming.

Proposition 6.3.3. *Given a set of acceptable random payoffs* $\{W_k\}_{k=1}^{K}$ *and any random payoff* $G(z)$*, the value* $\sup_{\mathcal{A}:\, \rho_{\mathcal{A}} \in \mathcal{R}_2} \rho_{\mathcal{A}}(G(z))$ *with* $\varSigma_0 \triangleq$

$\{(W_k,\, 0)\}_{k=1}^K$ and $\mathcal{R}_2 \neq \varnothing$ is obtained as the optimal value of the following linear program:

$$\begin{aligned} \min_{z,\, m,\, \theta}\ & m \\ s.t.\, & G(z) + m\, e_1 \geq \sum_{k=1}^K \theta_k W_k, \\ & \theta_k \geq 0,\ \forall k, \end{aligned}$$

where $m \in \mathbb{R}$, $\theta \in \mathbb{R}^K$.

Proof. Following Burgert and Rschendorf (2006), all convex risk measures can be equivalently defined in terms of acceptance sets. We thus focus on the set of acceptance set candidates

$$\mathbb{A} := \{\mathcal{A} : \rho_{\mathcal{A}} \in \mathcal{R}_2\},$$

and we wish to evaluate $\sup_{\mathcal{A}:\, \rho_{\mathcal{A}} \in \mathcal{R}} \rho_{\mathcal{A}}(X)$. We also introduce

$$\mathcal{A}^* := \left\{ Z : \exists \theta,\, Z \geq \sum_{k=1}^K \theta_k W_k,\, \theta \geq 0 \right\}.$$

We start by showing that $\mathcal{A}^* \subseteq \mathcal{A}$ for all $\mathcal{A} \in \mathbb{A}$. $\mathcal{A}^*$ is the set of points that dominate some convex combinations of the set $0 \cup \{W_k\}_{k=1}^K$ and must therefore be included in any convex monotone set containing $0 \cup \{W_k\}_{k=1}^K$. Note that zero vector is implicitly acceptable since $\rho(0) = 0$. It can be noted that for any two sets $\mathcal{A}_1,\, \mathcal{A}_2 \in \mathbb{A}$, if $\mathcal{A}_1 \subseteq \mathcal{A}_2$, then $\rho_{\mathcal{A}_1}(G(z)) \geq \rho_{\mathcal{A}_2}(G(z))$. Therefore, $\rho_{\mathcal{A}^*}(G(z)) \geq \sup_{\mathcal{A} \in \mathbb{A}} \rho_{\mathcal{A}}(G(z))$.

We then show that $\mathcal{A}^* \in \mathbb{A}$. In other words, we verify that: a) $\mathcal{A}^*$ contains the points $\{W_k\}_{k=1}^K$; b) $\mathcal{A}^*$ is convex and monotone; c) $\rho_{\mathcal{A}^*}(0) = 0$. a) can be seen by the definition of $\mathcal{A}^*$.

For b), given $X_1,\, X_2 \in \mathcal{A}^*$, we have two convex combinations ζ^1 and ζ^2 such that $X_j \geq \sum_{k=1}^K \zeta_i^j W_k$ when $j = 1,\, 2$. Therefore, given any $0 \leq \alpha \leq 1$, we have

$$\alpha X_1 + (1-\alpha) X_2 \geq \alpha \sum_{k=1}^{K} \zeta_i^1 W_k + (1-\alpha) \sum_{k=1}^{K} \zeta_i^2 W_k$$
$$= \sum_{k=1}^{K} \left(\alpha \zeta_i^1 + (1-\alpha) \zeta_i^2\right) W_k,$$

thus $\alpha X_1 + (1-\alpha) X_2 \in \mathcal{A}^*$. Also, given $X_1 \in \mathcal{A}^*$, there exists a convex combination ζ^1 such that $X_1 \geq \sum_{k=1}^{K} \zeta_i^1 W_k$. With the same ζ^1, for any $X_2 \geq X_1$ we have $X_2 \geq X_1 \geq \sum_{k=1}^{K} \zeta_i^1 W_k$ and $X_2 \in \mathcal{A}^*$.

For c), by construction $\rho_{\mathcal{A}^*}(0) \leq 0$ since $0 \in \mathcal{A}^*$. Also, since $\mathcal{A}^* \subseteq \mathcal{A}$ for all $\mathcal{A} \in \mathbb{A}$, then $\rho_{\mathcal{A}^*}(0) \geq \rho_{\mathcal{A}}(0) = 0$.

Since $\mathcal{A}^* \in \mathbb{A}$, then $\rho_{\mathcal{A}^*}(G(z)) \leq \sup_{\mathcal{A} \in \mathbb{A}} \rho_{\mathcal{A}}(G(z))$. Therefore,

$$\rho_{\mathcal{A}^*}(G(z)) = \sup_{\mathcal{A} \in \mathbb{A}} \rho_{\mathcal{A}}(G(z)).$$

□

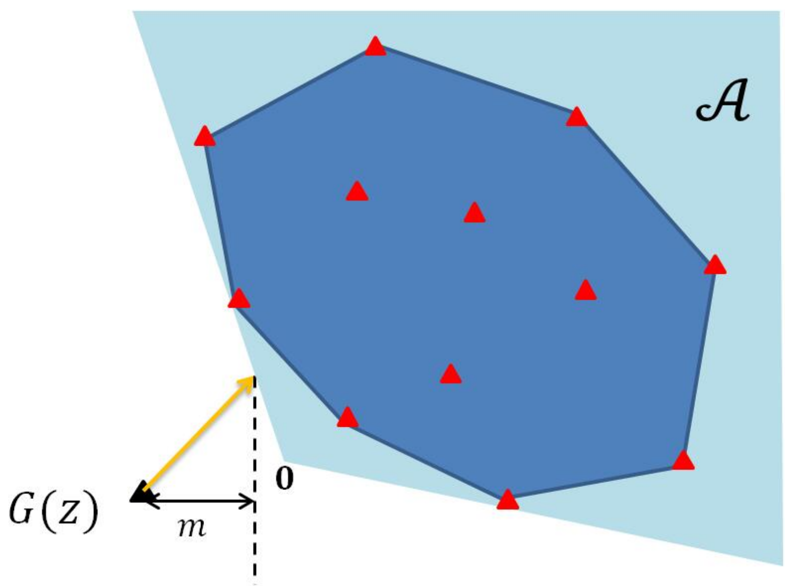

Fig. 6.1. Illustration of acceptance set method.

As shown in Figure 6.1, dark blue area is the convex hull of all the acceptable points; light blue area is the acceptance set, which is the smallest convex cone that contains all the acceptable points. Proposition 3.2 in

Delage and Li (2015) shows a similar proof which covers the univariate case. It is apparent to see the intuition here. The acceptable points are determined by the elicited comparisons, that is, those prospects $\{W_k\}_{k=1}^K$ with value greater than 0 are accepted. For all the candidate acceptance sets corresponding to $\mathcal{R}_{icx} \cap \mathcal{R}_{eq}$ and the above-mentioned acceptable points, the worst-case acceptance set is the smallest monotone polyhedron that contains the convex hull of all the acceptable points.

So far we have discussed the case where the acceptance set is determined by the elicited comparisons between prospects and 0. A more practical and interesting problem is when the elicited comparisons are made between non-zero prospects. With the worst-case acceptance set method, we can extend the results to this new case. The key thing here is still how to determine the worst-case acceptance set based on the elicited comparisons. In particular, we are interested in the set of risk measures $\mathcal{R}_3 = \mathcal{R}_{icx} \cap \mathcal{R}_{eq} \cap \mathcal{R}_{elc}(\Sigma)$ with $\Sigma \triangleq \{(W_k, Y_k)\}_{k=1}^K$. In addition, we still list the realizations of the prospect component-wisely.

Proposition 6.3.4. *Given a set of acceptable random payoffs* $\{W_k, Y_k\}_{k=1}^K$ *and any random payoff* $G(z)$, *let* $\{X_j\}_{j=1}^J = 0 \cup \{W_k, Y_k\}_{k=1}^K$ *be the set of all elements in the elicited comparisons and the zero prospect and we denote it as* $\mathbb{X}$. *The value* $\sup_{\mathcal{A}:\, \rho_{\mathcal{A}} \in \mathcal{R}_3} \rho_{\mathcal{A}}(G(z))$ *with* $\Sigma \triangleq \{(W_k, Y_k)\}_{k=1}^K$ *and* $\mathcal{R}_3 \neq \varnothing$ *is obtained as the optimal value of the following linear program:*

$$\min_{z,\, m,\, \theta} m \tag{6.11}$$

$$s.t.\; G(z) + m\, e_1 \geq \sum_{j=1}^{J} \theta_j X_j + \theta_j^{\mathrm{T}} \bar{\delta}_j e_1, \tag{6.12}$$

$$\theta_j \geq 0,\, \forall j, \tag{6.13}$$

where $m \in \mathbb{R}$, $\theta \in \mathbb{R}^J$, *and where* $\bar{\delta} \in \mathbb{R}^J$ *is the optimal solution of the* *linear programming:*

$$\min_{\delta,\,\{y_j\}_{j=1}^J} \sum_{j=1}^{J} \delta_j \tag{6.14}$$

$$s.t.\, \delta_i \le \delta_j,\, \forall (i,\, j) \in \bar{\Sigma} \tag{6.15}$$

$$(X_i - X_j)^{\mathrm{T}} y_j + \delta_i e_1 - \delta_j e_1 \ge 0,\, \forall i \ne j, \tag{6.16}$$

$$\theta_j \ge 0,\, \forall j, \tag{6.17}$$

$$\delta_1 = 0, \tag{6.18}$$

where each $y_j \in \mathbb{R}^n$ *and* $\bar{\Sigma}$ *is the set of pairs in the partial ordering of* $\{X_j\}_{j=1}^J$ *described by the elicited comparisons:*

$$\bar{\Sigma} \coloneqq \{(i,\, j) \in \{1, 2, ..., J\}^2 : (X_i,\, X_j) \in \Sigma\}.$$

Since the proof is complicated, we just present the general idea and steps here. A similar proof with details can be seen in Delage and Li (2015).

In order to prove the above mentioned proposition, we need the following three steps. Firstly, we need the fact that $\rho_{\mathcal{A}(\mathbb{X},\,\delta)}\,(G(z))$ is non-decreasing in δ, where $\mathcal{A}\,(\mathbb{X},\,\delta)$ represents the convex hull of all the points $\{X_j + \delta_j\}_{j=1}^J$:

$$\mathcal{A}\,(\mathbb{X},\,\delta) = \left\{ Z \in \mathbb{R}^n : \exists \theta \in \mathbb{R}^J,\, Z \ge \sum_{j=1}^{J} \theta_j \left(X_j + \delta_j e_1\right),\, \theta \ge 0 \right\}.$$

Next, we need the fact that the problem $\max_{\delta\in\Delta} \sum_{j=1}^J \delta_j$ is equivalent to Problem (6.14)-(6.18). In particular, let $\bar{\delta} \in \mathbb{R}^J$ be its optimal solution, then each $\bar{\delta}_i$ is the optimal solution of $\max_{\delta\in\Delta} \delta_j$, where the set Δ is denoted by

$$\Delta \coloneqq \left\{ \delta \in \mathbb{R}^J : \rho_{\mathcal{A}(\mathbb{X},\,\delta)}\,(X_j + \delta_j e_1) \ge 0,\, i = 1,\, ...,\, J,\, \delta_1 = 0,\, \delta_i \le \delta_j,\, \forall (i,\, j) \in \bar{\Sigma} \right\}.$$

It follows that for any prospect $G(z)$, the worst-case risk measure of $G(z)$ in the set of $\mathcal{R}_3$ is the optimal value of the problem:

$$\varrho_{\mathcal{R}_3}\left(G(z)\right) = \max_{\delta\in\Delta} \rho_{\mathcal{A}(\mathbb{X},\,\delta)}\left(G(z)\right),$$

Together with the three steps, we can show that

$$\max_{\delta\in\Delta} \rho_{\mathcal{A}(\mathbb{X},\,\delta)}\left(G(z)\right) = \rho_{\mathcal{A}\left(\mathbb{X},\,\bar{\delta}\right)}\left(G(z)\right).$$

The idea here is similar to the case where we make elicited comparisons between prospects and 0. For a given prospect, the acceptance set in the worst-case is determined by the set of the feasible risk values $\delta \in \Delta$ for the prospects in the elicited comparisons $\{X_j\}_{j=1}^{J}$. Then, the worst-case risk measure is obtained by determining the monotone convex hull of the points $\{X_j + \delta_j\}_{j=1}^{J}$.

Chapter 7

Conclusion and Future Research

The fundamental goal of this thesis is to build optimization frameworks for decision makers with unknown preferences under different conditions. In brief, we answer the questions raised in Section 1.2 as follows:

1. We have proposed a new SD relationship in which utility functions are weighted against a reference utility for risk-averse decision makers.

2. We have developed an optimization framework that introduces a set of constraints of the above mentioned stochastic dominance relationship.

3. We have extended the concept of ASD to log-normal probability distributions and applied the results to mean-variance analysis and MGM strategy.

4. We provided a more general optimization framework that considers the four considerations we have mentioned.

In Chapter 4, we proposed weighted almost stochastic dominance for risk-averse decision makers, which seeks to reveal unanimous preference between two prospects by all decision makers with non-decreasing concave utility function whose marginal utility does not deviate from the reference by a maximum factor $\left(\frac{1}{\varepsilon} - 1\right)$. We also formulated our proposed WRASD in the constraints of our optimization framework. Specifically, we introduced a cut generation problem with linear programming and used piece-wise linear utility functions by interpolation with monotone concave functions. In addition, we imposed boundaries on the subgradients of utility functions so that the deviation from the reference utility was limited. The optimal solution of the cut generation problem could be derived by solving its dual problem once. In our optimization framework of robust certainty equivalent maximization problem, we solved a series of cut generation problems where a scalar target was deemed the benchmark in each iteration. The robust certainty equivalent solution was determined via a binary search. The framework in this chapter is a non-parametric representation that leads to efficient tractable linear programming.

In Chapter 5, we applied the concept of weighted almost stochastic dominance to prospects with log-normal probability distribution. In particular, we addressed the gap between the preference for the MGM strategy in the long run from the perspectives of the law of large numbers and SD. Besides adding to the theoretic debate regarding the superiority of the MGM strategy in the long run, this work also adds to the SD literature by providing additional support for the use of log-weighted almost stochastic dominance to explain clear preferences between risky prospects in practice for a wide range of decision makers. This work also presents an alternative to the mean-variance framework proposed by Markowitz (1952), which is often criticized for assumptions on normality and quadratic utility. For investments with log-normal returns, which are reasonable for investors with a sufficiently long investment horizon, we show that a geometric-mean-geometric-standard-deviation framework is more appropriate.

In Chapter 6, we firstly proposed the subgradient characterization approach to tackle risk measures. The set of risk measures was constructed by convex function interpolation as well as subgradient characterization. The ambiguity in risk preferences was tackled by the elicited comparison information. Then we aim to solve the robust optimization problem by minimizing the worst-case of the constructed risk measures. The acceptance set approach is an alternative technique. In this case, the set of risk measures consists of the positions that could get rid of risk induced by the prospect (i.e. to make it accepted) by adding to it. We showed that the acceptance set is the smallest monotone polyhedron that contains the convex hull of all the acceptable points.

So far we provided mathematical optimization models that could deal with the ambiguity in risk preferences for decision makers. On the whole some contributions have been achieved in this thesis. Nevertheless, some further research is necessary to extend our work. Some possible topics for future extension are as follows.

In Chapters 4 and 6, we restrict our attention to risk-averse decision makers and the corresponding optimization problems that can be formulated or approximately solved by LP models. The risk-averse restriction, which corresponds to concavity in utility function and convexity in risk measure, is reasonable for both explicit quantitative analysis and in real life. In contrast, the LP model assumption is made from a computational perspective. Hence, the extension from this LP model to more general and complicated optimization models (e.g., nonlinear problems) could be one of the future research directions. However, the computation and optimization would be more difficult. As a result, new analytic or approximation techniques are required for this extension.

In Chapter 5, we note that a relationship exists between the difference of means and the difference of standard deviations when we compare two log-normally distributed random variables. This relationship is useful when we aim to choose the prospect that is not dominated by all other choices in a feasible set. Incorporating this constraint in the optimization model would be of interest and significance. This will also provide a new perspective of mean-variance analysis.

In Chapter 6, we restrict our interest in scalar-valued risk measures. It is sometimes more natural to use vector-valued risk functions when dealing with multivariate prospects (Ararat et al., 2014; Molchanov and Cascos, 2016). In the case of scalar-valued risk measures, we deem that investors are able to aggregate the assets in their portfolio. However, it is not available for investors to aggregate the assets in their portfolio under some circumstances. In addition, treating each asset in the portfolio and allocating reserves separately are not beneficial to the financial agents. Therefore, it would be interesting to investigate vector-valued risk measures. The assumption of convexity could ensure computational tractability.

Bibliography

C. Ararat, A. H. Hamel, and B. Rudloff. Set-valued shortfall and divergence risk measures. *arXiv preprint arXiv:1301.1496*, 2014.

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

B. Armbruster and J. Luedtke. Models and formulations for multivariate dominance constrained stochastic programs. *IIE Transactions*, 47:1–14, 2015.

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

T. G. Bali, K. O. Demirtas, H. Levy, and A. Wolf. Bonds versus stocks: Investors age and risk taking. *Journal of Monetary Economics*, 56(6): 817830, 2009.

T. G. Bali, S. J. Brown, and K. O. Demirtas. Investing in stock market anomalies. *http://ssrn.com/abstract=1806137*, 2011.

A. Ben-Tal and M. Teboulle. An old-new concept of convex risk measures: The optimized certainty equivalent. *Mathematical Finance*, 17(3):449–476, 2007.

D. Bernoulli. Exposition of a new theory on the measurement of risk. *Econometrica*, 22(1):23–36, 1954.

S. Boyd and L. Vandenberghe. *Convex optimization.* Cambridge University Press, 2004.

L. Breiman. Investment policies for expanding businesses optimal in a long-run sense. *Naval Research Logistics Quarterly*, 7(4):647–651, 1960.

S. Brenner. The risk preferences of U.S. executives. *Management Science*, 61(6):1344–1361, 2015.

D. B. Brown and M. Sim. Satisficing measure for analysis of risky positions. *Management Science*, 55(1):71–84, 2009.

D. B. Brown, E. D. Giorgi, and M. Sim. Aspirational preferences and their representation by risk measures. *Management Science*, 58(11): 2095–2113, 2012.

C. Burgert and L. Rschendorf. Consistent risk measures for portfolio vectors. *Insurance: Mathematics and Economics*, 38(2):289–297, 2006.

I. Cascos and I. Molchanov. Multivariate risks and depth-trimmed regions. *Finance and Stochastics*, 11(3):373–397, 2007.

A. Chen, J. Kim, S. Lee, and Y Kim. Stochastic multi-objective models for network design problem. *Expert Systems with Applications*, 37:1608–1619, 2010.

E. Delage and J. Y. Li. Minimizing risk exposure when the choice of a risk measure is ambiguous. *GERAD*, 2015.

D. Dentcheva and A. Ruszczyński. Optimization with stochastic dominance constraints. *Society of Industrial and Applied Mathematics Journal of Optimization*, 14(2):548–566, 2003.

D. Dentcheva and A. Ruszczyński. Optimality and duality theory for stochastic optimization problems with nonlinear dominance constraints. *Mathematical Programming*, 99:329–350, 2004.

D. Dentcheva and A. Ruszczyński. Stochastic dynamic optimization with discounted stochastic dominance constraints. *Society of Industrial and Applied Mathematics Journal of Control and Optimization*, 47(5):2540–2556, 2008.

D. Dentcheva and A. Ruszczyński. Optimization with multivariate stochastic dominance constraints. *Mathematical Programming*, 117:111–127, 2009.

C. C. Eckel and P. J. Grossman. Men, women and risk aversion: Experimental evidence. *Handbook of Experimental Economics Results*, 1: 1061–1073, 2008.

H. Follmer and A. Schied. Convex measures of risk and trading constraints. *Finance and Stochastics*, 6:429–447, 2002.

J. Goh and M. Sim. Distributionally robust optimization and its tractable approximations. *Operations Research*, 58(4):902–917, 2010.

X. Guo, T. Post, W. K. Wong, and L. X. Zhu. Moment conditions for almost stochastic dominance. *MPRA Paper 51725, University Library of Munich, Germany.*, 2013.

X. Guo, T. Post, W. K. Wong, and L. X. Zhu. A note on almost stochastic dominance. *Economics Letters*, 121(2):252–256, 2013a.

X. Guo, X. H. Zhu, W. K. Wong, and L. X. Zhu. Almost stochastic dominance for risk-averse and risk-seeking investors. *MPRA Paper 51744, University Library of Munich, Germany*, 2013b.

W. J. Gutjahr and A. Pichler. Stochastic multi-objective optimization: a survey on non-scalarizing methods. *Annals of Operations Research*, 236: 475–499, 2013.

J. Hadar and W. R. Russell. Rules for ordering uncertain prospects. *The American Economic Review*, 59(1 (1969)):25–34, 1969.

A. H. Hamel and F Heyde. Duality for set-valued measures of risk. *SIAM Journal on Financial Mathematics*, 1(1):66–95, 2010.

A. H. Hamel, F. Heyde, and R. Birgit. Set-valued risk measures for conical market models. *Math Finan Econ*, 5:1–28, 2011.

J. S. Hammond. Simplifying the choice between uncertain prospects where preference is nonlinear. *Management Science*, 20(7):1047–1072, 1974.

G. Hanoch and H. Levy. The efficiency analysis of choices involving risk. *The Review of Economic Studies*, 36(3 (Jul., 1969)):335–346, 1969.

W. B. Haskell and R. Jain. Stochastic dominance-constrained Markov decision processes. *Society of Industrial and Applied Mathematics Journal of Control and Optimization*, 2013.

W. B. Haskell, J. Shanthikumar, and Z. Shen. Optimization with a class of multivariate integral stochastic order cconstraints. *Annals of Operations Research*, 51(1):273–303, 2013.

W. B. Haskell, J. Shanthikumar, and Z. Shen. Aspects of optimization with stochastic dominance. *Submitted for publication*, 2014.

W. B. Haskell, Z. Yang, and C. H. Tan. Robust risk-aware optimization for multivariate prospects. *Working paper*, 2016.

T. Homem-de Mello and S. Mehrotra. A cutting surface method for uncertain linear programs with polyhedral stochastic dominance constraints. *Society of Industrial and Applied Mathematics Journal of Optimization*, 20(3):1250–1273, 2009.

J. Hu, T. Homem-de Mello, and S. Mehrotra. Sample average approximation of stochastic dominance constrained programs. *Mathematical Programming*, 133:171–201, 2012.

E. Jouini, M. Meddeb, and N. Touzi. Vector-valued coherent risk measures. *Finance and Stochastics*, 8:531–552, 2004.

D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–292, 1979.

J. L. Kelly. A new interpretation of information rate. *Bell System Tech Journal*, 35(4):917–926, 1956.

E. L. Lehmann. Ordered families of distributions. *Annals of Mathematical Statistics*, 26:399–419, 1955.

M. Leshno and H. Levy. Preferred by "all" and preferred by "most" decision makers: Almost stochastic dominance. *Management Science*, 48(8): 1074–1085, August 2002.

H. Levy. Stochastic dominance among log-normal prospects. *International Economic Review*, 14(3):601–614, 1973.

H. Levy. Stochastic dominance and expected utility: Survey and analysis. *Management Science*, 38.(4):555–593, 1992.

H. Levy. Aging population, retirement, and risk taking. *Management Science*, 62(5):1415–1430, 2016.

H. Levy, M. Leshno, and B. Leibovitch. Economically relevant preferences for all observed epsilon. *Annals of Operations Research*, 176(1):153–178, 2010.

M. Levy. Almost stochastic dominance and stocks for the long run. *European Journal of Operational Research*, 194:250–257, 2009.

C. K. Li and W. K. Wong. Extension of stochastic dominance theory to random variables. *RAIRO Operations Research*, 33:509–524, 1999.

A. Liefooghe, M. Basseur, L. Jourdan, and E. Talbi. Combinatorial optimization of stochastic multi-objective problems: an application to the flow-shop scheduling problem. *International Conference on Evolutionary Multi-Criterion Optimization*, pages 457–471, 2007.

A. Lizyayev and A. Ruszczyński. Tractable almost stochastic dominance. *European Journal of Operational Research*, 218(2):448–455, April 2012.

J. Luedtke. New formulations for optimization under stochastic dominance constraints. *Society of Industrial and Applied Mathematics Journal of Optimization*, 19(3):1433–1450, 2008.

H. B. Mann and D. R. Whitney. On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, 18:50–60, 1947.

H. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

H. Markowitz. Investment for the long run: New evidence for an old rule. *The Journal of Finance*, 31(5):1273–1286, 1976.

R. C. Merton and P. A. Samuelson. Fallacy of the log-normal approximation to optimal portfolio decision-making over many periods. *Journal of Financial Economics*, 1(1):67–94, 1974.

I. Molchanov and I. Cascos. Multivariate risk measures: a constructive approach based on selections. *Mathematical Finance*, 26(4):867–900, 2016.

J. P. Quirk and R. Saposnik. Admissibility and measurable utility functions. *Review of Economic Studies*, 29:140–146, 1962.

M. O. Rieger, M. Wang, and T. Hens. Risk preferences around the world. *Management Science*, 61(3):637–648, June 2015.

M. Rothschild and J. E. Stiglitz. Increasing risk: I. a definition. *Journal of Economic Theory*, 2(1970):225–243, 1970.

A. Ruszczyński and A. Shapiro. Optimization of convex risk functions. *Mathematics of Operations Research*, 31(3):433–452, 2006.

C. H. Tan. Weighted almost stochastic dominance: Revealing the preferences of most decision makers in the St. Petersburg paradox. *Decision Analysis*, 12(2):74–80, June 2015.

I. Tsetlin, R. L. Winkler, R. J. Huang, and L. Y. Tzeng. Generalized almost stochastic dominance. *Operations Research*, 63(2):363–377, 2015.

L. Y. Tzeng, R. J. Huang, and P-T. Shih. Revisiting almost second-degree stochastic dominance. *Management Science*, 59(5 (May, 2013)):1250–1254, 2013.

P. P. Wakker. Explaining the characteristics of the power (CRRA) utility family. *Health Economics*, 17:1329–1344, 2008.

G. A. Whitmore. Third-degree stochastic dominance. *The American Economic Review*, 60(3 (Jun., 1970)):457–459, 1970.

W. Wiesemann, D. Kuhn, and M. Sim. Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376, 2014.

W. K. Wong. Stochastic dominance and mean-variance measures of profit and loss for business planning and investment. *European Journal of Operational Research*, 182:829–843, 2007.

Appendix A

Proof of Theorem 4.2.1

Proof. Firstly, we show the necessity. For any differentiable real-valued function $u(x)$ and distribution $F_X(x)$, by applying integration by parts twice:

$$\begin{aligned}
\mathbb{E}\left[u\left(X\right)\right] &= \int_{x_{min}}^{x_{max}} u\left(\xi\right) dF_X\left(\xi\right) \\
&= u\left(x_{max}\right) - \int_{x_{min}}^{x_{max}} u'\left(\xi\right) F_X\left(\xi\right) d\xi \\
&= u\left(x_{max}\right) - u'\left(x_{max}\right) \int_{x_{min}}^{x_{max}} F_X\left(\xi\right) d\xi \\
&\quad - \int_{x_{min}}^{x_{max}} \left(\int_{x_{min}}^{\xi} F_X\left(\eta\right) d\eta \right) \left(-u''\left(\xi\right)\right) d\xi.
\end{aligned}$$

Define $S_1 = \{x \in \mathcal{X} \mid F_X(x) > F_Y(x)\}$ and denote the complement of S_1 as S_1^C. Hence, for any $u \in U_w\left(m, \varepsilon\right)$:

$$\begin{aligned}
& \mathbb{E}\left[u\left(X\right)\right] - \mathbb{E}\left[u\left(Y\right)\right] \\
= \; & u'\left(x_{max}\right) \int_{x_{min}}^{x_{max}} \left[F_Y\left(\xi\right) - F_X\left(\xi\right)\right] d\xi \\
& + \int_{x_{min}}^{x_{max}} \left(\int_{x_{min}}^{\xi} \left[F_Y\left(\eta\right) - F_X\left(\eta\right)\right] d\eta \right) \left(-u''\left(\xi\right)\right) d\xi \\
= \; & u'\left(x_{max}\right) \int_{x_{min}}^{x_{max}} \left[F_Y\left(\xi\right) - F_X\left(\xi\right)\right] d\xi
\end{aligned}$$

$$
\begin{aligned}
& -\int_{S_1}\left(\int_{x_{min}}^{\xi}\left[F_X(\eta)-F_Y(\eta)\right]d\eta\right)\left(-u''(\xi)\right)d\xi \\
& +\int_{S_1^C}\left(\int_{x_{min}}^{\xi}\left[F_Y(\eta)-F_X(\eta)\right]d\eta\right)\left(-u''(\xi)\right)d\xi \\
\geq\ & u'(x_{max})\int_{x_{min}}^{x_{max}}\left[F_Y(\xi)-F_X(\xi)\right]d\xi \\
& -\int_{S_1}\left(\int_{x_{min}}^{\xi}\left[F_X(\eta)-F_Y(\eta)\right]d\eta\right)\left(-u''(\xi)\right)d\xi \qquad \text{(A.1)} \\
\geq\ & u'(x_{max})\int_{x_{min}}^{x_{max}}\left[F_Y(\xi)-F_X(\xi)\right]d\xi \\
& -\max_{x\in S_1}\left\{\int_{x_{min}}^{x}\left[F_X(\eta)-F_Y(\eta)\right]d\eta\right\}\cdot\int_{S_1}\left(-u''(\xi)\right)d\xi \\
\geq\ & u'(x_{max})\int_{x_{min}}^{x_{max}}\left[F_Y(\xi)-F_X(\xi)\right]d\xi \\
& -\left[u'(x_{min})-u'(x_{max})\right]\max_{x\in S_1}\left\{\int_{x_{min}}^{x}\left[F_X(\eta)-F_Y(\eta)\right]d\eta\right\} \qquad \text{(A.2)} \\
=\ & u'(x_{max})\int_{x_{min}}^{x_{max}}\left[F_Y(\xi)-F_X(\xi)\right]d\xi \\
& -\left[u'(x_{min})-u'(x_{max})\right]\max_{x\in[x_{min},x_{max}]}\left\{\int_{x_{min}}^{x}\left[F_X(\eta)-F_Y(\eta)\right]d\eta\right\} \\
\geq\ & \left(\frac{1}{\varepsilon}-1\right)^{-0.5}m(x_{max})\int_{x_{min}}^{x_{max}}\left[F_Y(\xi)-F_X(\xi)\right]d\xi \\
& -\left[\left(\frac{1}{\varepsilon}-1\right)^{0.5}m(x_{min})-\left(\frac{1}{\varepsilon}-1\right)^{-0.5}m(x_{max})\right] \\
& \cdot\max_{x\in[x_{min},x_{max}]}\left\{\int_{x_{min}}^{x}\left[F_X(\eta)-F_Y(\eta)\right]d\eta\right\} \qquad \text{(A.3)} \\
\geq\ & 0. \qquad \text{(A.4)}
\end{aligned}
$$

Equation (A.1) follows that $\int_{S_1^C}\left(\int_{x_{min}}^{\xi}\left[F_Y(\eta)-F_X(\eta)\right]d\eta\right)\left(-u''(\xi)\right)d\xi$ is non-negative. Equation (A.2) follows because u is concave. Equation (A.3) follows from the fact that $u'(x_{min})\leq\left(\frac{1}{\varepsilon}-1\right)^{0.5}m(x_{min})$, and $u'(x_{max})\geq\left(\frac{1}{\varepsilon}-1\right)^{-0.5}m(x_{max})$. Equation (A.4) is true by assumption.

Next, we show sufficiency by contradiction. Suppose that $\mathbb{E}\left[u(X)\right]\geq$

$\mathbb{E}\left[u\left(Y\right)\right], \forall u \in U_w\left(m, \varepsilon\right)$ and

$$\begin{aligned}&\left[\left(\frac{1}{\varepsilon}-1\right)^{0.5} m\left(x_{min}\right)-\left(\frac{1}{\varepsilon}-1\right)^{-0.5} m\left(x_{max}\right)\right]\\&\cdot \max_{x\in[x_{min},\, x_{max}]}\left\{\int_{x_{min}}^{x}\left[F_X(t)-F_Y(t)\right]dt\right\}\\>&\left(\frac{1}{\varepsilon}-1\right)^{-0.5} m\left(x_{max}\right)\int_{x_{min}}^{x_{max}}\left[F_Y(t)-F_X(t)\right]dt.\end{aligned}$$

Consider a differentiable concave utility function $u \in U_w\left(m, \varepsilon\right)$ such that if $x_{min} \leq x < x_m - \delta$,

$$u'(x) = \left(\frac{1}{\varepsilon}-1\right)^{0.5} m\left(x_{min}\right),$$

if $x_m - \delta \leq x < x_m + \delta$,

$$u'(x) = \frac{\left(\frac{1}{\varepsilon}-1\right)^{0.5} m\left(x_{min}\right)-\left(\frac{1}{\varepsilon}-1\right)^{-0.5} m\left(x_{max}\right)}{2\delta}\left(x_m-\delta-x\right)+\left(\frac{1}{\varepsilon}-1\right)^{0.5} m\left(x_{min}\right),$$

if $x_m + \delta \leq x \leq x_{max}$,

$$u'(x) = \left(\frac{1}{\varepsilon}-1\right)^{-0.5} m\left(x_{max}\right),$$

where $[x_m-\delta,\, x_m+\delta]$ is a small neighborhood of $x_m = \arg\max_{x\in[x_{min},\, x_{max}]}$ $\left\{\int_{x_{min}}^{x}\left[F_X\left(\eta\right)-F_Y\left(\eta\right)\right]d\eta\right\}$.

Using integration by parts, we obtain:

$$\begin{aligned}&\mathbb{E}\left[u\left(X\right)\right]-\mathbb{E}\left[u\left(Y\right)\right]\\=\;&u'\left(x_{max}\right)\int_{x_{min}}^{x_{max}}\left[F_Y\left(\xi\right)-F_X\left(\xi\right)\right]d\xi\\&+\int_{x_{min}}^{x_{max}}\left(\int_{x_{min}}^{\xi}\left[F_Y\left(\eta\right)-F_X\left(\eta\right)\right]d\eta\right)\left(-u''\left(\xi\right)\right)d\xi\\=\;&\left(\frac{1}{\varepsilon}-1\right)^{-0.5} m\left(x_{max}\right)\int_{x_{min}}^{x_{max}}\left[F_Y\left(\xi\right)-F_X\left(\xi\right)\right]d\xi\\&-\int_{x_m-\delta}^{x_m+\delta}\left(\int_{x_{min}}^{\xi}\left[F_X\left(\eta\right)-F_Y\left(\eta\right)\right]d\eta\right)\left(-u''\left(\xi\right)\right)d\xi \qquad \text{(A.5)}\end{aligned}$$

$$
\begin{aligned}
&= \left(\frac{1}{\varepsilon}-1\right)^{-0.5} m\left(x_{max}\right) \int_{x_{min}}^{x_{max}} \left[F_Y\left(\xi\right)-F_X\left(\xi\right)\right] d\xi \\
&\quad - \int_{x_m-\delta}^{x_m+\delta} \max_{x\in\left[x_{min},\, x_{max}\right]} \left\{\int_{x_{min}}^{x} \left[F_X\left(\eta\right)-F_Y\left(\eta\right)\right] d\eta\right\} \\
&\quad \left(-u''\left(\xi\right)\right) d\xi + o\left(\delta\right) \qquad \text{(A.6)} \\
&= \left(\frac{1}{\varepsilon}-1\right)^{-0.5} m\left(x_{max}\right) \int_{x_{min}}^{x_{max}} \left[F_Y\left(\xi\right)-F_X\left(\xi\right)\right] d\xi \\
&\quad - \frac{\left(\frac{1}{\varepsilon}-1\right)^{0.5} m\left(x_{min}\right) - \left(\frac{1}{\varepsilon}-1\right)^{-0.5} m\left(x_{max}\right)}{2\delta} \cdot 2\delta \\
&\quad \cdot \max_{x\in\left[x_{min},\, x_{max}\right]} \left\{\int_{x_{min}}^{x} \left[F_X\left(\eta\right)-F_Y\left(\eta\right)\right] d\eta\right\} + o\left(\delta\right) \\
&= \left(\frac{1}{\varepsilon}-1\right)^{-0.5} m\left(x_{max}\right) \int_{x_{min}}^{x_{max}} \left[F_Y\left(\xi\right)-F_X\left(\xi\right)\right] d\xi \\
&\quad - \left[\left(\frac{1}{\varepsilon}-1\right)^{0.5} m\left(x_{min}\right) - \left(\frac{1}{\varepsilon}-1\right)^{-0.5} m\left(x_{max}\right)\right] \\
&\quad \cdot \max_{x\in\left[x_{min},\, x_{max}\right]} \left\{\int_{x_{min}}^{x} \left[F_X\left(\eta\right)-F_Y\left(\eta\right)\right] d\eta\right\} + o\left(\delta\right)
\end{aligned}
$$

Equation (A.5) follows from the fact that $u''\left(x\right)=0$ for all $x \notin \left[x_m-\delta,\, x_m+\delta\right]$. Equation (A.6) follows from Equation (A.5) by the substitution

$$
\begin{aligned}
o\left(\delta\right) = \int_{x_m-\delta}^{x_m+\delta} &\left(\max_{x\in\left[x_{min},\, x_{max}\right]} \left\{\int_{x_{min}}^{x} \left[F_X\left(\eta\right)-F_Y\left(\eta\right)\right] d\eta\right\}\right. \\
&\left. - \int_{x_{min}}^{\xi} \left[F_X\left(\eta\right)-F_Y\left(\eta\right)\right] d\eta\right) \left(-u''\left(\xi\right)\right) d\xi
\end{aligned}
$$

By our assumption,

$$
\begin{aligned}
&\left[\left(\frac{1}{\varepsilon}-1\right)^{0.5} m\left(x_{min}\right) - \left(\frac{1}{\varepsilon}-1\right)^{-0.5} m\left(x_{max}\right)\right] \\
&\cdot \max_{x\in\left[x_{min},\, x_{max}\right]} \left\{\int_{x_{min}}^{x} \left[F_X(t)-F_Y(t)\right] dt\right\} \\
> &\left(\frac{1}{\varepsilon}-1\right)^{-0.5} m\left(x_{max}\right) \int_{x_{min}}^{x_{max}} \left[F_Y(t)-F_X(t)\right] dt.
\end{aligned}
$$

In addition, observe that $o\left(\delta\right) \to 0$ when $\delta \to 0$. Therefore, $\mathbb{E}\left[u\left(X\right)\right] < \mathbb{E}\left[u\left(Y\right)\right]$ for a sufficiently small δ which contradicts our assumption that $\mathbb{E}\left[u\left(X\right)\right] \geq \mathbb{E}\left[u\left(Y\right)\right], \forall u \in U_w\left(m,\, \varepsilon\right)$. □

Appendix B

Proof of Theorem 4.3.2

Proof. (i) The function $\psi\left(X;\, U_w\left(m_0,\, \varepsilon\right),\, Y\right)$ is explicitly (recall $v_1 = 0$)

$$\min_{v,\, s,\, a,\, b} \sum_{\omega \in \Omega} P\left(\{\omega\}\right)\left(a_\omega X\left(\omega\right) + b_\omega\right) - v_t \tag{B.1}$$

$$\text{s.t. } v_{k+1} = s_k\left(\theta_{k+1} - \theta_k\right) + v_k,\ \forall k = 1, \ldots, K-1, \tag{B.2}$$

$$s_k \geq s_{k+1},\ \forall k = 1, \ldots, K-2 \tag{B.3}$$

$$a_\omega \theta_k + b_\omega \geq v_k,\ \forall k = 1, \ldots, K,\ \forall \omega \in \Omega, \tag{B.4}$$

$$\left(\frac{1}{\varepsilon} - 1\right)^{-0.5} \beta_k \leq s_k \leq \left(\frac{1}{\varepsilon} - 1\right)^{0.5} \beta_k,\ \forall k = 1, \ldots, K-1, \tag{B.5}$$

$$a,\ s \geq 0. \tag{B.6}$$

The Lagrangian for Problem (B.1) - (B.6) is

$$
\begin{aligned}
& L\left(v,\, s,\, a,\, b,\, \lambda,\, \mu,\, \delta,\, \gamma\right) \\
= & \sum_{\omega\in\Omega} P\left(\{\omega\}\right)\left(a_\omega X\left(\omega\right)+b_\omega\right)-\sum_{k=1}^{K}\mathbf{1}_t\left(\theta_k\right)v_k \\
+ & \sum_{k=1}^{K-1}\lambda_k\left(v_{k+1}-s_k\left(\theta_{k+1}-\theta_k\right)-v_k\right) \\
+ & \sum_{k=1}^{K-2}\mu_k\left(s_{k+1}-s_k\right)+\sum_{k=1}^{K}\sum_{\omega\in\Omega}\delta_{k\omega}\left(v_k-a_\omega\theta_k-b_\omega\right) \\
+ & \sum_{k=1}^{K-1}\gamma_k^l\left(\left(\frac{1}{\varepsilon}-1\right)^{-0.5}\beta_k-s_k\right)+\sum_{k=1}^{K-1}\gamma_k^u\left(s_k-\left(\frac{1}{\varepsilon}-1\right)^{0.5}\beta_k\right).
\end{aligned}
$$

Rearranging terms gives

$$
\begin{aligned}
& L\left(v,\, s,\, a,\, b,\, \lambda,\, \mu,\, \delta,\, \gamma\right) \\
= & \sum_{k=1}^{K-1}\left(\frac{1}{\varepsilon}-1\right)^{-0.5}\beta_k\gamma_k^l-\sum_{k=1}^{K-1}\left(\frac{1}{\varepsilon}-1\right)^{0.5}\beta_k\gamma_k^u \\
+ & \sum_{k=1}^{K}v_k\left(-\mathbf{1}_t\left(\theta_k\right)+\lambda_{k-1}-\lambda_k+\sum_{\omega\in\Omega}\delta_{k\omega}\right) \\
+ & \sum_{k=1}^{K-1}s_k\left(-\lambda_k\left(\theta_{k+1}-\theta_k\right)+\mu_{k-1}-\mu_k-\gamma_k^l+\gamma_k^u\right) \\
+ & \sum_{\omega\in\Omega}a_\omega\left(P\left(\{\omega\}\right)X\left(\omega\right)-\sum_{k=1}^{K}\theta_k\delta_{k\omega}\right) \\
+ & \sum_{\omega\in\Omega}b_\omega\left(P\left(\{\omega\}\right)-\sum_{k=1}^{K}\delta_{k\omega}\right).
\end{aligned}
$$

The desired form of the dual follows, which we substitute into Problem (4.5).

(ii) All of the terms in Problem (4.9) - (4.14) are linear except for the objective $\mathbb{E}\left[G\left(z\right)\right]$ and the constraints $P\left(\{\omega\}\right)\left[G\left(z\right)\right]\left(\omega\right)-\sum_{k=1}^{K}\theta_k\delta_{k\omega}\geq 0$. The objective is concave and the function $\mathrm{P}(\{\omega\})\left[G\left(z\right)\right]\left(\omega\right)$ is concave by assumption on G. □

Appendix C

Proof of Corollary 4.3.1

Proof. (i) Follows immediately from the assumption of Lipschitz continuity of m, for $x \in [\theta_k,\, \theta_{k+1})$ we have

$$|m_0(x) - m(x)| = |m(\theta_k) - m(x)| \le L\,|x - \theta_k| < \epsilon\, L.$$

(ii) For $u \in \mathcal{U}_w(m,\, \varepsilon)$, we have

$$\left(\frac{1}{\varepsilon} - 1\right)^{-0.5} m(x) \le u'(x) \le \left(\frac{1}{\varepsilon} - 1\right)^{0.5} m(x)\,,\ \forall x \in \mathcal{X}.$$

By part (i),

$$\left(\frac{1}{\varepsilon} - 1\right)^{-0.5} (m_0(x) - \epsilon\, L) \le u'(x) \le \left(\frac{1}{\varepsilon} - 1\right)^{0.5} (m_0(x) + \epsilon\, L)\,,\ \forall x \in \mathcal{X}.$$

Now set

$$\hat{u}'(x) = \begin{cases} u'(x)\,, & \left(\frac{1}{\varepsilon} - 1\right)^{-0.5} m_0(x) \le u'(x) \le \left(\frac{1}{\varepsilon} - 1\right)^{0.5} m_0(x)\,, \\ \left(\frac{1}{\varepsilon} - 1\right)^{-0.5} m_0(x)\,, & u'(x) < \left(\frac{1}{\varepsilon} - 1\right)^{-0.5} m_0(x)\,, \\ \left(\frac{1}{\varepsilon} - 1\right)^{0.5} m_0(x)\,, & u'(x) > \left(\frac{1}{\varepsilon} - 1\right)^{0.5} m_0(x)\,. \end{cases}$$

If we define $\hat{u}(x) = \int_{x_{\min}}^{x} \hat{u}'(\xi)\, d\xi$, then $\hat{u} \in U_w(m_0,\, \varepsilon)$. We have $\|u' -$

$\hat{u}'\|_{\mathcal{C}(\mathcal{X})} \leq \left(\frac{1}{\varepsilon} - 1\right)^{-0.5} \epsilon\, L$ by construction, and integration gives the desired result. $\square$

Appendix D

Proof of Theorem 5.3.1

Proof. First, we present four technical lemmas which will be used in the proof of Theorem 5.3.1.

Lemma D.1. *For a normally distributed prospect with mean μ, standard deviation σ and cumulative distribution function F:*

$$\int F(t)dt = \frac{t}{2} + \frac{1}{2}\left[(t-\mu)erf\left(\frac{t-\mu}{\sqrt{2}\sigma}\right) + \frac{\sqrt{2}\sigma}{\sqrt{\pi}}e^{-\left(\frac{t-\mu}{\sqrt{2}\sigma}\right)^2}\right] + const.$$

Proof of Lemma D.1.

$$\begin{aligned}
F(t) &= \frac{1}{2}\left[1 + erf\left(\frac{t-\mu}{\sqrt{2}\sigma}\right)\right] \\
\int F(t)dt &= \frac{1}{2}\int 1 + erf\left(\frac{t-\mu}{\sqrt{2}\sigma}\right)dt \\
&= \frac{1}{2}\left[t + \sqrt{2}\sigma\left(\frac{t-\mu}{\sqrt{2}\sigma}erf\left(\frac{t-\mu}{\sqrt{2}\sigma}\right) + \frac{1}{\sqrt{\pi}}e^{-\left(\frac{t-\mu}{\sqrt{2}\sigma}\right)^2}\right)\right] + const \\
&= \frac{t}{2} + \frac{1}{2}\left[(t-\mu)erf\left(\frac{t-\mu}{\sqrt{2}\sigma}\right) + \frac{\sqrt{2}\sigma}{\sqrt{\pi}}e^{-\left(\frac{t-\mu}{\sqrt{2}\sigma}\right)^2}\right] + const. \square
\end{aligned}$$

Lemma D.2. *If $t_0 = \frac{\mu_X\sigma_Y - \mu_Y\sigma_X}{\sigma_Y - \sigma_X}$ then $\frac{t_0-\mu_X}{\sigma_X} = \frac{t_0-\mu_Y}{\sigma_Y}$.*

Proof of Lemma D.2. First, observe that it follows from the lemma condition that:

$$\frac{t_0 - \mu_X}{\sigma_X} = \frac{\frac{\mu_X\sigma_Y - \mu_Y\sigma_X}{\sigma_Y - \sigma_X} - \mu_X}{\sigma_X} = \frac{\mu_X - \mu_Y}{\sigma_Y - \sigma_X}$$

Next, observe that it follows from the lemma condition that:

$$\frac{t_0 - \mu_Y}{\sigma_Y} = \frac{\frac{\mu_X\sigma_Y - \mu_Y\sigma_X}{\sigma_Y - \sigma_X} - \mu_Y}{\sigma_Y} = \frac{\mu_X - \mu_Y}{\sigma_Y - \sigma_X}$$

It follows from the above two equations that $\frac{t_0-\mu_X}{\sigma_X} = \frac{t_0-\mu_Y}{\sigma_Y}$. □

Lemma D.3. *If* $t_0 = \frac{\mu_X\sigma_Y - \mu_Y\sigma_X}{\sigma_Y - \sigma_X}$ *then* $\frac{\mu_X - t_0}{\sigma_X} = \frac{\mu_X - \mu_Y}{\sigma_X - \sigma_Y}$.

Proof of Lemma D.3.

$$\frac{\mu_X - t_0}{\sigma_X} = \frac{\mu_X - \frac{\mu_X\sigma_Y - \mu_Y\sigma_X}{\sigma_Y - \sigma_X}}{\sigma_X} = \frac{\mu_Y - \mu_X}{\sigma_Y - \sigma_X} = \frac{\mu_X - \mu_Y}{\sigma_X - \sigma_Y}. \ \square$$

Lemma D.4. *Suppose* $X \sim N(\mu_X, \sigma_X^2)$*,* $Y \sim N(\mu_Y, \sigma_Y^2)$*,* $\mu_X \geq \mu_Y$ *and* $\sigma_X \neq \sigma_Y$*. X dominates Y with* $(1, \varepsilon)$*-WASD if and only if:*

$$\varepsilon \geq \frac{1}{2}\left[1 - \frac{\phi\sqrt{\pi}}{\phi\sqrt{\pi} \cdot erf(\frac{\phi}{\sqrt{2}}) + \sqrt{2}e^{-\frac{\phi^2}{2}}}\right],$$

where $\phi = \frac{\mu_X - \mu_Y}{|\sigma_X - \sigma_Y|}$ *and* $erf(\cdot)$ *denotes the Gauss error function.*

Proof of Lemma D.4. Under the lemma conditions, there is exactly one intersection point $t_0 = \frac{\mu_X\sigma_Y - \mu_Y\sigma_X}{\sigma_Y - \sigma_X}$ between F and G (Levy 2006). Hence, there are the following two possible cases to consider:

- $S_1 = \{t : -\infty < t \leq t_0\}$
- $S_1 = \{t : t_0 < t < \infty\}$

First, we consider the case where $S_1 = \{t : -\infty < t \le t_0\}$. It follows from Lemma D.1 that:

$$
\begin{aligned}
& 2\int_{S_1} [F(t) - G(t)]\, dt \\
= \;& 2\int_{-\infty}^{t_0} [F(t) - G(t)]\, dt \\
= \;& (t_0 - \mu_X) erf\left(\frac{t_0 - \mu_X}{\sqrt{2}\sigma_X}\right) + \frac{\sqrt{2}\sigma_X}{\sqrt{\pi}} e^{-\left(\frac{t_0-\mu_X}{\sqrt{2}\sigma_X}\right)^2} \\
& - \lim_{t\to-\infty}\left[(t - \mu_X) erf\,(t) + \frac{\sqrt{2}\sigma_X}{\sqrt{\pi}} e^t\right] - (t_0 - \mu_Y) erf\left(\frac{t_0 - \mu_Y}{\sqrt{2}\sigma_Y}\right) \\
& - \frac{\sqrt{2}\sigma_Y}{\sqrt{\pi}} e^{-\left(\frac{t_0-\mu_Y}{\sqrt{2}\sigma_Y}\right)^2} + \lim_{t\to-\infty}\left[(t - \mu_Y) erf\,(t) + \frac{\sqrt{2}\sigma_Y}{\sqrt{\pi}} e^t\right] \\
= \;& t_0\left[erf\left(\frac{t_0 - \mu_X}{\sqrt{2}\sigma_X}\right) - erf\left(\frac{t_0 - \mu_Y}{\sqrt{2}\sigma_Y}\right)\right] \\
& - \left[\mu_X erf\left(\frac{t_0 - \mu_X}{\sqrt{2}\sigma_X}\right) - \mu_Y erf\left(\frac{t_0 - \mu_Y}{\sqrt{2}\sigma_Y}\right)\right] \\
& + \frac{\sqrt{2}}{\sqrt{\pi}}\left[\sigma_X e^{-\left(\frac{t_0-\mu_X}{\sqrt{2}\sigma_X}\right)^2} - \sigma_Y e^{-\left(\frac{t_0-\mu_Y}{\sqrt{2}\sigma_Y}\right)^2}\right] \\
& - \lim_{t\to-\infty}\left[(\mu_Y - \mu_X) erf\,(t) + \frac{\sqrt{2}\,(\sigma_X - \sigma_Y)}{\sqrt{\pi}} e^t\right] \\
= \;& (\mu_Y - \mu_X) erf\left(\frac{t_0 - \mu_X}{\sqrt{2}\sigma_X}\right) + \frac{\sqrt{2}(\sigma_X - \sigma_Y)}{\sqrt{\pi}} e^{-\left(\frac{t_0-\mu_X}{\sqrt{2}\sigma_X}\right)^2} \\
& - \lim_{t\to-\infty}\left[(\mu_Y - \mu_X) erf\,(t) + \frac{\sqrt{2}\,(\sigma_X - \sigma_Y)}{\sqrt{\pi}} e^t\right] \qquad \text{(D.1)} \\
= \;& (\mu_Y - \mu_X) erf\left(\frac{t_0 - \mu_X}{\sqrt{2}\sigma_X}\right) - \frac{\sqrt{2}}{\sqrt{\pi}}(\sigma_Y - \sigma_X) e^{-\left(\frac{t_0-\mu_X}{\sqrt{2}\sigma_X}\right)^2} + (\mu_Y - \mu_X).
\end{aligned}
$$

Equation (D.1) follows from Lemma D.2. The last equality follows from the fact that $\lim_{t\to-\infty} erf\,(t) = -1$ and $\lim_{t\to-\infty} e^t = 0$.

In a similar fashion, it can be shown that:

$$
\begin{aligned}
& 2\int_{S_1^C} [G(t)-F(t)]\,dt \\
= \; & 2\int_{t_0}^{\infty} [G(t)-F(t)]\,dx \\
= \; & (\mu_Y-\mu_X)erf\left(\frac{t_0-\mu_X}{\sqrt{2}\sigma_X}\right) - \frac{\sqrt{2}}{\sqrt{\pi}}(\sigma_Y-\sigma_X)e^{-\left(\frac{t_0-\mu_X}{\sqrt{2}\sigma_X}\right)^2} - (\mu_Y-\mu_X).
\end{aligned}
$$

Therefore:

$$
\begin{aligned}
& \frac{\int_{S_1}[F(t)-G(t)]\,dt}{\int_{-\infty}^{\infty}|F(t)-G(t)|dt} \\
= \; & \frac{\int_{S_1}[F(t)-G(t)]\,dt}{\int_{S_1}[F(t)-G(t)]\,dt + \int_{S_1^C}[G(t)-F(t)]\,dt} \\
= \; & \frac{(\mu_Y-\mu_X)erf\left(\frac{t_0-\mu_X}{\sqrt{2}\sigma_X}\right) - \frac{\sqrt{2}}{\sqrt{\pi}}(\sigma_Y-\sigma_X)e^{-\left(\frac{t_0-\mu_X}{\sqrt{2}\sigma_X}\right)^2} + (\mu_Y-\mu_X)}{2\left\{(\mu_Y-\mu_X)erf\left(\frac{t_0-\mu_X}{\sqrt{2}\sigma_X}\right) - \frac{\sqrt{2}}{\sqrt{\pi}}(\sigma_Y-\sigma_X)e^{-\left(\frac{t_0-\mu_X}{\sqrt{2}\sigma_X}\right)^2}\right\}} \\
= \; & \frac{1}{2} + \frac{\mu_Y-\mu_X}{2\left\{(\mu_Y-\mu_X)erf\left(\frac{t_0-\mu_X}{\sqrt{2}\sigma_X}\right) - \frac{\sqrt{2}}{\sqrt{\pi}}(\sigma_Y-\sigma_X)e^{-\left(\frac{t_0-\mu_X}{\sqrt{2}\sigma_X}\right)^2}\right\}} \\
= \; & \frac{1}{2} + \frac{\frac{\sqrt{\pi}(\mu_Y-\mu_X)}{\sqrt{2}(\sigma_Y-\sigma_X)}}{2\left\{\frac{\sqrt{\pi}(\mu_Y-\mu_X)}{\sqrt{2}(\sigma_Y-\sigma_X)}erf\left(\frac{t_0-\mu_X}{\sqrt{2}\sigma_X}\right) - e^{-\left(\frac{t_0-\mu_X}{\sqrt{2}\sigma_X}\right)^2}\right\}} \\
= \; & \frac{1}{2} - \frac{\frac{\sqrt{\pi}(\mu_X-\mu_Y)}{\sqrt{2}(\sigma_X-\sigma_Y)}}{2\left\{\frac{\sqrt{\pi}(\mu_X-\mu_Y)}{\sqrt{2}(\sigma_X-\sigma_Y)}erf\left(\frac{\mu_X-t_0}{\sqrt{2}\sigma_X}\right) + e^{-\left(\frac{\mu_X-t_0}{\sqrt{2}\sigma_X}\right)^2}\right\}} \qquad \text{(D.2)} \\
= \; & \varepsilon \geq \frac{1}{2}\left[1 - \frac{\phi\sqrt{\pi}}{\phi\sqrt{\pi}\cdot erf(\frac{\phi}{\sqrt{2}}) + \sqrt{2}e^{-\frac{\phi^2}{2}}}\right]. \qquad \text{(D.3)}
\end{aligned}
$$

Equation (D.2) follows from the observation that $erf(-t) = -erf(t)$. Equation (D.3) follows from Lemma D.3 and the observation that $\sigma_X > \sigma_Y$ when $S_1 = \{t : -\infty < t \leq t_0\}$. When $S_1 = \{t : -\infty < t \leq t_0\}$, it follows from Equation 2.2 and Equation (D.3) that X dominates Y with $(1, \varepsilon)$-

WASD if and only if:

$$\varepsilon \geq \frac{1}{2}\left[1 - \frac{\phi\sqrt{\pi}}{\phi\sqrt{\pi}\cdot erf(\frac{\phi}{\sqrt{2}}) + \sqrt{2}e^{-\frac{\phi^2}{2}}}\right].$$

In a similar fashion, it can be shown that, when $S_1 = \{t : t_0 \leq t < \infty\}$, X also dominates Y with ε-ASD if and only if:

$$\varepsilon \geq \frac{1}{2}\left[1 - \frac{\phi\sqrt{\pi}}{\phi\sqrt{\pi}\cdot erf(\frac{\phi}{\sqrt{2}}) + \sqrt{2}e^{-\frac{\phi^2}{2}}}\right].$$

Hence, X dominates Y with ε-ASD if and only if:

$$\varepsilon \geq \frac{1}{2}\left[1 - \frac{\phi\sqrt{\pi}}{\phi\sqrt{\pi}\cdot erf(\frac{\phi}{\sqrt{2}}) + \sqrt{2}e^{-\frac{\phi^2}{2}}}\right]. \square$$

Proof of Theorem 5.3.1. Since $X \sim \ln N(\mu_X, \sigma_X^2)$, $Y \sim \ln N(\mu_Y, \sigma_Y^2)$, we have $\ln(X) \sim N(\mu_X, \sigma_X^2)$, $\ln(Y) \sim N(\mu_Y, \sigma_Y^2)$. It follows from Lemma D.4 that $\varepsilon \geq \frac{1}{2}\left[1 - \frac{\phi\sqrt{\pi}}{\phi\sqrt{\pi}\cdot erf(\frac{\phi}{\sqrt{2}})+\sqrt{2}e^{-\frac{\phi^2}{2}}}\right]$ if and only if:

$$\mathbb{E}[u(\ln(X))] \geq \mathbb{E}[u(\ln(Y))], \forall u \in U^*(1, \varepsilon),$$

which is equivalent to:

$$\mathbb{E}[v(X)] \geq \mathbb{E}[v(Y)], \forall v \in U^*\left(\frac{1}{t}, \varepsilon\right). \tag{D.4}$$

by setting $v(t) = u(\ln(t))$ and invoking chain rule for computing the derivative of the composition of two functions.

Since Equation (D.4) is necessary and sufficient for $\left(\frac{1}{t}, \varepsilon\right)$-WASD, $\varepsilon \geq \frac{1}{2}\left[1 - \frac{\phi\sqrt{\pi}}{\phi\sqrt{\pi}\cdot erf(\frac{\phi}{\sqrt{2}})+\sqrt{2}e^{-\frac{\phi^2}{2}}}\right]$ is also necessary and sufficient for $\left(\frac{1}{t}, \varepsilon\right)$-WASD. $\square$

$\square$