UNIVERSITY OF SOUTHAMPTON

Stochastic Programming Models and Methods for Portfolio Optimization and Risk Management

by

Rudabeh Meskarian

A thesis submitted in partial fulfillment for the degree of Doctor of Philosophy

in the

Faculty of Social and Human Sciences

School of Mathematics

November 2012

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL AND HUMAN SCIENCES
SCHOOL OF MATHEMARICS

Doctor of Philosophy

by Rudabeh Meskarian

This project is focused on stochastic models and methods and their application in portfolio optimization and risk management. In particular it involves development and analysis of novel numerical methods for solving these types of problem. First, we study new numerical methods for a general second order stochastic dominance model where the underlying functions are not necessarily linear. Specifically, we penalize the second order stochastic dominance constraints to the objective under Slater's constraint qualification and then apply the well known stochastic approximation method and the level function methods to solve the penalized problem and present the corresponding convergence analysis. All methods are applied to some portfolio optimization problems, where the underlying functions are not necessarily linear all results suggests that the portfolio strategy generated by the second order stochastic dominance model outperform the strategy generated by the Markowitz model in a sense of having higher return and lower risk. Furthermore a nonlinear supply chain problem is considered, where the performance of the level function method is compared to the cutting plane method. The results suggests that the level function method is more efficient in a sense of having lower CPU time as well as being less sensitive to the problem size. This is followed by study of multivariate stochastic dominance constraints. We propose a penalization scheme for the multivariate stochastic dominance constraint and present the analysis regarding the Slater constraint qualification. The penalized problem is solved by the level function methods and a modified cutting plane method and compared to the cutting surface method proposed in [70] and the linearized method proposed in [4]. The convergence analysis regarding the proposed algorithms are presented. The proposed numerical schemes are applied to a generic budget allocation problem where it is shown that the proposed methods outperform the linearized method when the problem size is big. Moreover, a portfolio optimization problem is considered where it is shown that the a portfolio strategy generated by the multivariate second order stochastic dominance model outperform the portfolio strategy generated by the Markowitz model in sense of having higher return and lower risk. Also the performance of the algorithms is investigated with respect to the computation time and the problem size. It is shown that the level function method and the cutting plane method outperform the cutting surface method in a sense of both having lower CPU time as well as being less sensitive to

the problem size. Finally, reward-risk analysis is studied as an alternative to stochastic dominance. Specifically, we study robust reward-risk ratio optimization. We propose two robust formulations, one based on mixture distribution, and the other based on the first order moment approach. We propose a sample average approximation formulation as well as a penalty scheme for the two robust formulations respectively and solve the latter with the level function method. The convergence analysis are presented and the proposed models are applied to Sortino ratio and some numerical test results are presented. The numerical results suggests that the robust formulation based on the first order moment results in the most conservative portfolio strategy compared to the mixture distribution model and the nominal model.

Statement of Authorship

Declaration of Authorship I, Rudabeh Meskarian, declare that the thesis entitled "Stochastic Programming Models and Methods for Portfolio Optimization and Risk Management" and the work presented in it are my own. I confirm that:

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

Signed: ..... Date: .....

List of Publications

1. R. Meskarian, H. Xu, J. Fliege, Numerical methods for stochastic programs with second order stochastic dominance constraints with applications to portfolio optimization, European Journal of Operational Research, Vol. 216, pp. 376-385, 2012.

2. H. Sun, H. Xu, R. Meskarian, Exact penalization, level function method and modified cutting-plane method for stochastic programs with second order stochastic dominance constraints, SIAM Journal of Optimization, 2012. (Accepted)

3. R. Meskarian, J. Fliege, H. Xu, Stochastic programming with multivariate second order stochastic dominance constraints with applications in portfolio optimization, Applied Mathematics and Optimization, 2012. (Under review)

Contents

List of Figures

List of Tables

Acknowledgements

I owe my deepest gratitude to my supervisor, Dr Huifu Xu and my advisor, Prof. Joerg Fliege for giving me the opportunity to do this PhD. Their critical insights and thoughtful guidance have helped me immensely during this dissertation work. Their enthusiasm and inspiration has always stimulated me to take steps forward. I am forever grateful and indebted to them for everything, and it has been a privilege to study under their supervision.

I am also grateful to my fellow CORMSIS students and staff, without whom this stimulating scientific environment would not exists and many of whom have become good friends. I am grateful to all who in one way or another helped me in carrying out this research.

I would also like to thank my family, my parents Mohammadreza Meskarian and Tahmineh Tehrani along with my brother and sister for always believing in me and supporting me in all my pursuits. I am forever grateful for all their support throughout my life.

For his support, encouragement and belief in me, especially when it was most needed. For his love and eternal patience I would like to thank my dear husband, Behrad from the bottom of my heart. Without him, this thesis would not exists.

To My Dear Husband,
Behrad

Chapter 1

Introduction

The focus of this thesis is on the development of stochastic programming models and methods for financial portfolio optimization and risk management problems. In this chapter, we give an introduction to the problems considered and discuss the methodologies used in this thesis. An outline of this thesis is provided at the end.

1.1 Motivation

The recent financial crisis around the world, the globallization of financial markets, deregulation and the increasing complexity of financial products, have raised the importance of effective financial portfolio optimization and risk management techniques. The recent changes in the financial and banking industry has resulted in significant adjustment in the risk profile of both financial and non-financial institutions. The important role of these institutions in both economic growth and financial stability has made financial portfolio optimization and risk management more important than ever.

Consequently, in this thesis we are going to study various stochastic programming methods and models, focusing on numerical, as well as the underlying theories with specific applications in financial portfolio optimization and risk management.

1.2 Financial Portfolio Optimization and Risk Management

The problem of optimizing a portfolio of finitely many assets is a classical problem in theoretical and computational finance. Since the pioneering work of Markowitz [98] it is agreed that portfolio performance should be measured in two distinct dimensions: the

expected rate of return (mean), and the risk which measures the uncertainty of the rate of return.

Portfolio selection has always been the main problem in finance, due to uncertainty about future returns. To choose amongst random variables, there are three main models that can be used, these include mean-risk models, expected utility maximization, and stochastic dominance models (for a detailed review see [118]).

Markowitz developed the mean-risk model for portfolio selection problems, where preferences among return distributions are classified using a trade-off between mean and risk [98, 99, 100].

Von Neumann and Morgenstern [141] introduced the expected utility concept to decision theory. In expected utility theory, the expected utility as a single scalar value is attached to each random variable. Preference is then defined by comparing expected utilities with a larger value preferred.

The mean-risk model has been criticized by extensive studies both in theoretical and empirical aspects. The main reason is that, there is only a limited set of circumstances under which the mean-risk model applies. The first one, is the quadratic utility functions assumption which implies that beyond some level of return, the investor's marginal utility for wealth becomes negative as their risk aversion increases with wealth. Secondly, it requires the return distribution to belong to a certain class such as normal, lognormal, exponential and uniform distributions (see, Bawa [8]). Fama [53], Breen and Savage [25] have shown that the distribution of stock process changes is inconsistent with the assumption of normal probability functions.

The above offers the motivation to search for alternative models. In this research we focus on the stochastic dominance model which has a better theoretical basis as compared to the other two models mentioned above. The stochastic dominance is a non-parametric method, which does not model an explicit utility function, but takes into account all possible forms of this function which conform to a set of restrictions. The stochastic dominance notion accounts for the entire probability distribution and employs some general condition for decision maker's risk preferences.

In the stochastic dominance theory, the first order stochastic dominance was developed by Quirk and Saposnik [111]. It requires only that the first derivative of the utility function be positive throughout or monotone increasing; therefore, it allows for risk preference, risk indifference, or risk aversion. Hadar and Russell in 1969 [68] tried to find a set of rules to make predictions about the preferences possible; as a result, second order stochastic dominance was brought into the field of economics. They eliminates risk preference by adding the restriction that the second derivative of the utility function be everywhere non-positive. Consequently, second order stochastic dominance adds assumption of global risk aversion; thus utility is everywhere concave. Second order

stochastic dominance is known as an important choice criterion in portfolio selection, as it closely represents the model of risk-averse economic behavior. However, until recently due to computational difficulties it has not been used as an alternative to other portfolio construction models. This motivates us to focus on development of efficient numerical methods for stochastic programming problems with second order stochastic dominance.

1.3 Stochastic Programming

Stochastic programming has been one of the main approaches for decision making problems in finance including portfolio optimization and risk management. Stochastic programming is mathematical programming with random parameters. Initiated in the late fifties by Dantzig and Madansky, stochastic programming provides a paradigm to include uncertainty into optimization-based decision models [26, 77]. It makes modeling possible in case the parameters needed are random, i.e. the value could be from sets, continuous or discrete. We still need to know what set it is and the corresponding behavior (probability distribution over this set). A big assumption of general stochastic programming is that the probability distributions of random parameters are known. In most cases, we can use historical data or do simulation with assumptions upon the statistical parameters.

This research, studies various modeling techniques and numerical methods based on stochastic programming that help seek optimal strategies in financial risk management.

1.4 Outline of the Thesis

In Chapter 2, we provide the literature review with regard to the portfolio optimization, development of the stochastic dominance theory and optimization methodologies. In Chapter 3, we consider a stochastic programming problem with second order stochastic dominance constraints where the underlying functions could be linear and/or nonlinear. We apply an exact penalization technique and move the constraints to the objective. We propose a series of algorithms for solving the penalized problem including: the stochastic approximation method and the level function methods. Moreover, we apply the model to a portfolio optimization problem and report series of numerical tests and compare the performance to the Markowitz model. Finally, we propose a nonlinear supply chain problem and investigate the performance of the proposed algorithms. In Chapter 4, we focus on a stochastic programming problem with multivariate stochastic dominance constraints. The multivariate stochastic dominance refers to the stochastic ordering of random vectors [40]. We discuss the Slater constraint qualification and propose an exact penalization scheme for this type of problem. We solve the penalized problem with the stochastic approximation method, the level function methods, a modified cutting plane

method and compare their performance to the cutting surface algorithm proposed in [70]. Moreover, we carry out some numerical tests and report the results. In Chapter 5, we consider a robust optimization of financial performance ratio with a focus on one-sided variability measures such as Sortino-Satchell ratio [134]. We focus on robust optimization of the problem based on the mixture distribution and the first order moment approach followed with an exact penalization technique. Moreover, we carry out some numerical test and report the results. In Chapter 6, we provide some conclusions and discuss future research directions.

Chapter 2

Literature Review

2.1 Financial Optimization and Risk Management

Most significant financial problems involve decision making under uncertainty. One example is the portfolio optimization problem. In this situation, the decision maker has numerous plausible choices, wide outcome uncertainty, and large financial implications. As discussed, there are three well established models for portfolio optimization: mean-risk models, expected utility maximization and stochastic dominance. In this chapter we provide the literature review of the optimization methods in finance including stochastic dominance and related risk measures followed by a review of stochastic optimization techniques, as well as robust optimization methods.

2.1.1 Expected utility maximization in decision theory

The concept of comparing random variables using their expected utilities dates back to 1738 [13]. However, it was only in the last century and in an economic context, that expected utility theory was extensively used. Most significantly, Von Neumann and Morgenstern [141] introduced the expected utility concept to decision theory. In expected utility theory, the expected utility as a single scalar value is attached to each random variable. Preference is then defined by comparing expected utilities with a larger value preferred.

Expected utility theory provides basis for extending a utility function defined on real numbers (outcomes) to a utility function defined on random variable. A utility value is assigned to each random variable in terms of the utility values of its outcomes and the probabilities associated with these outcomes: Given a utility function U, the expected

utility of a random variable $f(x, \xi)$ is:

$$\mathbb{E}[U(f(x,\xi))] = \int_{-\infty}^{\infty} U(f(x,\alpha))dF(\alpha),$$

where F is the distribution function of $f(x, \xi)$. Additionally, if the distribution is of a discrete nature and the random variable ξ has the outcomes $\xi_1, \ldots, \xi_m$ with probabilities $p_1, \ldots, p_m$, the expected utility of $f(x, \xi)$ is defined as:

$$\mathbb{E}[U(f(x,\xi))] = p_1 U(f(x,\xi_1)) + \cdots + p_m U(f(x,\xi_m)).$$

The main shortfall of a utility function is the assumption that they reflect the behavior of investors. The first assumption states that all the investors are rational (prefer more wealth), therefore the utility function is assumed to be nondecreasing. The second assumption is that the investors are risk averse which means that, as wealth increases, each additional growth is less valuable than the previous one. There are two attitudes categorized as risk neutral and risk seeking, which are not being considered [58, Chapter 2].

Once a utility function is constructed, one could find the efficient portfolios with respect to utility criterion by solving the following optimization problem:

$$\begin{aligned} \max \quad & \mathbb{E}[U(f(x,\xi))] \\ \text{s.t.} \quad & x \in X, \end{aligned}$$

where $x \in X$ is a decision vector with X being a nonempty convex subset of $\mathbb{R}^n$.

Overall, The expected utility maximization requires the specification of the utility function, which is a subjective task; for example, consider two utility functions belonging to the same class (nondecreasing and concave), the maximization problem could lead to a different ranking of random variables.

While it is not easy to find out the precise utility function for each investor, we can order utilities of portfolios by preference instead. Stochastic dominance manages to rank portfolios consistent with general utility functions. In Chapter 3 and 4, we will show how stochastic dominance can be used to control and manage market risk so as to construct an optimal portfolio strategy.

2.1.2 The mean-risk model

Mean-risk models were developed in early fifties for the portfolio selection problem. Under mean-risk models, two scalars are attached to each random variable: the expected return and the associated risk measure. Preference is then defined using a trade-off between the mean where a larger value is desirable and the risk where a smaller value is desirable.

Consider two portfolios with returns $f(x,\xi)$ and $f(y,\xi)$ and risks measures indicated by $\rho(f(x,\xi))$ where x is the vector containing the proportion of wealth invested in each asset and ξ is the random return of each asset. The random variable $f(x,\xi)$ is efficient if and only if there is no other portfolio, such as $f(y,\xi)$ that has higher expected return as well as lower risk. If the mentioned condition is met then $f(x,\xi)$ is an efficient portfolio. Therefore, an efficient portfolio has the lowest level of risk for a given return. The efficient portfolios are obtained by solving optimization problems; the most common formulation is to set a minimum target on the portfolios return while minimizing the risk [78]:

$$\begin{aligned} \min \quad & \rho(f(x,\xi)) \\ \text{s.t.} \quad & \mathbb{E}[f(x,\xi)] \geq R_b, \\ & x \in X, \end{aligned}$$

where $x \in X$ is a decision vector with X being a nonempty convex subset of $\mathbb{R}^n$ and R_b represent the benchmark of expected return of the portfolio set by the investor. Solving the above for different values of R_b would eventually give us a set of minimum risk portfolios (efficient portfolios) for each value of R_b.

An alternative formulation of the above problem is the one which explicitly trades risk against the return in the objective function:

$$\begin{aligned} \max \quad & \mathbb{E}[f(x,\xi)] - \lambda\rho(f(x,\xi)) \\ \text{s.t.} \quad & \lambda \geq 0, \\ & x \in X. \end{aligned}$$

Repeatedly solving the above and varying the trade-off coefficient λ would result in portfolios constructing the efficient frontier.

Mean-risk models are convenient from a computational point of view. However, depending on the risk measure used, they may lack a rational and theoretical basis for making a choice. Moreover, they use only two statistics to characterize a distribution, and thus may ignore important information. One approach is to construct mean-risk models that are consistent with expected utility maximization / stochastic dominance; this has been the research subject of several recent papers [104, 105, 106, 148] .

Markowitz [98] proposed variance as a risk measures. Variance, as one of the key statistical parameters, has been used to measure market risk in mean-risk model. Its application, however, has several drawbacks. A straightforward explanation is that variance considers extremely high and extremely low returns equally undesirable. Besides downside risk, variance also takes upside variability as risk. The analysis of pros and cons of variance can be found in [99, Chapter 9]. From the risk measures perspective, variance is not coherent. We will give the definition of coherent risk measure with examples of such measures in the next section.

2.1.3 Coherent risk measures

2.1.3.1 Coherent risk measures characteristics

Many people think of the risk inherent in a financial random variable as tied entirely to the uncertainty (inconstancy) in that variable. Risk measures, most notably the coherent measures of risk were introduced by Artzner et al. [5, 6]. They defined a coherent measure of risk as follows:

Definition 2.1. A risk measure $\rho(\xi)$, where $\xi \in \Omega$ is a random variable (future value of a portfolio) defined on probability space $(\Omega, \mathcal{F}, P)$, is called coherent if it satisfies the following conditions:

1. Translation invariance: for all $\xi \in \Omega$, and all real numbers α, we have $\rho(\xi + \alpha) = \rho(\xi) - \alpha$;
2. Subadditivity: for all ξ_1 and $\xi_2 \in \Omega$, $\rho(\xi_1 + \xi_2) \leq \rho(\xi_1) + \rho(\xi_2)$.
3. Positive homogeneity: for all $\lambda \geq 0$ and all $\xi \in \Omega$, $\rho(\lambda\xi) = \lambda\rho(\xi)$.
4. Monotonicity: for all ξ_1 and $\xi_2 \in \Omega$, with $\xi_1 \leq \xi_2$, we have $\rho(\xi_2) \leq \rho(\xi_1)$.

Translation invariance implies that by adding an amount α to the portfolio, the risk will be reduced by α because the future value of the portfolio will increase by α. Subadditivity demonstrates the diversification of the portfolio. Positive homogeneity holds, because multiplying the same position cannot lead to diversification. Monotonicity is natural. Ruszczyński and Shapiro [122] exploited representation of coherence using convex analysis from a topology perspective, they generalize the dual theorem given in Artzner et al. [6], Cheridito et al. [28], Delbaen [34], Füllmer and Schied [59] and Rockafellar et al. [116].

Several risk measures have been proved to be coherent, including Value at Risk (VaR) and Conditional Value at Risk (CVaR). In the following parts, we will discuss VaR and CVaR which have been attracting significant attention from the financial industry.

2.1.3.2 Value at Risk and Conditional Value at Risk

Value at risk [2, 83, 85, 115, 140] is the best known tail risk measure as it only takes into account the left tail of distributions, which corresponds to the largest losses. it describes the maximum loss with a specified confidence level. Value at Risk has been accepted and used in a lot of financial institutions. Let $f(x, \xi)$ denote the measure of performance (or the loss), where $x \in X$ is a decision vector with X being a subset of $\mathbb{R}^n$, and $\xi \in \mathbb{R}^m$ is a random vector. For each x, the loss $f(x, \xi)$ is a random variable

having a distribution $\mathbb{R}$ induced by that of ξ. Let $p(\xi)$ denote the probability density of ξ. The probability of $f(x,\xi)$ not exceeding a threshold α is then given by

$$\Phi(x,\alpha) = \int_{f(x,\xi)\leq\alpha} p(\xi)d\xi.$$

As a function of α for fixed x, $\Phi(x,\alpha)$ is the cumulative distribution function for the loss associated with x. It completely determines the behavior of this random variable and is fundamental in defining VaR and CVaR. In general, it is nondecreasing with respect to α.

The β-VaR for the loss random variable associated with x and specified probability level $\beta \in (0,1)$, is defined as:

$$\alpha_\beta(x) = \min\{\alpha \in \mathbb{R} : \Phi(x,\alpha) \geq \beta\}.$$

It can be seen that, $\alpha_\beta(x)$ comes out as the left endpoint of the nonempty interval consisting of the values α such that $\Phi(x,\alpha) = \beta$. This follows from $\Phi(x,\alpha)$ being continuous and nondecreasing with respect to α. In short, β-VaR gives the lowest amount of loss α that will not be exceeded with probability β, i.e.

$$\Phi(x,\alpha_\beta(x)) \geq \beta. \tag{2.1.1}$$

Similarly, β-CVaR is defined as:

$$\phi_\beta(x) = (1-\beta)^{-1} \int_{f(x,\xi)\geq\alpha_\beta(x)} f(x,\xi)p(\xi)d\xi.$$

The β-CVaR is the conditional expectation of the loss associated with x relative to that loss being $\alpha_\beta(x)$ or greater. It can be seen that the probability that $f(x,\xi) \geq \alpha_\beta(x)$ is equal to $1-\beta$.

VaR and CVaR have been widely applied in portfolio selection problems. The theory of probabilistic functions and percentiles was introduced in [140]. The problem with CVaR constraints was translated to L-shape and solved efficiently in [85]. A decomposition framework handling CVaR objectives and constraints in two-stage stochastic models was discussed in [52].

Although VaR is widely used, but there are some disadvantages about its properties. The main problem with VaR is that, it actually is not coherent because the subadditivity condition is not satisfied. This implies that the VaR of a portfolio with two assets may be greater than the sum of individual VaRs of the two assets, i.e. $\alpha_\beta(x+y) \geq \alpha_\beta(x)+\alpha_\beta(y)$, for more detail see [6]. On the other hand, CVaR is a coherent measure of risk, and has better properties compared to VaR.

It is difficult to handle CVaR because of the VaR function $\alpha_\beta(x)$ involved in the definition, unless we have an analytical representation for VaR. Rockafellar and Uryasev [115] characterized $\phi_\beta(x)$ in temrs of the function F_β defined by

$$F_\beta(x,\alpha) = \alpha + (1-\beta)^{-1} \int_{f(x,\xi)\geq\alpha} [f(x,\xi)-\alpha]_+ p(\xi) d\xi,$$

where $[f(x,\xi)-\alpha]_+ = \max(f(x,\xi)-\alpha, 0)$. The $F_\beta(x,\alpha)$ is convex and continuously differentiable with respect to α. The β-CVaR of loss associated with any $x \in X$ can be determined from the formula

$$\phi_\beta(x) = \min_{\alpha\in\mathbb{R}} F_\beta(x,\alpha).$$

Consequently

$$\min_{x\in X} \phi_\beta(x) = \min_{(\alpha,x)\in X\times\mathbb{R}} F_\beta(x,\alpha),$$

where a pair (x^*, α^*) achieves the right-hand side minimum if and only if x^* achieves the first minimum and α^* is the corresponding VaR. This is proved in [115, Theorem 1].

Dentcheva and Ruszczyński [38] showed that there is a fundamental relationship between the concept of CVaR and the second order stochastic dominance constraints. Specifically, they showed that second order stochastic dominance can be interpreted as a series of CVaR constraints for various threshold values. In the next section we will define and discuss the stochastic dominance and relation of VaR and CVaR to the second order stochastic dominance.

2.2 Stochastic Dominance

Stochastic dominance is based on an axiomatic model of risk-averse preferences [56]. It originated in the majorization theory [69] for the discrete case and was later extended to general distributions [67, 119]. Since then it has been widely used in economics and finance (see [90] for numerous references); Quirk and Saposnik [111] considered the first order stochastic dominance relation and demonstrated the connection to utility functions. Second order stochastic dominance was brought to economics by Hardar and Russel [68] and third order stochastic dominance by Whitmore [144]. A detailed discussion is given in [84].

The main difference of stochastic dominance to other portfolio selection models is that it takes into account the entire distribution of a random variables. Furthermore, it is linked to the expected utility theory [141]. However, it does not require explicitly specifying a utility function [111].

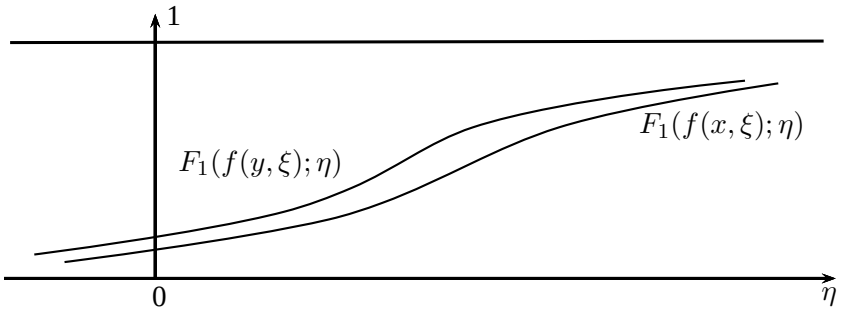

FIGURE 2.2.1: Fist order stochastic dominance [75, Chapter 9].

With stochastic dominance relations, the random variables are ranked under assumptions about general characteristics of utility functions that follow from prevalent modes of economic behavior. Stochastic dominance ensures that all individuals, whose utility functions are in the same class, rank choices in the same way.

2.2.1 Definition of stochastic dominance

In the stochastic dominance approach random variables are compared by pointwise comparison of some performance functions constructed from their distribution functions. Let $F_1(f(x,\xi);p)$ and $F_1(f(y,\xi);p)$ denote the cumulative distribution function of $f(x,\xi)$ and $f(y,\xi)$, respectively. It is said that $f(x,\xi)$ stochastically dominate $f(y,\xi)$ in the first order, denoted by $f(x,\xi) \succeq_{(1)} f(y,\xi)$, if

$$F_1(f(x,\xi);\eta) \leq F_1(f(y,\xi);\eta),\ \forall \eta \in \mathbb{R}.$$

This is illustrated in Figure 2.2.1.

Similarly, $f(x,\xi)$ stochastically dominates $f(y,\xi)$ in the second order (Figure 2.2.2), denoted by $f(x,\xi) \succeq_{(2)} f(y,\xi)$, if

$$F_2(f(x,\xi);\eta) \leq F_2(f(y,\xi);\eta),\ \forall \eta \in \mathbb{R},$$

where

$$F_2(f(x,\xi);\eta) := \int_{-\infty}^{\eta} F_1(f(x,\xi);\alpha)d\alpha,$$

see Hardar and Russell [68] and Rothschild and Stiglitz [119]. The function $F_2(f(x,\xi);\eta)$ can be expressed as the expected shortfall [104]: for each target value η we have

$$F_2(f(x,\xi);\eta) = \mathbb{E}[(\eta - f(x,\xi))_+], \tag{2.2.2}$$

where $(\eta - f(x,\xi))_+ = \max(\eta - f(x,\xi), 0)$. The function $F_2(f(x,\xi);\cdot)$ is continuous, convex, nonnegative, and nondecreasing. It is well defined for all random variables $f(x,\xi)$ with finite expected value. Due to this representation, the second order stochastic

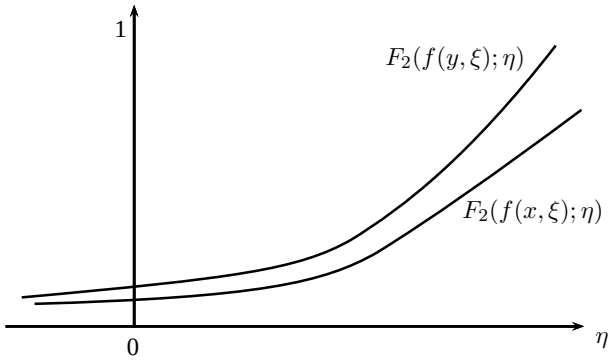

FIGURE 2.2.2: Second order stochastic dominance [75, Chapter 9].

dominance relation can be equivalently characterized by the following infinite system of inequalities:

$$\mathbb{E}[(\eta - f(x,\xi))_+] \leq \mathbb{E}[(\eta - f(y,\xi))_+],\ \forall \eta \in \mathbb{R}. \tag{2.2.3}$$

Stochastic dominance relations are of crucial importance for decision theory. It is known that $f(x,\xi) \succeq_{(1)} f(y,\xi)$ if and only if

$$\mathbb{E}[U(f(x,\xi))] \geq \mathbb{E}[U(f(y,\xi))], \tag{2.2.4}$$

for any nondecreasing function $U(\cdot)$ for which these expected values are finite. Furthermore, $f(x,\xi) \succeq_{(2)} f(y,\xi)$ if and only if (2.2.4) holds true for every nondecreasing and concave $U(\cdot)$ for which these expected values are finite [101].

A survey of stochastic dominance and utility theory can be found in [90]. The proof of consistency of the stochastic dominance with utility theory and further analysis in this topic can be found in [67, 139].

Unfortunately, application of second order stochastic dominance as a criteria of choice proves to be difficult. Generally, comparing two random variables with respect to second order stochastic dominance involves an infinite number of comparisons. Some models that use second order stochastic dominance have been proposed in the literature.

Dentcheva and Ruszczyński [36] showed that the second order stochastic dominance can be incorporated in the form of a set of linearized constraints. They proved that the second order stochastic dominance constraints construct a convex and closed set. Additionally, the optimality and duality conditions were also discussed in this paper. An alternative approach to mean-risk portfolio models was provided by using stochastic dominance. Moreover, it was shown that the Lagrange multiplier associated with the dominance constraint can be identified with a certain concave and nondecreasing utility

function. An application to static portfolio selection with a utility function constructed based on the methodology of Lagrange can be found in [38].

Optimization problems involving nonlinear stochastic dominance constraints, where stochastic dominance is used to compare nonlinear functions of random factors, were considered in [37]. Their newly developed optimality and duality theory for this special class of problems also allows the creation of a decomposition approach to the problem, which they illustrated with a portfolio example.

Roman et al. [117] proposed a multi-objective portfolio selection model with second order stochastic dominance constraints to track or outperform a reference point, while Fábián et al. [52] developed an efficient method to solve this model based on a cutting plane scheme.

The application of stochastic dominance in energy planning and decision problems, where the decision variables are integer has been discussed in [63, 64, 62] in the form of a mixed integer problem, including both first order and second order stochastic dominances. Stability and structural properties of the integer problems with dominance constraints were analysed in these papers. The authors applied a branch and bound decomposition algorithm to solve the problems.

In a more recent development, Dentcheva and Ruszczyński [40] introduced the concept of positive linear multivariate stochastic dominance and obtained necessary conditions of optimality for non-convex problems. Furthermore, Homem-de-Mello et al. [70] proposed a sample average cutting-surface algorithm for optimization problems with multidimensional polyhedral second-order stochastic dominance constraints. More recently, Hu et al. [72] proposed a new concept of stochastically weighted dominance, in which they treat the vector of weights as a random vector. They showed that such an approach is much less restrictive than the deterministic weighted approach.

2.2.2 Relation of VaR and CVaR to stochastic dominance

We first need to discuss the inverse stochastic dominance relation, which compares the Lorenz curves of two random variables and it is referred to as Lorenz dominance. For a random variable $f(x,\xi)$, we define the left-continuous inverse of the cumulative distribution function $F_1(f(x,\xi);\cdot)$ as follows:

$$F_{(-1)}(f(x,\xi);p) = \inf\{\eta : F_1(f(x,\xi);\eta) \geq p\}, \text{ for } 0 < p < 1.$$

Consequently, first order stochastic dominance can be characterized equivalently as:

$$F_{(-1)}(f(x,\xi);p) \geq F_{(-1)}(f(y,\xi);p) \ \forall p \in (0,1). \tag{2.2.5}$$

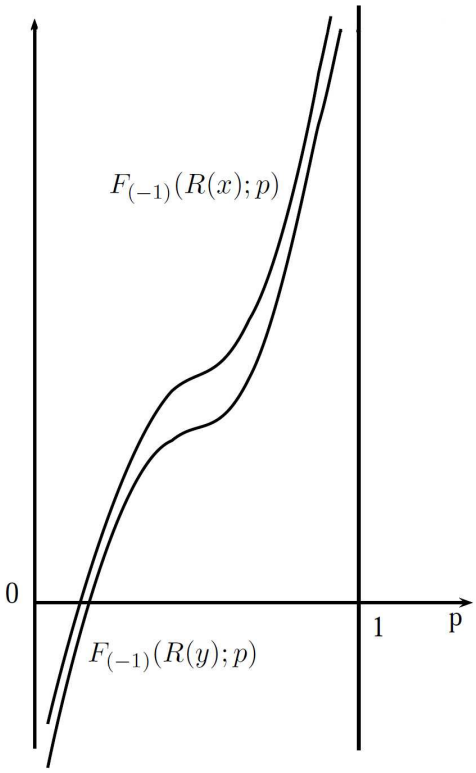

FIGURE 2.2.3: First order stochastic dominance : in the inverse form [75, Chapter 9].

This is shown in Figure 2.2.3. The first order stochastic dominance constraint can be interpreted as a continuum of chance constraints in stochastic optimization (see, [129]). Similarly, we can characterized the second order stochastic dominance by using the Lorenz function as follows:

$$F_{(-2)}(f(x,\xi);p) \geq F_{(-2)}(f(y,\xi);p)\ \forall p \in [0,1], \tag{2.2.6}$$

where

$$F_{(-2)}(f(x,\xi);p) := \int_0^p F_{(-1)}(f(x,\xi);\alpha)d\alpha.$$

This is well known from the work by Orgyczak and Ruszczyński [106]. The second order relation is illustrated in Figure 2.2.4.

Dentcheva and Ruszczyński, showed that the infinite set of inequalities (2.2.5) and (2.2.6) have relations to the concepts of VaR and CVaR, which are fundamental characteristics of portfolio return. The VaR constraint can be formulated as follows. Let $L(x,\xi) = -f(x,\xi)$. Let ω_p denote the maximum fraction of initial capital allowed for risk exposure at risk level $p \in (0,1)$. We require that

$$\mathbb{P}[L(x,\xi) \leq \omega_p] \geq 1-p.$$

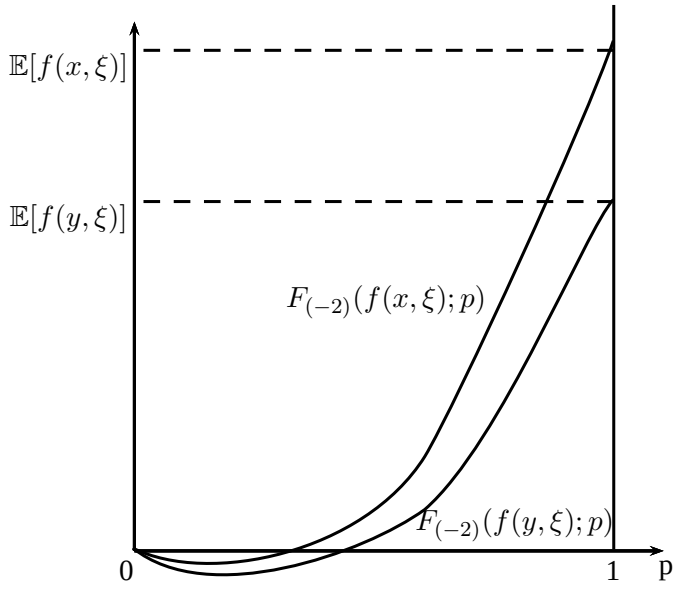

FIGURE 2.2.4: Second order stochastic dominance: in the inverse form [75, Chapter 9].

Denote the left $(1-p)$ quantile of the random variable $L(x,\xi)$ by $\mathrm{VaR}_p(L(x,\xi))$, then we can formulate VaR constraint as

$$\mathrm{VaR}_p(L(x,\xi)) \leq \omega_p.$$

Consequently, it can be seen that the first order stochastic dominance relation (2.2.5) is equivalent to the continuum of VaR constraints [75, Chapter 9]. Portfolio x dominates another portfolio y in the first order, if

$$\mathrm{VaR}_p(L(x,\xi)) \leq \mathrm{VaR}_p(L(y,\xi)),\ \forall p \in (0,1).$$

Furthermore, the CVaR at level p for continuous distribution is given by

$$\mathrm{CVaR}_p(L(x,\xi)) = \mathbb{E}[L(x,\xi)|L(x,\xi) \geq \mathrm{VaR}_p(L(x,\xi))].$$

Rockafellar and Uryasev [115] used extremal properties of quantiles to equivalently represented CVaR_p as

$$\mathrm{CVaR}_p(L(x,\xi)) = \inf_{\eta}\left\{\frac{1}{p}\mathbb{E}[\eta - f(x,\xi))_+] - \eta\right\}.$$

Notice that

$$\mathrm{CVaR}_p(L(x,\xi)) = -\frac{1}{p}F_{(-2)}(f(x,\xi),p), \tag{2.2.7}$$

using (2.2.7) and (2.2.6), it can be seen that the second order stochastic dominance is equivalent to the continuum of CVaR constraints:

$$\mathrm{CVaR}_p(L(x,\xi)) \leq \mathrm{CVaR}_p(L(x,\xi)),\ \forall p \in (0,1]. \tag{2.2.8}$$

2.3 Financial Performance Ratio

There are two basic approaches to the problem of portfolio selection under uncertainty. One of them as discussed, is the stochastic dominance approach. The other is the reward-risk analysis in which the portfolio choice is made with respect to the expected portfolio return and the associated risk. A portfolio with higher return and lower risk is preferred.

Related to reward-risk analysis is the reward-risk ratio optimization. Since the publication of the well-known *Sharpe Ratio* [132] which is based on mean-variance analysis, other performance ratios like STARR ratio, Minimax measure, Sortino ratio, Farinelli-Tibiletti ratio and most recently, Rachev ratio and the Generalized Rachev ratio have been proposed. For detailed discussion and comparison see Biglova et. al. [22], Rachev et. al. [124], and references therein. These new measures take into account the phenomena that the assets returns distributions are fat-tailed and skewed, by incorporation proper reward and risk measures.

2.3.1 Sharpe ratio

The well-known Sharpe ratio [132] of a portfolio with return $\mu(x,\xi)$ and a benchmark $Y(\xi)$ can be calculated as:

$$\Phi_{Sharpe}(\mu(x,\xi), Y(\xi) = \frac{\mathbb{E}[\mu(x,\xi) - Y(\xi)]}{\sigma(\mu(x,\xi) - Y(\xi))},$$

where σ denotes the standard deviation. Sharpe ratio quantifies reward and risk through two-sided type measures, consequently positive and negative deviations from the benchmark are weighted in the same manner. Although that may be correct in some circumstances, such as if we are aiming at capturing the "stability" around a "central tendency", that could be misleading if we are interested in keeping under control the over performance and/or the under performance. This drawback could get worse if we deal with skewed and fat tailed returns. In fact, evidence shows that investors do not share a unilateral risk aversion in rewarding and losing. A way of separately measuring reward and loss is the use of one-sided type parameter-dependent measures; that is the case of Sortino, Farinelli-Tibiletti and Rachev ratios.

2.3.2 Sortino ratio

Sortino ratio is a measure developed by F. A. Sortino to differentiate between good and bad volatility in the Sharpe ratio. This differentiation of upwards and downwards volatility allows the calculation to provide a risk-adjusted measure of a security or fund's performance without penalizing it for upward price changes. Let $\mu(x, \xi)$ be the portfolio return, then the Sortino ratio is calculated as follows:

$$\Phi_{SS}(\mu(x,\xi), Y(\xi)) = \frac{\mathbb{E}[\mu(x,\xi), Y(\xi)]}{\mathbb{E}[((Y(\xi) - \mu(x,\xi))_+)^q]^{1/q}},$$

where $Y(\xi)$ is the benchmark, $q > 0$ and denote the left orders of the performance ratio.

The Sortino ratio is similar to the Sharpe ratio, except it uses downside deviation for the denominator instead of standard deviation, the use of which does not discriminate between up and down volatility.

2.4 Stochastic Programming

Stochastic programming is the study of procedures for decision making under uncertainty over time. The uncertainty can be in the models parameters or in the model itself. Parameters may be uncertain because of lack of reliable data, future and unobservable events. The uncertainty of events, details of the problem structures and constraints and the risky payoff of decisions are modeled in an optimization framework. High performance PCs are used to enable exact and approximate algorithms to determine robust decisions that hedge against future uncertainty.

Stochastic programming provides a general purpose-modeling framework, which captures the real-world features such as turnover constraints, transaction costs, risk aversion, limits on groups of assets and other consideration. Stochastic programming models have been proposed and studied since late 1950s by Dantzing [32, 31], Beale [9], Charnes and Cooper [27] and others. They proposed a stochastic view to replace the deterministic one, where the unknown coefficients or parameters are random with assumed probability distribution that is independent of the decision variables. Stochastic program can be presented as:

$$\min_{x \in X} \{f(x) := \mathbb{E}[F(x,\xi)]\},$$

where ξ is a random variable vector having probability distribution P and X is a finite set. $F(x, \xi)$ is a real valued function of two (vector) variables x and ξ, and

$$\mathbb{E}[F(x,\xi)] = \int F(x,\xi) P(d\xi),$$

is the corresponding expected value. We assume that the expected value function $f(x)$ is well defined.

The main issue in stochastic programming besides modeling comes from the solution techniques. The main challenge in solving stochastic programming is the size of the model. It can easily grow with the increase of time horizons and the set of random parameters values. There are methods and algorithms, like Monte Carlo sampling methods [127] for solving large scale problems and stochastic approximation (SA) method which can be traced back to the pioneering work of Robbins and Monro [112].

2.4.1 Sample average approximation methods

For numerical problems in a large number of dimensions, sample average approximation methods, also known as Monte Carlo methods are often more efficient than conventional numerical methods. However, implementation of the Monte Carlo method requires sampling from high dimensional probability distributions and this may be very difficult and expensive in analysis and computer time.

Suppose that we can generate a sample of N replications of the random vector ξ. In the Monte Carlo sampling method this is accomplished by generating a random sequence $U^1, U^2, \ldots$ of numbers independent of each other and uniformly distributed on the interval $[0, 1]$, and then constructing a sample of ξ by an appropriate transformation. We can consider the sequence $\omega := \{U^1, U^2, \ldots\}$ as an element of the probability space equipped with the corresponding (product) probability measure, and the sample $\xi^i = \xi^i(\omega)$, $i = 1, 2, \ldots$ as a function of ω. Further, we could view the generated sample $\xi^1, \xi^2, \ldots$ as a sequence of random vectors, each having the same probability distribution as ξ. If the generated random vectors are (stochastically) independent of each other, we say that the sample is independent identically distributed. By $\xi^1, \xi^2, \ldots$ we denote a particular realization of the considered random sample. With the generated sample $\xi^1, \ldots, \xi^N$ we associate the sample average function

$$\hat{f}_N(x) := \frac{1}{N} \sum_{i=1}^{N} F(x, \xi^i). \tag{2.4.9}$$

Since each ξ^i has the same probability distribution as ξ , we have that for any $x \in X$,

$$\mathbb{E}[F(x, \xi^i)] = f(x), \tag{2.4.10}$$

and hence

$$\mathbb{E}[\hat{f}_N(x)] = f(x). \tag{2.4.11}$$

That is, $\hat{f}_N(x)$ is an unbiased estimator of $f(x)$. Moreover, the Law of Large Numbers (LLN) can be applied with the implication that $\hat{f}_N(x)$ converges to $f(x)$ w.p.1. uniformly

as $N \to \infty$ [121, Proposition 19]. Therefore we can say that $\hat{f}_N(x)$ is a consistent estimator of (x). This certainly holds true if the sample is independent identically distributed.

Note that the Monte Carlo method is not an algorithm, the obtained problem still has to be solved by an appropriate numerical procedure. Recent theoretical studies [82, 121, 131] and numerical experiments (see [92, 97]) show that the Monte Carlo method coupled with a good (deterministic) algorithm could be reasonably efficient for solving certain classes of stochastic programming problems.

2.4.2 Stochastic approximation methods

It is quite often that an optimization problem can be reduced to finding zeros (roots) of an unknown function $f(\cdot)$, which can be observed but the observation may be corrupted by error. This is in short the topic of stochastic approximation (SA). The error source may be observation noise, but may also come from structural inaccuracy of the observed function.

The SA method can be traced back to the pioneering work of Robbins and Monro [112] and Keifer and Wolfowitz [79]. They introduced the basic recursive algorithm for finding roots of an unknown function on the basis of noisy observations.

The Robbins-Monro algorithm and the Kiefer-Wolfowitz algorithm are the two most commonly used algorithms for unconstrained stochastic optimization. They differ in how they estimate the gradient of the objective function. The Robbins-Monro algorithm estimates the gradient directly, whereas the Kiefer-Wolfowitz algorithm uses finite differences to estimate the gradient.

Since then the SA algorithm has become widely used in stochastic optimization (see, [7, 47, 48, 49, 51, 86, 50], and references therein). This is due to the large number of applications and the interesting theoretical issues in the analysis of "dynamically defined" stochastic processes. The basic idea is a stochastic difference equation such as $\theta_{n+1} = \theta_n + \epsilon_n Y_n$, where θ_n takes its value in some Euclidean space, Y_n is a random variable, and the "step size" $\epsilon_n > 0$ is small and goes to zero as $n \to \infty$. In its simplest form, θ is a parameter of a system and the random vector Y_n is a function of noisy observations taken on the system when the parameter is set to θ_n. One recursively adjusts the parameters so that some goal is met asymptotically.

2.5 Robust Optimization

Today, stochastic programming has established itself as a powerful modeling tool when an accurate probabilistic description of the randomness is available; however, in many

real-life applications the decision-maker does not have this information, for instance when it comes to estimating financial stock returns. The need for an alternative, non-probabilistic, theory of decision-making under uncertainty has become pressing in recent years because of volatile market conditions and unstable economical states, which reduce the amount of reliable information available and make it obsolete more quickly.

Traditional models of decision making under uncertainty assume perfect information, i.e. accurate values for the system parameters and specific probability distributions for the random variables. However, such information is rarely available in practice. Soyster addressed this issue in his work [135] in the early 1970s, where every uncertain parameter in convex programming problems was taken equal to its worst-case value within a set. While this achieved the desired effect of immunizing the problem against parameter uncertainty, it was widely considered too conservative for practical implementation. Ben-Tal and Nemirovski [10, 11, 12] and El-Ghaoui and Lebret [45, 46] addressed the issue of over conservatism by restricting the uncertain parameters to belong to ellipsoidal uncertainty sets, which removes the most unlikely outcomes from consideration and yields tractable mathematical programming problems. A drawback of this method is that it increases the complexity of the problem considered, e.g., the robust counterpart of a linear programming problem is a second-order cone problem. More recently, Bertsimas and Sim [20, 21] and Bertsimas et. al. [17] have proposed a robust optimization approach based on polyhedral uncertainty sets, which preserves the class of problems under analysis, e.g., the robust counterpart of a linear programming problem remains a linear programming problem, and thus has advantages in terms of tractability in large-scale settings. It can also be connected to the decision maker's attitude towards uncertainty, providing guidelines to construct the uncertainty set from the historical realizations of the random variables using data-driven optimization [15].

2.5.1 Problem of moments

Given historical data, it is easier to estimate moment information of random parameters than to derive their probability distributions. This motivates the use of moment information in developing uncertainty models for random parameters. The problem of moments and its variations have been extensively studied and applied to many optimization problems in the literature.

The problem of moment has been studied by Stieltjes [136] in the ninetheenth century. The problem is related to the characterization of a feasible sequence of moments. Schmudgen [126], Putinar [110], and Curto and Fialkow [30] derived necessary and sufficient conditions sequences of moments with different settings. The problem of moments is also related to optimization over polynomials (the dual theory of moment). Lasserre [87] and Parrilo [107] among others proposed relaxation hierarchies for optimization over polynomials using moment results.

Bertsimas and Popescu [19] further studied the optimal inequalities given moment information. Moment problems in finance such as option pricing problems have been investigated in the literature (see [18, 94, 23]).

Chapter 3

Stochastic Programs with Second Order Stochastic Dominance Constraints

3.1 Overview

Inspired by the successful applications of the stochastic optimization with second order stochastic dominance (SSD) model in portfolio optimization, we study new numerical methods for a general SSD model where the underlying functions are not necessarily linear. Specifically, we penalize the SSD constraints to the objective and then apply the well known stochastic approximation (SA) method and the level function methods to solve the penalized problem. Both methods are iterative: the former requires the calculation of only one approximate subgradient per iteration and can be applied to the case when the underlying functions are highly nonlinear and/or non-smooth, and the distribution of the random variable may be unknown.

The main contribution of this chapter can be summarized as follows:

- We exploit a recently developed exact penalization scheme for stochastic programming models with SSD constraints and apply the stochastic approximation method and the level function methods to solve the penalized problem.
- We apply the penalization scheme and the numerical methods to some portfolio problems where the underlying return functions are not necessarily linear and present some test results. Moreover, we use real world test data to set up both backtest and out-of-sample test for investigating the performance of the portfolio based on the SSD model in comparison with the Markowitz model. Furthermore, a nonlinear supply chain problem is introduced, and the performance of the level

function methods along with the cutting plane method discussed in [81, 52] is investigated.

Throughout this chapter, we use the following notation. Let $x^T y$ denotes the scalar products of two vectors x and y, and let $\|\cdot\|$ denotes the Euclidean norm. For a real valued smooth function $h(x)$, we use $\nabla h(x)$ to denote the gradient of h at x. Let "conv" denotes the convex hull of a set.

The rest of this chapter is organized as follows. In Section 3.2, we introduce the optimization problem and discuss preliminaries needed throughout the chapter. In Section 3.3 we discuss the stochastic quasi-gradient algorithm and the level function algorithms and analyze the convergence of optimal solutions. In Section 3.4, we apply the proposed methods to portfolio optimization problems, a supply chain problem and report some numerical test results. Finally, in Section 3.5 we present some conclusions.

3.2 Stochastic Optimization Problem with SSD Constraints

3.2.1 Introduction

The notion of stochastic dominance as a constraint for optimization problems was introduced by Dentcheva and Ruszczyński [36]. The concept of stochastic dominance is fundamental when comparing two random variables, it allows one to define preference among random variables. This concept has been playing an important role in portfolio optimization. Let $g(x,\xi)$ be a concave function, with decision vector x and random variable ξ. Let $F(g(x,\xi);\eta)$ denote the cumulative distribution function of $g(x,\xi)$. We say that $g(x,\xi)$ stochastically dominates $g(y,\xi)$ in the first order, denoted by $g(x,\xi) \succeq_{(1)} g(y,\xi)$, if

$$F(g(x,\xi);\eta) \leq F(g(y,\xi);\eta),\ \forall \eta \in \mathbb{R}.$$

Similarly, $g(x,\xi)$ stochastically dominates $g(y,\xi)$ in the second order, denoted by $g(x,\xi) \succeq_{(2)} g(y,\xi)$, if

$$\int_{-\infty}^{\eta} F(g(x,\xi);\alpha)d\alpha \leq \int_{-\infty}^{\eta} F(g(y,\xi);\alpha)d\alpha,\ \forall \eta \in \mathbb{R}.$$

Consider the following optimization problem with second order stochastic dominance constraints:

$$\begin{array}{ll} \max\limits_{x} & \mathbb{E}[f(x,\xi)] \\ \text{s.t.} & g(x,\xi) \succeq_{(2)} g(y,\xi), \\ & x \in \mathcal{X}, \end{array} \tag{3.2.1}$$

where $f : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$, $g : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$, are concave continuous functions both in x and ξ, $x \in \mathcal{X}$ is a decision vector with $\mathcal{X}$ being a nonempty convex subset of $\mathbb{R}^n$, $y \in \mathcal{X}$

is a predefined vector, and $\xi : \Omega \to \Xi \subset \mathbb{R}^k$ is a random vector defined on probability space $(\Omega, \mathcal{F}, P)$ with support Ξ, $\mathbb{E}[\cdot]$ denotes the expected value w.r.t. the probability distribution of ξ.

Dentcheva and Ruszczyński analyzed several aspects of the stochastic dominance model including optimality and duality [38, 39], as well as numerical methods [36]. Roman et al. [117] proposed a multi-objective portfolio selection model with second order stochastic dominance constraints, and Fábián et al. [52] developed an efficient method to solve this model based on a cutting plane scheme.

It is well known [104, 145] that the second order stochastic dominance constraints in (3.2.1) can be reformulated as

$$\mathbb{E}[(\eta - g(x,\xi))_+] \leqslant \mathbb{E}[(\eta - g(y,\xi))_+],\ \forall \eta \in \mathbb{R},$$

where $(\eta - g(x,\xi))_+ = \max(\eta - g(x,\xi), 0)$. Consequently, problem (3.2.1) can be formulated as a stochastic semi-infinite programming problem:

$$\begin{array}{ll} \min\limits_{x} & -\mathbb{E}[f(x,\xi)] \\ \text{s.t.} & G(x,\eta) := \mathbb{E}[(\eta - g(x,\xi))_+] - \mathbb{E}[(\eta - g(y,\xi))_+] \leqslant 0,\ \forall \eta \in \mathbb{R}, \\ & x \in \mathcal{X}. \end{array} \tag{3.2.2}$$

To overcome serious technical difficulties associated with the dominance constraint, a so-called *relaxed* form of the program is proposed:

$$\begin{array}{ll} \min\limits_{x} & -\mathbb{E}[f(x,\xi)] \\ \text{s.t.} & G(x,\eta) \leqslant 0,\ \forall \eta \in [a,b], \\ & x \in \mathcal{X}, \end{array} \tag{3.2.3}$$

where $[a,b]$ is a closed interval in $\mathbb{R}$. Dentcheva and Ruszczyński [36] showed that, if ξ has uniformly bounded distribution, problem (3.2.3) is equivalent to problem (3.2.2) for some appropriate interval $[a,b]$. However, under general conditions, (3.2.3) is a relaxation of (3.2.2) in the sense that (3.2.3) has a larger set of feasible solutions and subsequently its optimal value gives a lower bound for the problem (3.2.2). Furthermore, the relaxed problem (3.2.3) is more likely to satisfy the Slater condition which is closely related to numerical stability.

3.2.2 Clarke's subgradient and exact penalization method

The focus of this chapter is on numerical methods for solving the relaxed SSD problem (3.2.3). There are three issues to deal with: (a) the expectation of random functions in both the objective and constraints, (b) the infinite number of constraints, (c) the non-smoothness resulting from the max functions.

An exact penalization scheme for problem (3.2.3) is used to move the infinite number of constraints to the objective. The penalty function method is well known [66] and has recently been used by Liu and Xu [93] for (3.2.3). Specifically, we consider the following problem:

$$\begin{array}{ll} \min\limits_{x} & \varphi(x,\rho) = -\mathbb{E}[f(x,\xi)] + \rho\vartheta(x) \\ \text{s.t.} & x \in \mathcal{X}, \end{array} \tag{3.2.4}$$

where $\rho > 0$ is a penalty parameter and

$$\vartheta(x) := \max_{\eta\in[a,b]} P(x,\eta), \tag{3.2.5}$$

where

$$P(x,\eta) := \max(G(x,\eta), 0). \tag{3.2.6}$$

Liu and Xu [93] established the equivalence between problem (3.2.2) and penalized problem (3.2.4) in the sense of optimal solutions under some moderate conditions. Penalty methods for stochastic programs have also been discussed by Branda [24] and Dupačová et al. [42].

Definition 3.1. Problem (3.2.3) is said to satisfy strong Slater condition, if there exists a positive number μ such that for any feasible point x satisfying $G(x,\eta) = 0$ for some $\eta \in [a,b]$ there exists a point x^* with $G(x^*,\eta) < 0$ for all $\eta \in T$ and

$$\|x - x^*\| \le \mu \min_{\eta\in T}(-G(x^*,\eta)). \tag{3.2.7}$$

Definition 3.2. Problem (3.2.3) is said to satisfy Slater condition, if there exists a positive number $\bar{\delta}$ and a point $\bar{x} \in \mathcal{X}$ such that

$$\max_{\eta\in T} G(x,\eta) \le -\bar{\delta}.$$

Since $\mathcal{X}$ is a compact, the Slater condition implies the strong Slater condition and then the positive number μ in Definition 3.1 can be estimated by

$$\mu := \sup_{x\in\mathcal{X}} \frac{\|x - \bar{x}\|}{\min\limits_{\eta\in T} -G(\bar{x},\eta)}. \tag{3.2.8}$$

See [65, Proposition 1 and 2] and [93] for details about the relationship.

Theorem 3.3. *[93, Theorem 2.3] Assume that problem (3.2.3) satisfies the Slater condition, that is, there exists a positive number δ and a point $\bar{x} \in \mathcal{X}$ such that*

$$\max_{\eta\in T} G(\bar{x},\eta) < -\delta. \tag{3.2.9}$$

Assume also that $\mathcal{X}$ is a compact set, and $f(\cdot,\xi)$ and $g(\cdot,\xi)$ are locally Lipschitz continuous w.r.t. x and their Lipschitz modulus are bounded by an integrable function $\kappa(\xi) > 0$.

Then there exists a positive constant $\hat{\rho}$ such that for any $\rho > \hat{\rho}$, the set of optimal solutions of the problems (3.2.2) and (3.2.4) coincide.

Proof. Let $d(x, D) := \inf_{x' \in D} \|x - x'\|$ denotes the distance from a point x to a set D. Under the Slater condition, it follows by [93, Lemma 2.5] that there exists a constant $\beta > 0$ such that

$$d(x, \mathcal{F}) \leq \beta \left\| \mathbb{E}[G(x, \eta)]_+ \right\|_\infty, \ \forall x \in \mathcal{X}, \tag{3.2.10}$$

where $\mathcal{F}$ denotes the feasible set of problem (3.2.4). Let C denotes the Lipschitz modulus of $\mathbb{E}[f(x, \xi)]$. By [29, Proposition 2.4.3] for any $\rho > \beta C$, the two optimal solutions of problems (3.2.2) and (3.2.4) coincide. Note that under the Slater condition assumption, we can set $C = \mathbb{E}[\kappa(\xi)]$. This shows the existence of a positive constant $\bar{\rho} := \beta C$. The proof is complete. □

In what follows, we focus on development of numerical methods for solving penalized optimization problem (3.2.4).

Let $v : \mathbb{R}^n \to \mathbb{R}^m$ be a locally Lipschitz continuous function. Recall that *Clarke generalized derivative* of v at point x in direction d is defined as

$$v^o(x, d) := \limsup_{y \to x, t \downarrow 0} \frac{v(y + td) - v(y)}{t}.$$

The function v is said to be *Clarke regular* at x if the usual one sided directional derivative, denoted by $v'(x, d)$, exists for all $d \in \mathbb{R}^n$ and $v^o(x, d) = v'(x, d)$. The *Clarke generalized gradient* (also known as Clarke subdifferential) is defined as

$$\partial v(x) := \{\zeta : \zeta^T d \leqslant v^o(x, d)\}.$$

see [29, Chapter 2].

Proposition 3.4. *Let $G(x, \eta)$ be defined as in (3.2.2). Assume that $g(x, \xi)$ is continuous w.r.t. x and ξ is Lipschitz continuous w.r.t. x with integrably bounded Lipschitz modulus $\kappa(\xi)$. Let $T = [a, b]$*

$$P(x, \eta) := \max(G(x, \eta), 0), \tag{3.2.11}$$

and

$$\vartheta(x) := \max_{\eta \in T} P(x, \eta). \tag{3.2.12}$$

For any fixed $x \in \mathcal{X}$, let $T^*(x)$ denotes the set of $\bar{\eta} \in T$ such that $P(x, \bar{\eta}) = \max_{\eta \in T} P(x, \eta)$. Then

$$\partial_x P(x, \eta) = \begin{cases} \{0\}, & \text{if } G(x, \eta) < 0, \\ \operatorname{conv}\{0, \partial_x G(x, \eta)\}, & \text{if } G(x, \eta) = 0, \\ \partial_x G(x, \eta), & \text{if } G(x, \eta) > 0. \end{cases} \tag{3.2.13}$$

Moreover, $\vartheta(x)$ is Lipschitz continuous with Lipschitz modulus $\mathbb{E}[\kappa(\xi)]$ and

$$\partial\vartheta(x) = \operatorname{conv}\left\{\bigcup_{\eta\in T^*(x)} \partial_x P(x,\eta)\right\}. \tag{3.2.14}$$

Proof. Since $g(x,\xi)$ is concave, then $G(x,\eta)$ is convex in x and hence it is Clarke regular, see [29, Proposition 2.3.6]. By [29, Proposition 2.3.12],

$$\partial_x[G(x,\eta)]_+ = \begin{cases} \{0\}, & \text{if } G(x,\eta) < 0, \\ \operatorname{conv}\{0, \partial_x G(x,\eta)\}, & \text{if } G(x,\eta) = 0, \\ \partial_x G(x,\eta), & \text{if } G(x,\eta) > 0. \end{cases} \tag{3.2.15}$$

The verification of Lipschitzness of $\vartheta(x)$ is straightforward. Applying the Levin-Valadier theorem (see [121, Section 2, Theorem 51]) to $\vartheta(x)$, we obtain (3.2.14). □

Remark 3.5. Under the conditions of Theorem 3.3, problem (3.2.4) is a convex minimization problem with the objective function $\varphi(x,\rho)$ being Lipschitz continuous. The optimality condition of the problem can be written as

$$0 \in -\mathbb{E}[\nabla f(x,\xi)] + \rho\partial\vartheta(x) + \mathcal{N}_{\mathcal{X}}(x), \tag{3.2.16}$$

where $\mathcal{N}_{\mathcal{X}}(x)$ denotes the normal cone to $\mathcal{X}$ at point x in the sense of convex analysis [114]. Let $P_{\mathcal{X}}(x) = \arg\min_{y\in\mathcal{X}} \|x-y\|$ denote the orthogonal projection of x on $\mathcal{X}$. Then the optimality condition (3.2.16) can be stated as follows: there exists $w \in \partial\vartheta(x)$ such that

$$P_{\mathcal{X}}(x + \mathbb{E}[\nabla f(x,\xi)] - \rho w) = x. \tag{3.2.17}$$

We will use this in Section 3.3.

3.3 Solution Methods

In this section we are concerned with numerical methods for solving problem (3.2.4). Specifically, we propose three methods: the stochastic approximation (SA) method and the level function methods to solve problem (3.2.4).

3.3.1 Stochastic approximation algorithm

In this section we discuss the stochastic approximation method for solving the penalized problem (3.2.4). One of the main reasons that we apply this method is that the objective function of (3.2.4) is non-smooth.

The stochastic approximation (SA) method can be traced back to the pioneering work of Robbins and Monro [112]. Since then the SA algorithm has become widely used in stochastic optimization (see, [7, 47, 48, 49, 51, 86, 50], and reference therein). In this section, we focus on a stochastic quasi-gradient method (SQG) which generalizes the SA method. The SQG method is a stochastic algorithmic procedure for solving general constrained optimization problems with non-differentiable, non-convex functions. Poljak [109] proposed techniques for investigating the local convergence of stochastic optimization processes and proved some results concerning differentiable optimization. A formal investigation of the asymptotic rate of convergence of SQG procedures was also carried out by Poljak [109].

Let $x_k \in \mathcal{X}$ be an approximate solution of (3.2.4). The SQG method calculates a quasi-gradient, denoted by ζ_k, of $\varphi(x,\rho)$ at x_k such that

$$\mathbb{E}[\zeta_k / \{x_0,\ldots,x_k\}] \in -\mathbb{E}[\nabla f(x_k,\xi)] + \rho\partial_x\vartheta(x_k) + \nu_k, \tag{3.3.18}$$

where ν_k is a controlled error, and by Proposition 3.4

$$\partial_x\vartheta(x_k) = \operatorname{conv}\left\{\bigcup_{\eta\in T^*(x_k)} \partial_x P(x_k,\eta)\right\}, \tag{3.3.19}$$

where $T^*(x_k)$ is the set of solutions to (3.2.12) for $x = x_k$. In order to calculate an element of $\partial_x\vartheta(x_k)$, we need to find an $\eta \in T^*(x_k)$. This amounts to solving optimization problem (3.2.12) w.r.t. η. Note that $P(x,\eta) := \max(G(x,\eta),0)$, and $G(x,\eta) = \mathbb{E}[(\eta - g(x,\xi)_+] - \mathbb{E}[(\eta - g(y,\xi)_+]$. Obviously for a fixed x, $G(x,\eta)$ is the difference of two convex functions in η, which means $P(x,\eta)$ is not a convex function in η. Homem-de-Mello et al. [70] tackled this type of challenge with a branch and cut method: reformulating the problem as a DC-programming problem and then solving it with branch and cut algorithm. Here, we propose to approximate this subgradient through sampling. Let $\xi^1,\ldots,\xi^N$ be a sampling of ξ, and w_{k_i} be a subgradient of $\vartheta(x)$ at x_k. Then we may choose

$$\zeta_k = \frac{1}{N}\sum_{i=1}^{N}\left(-\nabla f(x_k,\xi^i) + \rho w_{k_i}\right).$$

Let us explain how to calculate the w_{k_i}. By [36, Proposition 3.2], we reformulate the constraints

$$\mathbb{E}[(\eta - g(x,\xi))_+] \leqslant \mathbb{E}[(\eta - g(y,\xi))_+], \quad \forall \eta \in [a,b],$$

as

$$\mathbb{E}[(\eta_i - g(x,\xi))_+] \leqslant \mathbb{E}[(\eta_i - g(y,\xi))_+], \quad i = 1,\ldots,N,$$

where $\eta_i = g(x,\xi^i)$, $i = 1,\ldots,N$. Assume that $g(x,\xi)$ is bounded. Then we may choose the interval $[a,b]$ such that $g(x,\xi) \in [a,b]$ for all $x \in \mathcal{X}$, $\xi \in \Xi$, which means $\eta_i \in [a,b]$,

$i = 1, \ldots, N$. Consequently we can reformulate problem (3.2.12) as follows:

$$\max_{\eta_i} \quad \frac{1}{N}\sum_{i=1}^{N}(\eta_i - g(x_k, \xi^i))_+ - \frac{1}{N}\sum_{i=1}^{N}(\eta_i - g(y, \xi^i))_+. \tag{3.3.20}$$

Based on the discussions above, we present a stochastic quasi-subgradient algorithm for solving problem (3.2.4).

Algorithm 3.1 (Stochastic quasi-subgradient algorithm)

Step 1. Set a sequence of stepsizes $\{\lambda_k\}$ satisfying

$$\sum_{k=0}^{\infty}\lambda_k^2 < \infty, \quad \sum_{k=0}^{\infty}\lambda_k = \infty, \quad \lambda_k \geqslant 0. \tag{3.3.21}$$

Choose an initial vector $x_0 \in \mathcal{X}$, set $k = 0$.

Step 2. At x_k calculate an approximated subgradient of $\varphi(x, \rho)$, denoted by ζ_k, that is

$$\mathbb{E}[\zeta_k / x_0, \ldots, x_k] \in -\mathbb{E}[\nabla f(x_k, \xi)] + \rho\partial_x\vartheta(x_k) + \nu_k, \tag{3.3.22}$$

where $\partial_x\vartheta(x)$ is as in Proposition 3.4 and ν_k is a controlled error satisfying

$$\sum_{k=0}^{\infty}\mathbb{E}[\lambda_k \|\nu_k\| + \lambda_k^2 \|\zeta_k\|^2] < \infty. \tag{3.3.23}$$

Step 3. Set

$$x_{k+1} := P_{\mathcal{X}}(x_k - \lambda_k\zeta_k), \tag{3.3.24}$$

where $P_{\mathcal{X}}(x)$ is the orthogonal projection of x on $\mathcal{X}$.

Step 4. If $x_{k+1} = x_k$ and $\nu_k = 0$, stop. Otherwise, set $k := k + 1$, go to Step 2.

Let us make a comment on the stopping rule. In the case when $x_{k+1} = P_{\mathcal{X}}(x_k - \lambda_k\zeta_k) = x_k$, we have

$$-\lambda\zeta_k \in \mathcal{N}_{\mathcal{X}}(x_k),$$

and hence

$$-\zeta_k \in \mathcal{N}_{\mathcal{X}}(x_k). \tag{3.3.25}$$

Since $\nu_k = 0$, then

$$0 \in -\mathbb{E}[\nabla f(x_k, \xi)] + \rho\partial_x\vartheta(x_k) + \mathcal{N}_{\mathcal{X}}(x_k), \tag{3.3.26}$$

which, by Remark 3.5, implies that x_k is an optimal solution of (3.2.4).

Let us now consider the case that $x_k = x_{k_0}$ for $k \geq k_0$ but $\nu_k \neq 0$. By (3.3.25),

$$0 \in -\mathbb{E}[\nabla f(x_k, \xi)] + \rho\partial_x\vartheta(x_k) + \nu_k + \mathcal{N}_{\mathcal{X}}(x_k).$$

Under (3.3.23), $\nu_k \to 0$ as $k \to \infty$. By taking a limit on the equation above, we have

$$0 \in -\mathbb{E}[\nabla f(x_{k_0}, \xi)] + \rho \partial_x \vartheta(x_{k_0}) + \mathcal{N}_{\mathcal{X}}(x_{k_0}),$$

which implies that x_{k_0} satisfies the first order optimality condition and hence x_{k_0} is an optimal solution.

In what follows, we study the convergence of the general case.

Definition 3.6. Let $\|\cdot\|$ denotes the Euclidean norm. A random process $\{x_k\}$ valued in $\mathbb{R}^n$ and adapted to the filtration $\mathcal{F}_k$ is called a random quasi-Feyer sequence w.r.t. a set $S \subseteq \mathbb{R}^n$, if $\mathbb{E}[\|x_0\|] < \infty$, and for any $s \in S$,

$$\mathbb{E}[\|x_{k+1} - s\| / \mathcal{F}_k] \leqslant \|x_k - s\| + \sigma_k,$$

and

$$\sum_{k=0}^{\infty} \mathbb{E}[\sigma_k] < \infty, \ \sigma_k \geqslant 0,$$

where σ_k is an error.

Lemma 3.7. *[48, page 98] Let $\{x_k\}$ be a stochastic quasi-Feyer sequence w.r.t. Z. Then the following assertions hold.*

(i) The sequence $\left\{\|z - x_k\|^2\right\}$ converges w.p.1. for any $z \in Z$, and $\mathbb{E}[\|z - x_k\|^2] < C < \infty$ for some constant C.

(ii) The set of accumulation points of $\{x_k\}$ is not empty. Suppose that an accumulation point of $\{x_k\}$ belongs to Z. Then $\{x_k\}$ has only one limiting point.

We are now ready to present our main results.

Theorem 3.8. *Let $\{x_k\}$ be generated by the Algorithm 3.1 and let $\mathcal{X}^*$ denote the set of optimal solutions of (3.2.4). Assume: (a) $f(x, \xi)$ and $g(x, \xi)$ are concave for almost every ξ and continuous w.r.t. both x and ξ, (b) $\mathcal{X}$ is a convex compact set, (c) there exists a constant $C > 0$ such that $\mathbb{E}[\|\zeta_k\|^2 / \mathcal{F}_k] \leqslant C$, ζ_k satisfy (3.3.22), $\{\lambda_k\}$, and ν_k satisfy conditions (3.3.21) and (3.3.23) w.p.1. Then there is a subsequence $\{x_{k_i}\}$ such that $\{x_{k_i}\} \to x^*$ and $\varphi(x_{k_i}, \rho) \to \varphi(x^*, \rho)$, where $x^* \in \mathcal{X}^*$.*

Proof. Let $\{x_k\}$ be generated by (3.3.24) and $x^* \in \mathcal{X}^*$. By definition

$$\begin{aligned} \|x^* - x_{k+1}\|^2 &= \|x^* - P_X(x_k - \lambda_k \zeta_k)\|^2 \\ &\leqslant \|x^* - x_k + \lambda_k \zeta_k\|^2 \\ &= \|x^* - x_k\|^2 + 2\lambda_k \zeta_k^T (x^* - x_k) + \lambda_k^2 \|\zeta_k\|^2 . \end{aligned} \tag{3.3.27}$$

Let $\mathcal{F}_k = \{x_1, \ldots, x_k\}$. Taking the conditional expectation on both sides of the above inequality w.r.t. $\mathcal{F}_k$, we have

$$\mathbb{E}[\|x^* - x_{k+1}\|^2 / \mathcal{F}_k] \leqslant \mathbb{E}[\|x^* - x_k\|^2] + 2\lambda_k \mathbb{E}[\zeta_k / \mathcal{F}_k]^T (x^* - x_k) + \lambda_k^2 \mathbb{E}[\|\zeta_k\|^2 / \mathcal{F}_k].$$

Observe from (3.3.22) that $\mathbb{E}[\zeta_k/\mathcal{F}_k] - \nu_k \in \partial_x \varphi(x_k, \rho)$. By the convexity of $\varphi(x, \rho)$, we have

$$\mathbb{E}[\zeta_k/\mathcal{F}_k]^T (x^* - x_k) - \nu_k^T (x^* - x_k) \leqslant \varphi(x^*, \rho) - \varphi(x_k, \rho). \tag{3.3.28}$$

Using conditions (b) and (c) and taking into account that $\varphi(x^*, \rho) - \varphi(x_k, \rho) \leqslant 0$, we obtain from the above two inequalities that

$$\mathbb{E}[\|x^* - x_{k+1}\|^2 / \mathcal{F}_k] \leqslant \|x^* - x_k\|^2 + \tilde{C}(\lambda_k \|\nu_k\| + \lambda_k^2 \|\zeta_k\|^2),$$

where $\tilde{C}$ is a constant.

In view of (3.3.23) and Definition 3.6, it is clear that $\{x_k\}$ is a stochastic quasi-Feyer sequence w.r.t. $\mathcal{X}^*$. Consequently, the sequence $\|x_k - x^*\|^2 \to 0$ w.p.1.. Furthermore, the set of accumulation points of $\{x_k\}$ is not empty. Consequently, if we show that one of the accumulation points belongs to $\mathcal{X}^*$, then from condition (c) it follows that $\{x_k\}$ converges w.p.1. to a point in $\mathcal{X}^*$ [48].

Referring back to (3.3.27) and taking expectations, we have

$$\mathbb{E}[\|x^* - x_{k+1}\|^2] \leqslant \mathbb{E}[\|x^* - x_1\|^2] + 2\sum_{i=1}^{k} \lambda_i \mathbb{E}[\zeta_i/\mathcal{F}_i]^T (x^* - x_i) + \sum_{i=1}^{k} \mathbb{E}[\lambda_i^2 \|\zeta_i\|^2 / \mathcal{F}_i],$$

through (3.3.28), this yields

$$\begin{aligned}
&\mathbb{E}[\|x^* - x_{k+1}\|^2] \leqslant \\
&\mathbb{E}[\|x^* - x_1\|^2] + 2\sum_{i=1}^{k} \lambda_i \mathbb{E}[\varphi(x^*, \rho) - \varphi(x_i, \rho) + \|\nu_i\| \|x_i - x^*\|] + \sum_{i=1}^{k} \mathbb{E}[\lambda_i^2 \|\zeta_i\|^2 / \mathcal{F}_i], \\
&\leqslant \mathbb{E}[\|x^* - x_1\|^2] + 2\sum_{i=1}^{k} \lambda_i \mathbb{E}[\varphi(x^*, \rho) - \varphi(x_i, \rho)] + \tilde{C} \sum_{i=1}^{k} \mathbb{E}[\lambda_i \|\nu_i\| + \lambda_i^2 \|\zeta_i/\mathcal{F}_i\|^2].
\end{aligned}$$

This and condition (3.3.23) implies

$$\sum_{i=1}^{\infty} \lambda_i \mathbb{E}[\varphi(x_i, \rho) - \varphi(x^*, \rho)] < \infty.$$

Since

$$\sum_{i=1}^{\infty} \lambda_i = \infty \text{ and } \varphi(x_i, \rho) - \varphi(x^*, \rho) \geqslant 0,$$

then there exists a subsequence x_{k_i} such that $\varphi(x_{k_i}, \rho) - \varphi(x^*, \rho) \to 0$, w.p.1. This shows there exists a subsequence such that $\|x_{k_i} - x^*\| \to 0$ w.p.1. and this completes the proof. □

Before concluding this section, we make a few general comments on stochastic quasi-subgradient method. The stochastic quasi-subgradient method [47, 48, 49] has been developed to solve stochastic problems with complicated functions. The main advantage of this method is that, at each iteration, the search direction is a stochastic subgradient of the objective function. Another advantage of stochastic approximation methods is that it allows working directly with the samples of random variables, rather than the full distributions. However, this advantage comes at a cost. One difficulty is the choice of the stepsize. In general, choosing the stepsize requires some experimentation, and there are no hard or fast rules for making the choice.

The SQG method uses a quasi-subgradient of the objective function at each iteration. However, it might be helpful to use the subgradient information at the previous iterate. This motivates us to resort to the level function method from non-smooth optimization proposed by Lemarechal et al. [89] and extended by Xu [146].

3.3.2 Level function algorithm

In this section, we consider level function method for solving (3.2.4). The fundamental idea of the method is to use a subgradient of the objective function at each iteration to construct a linear function and treat the minimizer of the maximum of the linear function as the next iterate.

Let us start with some basic definition of the method.

Let $\alpha \in \mathbb{R}$ be a scalar and $\varphi(x, \rho)$ be a general continuous function. We use

$$T_\varphi(\alpha) = \{x \in \mathcal{X} : \varphi(x, \rho) < \alpha\},$$

to denote the strict lower level set of φ. We discuss the case where the distribution of random variable ξ is known and a subgradient could be calculated based on the available scenarios. This will aid us in the calculation of a subgradient of the objective function at each iteration.

Definition 3.9. Let $\varphi(x, \rho)$ be continuous function and $x \in \mathcal{X}$, where $\mathcal{X}$ is a nonempty convex subset of $\mathbb{R}^n$. A function $\sigma : \mathbb{R}^n \to \mathbb{R}$ is called a level function of φ at x if it satisfies the following conditions:

(a) $\sigma(x) = 0$,

(b) σ is a continuous convex function,

(c) $T_\varphi(\varphi(x,\rho)) \subset T_\sigma(0)$.

From the definition, we can see that the minimizers of φ are contained in $T_\varphi(\varphi(x,\rho))$ and x is optimal if and only if $T_\varphi(\varphi(x,\rho)) = \emptyset$.

In what follows, we apply the level function method to (3.2.4). Let $f(x,\xi)$ and $g(x,\xi)$ be continuous convex and concave functions, respectively. Let $\zeta_k \in \partial_x \varphi(x_k,\rho)$, then

$$\sigma_{x_k}(x) = \zeta_k^T (x - x_k) / \|\zeta_k\|,$$

is a level function of $\varphi(x,\rho)$ at x_k.

Algorithm 3.2 (Scaled level function algorithm)

Step 1. Let $\epsilon > 0$ be a constant, select a starting point $x_0 \in \mathcal{X}$; set k:=0.

Step 2. Calculate $\zeta_k \in \partial_x \varphi(x_k,\rho)$. Define the functions $\sigma_{x_k}(x)$ and $\sigma_k(x)$ by

$$\sigma_{x_k}(x) = \zeta_k^T (x - x_k) / \|\zeta_k\|,$$

$$\sigma_k(x) = \max\{\sigma_{k-1}(x), \sigma_{x_k}(x)\},$$

where $\sigma_{-1}(x) \equiv -\infty$. Let

$$x_{k+1} \in \arg\min_{x\in\mathcal{X}} \sigma_k(x),$$

and

$$\Delta(k) = -\sigma_k(x_{k+1}).$$

Step 3. If $\Delta(k) \leqslant \epsilon$, stop. Otherwise, set $k := k+1$, go to Step 2.

It is important to note that here we need to calculate a subgradient of $\varphi(x,\rho)$ at each iterate. This is more demanding than the stochastic approximation method where only a quasi-subgradient is calculated at each iterate. However, in some practical instances, the random variable may have a finite distribution, in that case $\varphi(x,\rho)$, can be written as a sum of a finite number of deterministic functions. Calculating a subgradient of such a function might be numerically possible. In the case when we are not able to obtain a closed form of the expected value of the underlying functions, we may use the sample average approximation method [127] to approximate $\varphi(x,\rho)$ and reduce it to a finite sum.

Theorem 3.10. *Let $\varphi(x,\rho)$ be defined as in (3.2.4) and let the assumptions of Theorem 3.3 hold. Then $\lim_{k\to\infty} \Delta(k) = 0$ and there exists a subsequence of $\{x_k\}$ converging to a global minimizer of φ over $\mathcal{X}$.*

Proof. Under the assumptions, each of the level functions $\{\sigma_{x_k}(x)\}$ generated by Algorithm 3.2 is Lipschitz on $\mathcal{X}$. The conclusion follows from [146, Theorem 3.2].

□

The Algorithm 3.2 takes a minimizer of $\sigma_k(x)$ as the next iterate, the main drawback is that it is not possible to predict the maximum number of iterations required to reduce $\Delta(k)$ to a prescribed precision. To overcome this problem, Xu [146] modified the Algorithm 3.2 by updating an iterate using projection of the current point to a level set of $\sigma_k(x)$. This projection idea belongs to Lemarechal, Nemirovskii and Nesterov [89], who applied it to convex programming.

Algorithm 3.3 (Projected level function algorithm)

Step 1. Let $\epsilon > 0$ be a constant, and select a constant $\lambda \in (0, 1)$ and a starting point $x_0 \in \mathcal{X}$; set $k = 0$.

Step 2. Calculate a level function $\sigma_{x_k}(x)$ of φ at x, and set

$$\sigma_k(x) = \max\left\{\sigma_{k-1}(x), \sigma_{x_k}(x)\right\},$$

where $\sigma_{-1} = -\infty$. Let

$$\hat{x_k} = \arg\min\left\{\varphi(x_j, \rho) : j \in 0, \ldots, k\right\},$$

and

$$x_{k+1} \in P_{Q_k}(\hat{x_k}, Q_k),$$

where

$$Q_k := \left\{x \in \mathcal{X} : \sigma_k(x) \leqslant -\lambda\Delta(k)\right\}, \quad \Delta(k) = -\min_{x\in\mathcal{X}} \sigma_k(x),$$

and P_{Q_k} is the Euclidean projection of the point x on a set Q_k.

Step 3. If $\Delta(k) \leqslant \epsilon$, stop. Otherwise, set $k := k + 1$, go to Step 2.

Note that, when $\lambda = 1$, Q_k becomes the set of minimizers of σ_k over $\mathcal{X}$. Consequently, Algorithm 3.3 becomes identical to Algorithm 3.2. The following convergence results follow directly from [146, Theorem 3.3].

Theorem 3.11. *Let $\{x_k\}$ be generated by Algorithm 3.3. Assume the conditions of Theorem 3.10. Then*

$$\Delta(k) \leqslant \epsilon, \text{ for } k > M^2 d^2 \epsilon^{-2} \lambda^{-2} (1 - \lambda^2)^{-1},$$

where ϵ is specified as in Algorithm 3.3, M is the Lipschitz modulus of φ over $\mathcal{X}$, and d is the diameter of $\mathcal{X}$ defined as

$$d = \sup_{x,y}\left\{\|x - y\|, \ x, y \in \mathcal{X}\right\}.$$

For the proof refer to [146, Theorem 3.3].

3.4 Numerical Tests

We have carried out a number of numerical tests on the proposed algorithms by using MATLAB 7.10 built-in nonlinear programming solver "fmincon" installed on a Viglen PC with Windows XP operating system and 2.96 GB of RAM. The optimization subproblems within the Algorithm 3.2 and 3.3 are solved using IBM ILOG CPLEX Studio 12.4. In this section, we report the test results.

We consider primarily two portfolio optimization problems to examine the SSD model (3.2.2) and efficiency of our proposed numerical methods, that is, the penalization approach (3.2.4) and algorithms discussed in Section 3.3.

Suppose that we have a fixed capital to be invested in n assets. Let R_i, $i = 1, \ldots, n$, denotes the return of asset i. In practice, the return is often uncertain and we use a random variable ξ to describe the uncertainty. Specifically, we write R_i as $R_i(\xi)$ and in doing so we are assuming that all n assets have identical random factor.

To simplify the discussion, we normalize the capital to 1 and use x_i, $i = 1, \ldots, n$, to denote the fraction of capital to be invested in asset i. The portfolio return can then be formulated as:

$$g(x, \xi) := R_1(\xi)x_1 + R_2(\xi)x_2 + \cdots + R_n(\xi)x_n. \tag{3.4.29}$$

We apply the SSD model (3.2.2) to optimize our investment strategy. To ease the citation, we repeat the model:

$$\begin{array}{ll} \min\limits_{x} & \mathbb{E}[f(x, \xi)] \\ \text{s.t.} & g(x, \xi) \succeq_{(2)} g(y, \xi), \\ & x \in \mathcal{X}, \end{array} \tag{3.4.30}$$

where g is defined by (3.4.29). We need to specify $f(x, \xi)$ and $\mathcal{X}$. We will start with the simplest case of $f(x, \xi) = -g(x, \xi)$ and $\mathcal{X} := \{x : \sum_{i=1}^{n} x_i = 1, x_i \geq 0\}$ and then consider a variation, which allows f to include a quadratic term and x_i to take a negative value in order to address some practical need where investment in a particular asset is not too small and/or the short selling occurs. We will come to the details of the variations later on. Here y denotes a benchmark investment with $y_i = \frac{1}{n}$, for $i = 1, \ldots, n$.

To examine the appropriateness of the SSD model, we calculate the Conditional Value at Risk (CVaR) for random variables $g(x^*, \xi)$ and $g(y, \xi)$ where x^* is an approximate optimal solution obtained from solving (3.4.29). Recall, that by definition for a specified probability level α, the Value at Risk (VaR) of a portfolio is the lowest amount C such

that, with probability α, the profit does not fall below C. The CVaR_α is the conditional expectation of profit below C. In our context,

$$\text{CVaR}_\alpha(g(x^*,\xi)) = \sup_{\text{C}} \left\{ \text{C} - \frac{1}{\alpha}\mathbb{E}[(\text{C} - g(x^*,\xi))_+] \right\}, \tag{3.4.31}$$

where $\alpha \in (0,1)$ is a prespecified constant.

Dentcheva and Ruszczyński [38] showed that there is a fundamental relationship between the concept of CVaR and the second order stochastic dominance constraint. Specifically they showed that

$$g(x,\xi) \succeq_{(2)} g(y,\xi),$$

if and only if

$$\text{CVaR}_\alpha(g(x,\xi)) \geqslant \text{CVaR}_\alpha(g(y,\xi)),\ \forall \alpha \in (0,1],$$

which means that as the return of a portfolio increases the CVaR of that portfolio also increases. Three values of α are commonly considered: 0.90, 0.95, 0.99. However, in our analysis we focus on the case of $\alpha = 0.95$.

3.4.1 Numerical performance

Example 3.1. We consider a history of percentage returns, for $m = 6$ and $m = 10$ time periods, for a group of $n = 2$ and $n = 5$ assets

in Table 3.4.1 and Table 3.4.2, respectively.

	Returns % for period					
	January	February	March	April	May	June
Asset 1	1.2	1.3	1.4	1.5	1.1	1.2
Asset 2	1.3	1.0	0.8	0.9	1.4	1.3

TABLE 3.4.1: Monthly rates of return on two assets.

	Returns % for period									
	1	2	3	4	5	6	7	8	9	10
Asset 1	1.2	1.3	1.4	1.5	1.1	1.2	1.1	1.0	1.0	1.1
Asset 2	1.3	1.0	0.8	0.9	1.4	1.3	1.2	1.1	1.2	1.1
Asset 3	0.9	1.1	1.0	1.1	1.1	1.3	1.2	1.1	1.0	1.1
Asset 4	1.1	1.1	1.2	1.3	1.2	1.2	1.1	1.0	1.1	1.2
Asset 5	0.8	0.75	0.65	0.75	0.8	0.9	1.0	1.1	1.1	1.2

TABLE 3.4.2: Rates of return on five assets over ten periods.

Our aim is to find an optimal investment strategy for a fixed capital in the n assets which maximizes the expected profit subject to certain risk averse measures. Particularly we

consider the following model:

$$\begin{array}{ll} \min\limits_{x} & \mathbb{E}[f(x,\xi)] = -\mathbb{E}[g(x,\xi)] \\ \text{s.t.} & g(x,\xi) \succeq_{(2)} g(y,\xi), \\ & x \in \mathcal{X}. \end{array} \tag{3.4.32}$$

For the purpose of this example we set the upper and lower bound on the fraction of capital invested in each asset to 0.6 and 0, respectively. We do this to ensure diversification. Minimizing this function can be regarded as an attempt to get as close as possible to meeting requirements on both return and risk.

We apply the exact penalization as discussed in Section 3.2.2 to Examples 3.1 and set the penalty parameter $\rho = 1000$. We solve the reformulated problem with Algorithms 3.1, 3.2, and 3.3. For Algorithm 3.1, we use the step size:

$$\lambda_k = \frac{1}{k},$$

and the stopping rule:

$$\|x_{k+1} - x_k\| \leqslant \delta_x \|x_{k+1}\|, \ \|\varphi(x_{k+1},\rho) - \varphi(x_k,\rho)\| \leqslant \delta_\varphi \|\varphi(x_{k+1},\rho)\|,$$

where $\delta_x = 0.001$ and $\delta_\varphi = 0.001$ are specified precisions.

For Algorithms 3.2 and 3.3 we use $\epsilon = 0.001$ and $\lambda = 0.5$.

Consider Example 3.1, we apply the Algorithms 3.1, 3.2 and 3.3 to solve this problem. The optimal fractions of the invested capital from the starting point $x_0 = (1,0)^T$ are shown in Table 3.4.3.

In order to investigate the accuracy of the solution, we calculated the norm of the subgradient at this solution. The norm of subgradient at $x = (0.6, 0.4)$ is equal to 0.0068 which confirms that the solution is close to optimal.

Alg.	Iter.	Time	x	$\mathbb{E}[g(x,\xi)]$	CVaR
3.1	6	0.845	(0.600, 0.400)	1.217	0.9824
3.2	6	0.631	(0.599, 0.401)	1.218	0.9824
3.3	6	0.628	(0.599, 0.401)	1.218	0.9824

TABLE 3.4.3: Example 3.1 using data in Table 3.4.1. Time is in minutes, the expected return of the benchmark portfolio $\mathbb{E}[g(y,\xi)] = 1.200$ and its CVaR = 0.897.

In Table 3.4.3 and the rest of the tables "Iter" refers to the number of iterations, "Alg" is the short form for algorithm, and "S-sell" refers to short selling.

The results are obtained after 6 iterations by Algorithm 3.1 and 3.2 and are equivalent to the results obtained by Algorithm 3.3. Additionally, as it is expected both the return and its CVaR of the selected portfolio are higher than the benchmark return and CVaR.

Further, we use the data of 5 assets over 10 periods in Table 3.4.2 and examine Example 3.1. Table 3.4.4 shows the results of this example.

Alg.	Iter.	Time	x	$\mathbb{E}[g(x,\xi)]$	CVaR
3.1	115	5	(0.325, 0.231, 0.177, 0.266, 0)	1.147	1.004
3.2	7	0.8931	(0.322, 0.231, 0.177, 0.266, 0)	1.148	1.004
3.3	5	0.6355	(0.325, 0.231, 0.177, 0.266, 0)	1.148	1.004

TABLE 3.4.4: Example 3.1 using data in Table 3.4.2. Time is in minutes, the expected return of the benchmark portfolio $\mathbb{E}[g(y,\xi)] = 1.093$ and its CVaR $= 0.895$.

As it can be seen, all three algorithms result in a similar optimal portfolio. Also the selected portfolio dominates the benchmark portfolio in a sense that the CVaR of the selected portfolio is greater than that of the benchmark portfolio. However, the number of iterations is different and as a result the computation times differ. Algorithm 3.3 performed better than the other algorithms as it converged to the optimal portfolio with respect to the number of iteration and computation time.

Note that in Example 3.1, both f and g are linear. In what follows, we consider nonlinear portfolio optimization problems where either f or g or both are nonlinear. This is to demonstrate that the proposed algorithms can cope with both linear and nonlinear portfolio optimization problems.

Example 3.2. In Example 3.1 we considered an optimization problem where any fraction of capital between 0 and 0.6 was acceptable. However, due to transaction cost, investors do not like to invest very small amount of their capital in different assets. We now reformulate Example 3.1 into a slightly more complicated problem in which we do not want to invest very small amounts in an asset. We consider the following performance function:

$$f(x,\xi) = -g(x,\xi) - \sum_{i=1}^{n} x_i^2, \tag{3.4.33}$$

and incorporate (3.4.33) into the optimization problem (3.4.32). In the section we will consider two cases:

- Short-selling is allowed and upper and lower bounds on the fraction of capital invested in each asset are set to 2 and -1.
- Short-selling is prohibited and the bounds are set to 0.6 and 0 to ensure diversification.

The invested fractions which solve this problem using the discussed algorithms and the data in Table 3.4.1 are the same as the result obtained in Example 3.1. However, the results of the problem using data in Table 3.4.2 are shown in Table 3.4.5.

Furthermore, we collect 300 daily historical returns of 95 FTSE 100 assets prior to March 2011 and a cash account paying 0.5% interest. We used the first 200 daily returns to construct the portfolio strategy based on the performance function (3.4.33). We also used the FTSE 100 Index as the benchmark. The results are presented in Table 3.4.6.

Alg.	Problem	Iter.	Time	x	$\mathbb{E}[g(x,\xi)]$	CVaR
3.1	S-sell	684	16	(0.127, 0.495, 0.550, 0.380, -0.553)	1.230	1.045
	No S-sell	226	9	(0.600, 0, 0, 0.400, 0)	1.1740	1.007
3.2	S-sell	6	0.899	(0.400, 0.514, 0.314, 0.390, -0.500)	1.260	1.063
	No S-sell	6	0.768	(0.600, 0, 0, 0.400, 0)	1.1740	1.007
3.3	S-sell	4	0.649	(0.39, 0.527, 0.287, 0.3, -0.5)	1.260	1.063
	No S-sell	4	0.577	(0.600, 0, 0, 0.400, 0)	1.1740	1.007

TABLE 3.4.5: Example 3.2 using data in Table 3.4.2. Time is in minutes, the expected return of the benchmark portfolio $\mathbb{E}[g(y,\xi)] = 1.200$ and its CVaR $= 0.897$.

Alg.	Problem	Iter.	Time	No.Assets	$\mathbb{E}[g(x,\xi)]$	CVaR
3.1	S-sell	792	28.542	32	0.089	0.079
	No S-sell	685	25.43	27	0.036	0.025
3.2	S-sell	16	1.274	40	0.094	0.082
	No S-sell	13	0.973	26	0.037	0.026
3.3	S-sell	12	0.640	40	0.094	0.082
	No S-sell	10	0.561	26	0.037	0.026

TABLE 3.4.6: Example 3.2 using FTSE 100 historical return. Time is in minutes, "No. Assets" represent the number of assets in the optimal portfolio. The expected return of the benchmark portfolio $\mathbb{E}[g(y,\xi)] = -0.051$ and its CVaR $= 0.023$.

It can be seen that when short-selling is allowed the optimal portfolio has higher return compared to the case where short-selling is prohibited, however this higher return is associated with a higher risk. A rational risk-averse investor is expected to discourage short-selling as the excess return is not worth the extra risk. Additionally, the financial authorities in many countries including the U.K. and the U.S.A. restrict many financial institutions such as pension funds from the practice of short-selling.

Note that both of the level function algorithms converge to very similar portfolios. The portfolio returns are higher than the benchmark portfolio return and as it was expected the CVaR of the selected portfolios are higher than the CVaR of benchmark portfolio. Although the fraction of the capital invested in each asset differ from the results from stochastic approximation method (Algorithm 3.1), but the optimal portfolios return and risk are very close. The number of iterations in the level function method are much lower compared to the stochastic approximation method, consequently the optimization time is lower for the level function method. Furthermore, it could be seen that the projected level function algorithm converges to the optimal solution with fewer number of iterations compared to the scaled level function algorithm. This makes Algorithm 3.3 more attractive than Algorithm 3.2.

In next section, we demonstrate the advantage of taking stochastic dominance constraints into account using real world data for a portfolio optimization problem followed with a bakctest and a out-of-sample analysis.

3.4.2 Portfolio performance

In this section we focus on optimization problem (3.4.32) where $g(y, \xi)$ is set to be equal to FTSE 100 Index and draw some conclusions.

We use the FTSE 100 data collected and a cash account paying 0.5% interest. We used the first 200 observations to construct the portfolio strategy and the further 100 daily returns for an out-of-sample test. In practice there are many strict regulations imposed by authorities on short selling and as a result many financial institutions prohibit any short selling activity. Consequently, in this example we only consider the case where short selling is not allowed and set the upper and lower bounds on portion of capital invested at 0.6 and 0, in order to ensure diversification.

We solve the above optimization problem using the FTSE 100 data and compare the proposed algorithms. The results are presented in Table 3.4.7.

Algorithm	Iter.	Time	No.Assets	$\mathbb{E}[g(x,\xi)]$	CVaR
3.1	735	25.64	31	0.036	0.027
3.2	9	1.717	29	0.037	0.027
3.3	7	1.215	29	0.037	0.027

TABLE 3.4.7: Result of the problem using FTSE 100 data. Time is in minutes, No.Assets represents the number of assets in the optimal portfolio. The expected return of the benchmark portfolio $\mathbb{E}[g(y,\xi)] = -0.051$ and its CVaR = 0.023.

In the remainder of this section we concentrate on the investigating of the efficiency of selected portfolios by Algorithm 3.1, Algorithm 3.2, and Algorithm 3.3.

Furthermore, to investigate the dominance relationship we present the graph of cumulative distribution functions of portfolio return generated by the SSD model using the Algorithms 3.1-3.3, Markowitz model and the FTSE 100 Index in Figure 3.4.5. It is clear that the generated portfolio strategy dominates the benchmark portfolio. Moreover, to see the performance of the generated strategy out-of-sample we present graph of cumulative return of the of portfolio return generated by the SSD model using the Algorithms 3.1-3.3, Markowitz model and the FTSE 100 Index in Figure 3.4.6. It can be seen that the return generated by the portfolio strategy based on the SSD model is much higher compared to the Markowitz model and the benchmark portfolio.

To illustrate the benefit of using stochastic dominance constraints, we set up a backtest which is a key component of effective trading-system development in finance. It is accomplished by reconstructing, with historical data, trades that would have occurred

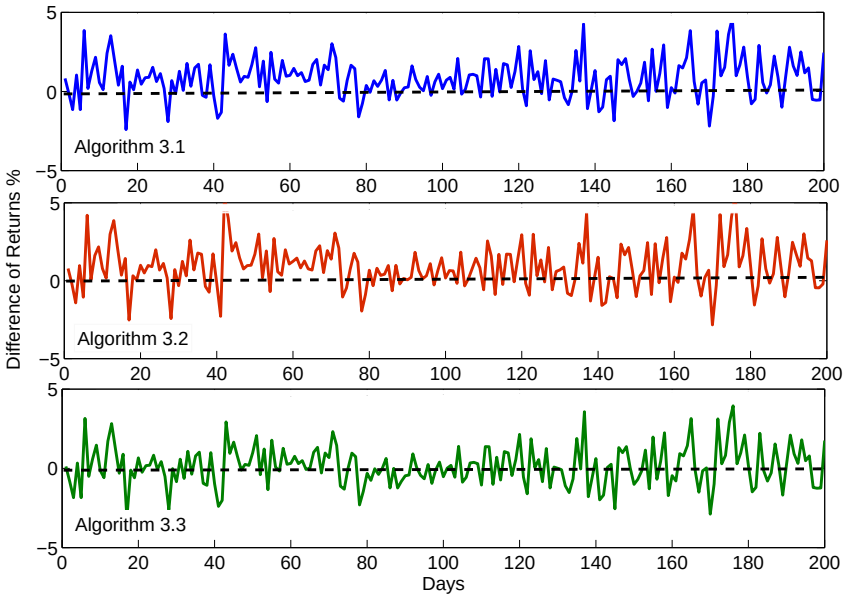

FIGURE 3.4.1: Difference between the portfolio return and the return of Markowitz model, in-sample.

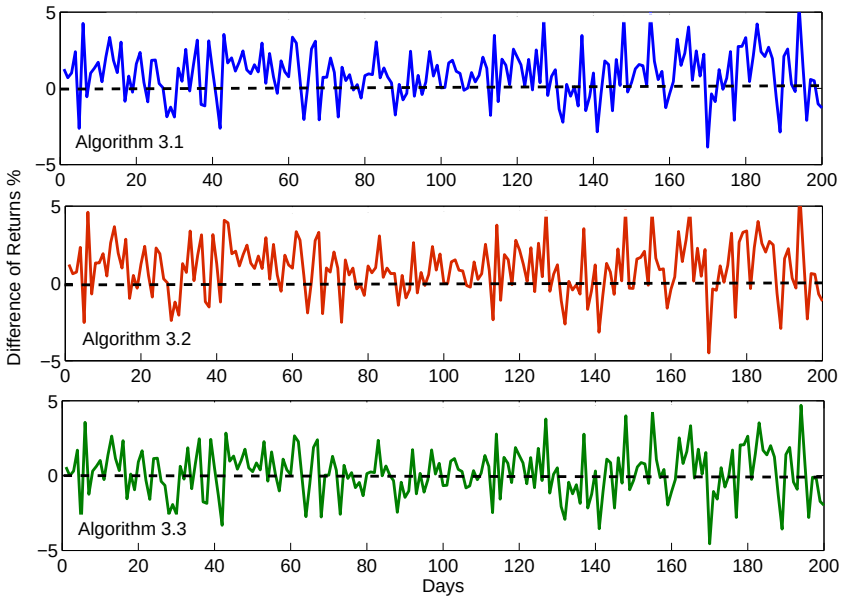

FIGURE 3.4.2: Difference between the portfolio return and the return of FTSE 100 index, in-sample.

in the past using rules defined by a given strategy. Furthermore, we set up an out-of-sample test to evaluate the performance of the selected portfolio over the remaining 100 samples. For the backtest the model finds the optimal portfolio weights from 200 historical market data, then the portfolio strategy is applied to the same data and daily portfolio return is calculated for each day (Figures 3.4.1 and 3.4.2). In the out-of-sample test, the same portfolio strategy is applied to the remaining data of 100 days and the portfolio return is again calculated for each day (Figures 3.4.3 and 3.4.4). In both tests the portfolio performance is compared with FTSE 100 Index and an investment strategy

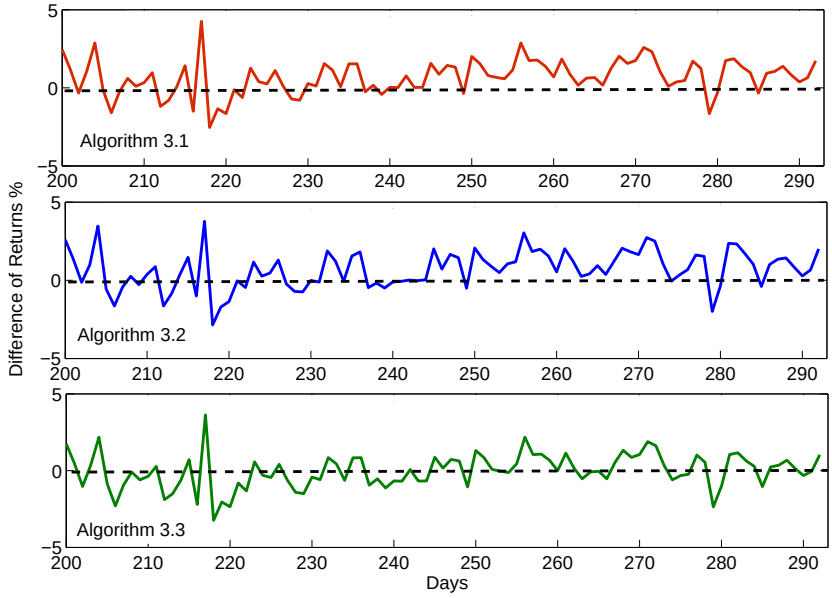

FIGURE 3.4.3: Difference between the portfolio return and the return of Markowitz model, out-of-sample.

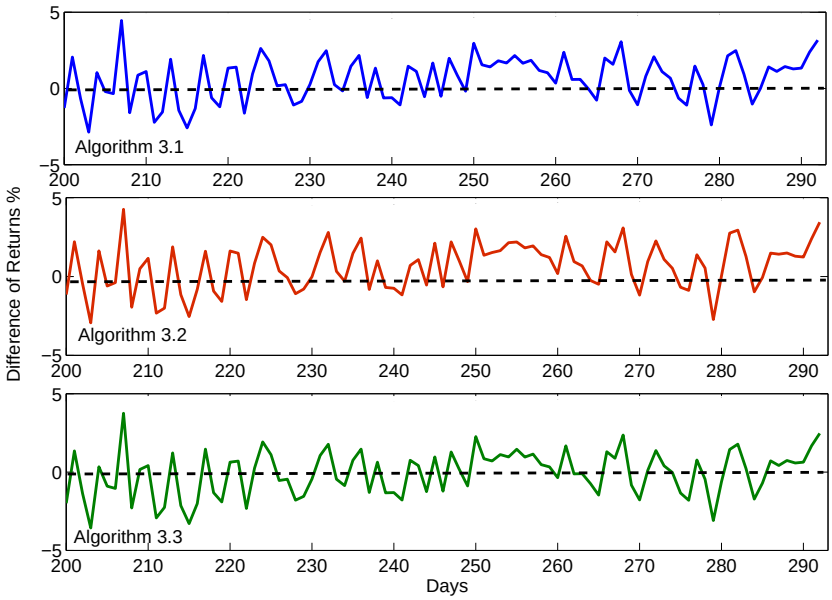

FIGURE 3.4.4: Difference between the portfolio return and the return of FTSE 100 index, out-of-sample.

generated by a Markowitz model as described below:

$$\begin{aligned} \max_{x\in\mathcal{X}} \quad & \mathbb{E}[g(x,\xi)] - \lambda\mathbb{E}[R(x,\xi)] \\ \text{s.t.} \quad & \mathbb{E}[g(x,\xi)] \geq R_b, \\ & \sum_{i=1}^{n} x_i = 1,\ x \geq 0,\ x \in \mathcal{X}, \end{aligned} \tag{3.4.34}$$

where $\lambda = 1$ is a fixed nonnegative number, $\mathbb{E}[R(x,\xi)]$ is the portfolio variance, R_b is

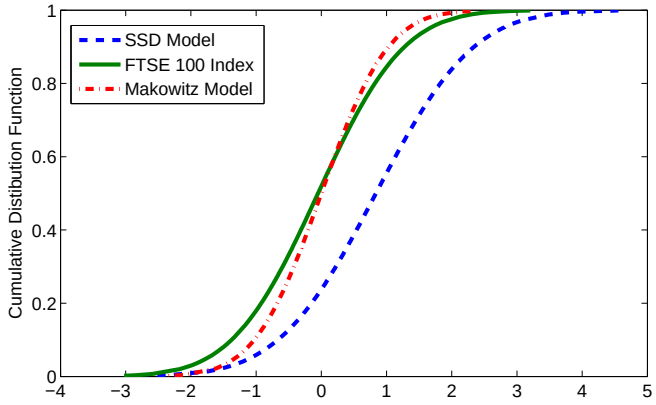

FIGURE 3.4.5: In-sample cumulative distribution functions for the generated portfolio strategy based on the SSD models, Markowitz model and FTSE 100 Index (benchmark).

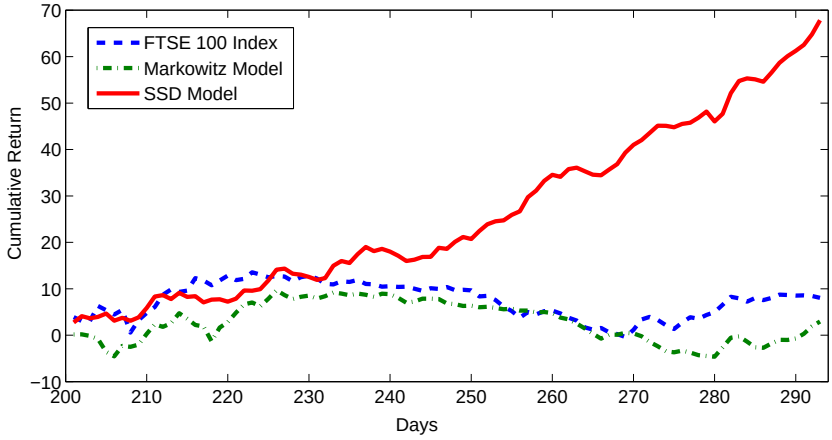

FIGURE 3.4.6: Out-of-sample cumulative return for the generated portfolio strategy based on the SSD models, Markowitz model and FTSE 100 Index (benchmark).

the benchmark return set equal to the FTSE 100 Index, and $\mathbb{E}[g(x,\xi)]$ is the return defined as in (3.4.29). The Markowitz model (3.4.34) assumes that portfolio can be characterized by their mean return and variance.

To compare the performance of the the two portfolios we use the Sortino ratio [134]. The Sortino ratio measures the risk-adjusted return of an investment asset, portfolio or strategy. It is a modification of the Sharpe ratio but penalizes only those returns falling below a user-specified target, or required rate of return, while the Sharpe ratio penalizes both upside and downside volatility equally. We used risk free rate (0.5%) and the benchmark portfolio (Index return) as the required rate of return. We calculated the Sortino ratio both at the 200th day and 300th day. The results are shown in Table 3.4.8. It can be seen that the portfolios constructed by the SSD model (3.4.32) and solved by proposed Algorithm 3.3 and Algorithm 3.1 perform better than the Markowitz model (3.4.34) and a FTSE 100 Index based on the backtest and out-of-sample test as well as the Sortino ratio.

Model	Required return	200^{th} day Sortino ratio	300^{th} day Sortino ratio
SSD Model	Benchmark	0.1902	0.0049
	Risk-free	0.1449	0.0545
Markowitz Model	Benchmark	0.0658	-0.0616
	Risk-free	-0.0066	0.0177

TABLE 3.4.8: Sortino ratio of the portfolio generated by optimization problem with SSD constraints and the Markowitz model.

Note that the two algorithms generate similar results, but their numerical efficiency differ significantly in terms of CPU time and the number of iterations. Figure 3.4.7 shows the CPU time of different numbers of assets for Algorithm 3.1 and 3.3.

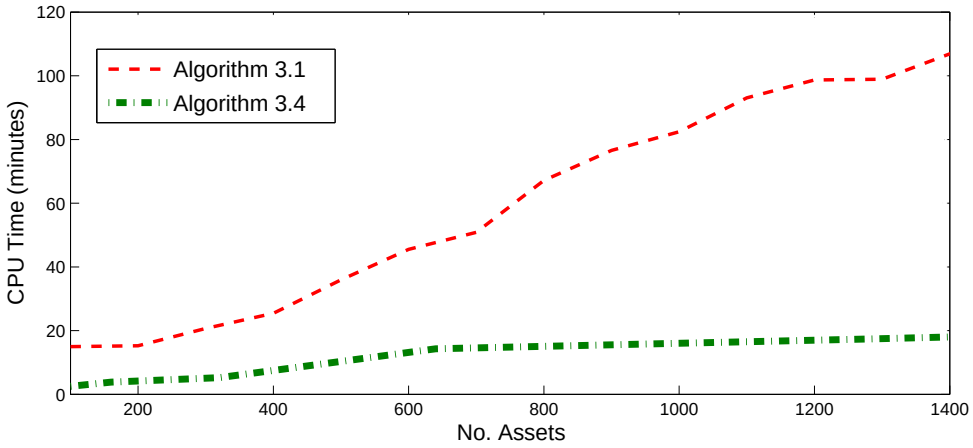

FIGURE 3.4.7: CPU time in minutes versus the number of instruments.

3.4.3 A supply chain problem

This example is a supply chain problem recently considered by Dentcheva and Martinez [35, Section 4]. Since, previous numerical analysis proved that the stochastic approximation method (Algorithm 3.1) is inefficient. In this example we only consider Algorithm 3.3 along with the *cutting plane method* introduced by [81] and applied to stochastic problems with second order stochastic dominance constraint in [52].

Example 3.3. *(Dentcheva and Martinez [35])* A company has a set F of factories that produce and supply perishable product to a set O of stores. Assume that the goods are supplied before the demand is observed. If the demand is not met, the customers buy the product elsewhere and the sales are lost. If the stock of the store is larger than the demand, then the remaining products need to be disposed of. Assume that the disposal cost is a deterministic quantity and that each factory has a limited capacity to produce goods. Furthermore, we assume that a benchmark of the acceptable cost distribution is available. The objective is to determine a production and shipping plan for each factory in order to minimize the expected cost of the company. Denote by x_{ij} the quantity of

goods delivered by factory i to store j, w_i is the quantity produced by factory i, z_j is the number of sales at store j, and y_j is the amount disposed at store j. The shipping cost from factory i to store j is denoted by c_{ij}; the production cost of one unit of product at store i is a_i; the capacity of store i is C_i. The disposal cost on site j is d_j and the price store j sets for the product is b_j.

Dentcheva and Martinez proposed a two-stage stochastic program with stochastic ordering constraint model for this problem, see [35, Section 4] for details. Here, we formulate the problem as a one-stage stochastic problem with SSD constraints as follows:

$$\begin{aligned} \min \quad & \sum_{i=1}^{F}\sum_{j=1}^{O} c_{ij}x_{ij} + \sum_{i=1}^{F} a_i w_i + \sum_{s=1}^{S} p_s Q(x, D^s) \\ \text{s.t.} \quad & -Q(x, D) \succeq_2 -Y, \\ & w_i = \sum_{j=1}^{O} x_{ij}, i = 1, \ldots, F, \\ & 0 \le w_i \le C_i, i = 1, \ldots, F, \\ & x_{ij} \ge 0, i = 1, \ldots, F; j = 1, \ldots, O, \end{aligned} \tag{3.4.35}$$

where $D := (D_1, \ldots, D_S)$, and

$$Q(x, D^s) := \sum_{j=1}^{O} (d_j(x_j - D_j^s)_+ - b_j(D_j^s - (D_j^s - x_j)_+)),$$

$x_j = \sum_{i=1}^{F} x_{ij}$ and Y is a benchmark with S scenarios.

We assume that each component D_j of the demand D satisfies the γ-distribution with parameters $(2, 3)$, $j = 1, \ldots, J$. Each data set is generated through i.i.d. sampling with size S and $p_s := \frac{1}{S}$, for $s = 1, \ldots, S$. The benchmark is constructed from solving the problem without SSD constraints. We have carried out a number of numerical tests on

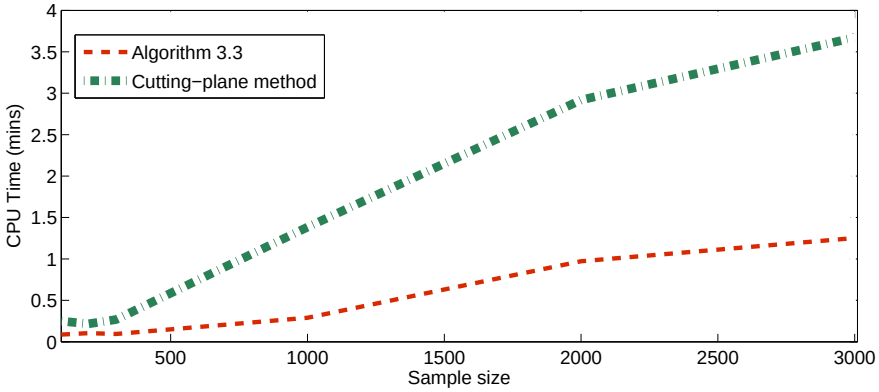

FIGURE 3.4.8: Computational time versus the number of observations for $F = 10$ and $O = 10$, Example 3.3

the two algorithms for problem (3.4.35).

Our tests are concentrated on CPU time against increment of sample size and problem size. There are two cases which may lead to the change of problem size: increase of the number of factories F and increase of the number of stores O. Figure 3.4.8 depicts CPU time of the two algorithms as the size increases from 10 to 3000, whereas Figure 3.4.9 and Figure 3.4.10 depict the sensitivity of CPU time against the change of F and O respectively.

We have made a few observations from the numerical tests. First, the two algorithms perform well as sample size increases particularly when the size goes beyond 2000, see Figure 3.4.8. This is primarily because the sample size does not increase the size of problem (3.4.35) albeit it increase the number of terms in both the objective and constraint functions. Second, Algorithm 3.3 performs better than the other algorithm in most cases.

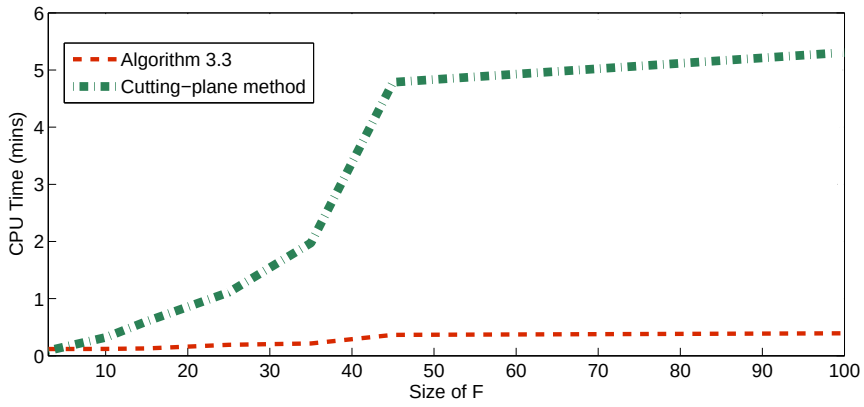

FIGURE 3.4.9: Computational time versus the number of observations for $O = 10$ and $S = 100$, Example 3.3

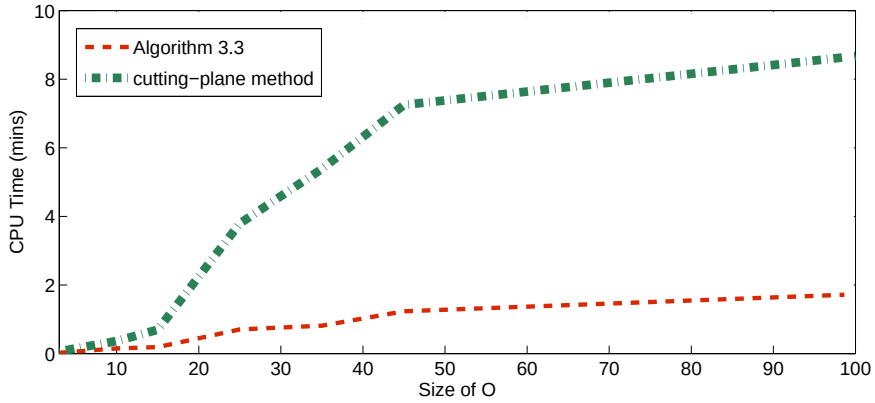

FIGURE 3.4.10: Computational time versus the number of observations for $F = 10$ and $S = 100$, Example 3.3

An underlying reason is that it constructs a single level function instead of adding two or more cutting planes as in the cutting-plane method. Third, increase of the number of stores O has more significant impact on on both the number of iterations and the CPU

time in the two algorithms than does the increase of the number of factories F. See Figures 3.4.9 and 3.4.10. This is primarily due to the fact that increasing O results in more nonsmooth terms in the constraint functions while increasing F does not have such consequence, see the composition of $Q(x, D^s)$. Finally, in comparison with Dentcheva and Martinez's test results, our algorithm is less sensitive to the increase of sample size because we don't introduce new variables to deal with plus functions, on the other hand, our algorithm are more sensitive to the increase of O whereas their algorithms deal with such a problem through introduction of a new variable per scenario to get around the nonsmoothness of the plus function.

3.5 Conclusions

Our preliminary numerical tests show that Algorithm 3.3 (projected level function) is numerically more efficient than Algorithm 3.1 (Stochastic approximation). However, Algorithm 3.1 has a unique advantage; that is at each iteration only one approximated subgradient of the objective function is calculated.

Furthermore, the portfolio optimization problem with SSD constraints performs better than the Markowitz model and it also outperforms the benchmark both in-sample and out-of-sample in sense of portfolio return, which was shown by the results from the back-test and out-of-sample test (Figures 3.4.1–3.4.4). This result was also confirmed with Sortino ratio calculation shown in Table 3.4.8, where the portfolio optimization problem with SSD constraints has higher risk adjusted-return compared to the Markowitz model.

Finally, the supply chain optimization problem with SSD constraints investigated the performance of the Algorithm 3.3 along with cutting plane method [81, 52] in a sense of CPU time versus the size of the problem. It was concluded that the Algorithm 3.3 is less sensitive to the sample size, however this algorithm was sensitive to changes in number of stores "O".

Chapter 4

Stochastic Programs with Multivariate Second Order Stochastic Dominance Constraints

4.1 Overview

In this chapter we study optimization problems with multivariate stochastic dominance constraints where the underlying functions are not necessarily linear. Stochastic dominance is used to compare the distribution of two random variables [101], thus providing a way to measure risk. Over the past few years there has been an increase in the discussion on stochastic programs with stochastic dominance constraints. Dentcheva and Ruszczyński [36, 37] first introduced optimization problems with stochastic dominance constraints. This is an attractive approach for managing risks in an optimization setting. While pursuing expected profits, one avoids high risks by choosing options that are preferable to a random benchmark.

Much of the work on optimization with stochastic dominance has focused on the case where the underlying random quantities being compared are unidimensional [38, 39, 95, 103]. More recently, Dentcheva and Ruszczyński [40] proposed the concept of positive linear second order stochastic dominance which is a special case of multivariate stochastic dominance and obtained necessary conditions of optimality for non-convex problems. The notion of multivariate stochastic dominance refers to the stochastic ordering of random vectors. It can be used as a tool for multicriterion decision making, since each component of vectors can be interpreted as the uncertain outcome of a given criterion.

Homem-de-Mello and Mehrota [70] expanded the definition of positive linear second order dominance to polyhedral second order dominance and called it P-dominance. They proposed a sample average cutting-surface algorithm for optimization problems with multidimensional polyhedral linear second order stochastic dominance constraints. Hu et al. [73] extended this work and presented a more general definition of stochastic dominance over random vectors as natural extension of the polyhedral linear stochastic dominance concept. More recently, Hu et al. [72] proposed a new concept of stochastically weighted dominance, in which they treat the vector of weights as a random vector to deal with large number of weights for bigger problems. They showed that such an approach is much less restrictive than the deterministic weighted approach. More recently, Armbruster and Luedtke [4] derived a linear formulation for multivariate second order stochastic dominance which can be solved with off the shelf linear programming solvers.

Inspired by notion of multivariate stochastic dominance and our earlier work on unidimensional second order stochastic dominance constraints particularly dealing with nonlinear underlying functions, we study stochastic optimization problems with multivariate second order stochastic dominance (SSD) constraints. Sun et al. [137] proposed an exact penalization scheme for scalar second order stochastic dominance. In this chapter we effectively extend the methods proposed in [137] to stochastic programs with multivariate second order stochastic dominance constraints. We propose an exact penalization scheme for such problems and solve the penalized problem by the level function method and a modified cutting plane method and compare them to the cutting surface method proposed by Homem-de-Mello et al. [70] and the linearized method proposed by Armbruster and Luedtke [4].

The main contribution of this chapter can be summarized as follows:

- We develop a penalization scheme for stochastic programs with multivariate second order stochastic dominance constraints. We do so by exploiting Clarke's exact penalty function theorem [29, Proposition 2.4.3] and Robinson's error bound [113]. We reformulate the multivariate stochastic dominance constraints and demonstrate that the reformulated problem satisfies the Slater Constraint Qualification under some moderate conditions. Furthermore, an exact penalization scheme based on L_∞-norm is derived. Based on the exact penalization formulations, we apply a well known level function method in nonsmooth optimization as discussed in Chapter 3 to the penalized problems. An obvious advantage of this approach is that we can effectively deal with excessive number of constraints, non-smoothness in the constraints and nonlinearity of the underlying functions.

- A modified cutting plane method is also proposed. This cutting plane method differs from those in the literature [120] in that it applies to the maximum of the constraint functions rather than each constraint function. Moreover, our modified

cutting plane method uses the cutting plane representation proposed in [81], so it differ from the methods proposed in [70, 73]. The idea of applying the cutting-plane method to the maximum of the constraint functions is similar to the idea in algorithm proposed by Fábián et al. [52]. However, their method is applied to linear models while our modified cutting plane method is also applicable to nonlinear case. Therefore we may regard our algorithm as an extension of theirs. Furthermore, the proposed numerical methods provides an alternative approach to the existing cutting surface method for multivariate stochastic dominance introduced by Homem-de-Mello and Mehrota [70] and the linearized method proposed by Armbruster and Luedtke [4].

- We examine the efficiency of the penalization scheme and the numerical methods by presenting an academic problem, a generic budget allocation problem, and a real world portfolio optimization problem. The budget allocation model is inspired by the homeland security application of Hu et al. [71] and the budget allocation example of Armbruster and Luedtke [4], in which a limited budget must be allocated to a set of possible projects, and the allocation must stochastically dominate a given benchmark. For the portfolio optimization problem, we use real world test data of three indices to set up backtest and out-of-sample test to inspect the performance of the generated portfolio and compare it to the benchmark portfolio and a portfolio strategy generated by Markowitz model.

Let us now present some of the notation that are used in the following sections. Let $x^T y$ denoted the scalar product of two vectors x and y, and let $\|\cdot\|$ denotes the Euclidean norm, while $\|\cdot\|_\infty$ denotes the maximum norm of continuous functions defined over a set T. Let $d(x, \mathcal{D}) := \inf_{x' \in \mathcal{D}} \|x - x'\|$ denote the distance from a point x to a set $\mathcal{D}$. For a real valued smooth function f, we use $\nabla f(x)$ to denote the gradient of f at x. The expected value operator is denoted by $\mathbb{E}$. The standard symbol $\mathcal{L}_1(\Omega, \mathcal{F}, P; \mathbb{R}^m)$ (shortly $\mathcal{L}_1^m$) denotes the space of all integrable mappings X from Ω to $\mathbb{R}^m$. If the values are taken in $\mathbb{R}$ the superscript m will be omitted.

The rest of this chapter is organized as follows. In Section 4.2, we define the optimization problem, discuss the Slater constraint qualification and present the exact penalization schemes. In Section 4.3, we discuss the solution method and correspondingly the algorithms. In Section 4.4, we apply the proposed method to some portfolio optimization problems and report some numerical test results. Finally, we present some conclusion in Section 4.5.

4.2 Stochastic Optimization Problem with Multivariate SSD Constraints

4.2.1 Introduction

The concept of stochastic ordering for scalar random variables has been introduced in statistics and further applied and developed in economics [57]. Let $g(x,\xi)$ be a concave function, with decision vector x and random variable ξ. It is said that $g(x,\xi)$ stochastically dominates a random variable $Y(\xi) \in \mathcal{L}_1$ in the first order, denoted by $g(x,\xi) \succeq_1 Y(\xi)$ if

$$F(g(x,\xi);\eta) \leq F(Y(\xi);\eta),\ \forall \eta \in \mathbb{R}, \tag{4.2.1}$$

where $F(g(x,\xi);\eta)$ and $F(Y(\xi);\eta)$ are the cumulative distribution functions of $g(x,\xi)$ and $Y(\xi)$, respectively. Let $F_2(g(x,\xi);\cdot)$ be defined as

$$F_2(g(x,\xi);\eta) = \int_{-\infty}^{\eta} F(g(x,\xi);\alpha)d\alpha \text{ for } \eta \in \mathbb{R}.$$

Similarly, we say that $g(x,\xi)$ dominates in the second order a random variable $Y(\xi) \in \mathcal{L}_1$ if

$$F_2(g(x,\xi);\eta) \leq F_2(Y(\xi);\eta),\ \forall \eta \in \mathbb{R}. \tag{4.2.2}$$

We denote the relation (4.2.2) as $g(x,\xi) \succeq_{(2)} Y(\xi)$.

Definition 4.1. A random vector $G(x,\xi) \in \mathcal{L}_1^m$ linearly dominates a random vector $Y(\xi) \in \mathcal{L}_1^m$ in positive linear second order, written as $G(x,\xi) \succeq_{(2)}^{Plin} Y(\xi)$, if

$$\nu^T G(x,\xi) \succeq_{(2)}^{Plin} \nu^T Y(\xi),\ \forall \nu \in \mathbb{R}_+^m. \tag{4.2.3}$$

In the same manner one can define the first and higher order linear dominance relations: $\nu^T G(x,\xi) \succeq_{(k)}^{Plin} \nu^T Y(\xi)$, $k = 1,2,\ldots$ provided that $(k-1)$-st moments of $G(x,\xi)$ and $Y(\xi)$ are finite.

It is clear that the set of scalarizing vectors ν in (4.2.3) can be truncated, by imposing the normalization constraint $\nu \in S$, where S is the simplex:

$$S = \left\{\nu \in \mathbb{R}_+^m : \nu_1 + \nu_2 + \cdots + \nu_m = 1\right\}. \tag{4.2.4}$$

This restriction does not change the relation $(\succeq_{(2)}^{Plin})$.

In this chapter, we consider the following optimization problem with multivariate second order stochastic dominance (SSD) constraints:

$$\begin{array}{ll} \min\limits_{x\in X} & \mathbb{E}[f(x,\xi)] \\ \text{s.t} & \nu^T G(x,\xi) \succeq_{(2)}^{Plin} \nu^T Y(\xi),\ \forall \nu \in S, \end{array} \tag{4.2.5}$$

where $f : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$ is convex continuous function and $G : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$ is concave continuous function, both in x and ξ , $x \in \mathcal{X}$ is a decision vector with X being a nonempty convex subset of $\mathbb{R}^n$ and $\xi : \Omega \to \Xi \subset \mathbb{R}^k$ is a random vector defined on probability space $(\Omega, \mathcal{F}, P)$ with support Ξ, $\mathbb{E}[\cdot]$ denotes the expected value w.r.t. the probability distribution of ξ. The random variable $Y(\xi)$ plays the role of a benchmark outcome. For example, one may consider $Y(\xi) = G(\bar{x}, \xi)$, where $\bar{x} \in \mathcal{X}$ is some reasonable value of the decision vector, which is currently employed in the system. We shall investigate the case when $G(x, \xi)$ and $Y(\xi)$ are m-dimensional random vectors, rather than scalar variables.

Using the properties of second order dominance [104, 145] and the definition of positive linear dominance, we reformulate the multivariate stochastic dominance constraint in (4.2.5) as,

$$\mathbb{E}[(\nu^T \eta - \nu^T G(x, \xi))_+] \leq \mathbb{E}[(\nu^T \eta - \nu^T Y(\xi))_+], \ \forall (\eta, \nu) \in \mathbb{R}^m \times S,$$

where $(\eta - \nu^T G(x, \xi))_+ = \max(\eta - \nu^T G(x, \xi), 0)$. Consequently, we reformulate the optimization problem (4.2.5) as a stochastic semi-infinite programming problem:

$$\begin{aligned} \min_{x \in X} \quad & \mathbb{E}[f(x, \xi)] \\ \text{s.t.} \quad & H(x, \eta, \nu) := \mathbb{E}[(\nu^T \eta - \nu^T G(x, \xi))_+] - \mathbb{E}[(\nu^T \eta - \nu^T Y(\xi))_+] \leq 0, \ \forall (\eta, \nu) \in \mathbb{R}^m \times S. \end{aligned} \tag{4.2.6}$$

Our focus is on numerical methods for solving the stochastic semi-infinite programming problem (4.2.6). There are three issues to deal with: (a) the expectation of random functions in both the objective and constraints, (b) the infinite number of constraints, (c) the non-smoothness resulting from the max functions.

Homem-de-Mello et al. [70] introduced a more general notion of dominance which includes positive linear dominance as a particular case. They considered the case where the set S is assumed to be a polyhedron. By using the polyhedral properties they proposed a cutting-surface algorithm. They dealt with the constraints by considering a cut generation, and solved the problem by a branch-and-cut algorithm. Although the proposed cutting-surface method is effective, it is computationally demanding. In particular, even for the case of second-order stochastic dominance, which induces a convex feasible region, their algorithm requires global optimization of a nonconvex problem as a subproblem. Furthermore, Armbruster and Luedtke [4] proposed to use a different notion of multivariate stochastic dominance as a constraint in a stochastic optimization model. They derived an LP formulation for an SSD constraint which could be solved using off-the-shelf linear programming solvers.

In the rest of the chapter, we consider problem (4.2.6) with the focus on the case when ξ has a discrete distribution, that is

$$\begin{aligned} \min_{x} \quad & \sum_{i=1}^{N} p_i f(x, \xi^i) \\ \text{s.t.} \quad & \sum_{i=1}^{N} p_i (\nu^T \eta - \nu^T G(x, \xi^i))_+ - \sum_{i=1}^{N} p_i (\nu^T \eta - \nu^T Y(\xi^i))_+ \leq 0, \ \forall (\eta, \nu) \in \mathbb{R}^m \times S, \\ & x \in \mathcal{X}, \end{aligned} \tag{4.2.7}$$

where the random variable ξ has a finite distribution, that is, $P(\xi = \xi^i) = p_i$, for $i = 1, \ldots, N$. In what follows, we investigate the Slater constraint qualification of the problem (4.2.7) and its reformulation, followed by an exact penalization scheme and numerical methods.

4.2.2 Slater constraint qualification

In the literature of stochastic programs with second order stochastic dominance constraints, Slater constraint quialification (SCQ) has been used as a key condition for deriving optimality conditions and exact penalization, see [36, 93].

Note that, the problem (4.2.7) is said to satisfies the SCQ if there exists $x_0 \in \mathcal{X}$ such that

$$\sum_{i=1}^{N} p_i (\nu^T \eta - \nu^T G(x_0, \xi))_+ - \sum_{i=1}^{N} (\nu^T \eta - \nu^T Y(\xi))_+ < 0, \ \forall (\eta, \nu) \in \mathbb{R}^m \times S. \tag{4.2.8}$$

Unfortunately, this kind of constraint qualification is not satisfied. To see this, as discussed in [137], for a fixed $\nu \in S$ let

$$Y(\Xi) := \{Y(\xi^i) : i = 1, \ldots, N\},$$

and

$$\eta^* := \min\{Y(\xi^1), \ldots, Y(\xi^N)\}.$$

For any $\eta \leq \eta^*$, it can be verified that $\mathbb{E}[(\nu^T \eta - \nu^T Y(\xi))_+] = 0$. For those η, the feasible constraint of problem (4.2.7) reduces to

$$\mathbb{E}[(\nu^T \eta - \nu^T G(x, \xi))_+] - \mathbb{E}[(\nu^T \eta - \nu^T Y(\xi))_+] = 0,$$

because the left hand term is non-negative. Consequently, there does not exist a feasible point $x_0 \in \mathcal{X}$ such that (4.2.8) holds.

More recently, Homem-de-Mello and Mehrota [70] proposed an alternative approach to deal with the SCQ issue, by considering ϵ-feasible solutions:

$$\begin{array}{ll} \min\limits_{x} & \mathbb{E}[f(x,\xi)] \\ \text{s.t.} & \mathbb{E}[(\nu^T\eta - \nu^T G(x,\xi))_+] \leq \mathbb{E}[(\nu^T\eta - \nu^T T(\xi))_+] + \epsilon, \forall(\eta,\nu) \in \mathbb{R}^m \times S, \end{array} \tag{4.2.9}$$

where ϵ is a small positive number. The relaxed problem (4.2.9) satisfies SCQ as long as the original problem is feasible. However, it must be shown that the feasible solution set of the relaxed problem approximates the feasible solution set of the original problem, which often in turn requires the original problem to satisfy certain regularity conditions such as lower semicontinuity of the feasible solution set of the relaxed problem.

In what follows, we propose an alternative way to tackle this problem by reformulating problem (4.2.6) using [36, Proposition 3.2] and [73] as follows:

$$\begin{array}{ll} \min\limits_{x} & \sum\limits_{i=1}^{N} p_i f(x,\xi^i) \\ \text{s.t.} & \sum\limits_{i=1}^{N} p_i(\nu^T\eta_j - \nu^T G(x,\xi^i))_+ \leq \gamma_j(\nu),\ \forall\nu \in S,\ j = 1,\cdots,N, \\ & x \in \mathcal{X}, \end{array} \tag{4.2.10}$$

where $\eta_j := Y(\xi^j)$ and $\gamma_j(\nu) := \sum_{i=1}^{N} p_i(\nu^T\eta_j - \nu^T Y(\xi^i))_+$. Note that, the reformulated problem also does not satisfy the SCQ.

Let us define the power set of $\{1,\ldots,N\}$, that is, a collection of all subsets of $\{1,\ldots,N\}$ including empty set and itself. Let $\mathcal{N}$ denote the power set excluding the empty set and for $j = 1,\ldots,N$,

$$\psi_j(x,\nu) := \max_{\mathcal{J}\in\mathcal{N}} \sum_{i\in\mathcal{J}} p_i(\nu^T\eta_j - \nu^T G(x,\xi^i)) - \gamma_j(\nu). \tag{4.2.11}$$

Consequently, we can reformulate problem (4.2.10) as

$$\begin{array}{ll} \min\limits_{x} & \sum\limits_{i=1}^{N} p_i f(x,\xi^i) \\ \text{s.t.} & \psi_j(x,\nu) \leq 0,\ \forall\nu \in S,\ j = 1,\cdots,N, \\ & x \in \mathcal{X}. \end{array} \tag{4.2.12}$$

In what follows, we show that problem (4.2.12) is equivalent to problem (4.2.10) but, under some circumstances, (4.2.12) satisfies the SCQ.

Lemma 4.2. *Let $\psi_j(x,\nu)$ be defined as in (4.2.11). Let*

$$\varphi_j(x,\nu) := \max_{\mathcal{J}\in\mathcal{N}} \sum_{i\in\mathcal{J}} p_i(\nu^T\eta_j - \nu^T G(x,\xi^i)).$$

Then

$$\sum_{i=1}^{N} p_i(\nu^T \eta_j - \nu^T G(x, \xi^i))_+ = \max\{\varphi_j(x, \nu), 0\}, \tag{4.2.13}$$

for each fixed $\nu \in S$.

Proof. Let $\nu \in S$ be fixed, we consider two cases. Case 1. $\varphi_j(x, \nu) \leq 0$, and Case 2. $\varphi_j(x, \nu) > 0$.

Case 1. $\varphi_j(x, \nu) \leq 0$ implies that $\max\{\varphi_j(x, \nu), 0\} = 0$ and $\nu^T \eta_j - \nu^T G(x, \xi^i) \leq 0$, for $j \neq \{1, \cdots, N\}$. This implies that

$$\sum_{i=1}^{N} p_i(\nu^T \eta_j - \nu^T G(x, \xi^i))_+ = 0,$$

hence (4.2.13) holds.

Case 2. Now consider case when $\varphi_j(x, \nu) > 0$, then there exists a nonempty subset $\mathcal{J} \subseteq \{1, \ldots, N\}$ such that

$$\varphi_j(x, \nu) = \sum_{i \in \mathcal{J}} p_i(\nu^T \eta_j - \nu^T G(x, \xi^i)) > 0.$$

It suffice to either show that

$$\sum_{i \in \mathcal{J}} p_i(\nu^T \eta_j - \nu^T G(x, \xi^i)) = \sum_{i=1}^{N} p_i(\nu^T \eta_j - \nu^T G(x, \xi^i))_+,$$

or equivalently $\mathcal{J}$ consists of every index i with

$$\nu^T \eta_j - \nu^T G(x, \xi^i) > 0.$$

Note that, if $\mathcal{J}$ does not include such an index, then adding it to $\mathcal{J}$ would increase the quantity $\sum_{\mathcal{J} \in \mathcal{N}} p_i(\nu^T \eta_j - \nu^T G(x, \xi^i))$. This contradicts the fact that $\varphi_j(x, \nu)$ is the maximum. Likewise, $\mathcal{J}$ does not consist of an index i with

$$\nu^T \eta_j - \nu^T G(x, \xi^i) < 0,$$

as removing the index will also increase the quantity $\sum_{\mathcal{J} \in \mathcal{N}} p_i(\nu^T \eta_j - \nu^T G(x, \xi^i))$. This completes the proof. □

We are now ready to state the main results.

Theorem 4.3. *Let* $G(x, \xi)$ *and* $Y(\xi)$ *be defined as in problem (4.2.7) and* $\psi_j(x, \nu)$ *be defined as in (4.2.11). Then*

(i) $\nu^T G(x,\xi) \succeq_{(2)} \nu^T Y(\xi)$ *for all* $\nu \in S$ *if and only if*

$$\bar{\psi}_j(x) := \max_{\nu \in S} \psi_j(x,\nu) \leq 0, \text{ for } j = 1, \cdots, N; \tag{4.2.14}$$

(ii) problems (4.2.10) and (4.2.12) are equivalent;

(iii) if there exists a feasible point $\bar{x}$ *such that* $\nu^T G(\bar{x},\xi) \succeq_{(1)} \nu^T Y(\xi)$ *and* $\nu^T G(\bar{x},\xi) > \nu^T Y(\xi)$ *for all* $\xi \in \Xi$*, then the system of inequalities (4.2.14) satisfies the SCQ.*

Proof. The proof is similar to that of [137, Theorem 2.1] except that we have to deal with parameter ν.

Part (i). By [36, Proposition 3.2], $\nu^T G(x,\xi) \succeq_{(2)} \nu^T Y(\xi)$ for all $\nu \in S$ if and only if

$$\max_{\nu \in S} \sum_{i=1}^{N} p_i (\nu^T \eta_j - \nu^T G(x,\xi^i))_+ - \gamma_j(\nu) \leq 0, \text{ for } j = 1, \cdots, N, \tag{4.2.15}$$

or equivalently for $j = 1, \ldots, N$,

$$\max_{\nu \in S} \max \left\{ \sum_{i=1}^{N} p_i (\nu^T \eta_j - \nu^T G(x,\xi^i))_+ - \gamma_j(\nu), 0 \right\} = 0.$$

By (4.2.13),

$$\max \left\{ \sum_{i=1}^{N} p_i (\nu^T \eta_j - \nu^T G(x,\xi^i))_+ - \gamma_j(\nu), 0 \right\} = \max \left\{ \max \left\{ \varphi_j(x,\nu), 0 \right\} - \gamma_j(\nu), 0 \right\}.$$

Note that for any value $a \in \mathbb{R}$ and $r > 0$, it is easy to verify that

$$\max\{\max\{a, 0\} - r, 0\} = \max\{a - r, 0\}. \tag{4.2.16}$$

Using (4.2.16), we have that

$$\max\{\max\{\varphi_j(x,\nu), 0\} - \gamma_j(\nu), 0\} = \max\{\varphi_j(x,\nu) - \gamma_j(\nu), 0\} = \max\{\psi_j(x,\nu), 0\}.$$

The last equality is due to the definition of ψ_j. The discussion above demonstrates that (4.2.15) is equivalent to (4.2.14) and hence the conclusion.

Part (ii) follows from Part (i) in that the feasible set of the two problems coincides, i.e.,

$$\left\{ x \in \mathcal{X} : \max_{\nu \in S} \sum_{i=1}^{N} p_i (\nu^T \eta_j - \nu^T G(x,\xi^i))_+ - \gamma_j(\nu) \leq 0 \right\} = \{ x \in \mathcal{X} : \bar{\psi}_j(x) \leq 0 \}.$$

Part (iii). Let $\hat{\gamma}(\nu) := \sum_{i=1}^{N} p_i(\nu^T\eta_0 - \nu^T Y(\xi^i))_+$, where

$$\nu^T\eta_0 := \min\{\nu^T Y(\xi^1), \ldots, \nu^T Y(\xi^N)\}.$$

By definition of $\nu^T\eta_0$ the $\sum_{i=1}^{N} p_i(\nu^T\eta_0 - \nu^T Y(\xi^i))_+ = 0$. Therefore, $\hat{\gamma}(\nu) := 0$. Likewise, the assumption $\nu^T G(\bar{x}, \xi) > \nu^T\eta_0$ for $\xi \in \Xi$ implies that

$$\max_{\nu\in S} \max_{\mathcal{J}\in\mathcal{N}} \sum_{i\in\mathcal{J}}^{N} p_i(\nu^T\eta_0 - \nu^T G(\bar{x}, \xi^i)) < 0.$$

This shows

$$\max_{\nu\in S} \left[\max_{\mathcal{J}\in\mathcal{N}} \sum_{i\in\mathcal{J}} p_i(\nu^T\eta_0 - \nu^T G(\bar{x}, \xi^i)) - \hat{\gamma}(\nu) \right] < 0. \tag{4.2.17}$$

For each fixed $\nu \in S$, let $\nu^T\eta_1, \ldots, \nu^T\eta_N$, where $\eta_j := T(\xi^j)$ denote the N elements in set $\nu^T Y(\Xi)$,

$$\nu^T\eta_1 \le \nu^T\eta_2 \le \cdots \le \nu^T\eta_N.$$

Then the inequality (4.2.17) means that

$$\bar{\psi}_1(\bar{x}) := \max_{\nu\in S} \left[\max_{\mathcal{J}\in\mathcal{N}} \sum_{i\in\mathcal{J}} p_i(\nu^T\eta_1 - \nu^T G(\bar{x}, \xi^i)) - \sum_{i=1}^{N} p_i(\nu^T\eta_1 - \nu^T Y(\xi^i))_+ \right] < 0.$$

In what follows, we show that

$$\bar{\psi}_j(\bar{x}) < 0, \text{ for } j = 2, \cdots, N.$$

By definition, for $j = 2, \ldots, N$

$$\begin{aligned}
\bar{\psi}_j(\bar{x}) &= \max_{\nu\in S} \left[\max_{\mathcal{J}\in\mathcal{N}} \sum_{i\in\mathcal{J}}^{N} p_i(\nu^T\eta_j - \nu^T G(\bar{x}, \xi^i)) - \sum_{i=1}^{N} p_i(\nu^T\eta_j - \nu^T Y(\xi^i))_+ \right] \\
&\le \max_{\nu\in S} \left[\max \left\{ \max_{\mathcal{J}\in\mathcal{N}} \sum_{i\in\mathcal{J}}^{N} p_i(\nu^T\eta_j - \nu^T G(\bar{x}, \xi^i)), 0 \right\} - \sum_{i=1}^{N} p_i(\nu^T\eta_j - \nu^T Y(\xi^i))_+ \right] \\
&= \max_{\nu\in S} \left[\max\{\varphi_j(x, \nu), 0\} - \gamma_j(\nu) \right] \\
&\overset{(4.2.13)}{=} \max_{\nu\in S} \left[\sum_{i=1}^{N} p_i((\nu^T\eta_j - \nu^T G(\bar{x}, \xi^i))_+ - (\nu^T\eta_j - \nu^T Y(\xi^i))_+) \right] \\
&= \max_{\nu\in S} \left[\int_{-\infty}^{\nu^T\eta_j} (F_1(\nu^T G(\bar{x}, \xi), \alpha) - F_1(\nu^T Y(\xi), \alpha)) d\alpha \right].
\end{aligned} \tag{4.2.18}$$

The equality (4.2.18) is due to the equivalent representation of second order stochastic dominance [40].

Assume without loss of generality that $\nu^T\eta_1 < \nu^T\eta_2$ (otherwise $\bar{\psi}_2(\bar{x}) = \bar{\psi}_1(\bar{x}) < 0$). Let

$\nu^T\eta_{min} \in (\nu^T\eta_1, \min\{\min_{\xi\in\Xi} G(\bar{x},\xi^i), \nu^T\eta_2\})$ for a given ν. Note that by assumption $\nu^T\eta_1 < \min\{\min_{\xi\in\Xi} G(\bar{x},\xi^i), \nu^T\eta_2\}$, $\nu^T\eta_{min}$ exists. Then

$$\begin{aligned}\max_{\nu\in S} & \left[\int_{-\infty}^{\nu^T\eta_j} (F_1(\nu^T G(x,\xi),\alpha) - F_1(\nu^T Y(\xi),\alpha))d\alpha\right] \\ = & \max_{\nu\in S}\left[\int_{-\infty}^{\nu^T\eta_{min}} (F_1(\nu^T G(\bar{x},\xi),\alpha) - F_1(\nu^T Y(\xi),\alpha))d\alpha\right. \\ & \left. + \int_{\eta_{min}}^{\nu^T\eta_j} (F_1(\nu^T G(x,\xi),\alpha) - F_1(\nu^T Y(\xi),\alpha))d\alpha\right].\end{aligned}$$

Note that

$$\max_{\nu\in S}\left[\int_{-\infty}^{\nu^T\eta_{min}} (F_1(\nu^T G(\bar{x},\xi),\alpha) - F_1(\nu^T Y(\xi),\alpha))d\alpha = 0 - p_1(\nu^T\eta_{min} - \nu^T\eta_1)\right] < 0,$$

where p_1 is the probability that $Y(\xi)$ takes value η_1. On the other hand, $\nu^T G(\bar{x},\xi) \succeq_{(1)} \nu^T Y(\xi)$ implies

$$\max_{\nu\in S}\left[\int_{\bar{\eta}}^{\nu^T\eta_j} (F_1(\nu^T G(\bar{x},\xi),\alpha) - F_1(\nu^T Y(\xi),\alpha))d\alpha\right] < 0.$$

This shows that

$$\max_{\nu\in S}\left[\int_{-\infty}^{\nu^T\eta_j} (F_1(\nu^T G(\bar{x},\xi),\alpha) - F_1(\nu^T Y(\xi),\alpha))d\alpha\right] < 0, \text{ for } j = 2,\cdots,N. \quad (4.2.19)$$

The conclusion follows by combining (4.2.17)–(4.2.19). □

Theorem 4.3 states that although problem (4.2.10) and (4.2.7) do not satisfy SCQ, the reformulated problem (4.2.12) may do under some circumstances. the reason behind this, is to do with plus function $(\cdot)_+$. Consdier a single variate function $a(x) = x$. It is easy to see that the single inequality $a(x) \leq 0$ satisfies SCQ but $(a(x))_+ \leq 0$ does not although the two inequalities represent the same set $(-\infty, 0]$. Clearly, the constraint qualification is closely related to the function which represents the feasible set. In problem (4.2.12), we give an alternative presentation of the feasible constraints of problem (4.2.10) and (4.2.7) without the plus function.

4.2.3 Exact penalization with L_∞-norm

In this section we apply an exact penalization technique to the problem (4.2.12). This scheme includes application of an exact penalty function method with L_∞-norm. Let us now present the technical results needed in the rest of this section.

Lemma 4.4. *[137, Lemma 3.1] Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuous function and $g : \mathbb{R}^n \to \mathbb{R}$ be continuous and convex. Let $X \subseteq \mathbb{R}^n$ be a compact and convex set. Consider the following constrained problem*

$$\begin{array}{ll} \min_x & f(x) \\ \text{s.t.} & g(x) \leq 0, \\ & x \in \mathcal{X}. \end{array} \tag{4.2.20}$$

(i) If $g(x)$ satisfies the SCQ, that is there exists a point x_0 and a real number $\delta > 0$ such that

$$\delta\mathcal{B} \subset g(x_0) + K,$$

and the feasible set, denoted by P, is bounded, then

$$d(x, P) \leq \delta^{-1} D d_\infty(0, g(x) + K),$$

where $\mathcal{B}$ is the closed unit ball in $\mathbb{R}^m$ and $K := [0, +\infty)^m$, and D denotes the diameter of P.

(ii) If $f(x)$ is Lipschitz continuous on $\mathcal{X}$ with modulus κ, then for

$$\rho > \kappa\delta^{-1} D,$$

the set of optimal solutions of (4.2.6) coincides with the set of optimal solutions of problem

$$\begin{array}{ll} \min & f(x) + \rho \|(g(x))_+\|_\infty \\ \text{s.t.} & x \in \mathcal{X}. \end{array} \tag{4.2.21}$$

Proof. Part(i) follows from Robinson's error bound for convex systems [113] and Part (ii) follows from Part(i) and Clarke's exact penalty function [29, Proposition 2.4.3]. □

One popular penalty scheme in optimization is based on the L_∞-norm. Here we consider such penalization scheme for (4.2.12) as follows:

$$\min_x \sum_{i=1}^N p_i f(x, \xi^i) + \rho \max_{j \in \{1, \cdots, N\}} (\max_{\nu \in S} \psi_j(x, \nu))_+, \tag{4.2.22}$$

and for problem (4.2.10)

$$\min_{x \in \mathcal{X}} \sum_{i=1}^N p_i f(x, \xi^i) + \rho \max_{j \in \{1, \cdots, N\}} (\max_{\nu \in S} \sum_{i=1}^N p_i (\nu^T \eta_j - \nu^T G(x, \xi^i))_+ - \gamma_j(\nu))_+. \tag{4.2.23}$$

In what follows, we show that the two penalty schemes are equivalent, and estimate the penalty parameter. This will effectively justify the exact penalization function (4.2.23) for problem (4.2.10), although it does not satisfy the SCQ.

Assumption 4.5. *$f(x,\xi^i)$ and $G(x,\xi^i)$ are continuously differentiable w.r.t. x for $i = 1,\ldots,N$. Moreover, they are globally Lipschitz over $\mathcal{X}$, that is, there exists $\kappa(\xi) < +\infty$ such that*

$$\max\left(\left\|\nabla_x f(x,\xi^i)\right\|, \left\|\nabla_x G(x,\xi^i)\right\|\right) \leq \kappa(\xi^i),\ i = 1,\cdots,N.$$

Theorem 4.6. *Assume that the problem (4.2.22) satisfies the SCQ and Assumption 4.5 holds; the feasible set of problem (4.2.12) is bounded. Then*

(i) problem (4.2.22) and (4.2.23) are equivalent;

(ii) there exist positive constants $\bar{\delta}$ and $\bar{D}$ such that when

$$\rho > \sum_{i=1}^{N} p_i \kappa(\xi^i)\bar{\delta}^{-1}\bar{D}, \tag{4.2.24}$$

the set of optimal solutions of (4.2.12) coincide with that of (4.2.22) and the set of optimal solutions of (4.2.10) coincides with that of (4.2.23).

Proof. Part (i). Using Lemma 4.2 and (4.2.16), the equivalence of the problem (4.2.22) and (4.2.23) can be verified as follows

$$\begin{aligned}\max_{j\in\{1,\ldots,N\}}(\max_{\nu\in S}\psi_j(x,\nu))_+ &= \max_{j\in\{1,\ldots,N\}}\left[\max_{\nu\in S}\max_{\mathcal{J}\in\mathcal{N}}\sum_{i\in\mathcal{J}} p_i(\nu^T\eta_j - \nu^T G(x,\xi^i)) - \gamma_j(\nu)\right]_+ \\ &= \max_{j\in\{1,\ldots,N\}}\left[\max_{\nu\in S}\sum_{i=1}^{N} p_i(\nu^T\eta_j - \nu^T G(x,\xi^i))_+ - \gamma_j(\nu)\right]_+.\end{aligned}$$

Part (ii). let $\mathcal{Q}$ denote the feasible set of problem (4.2.12) and define

$$\Psi(x,\nu) := (\psi_1(x),\cdots,\psi_N(x))^T.$$

Since $\mathcal{Q}$ is bounded, $\sum_{i=1}^{N} p_i f(x,\xi^i)$ is Lipschitz continuous with modulus $\sum_{i=1}^{N} p_i\kappa(\xi^i)$, problem (4.2.12) is convex and satisfies the SCQ. Using Lemma 4.4, there exists real number $\bar{\delta} > 0$ and $\bar{D} > 0$ such that

$$\bar{\rho} > \sum_{i=1}^{N} p_i \kappa(\xi^i)\bar{\delta}^{-1}\bar{D},$$

the set of optimal solutions of problem (4.2.10) coincides with that of (4.2.23). Moreover, since problem (4.2.12) and (4.2.10) are equivalent, and problem (4.2.22) and (4.2.23) are also equivalent, the set of optimal solutions of problem (4.2.12) coincides with that of (4.2.22). □

In the rest of this chapter, we apply the level function methods, and a modified cutting-plane method to solve the penalized optimization problem (4.2.22).

4.3 Solution Methods

In this section we discuss the solution methods used to solve the optimization problem (4.2.22). Specifically, we apply the following methods: the level function method and a modified cutting plane method and compare them to the cutting surface algorithm [70] and the linearized method proposed in [4].

4.3.1 Level function methods

Level function method is popular numerical approach for solving deterministic nonsmooth optimization problems. It is proposed by Lemaréchal et al. [89] for solving nonsmooth convex optimization problems and extended by Xu [146] for solving quasiconvex optimization problems. In Chapter 3, we applied the level function methods to stochastic programming problems with scalar second order stochastic dominance constraints where the distribution of ξ is discrete. In this chapter, we apply the level function methods (Algorithms 3.2 and 3.3) as discussed in Chapter 3 to the penalized problem (4.2.22). Let us define $\vartheta(x,\rho)$ as follows:

$$\vartheta(x,\rho) := \sum_{i=1}^{N} p_i f(x,\xi^i) + \rho \max_{j\in\{1,\cdots,N\}} (\max_{\nu\in S} \psi_j(x,\nu))_+.$$

Let $\zeta_k \in \partial_x \vartheta(x_k,\rho)$, then

$$\sigma_{x_k}(x) = \zeta_k^T (x - x_k)/\left\|\zeta_k\right\|,$$

is a level function of $\vartheta(x,\rho)$ at x_k.

To avoid confusion, in this chapter we refer to Algorithm 3.2 and Algorithm 3.3 as Algorithm 4.1 and Algorithm 4.2, respectively. In what follows, we present the theorem regarding the level function methods and discuss the method used in estimation of penalty parameter.

Theorem 4.7. *Let $\{x_k\}$ be generated by Algorithm 4.2. Assume that $f(x,\xi)$ and components of $G(x,\xi)$ are Lipschitz continuous functions with modulus $L_f(\xi)$ and $L_G(\xi)$ respectively, where $\mathbb{E}[L_f(\xi)] < +\infty$, $\mathbb{E}[L_G(\xi)] < +\infty$ and that the sequence of level functions $\{\sigma_{x_k}(x)\}$ is uniformly Lipschitz with constant M. Then*

$$\Delta(k) \le \epsilon, \text{ for } k > M^2\Upsilon^2\epsilon^{-2}\lambda^{-2}(1-\lambda^2)^{-1},$$

where Υ represents the diameter of the solution set $\mathcal{X}$, ϵ and λ are given in Algorithm 4.2.

Proof. It is easy to observe that the Lipschitz continuity of $G(x,\xi)$ w.r.t. x with modulus $L_G(\xi)$ implies the Lipschitz continuity of $\psi_j(x,\nu)$ with the same Lipschitz modulus $\mathbb{E}[L_G(\xi)]$. Along with the Lipschitzness of $f(x,\xi)$, this shows $\vartheta(x,\rho)$ is Lipschitz continuous with modulus $\mathbb{E}[L_f(\xi)]+\rho\mathbb{E}[L_G(\xi)]$. On the other hand, since $\vartheta(x,\rho)$ is convex, the function $\sigma_{x_k}(x)$ constructed at each iterate is a level function with modulus 1. The rest follows from Xu [146, Theorem 3.3]. □

In Algorithm 4.1 and Algorithm 4.2, a penalty parameter in $\vartheta(x,\rho)$ is fixed. In some cases, it might be difficult to estimate a good penalty parameter. One way to tackle this issue is to start with an estimate of penalty parameter and solve the resulting penalized problem with the above algorithms. The feasibility of the obtained solution is checked: if it is feasible the optimal solution is obtained, otherwise, the penalty parameter is increased the process is repeated. This kind of procedure in known as *Simple Penalty Function Method* in the literature of optimization, see for instance [138, Algorithm 10.2.3]. We describe the aforementioned procedure formally in the following algorithm for the penalized problem (4.2.22).

Algorithm 4.3 (Simple Penalty Function Method for penalized problem (4.2.22)).

Step 1. Let $\bar{\epsilon}$ be a positive number. Let ρ_0 be an intial estimate of the penalty parameter. Set $t := 0$.

Step 2. For $\bar{\rho} := \rho_t$, apply Algorithm 4.1 or 4.2 to solve problem (4.2.22). Let x_t denote the solution obtained from solving the problem.

Step 3. If $\max_{j\in\{1,\ldots,N\}}(\max_{\nu\in S}(\sum_{i=1}^{N} p_i(\nu^T\eta_j - \nu^T G(x_t,\xi^i))_+ - \gamma_j(\nu))_+) \leq \bar{\epsilon}$, stop; otherwise, set $x_{t+1} := x_t$, $\rho_{t+1} := 10\rho_t$ and $t := t+1$, go to step 2.

Algorithm 4.3 terminates in a finite number of iterations in that the exact penalty parameters for problem (4.2.22) is finite, see Theorems 4.6.

An alternative way to deal with the issue of penalty parameters is to solve the following problem

$$\min_{x\in X}\max_{j\in\{1,\ldots,N\}}(\max_{\nu\in S}(\sum_{i=1}^{N} p_i(\nu^T\eta_j - \nu^T G(x,\xi^i))_+ - \gamma_j(\nu))). \tag{4.3.25}$$

This can be achieved by applying Algorithm 4.1 or 4.2 directly. The optimal value of (4.3.25) effectively gives an upper bound for parameter $\bar{\delta}$ (see Theorems 4.6). Note that these parameters are dependent on the Slater condition of (4.2.12).

4.3.2 Modified cutting plane algorithm

Rodulf and Ruszczyński [120] and Fábián et al. [52] proposed cutting plane methods to solve stochastic program with unidimensional second order stochastic dominance constraints when the underlying random variable has finite distribution. This method is an extension of the cutting-plane method developed by Haneveld and Vlerk [81] for integrated chance constraints (ICC). In what follows, we consider a modification of the procedure where a cut is constructed.

Reformulate the optimization problem (4.2.12) as:

$$\begin{array}{ll} \min\limits_{x,z} & z \\ \text{s.t.} & \psi(x,\nu) := \max\limits_{j\in\{1,\dots,N\}} \bar{\psi}_j(x) \le 0, \\ & \sum\limits_{i=1}^{N} p_i f(x,\xi^i) - z \le 0, \\ & x \in \mathcal{X},\ z \in Z, \end{array} \tag{4.3.26}$$

where $\bar{\psi}_j(x) := \max_{\nu\in S} \psi_j(x,\nu)$, Z is a closed convex compact subset of $\mathbb{R}$ such that

$$\left\{ \sum_{i=1}^{N} p_i f(x,\xi^i) : x \in \mathcal{X} \right\} \subset Z.$$

Note that, the existence of set Z is due to the fact that $f(x,\xi^i)$, $i = 1,\dots,N$, is a continuous function and $\mathcal{X}$ is a compact set. Also the components of $G(x,\xi)$ are concave and $f(x,\xi)$ is convex w.r.t. x, which implies that $\psi(x,\nu)$ is convex w.r.t. x and $\sum_{i=1}^{N} p_i f(x,\xi^i) - z$ is convex w.r.t. (x,z). We apply the classical cutting-plane method [80] to both $\psi(x,\nu)$ and $\sum_{i=1}^{N} p_i f(x,\xi^i) - z$. Specifically, we propose the following algorithm.

Algorithm 4.4 (Modified cutting plane algorithm)
Define the optimization problem at iteration t as

$$\begin{array}{ll} \min\limits_{x,z} & z \\ \text{s.t.} & x \in \mathcal{X},\ z \in Z, \\ & (x,z) \in P_t := \left\{ (x,z) \in \mathcal{X}\times Z : a_l^T x \le b_l,\ d_l^T x + e_l z \le k_l,\ l = 1,\dots,t \right\}. \end{array} \tag{4.3.27}$$

Set $t := 0$, $P_0 := \mathcal{X} \times Z$. For each t, carry out the following.

Step 1. Solve the optimization problem (4.3.27), finding the optimal solution (x_t, z_t). If the problem (4.3.27) is infeasible, stop. Otherwise go to Step 2.

Step 2. Find the solution to $\max_{j\in\{1,\ldots,N\}} \bar{\psi}_j(x)$ and find optimal solution (η^*, ν^*), and set $\gamma(\nu_t^*) := \sum_{i=1}^N p_i(\nu_t^{*T}\eta_t^* - \nu_t^{*T}Y(\xi^i))_+$. If

$$\sum_{i=1}^N p_i(\nu^{*T}\eta^* - \nu^{*T}G(x,\xi^i))_+ \leq \gamma(\nu_t^*),$$

and

$$\sum_{i=1}^N p_i f(x_t,\xi^i) - z_t \leq 0,$$

stop: (x_t, z_t) is an optimal solution. Otherwise go to Step 3.

Step 3. Construct the set

$$\mathcal{J}_t := \left\{ i : (\nu^{*T}\eta^* - \nu^{*T}G(x,\xi^i)) > 0 \right\},$$

and the feasibility cuts $a_{t+1}^T x \leq b_{t+1}$, and $d_{t+1}^T x + e_{t+1} z \leq k_{t+1}$, where

$$a_{t+1} = -\sum_{i\in\mathcal{J}_t} p_i \nabla_x \nu^{*T} G(x_t,\xi^i),$$

$$b_{t+1} = \sum_{i\in\mathcal{J}_t} p_i(-\nabla_x \nu^{*T} G(x_t,\xi^i)^T x_t + \nu^{*T} G(x_t,\xi^i) - \eta^*) + \gamma(\nu^*),$$

$$d_{t+1} = -\nabla_x f(x,\xi),\ e_{t+1} = -1,\ k_{t+1} = -\nabla_x f(x_t,\xi)^T x_t + f(x_t,\xi).$$

and set

$$P_{t+1} = P_t \cap \left\{ (x,z) \in \mathcal{X}\times Z : a_{t+1}^T x \leq b_{t+1},\ d_{t+1}^T x + e_{t+1} z \leq k_{t+1} \right\}.$$

Proceed with iteration $t+1$.

Remark 4.8. We make a few comments about Algorithm 4.4.

(i) Algorithm 4.4 differs from the cutting-plane method discussed in [120, 52] in the way feasible cuts are constructed. In the former, N constraints/cuts are added at each iteration, these cuts are not necessarily the extreme support of $\psi(x,\nu)$ at x_t. In Algorithm 4.4, we exclude all those non-support constraints, instead we include a cut at the extreme support (to $\psi(x,\nu)$ at x_t) which we believe is the most effective and a single linear cut is adequate to ensure the convergence. All other non-support constraints/cuts may potentially reduce numerical efficiency. This approach is similar to the algorithm proposed by Fábián et al. [52]. Note that, Fábián's algorithm is applied to linear models while Algorithm 4.4 is applicable to the nonlinear case. Therefore, we may regard the latter as an extension of the former.

(ii) In Step 2 of the above algorithm, we solve the following DC-program

$$\max_{(\eta,\nu)\in[a,b]\times S} \psi(\eta,\nu) := \sum_{i=1}^{N} p_i[(\nu^T\eta - \nu^T G(x_t,\xi^i))_+ - (\nu^T\eta - \nu^T Y(\xi^i))_+],$$

where $[a,b] \supseteq \{\eta_1,\ldots,\eta_N\}$ and S is defined as in (4.2.4).

(iii) When f is linear w.r.t. x, there is no need to introduce additional variable z because the objective is linear.

We now present the convergence results.

Theorem 4.9. *Let $\{(x_t,z_t)\}$ be a sequence generated by the Algorithm 4.4. Let*

$$P := \{(x,z)\in\mathcal{X}\times Z : \psi(x,\nu)\le 0, \mathbb{E}[f(x,\xi)] - z \le 0\} \subset \mathcal{X}\times Z,$$

where $\psi(x,\nu)$ is defined in problem (4.3.26). Assume: (a) $f(x,\xi)$ and each of the component $g_i(x,\xi)$ of $G(x,\xi)$ are continuously differentiable and convex w.r.t. x for almost every ξ, (b) $\mathcal{X}\times Z \in \mathbb{R}^n$ is a compact set, (c) there exists a positive constant L such that the Lipschitz modulus of $\mathbb{E}[f(x,\xi)]$ and $\psi(x,\nu)$ are bounded by L on $\mathcal{X}\times Z$, (d) the set P is nonempty. Then, $\{(x_t,z_t)\}$ contains a subsequence which converges to a point $(x^*,z^*)\in P$, where (x^*,z^*) is the optimal solution.

Proof. The proof is similar to the results in [80]. Note that, at each iteration $t>0$, $a_{t+1}\in \partial_x\psi(x_t,\nu)$, $d_{t+1} = \nabla\mathbb{E}[f(x_t,\xi)]$, and $e_{t+1} = \nabla_z(\mathbb{E}[f(x_t,\xi)] - z_t = -1$. Then $a_{t+1}^T x - b_{t+1}$ and $d_{t+1}^T x + e_{t+1}z - k_{t+1}$ are the extreme support to the graphs of $\psi(x,\nu)$ and $\mathbb{E}[f(x,\xi)]-z$ at (x_t,z_t) respectively. By condition (a), $\psi(x,\nu)$ and $\mathbb{E}[f(x,\xi)]$ are convex and continuous w.r.t. (x,z). Consequently, if $(x_t,z_t)\in P$ and $\max\{\psi(x,\nu),\ \mathbb{E}[f(x,\xi)]\} \le 0$, then

$$\max\{a_{t+1}^T x_t - b_{t+1},\ d_{t+1}^T x_t + e_{t+1}z - k_{t+1}\} \le 0.$$

Further, for all $(x_t,z_t)\notin P$,

$$\max\{a_{t+1}^T x_t - b_{t+1},\ d_{t+1}^T x_t + e_{t+1}z - k_{t+1}\} = \max\{\psi(x_t,\nu),\ \mathbb{E}[f(x_t,\xi)] - z_t\} > 0.$$

Therefore, when $(x_t,z_t)\notin P$, the set P and the point (x_t,z_t) lie on opposite sides of the cutting angle $\max\{a_{t+1}^T x_t - b_{t+1},\ d_{t+1}^T x_t + e_{t+1}z - k_{t+1}\} = 0$.

Note that, from the definition of P_t and (x_t,z_t), we know that $P\subset P_t\subset P_{t-1}$, (x_t,z_t) minimizes z in P_t and $z_{t-1}\le z_t$. In the case when $(x_t,z_t)\in P$, it is easy to verify that (x_t,z_t) is the optimal solution of problem (4.3.26). Indeed, since (x_t,z_t) is an optimal solution, for every $(x,z)\in P_t$, we have $z\ge z_t$. Since $P\subset P_t$, then $z\ge z_t$ for $(x,z)\in P$, which implies optimality of (x_t,z_t) over P.

In what follows, we focus on the case when $(x_t,z_t)\notin P\ \forall t$. Since $\mathcal{X}\times Z$ is a compact set, the sequence $\{(x_t,z_t)\}$ contains a subsequence which converges to $(x^*,z^*)\in\mathcal{X}\times Z$.

Assume without loss of generality that $(x_t, z_t) \to (x^*, z^*)$. Let $P^* = \cap_t P_t$. Since P_t is compact and $P \subset P_t$, we have $P \subset P^*$ and $(x^*, z^*) \in P^*$. On the other hand, since

$$z \geq z_t, \forall (x, z) \in P_t,$$

then

$$z \geq z^*, \forall (x, z) \in P^*. \tag{4.3.28}$$

Indeed, if this is not true, then there exists $(\hat{x}, \hat{z}) \in P^*$ such that $\hat{z} < z^*$. Since $z_t \to z^*$, there exists some sufficiently large t such that $\hat{z} < z_t$. This is not possible because (x_t, z_t) is an optimal solution in P_t while $(\hat{x}, \hat{z}) \subset P^* \subset P_t$ is a feasible solutions. This shows that (4.3.28) holds. Since $P \subset P^*$, the inequality also holds for all $(x, z) \in P$, which implies (x^*, z^*) is an optimal solution of problem (4.3.27) if $(x^*, z^*) \in P$.

In what follows, we show that $(x^*, z^*) \in P$. Note that, (x_t, z_t) minimizes z in P_t, that is, it satisfies the inequalities:

$$a_{l+1}^T x - b_{l+1} \leq 0, \tag{4.3.29}$$

and

$$d_{l+1}^T x + e_{l+1} z - k_{l+1} \leq 0, \tag{4.3.30}$$

for $l = 0, \ldots, t-1$ and by condition (c), $\max\{\|a_{l+1}\|, \|d_{l+1}\|\} \leq L, \ \forall l$. Let $\{x_t, z_t\}$ denote the subsequence. We claim that $\{\max\{\psi(x_t, \nu), \ \mathbb{E}[f(x_t, \xi)] - z_t\}\}$ must converge to 0. Note that since

$$\begin{aligned} b_{l+1} &= \sum_{i \in \mathcal{J}_l} p_i (-\nabla_x \nu^{*T} G(x_l, \xi^i)^T x_l + \nu^{*T} G(x_l, \xi^i) - \eta^*) + \gamma(\nu^*), \\ &= a_{l+1}^T x_l - \psi(x_l, \nu^*), \\ &= a_{l+1}^T x_l - \psi(x_l, \nu), \end{aligned}$$

then (4.3.29) implies

$$\psi(x_l, \nu) + a_{l+1}^T (x - x_l) \leq 0.$$

Similarly, by the definition of e_{l+1}, k_{l+1}, we have from (4.3.30) that

$$\mathbb{E}[f(x_l, \xi)] + d_{l+1}^T (x - x_l) - z \leq 0.$$

Assume that the desired convergence does not occur. Then there exists an $r > 0$ independent of t such that

$$\begin{aligned} r &\leq \max\{\psi(x_l, \nu), \ \mathbb{E}[f(x_l, \xi)] - z_l\} \\ &\leq \max\{a_{l+1}^T (x_l - x_t), \ d_{l+1}^T (x_l - x_t) - (z_l - z_t)\}, \\ &\leq (L+1) \left\| (x_l, z_l) - (x_t, z_t) \right\|, \end{aligned}$$

for all $0 \leq l \leq t-1$, which shows that $\{(x_t, z_t)\}$ does not converge, a contradiction. This shows that

$$\{\max\{\psi(x_t, \nu),\ \mathbb{E}[f(x_t, \xi)] - z_t\}\}$$

converges to 0 and hence $(x_t, z_t) \in P$ is the optimal solution.

□

4.3.3 Cutting surface algorithm

In this section we give details of the cutting surface algorithm proposed by Homem-de-Mello et al. [70] to solve optimization problem with multivariate stochastic dominance constraints. They use cut-generation approach for solving the problem instead of adding all the constraints up front. In what follows, we restate the problem formulation and the cutting surface algorithm as discussed in [70].

Theorem 4.10. *[70, Theorem 1] Let P be a nonempty convex set. Let*

$$P_i := \left\{(\nu, y) | y_l \geq \nu^T (c^i - c^l),\ y_l \geq 0,\ \nu \in \tilde{P},\ l = 1, \ldots, r\right\},\ i = 1, \ldots, r. \tag{4.3.31}$$

where $\tilde{P} := \text{cl cone}(P) \cap \Delta$ [70, Proposition 1]. Then, the multivariate stochastic dominance constraints are equivalent to

$$\sum_{j=1}^{t} p_j (\nu^{ik^T} c^i - \nu^{ik^T} A^j x)_+ \leq \sum_{l=1}^{r} q_l (\nu^{ik^T} c^i - \nu^{ik^T} c^l)_+,\ i = 1, \ldots, r,\ k = 1, \ldots, \nu_i, \tag{4.3.32}$$

where ν^{ik} are the ν-component of the vertex solutions of P_i.

They define the following problem:

$$\begin{aligned} \min_{x} \quad & -\mathbb{E}[f(x, \xi)] \\ \text{s.t.} \quad & \sum_{j=1}^{t} p_j (\nu^{ik^T} c^i - \nu^{ik^T} A^j x)_+ \leq \sum_{l=1}^{r} q_l (\nu^{ik^T} c^i - \nu^{ik^T} c^l)_+,\ i = 1, \ldots, r,\ k = 1, \ldots, \nu_i. \end{aligned} \tag{4.3.33}$$

In the cut-generation approach they solve a sequence of linear relaxations of (4.3.33), over a subset of constraints in (4.3.33). At a solution $\hat{x}$ of a relaxed problem, they consider the subproblems

$$\begin{aligned} \min_{\nu, y} \quad & g(\nu, y) := \sum_{l=1}^{r} q_l y_l - \sum_{j=1}^{t} p_j (\nu^T c^i - \nu^T A^j \hat{x})_+ \\ \text{s.t.} \quad & (\nu, y) \in P_i. \end{aligned} \tag{4.3.34}$$

If all (4.3.34) have a non-negative objective value, we have a solution of (4.3.33). Otherwise, we have a vertex solution $\hat{\nu}$ of (4.3.34) with a negative objective value. Corresponding to this vertex, the constraint $\sum_{j=1}^{t} p_j(\nu^{ik^T} c^i - \nu^{ik^T} A^j x)_+ \leq \sum_{l=1}^{r} q_l(\nu^{ik^T} c^i - \nu^{ik^T} c^l)_+$, is a valid cut for $\hat{x}$.

Algorithm 4.5 outlines the basic steps.

Algorithm 4.5 (Cutting surface algorithm [70])

Step 0. Set $s := 0$, ν^0 :=an arbitrary vertex of $\mathcal{K}_i$, where $i \in \{1, \ldots, r\}$ is also chosen arbitrary. Set $\mathcal{V}$:= $(\nu^0, 0, i)$.

Step 1. Solve a linear programming reformulation of the problem

$$\begin{aligned} \min \quad & \mathbb{E}[f(x,\xi)] \\ \text{s.t.} \quad & \sum_{j=1}^{t} p_j(\nu^{ik^T} c^i - \nu^{ik^T} g(x,\xi^j))_+ \leq \sum_{l=1}^{r} q_l(\nu^{ik^T} c^i - \nu^{ik^T} Y^l)_+, \ (\nu, y, i) \in \mathcal{V}^s. \end{aligned} \tag{4.3.35}$$

If the problem is infeasible, stop; if it is unbounded, then let $\hat{x}$ and $\hat{h}$ be respectively a solution and a direction that generate a ray and go to Step 2. Otherwise, let $\hat{x}$ be an optimal solution (4.3.35) and go to Step 3.

Step 2. For each $j = 1, \ldots, t$, solve the linear program

$$\begin{aligned} \min \quad & \nu^T g(\hat{h}, \xi^j) \\ \text{s.t.} \quad & \nu \in \tilde{P}. \end{aligned} \tag{4.3.36}$$

If any of the problems (4.3.36) has negative objective value, let $\bar{\nu}$ be a vertex optimal solution to that problem and choose $i \in \{1, \ldots, r\}$ arbitrarily; let $\mathcal{V}^{s+1}$:=$\mathcal{V}^s \cup \{(\bar{\nu}, 0, i)\}$ and go to Step 5. Otherwise (i.e. of the problems (4.3.36) have non-negative objective values for all j), go to Step 3.

Step 3. Solve problems (4.3.34) to find one or more vertex solution(s) $(\nu, y) \in \mathcal{K}_i$, for some $i \in 1, \ldots, r$, such that

$$\sum_{l=1}^{r} q_l y_l - \sum_{j=1}^{t} p_j(\nu^T c^i - \nu^T g(\hat{x}, \xi^j)_+ < 0.$$

Let (ν^{ik}, y^{ik}), $k = 1, \ldots, k_i$ be these identified vertices.

Step 4. If no vertex solution is found in Step 3, stop; otherwise, let

$$\mathcal{V}^{s+1} := \mathcal{V}^s \cup \{(\nu^{ik}, y^{ik}, i), \ k = 1, \ldots, k_i\}.$$

Step 5. Set $s := s + 1$ and go to Step 1.

Theorem 4.11. *[70, Theorem 4] Algorithm 4.5 terminates after a finite number of steps with either an optimal solution to the true problem, or a proof of infeasibility or unboundedness of the true problem.*

In next section, we investigate the efficiency of the above algorithms along with the linearized method in [4].

4.4 Numerical Tests

We have carried out an academic test, a budget allocation example as well as a real world portfolio optimization problem on the proposed model and algorithms by using MATLAB 7.10 and IBM ILOG CPLEX 12.4 installed on a HP Notebook PC with Windows 7 operating system, and Intel Core i7 processor.

We consider primarily an academic test problem introduced in [70, Section 2.2] to examine the penalization approach and efficiency of our proposed methods. Additionally, for comparison purposes we consider a budget allocation problem as discussed in [4]. Moreover, we consider a portfolio optimization problem with real world test data to further investigate the efficiency of the proposed stochastic programming model with multivariate SSD constraint and compare it to the return generated by a Markowitz model and corresponding indices. Furthermore, to estimate the penalty parameter we have solved the optimization problem (4.3.25) using Algorithm 4.2 as discussed at the end of Section 4.3.1. Another approach is to integrate Algorithm 4.3 in Algorithm 4.1 and 4.2, to find a suitable penalty parameter. We solved the reformulated problem with Algorithms 4.1–4.5. For Algorithms 4.1 and 4.2 we use $\epsilon = 0.0001$ and $\lambda = 0.5$. In the rest of this section we report the corresponding results.

4.4.1 An academic example

Example 4.1. *Homem-de-Mello et al. [70] considered the following model using stochastic dominance:*

$$\begin{aligned} &\max_x \quad 3x_1 + 2x_2, \\ &\text{s.t.} \\ &\qquad -\begin{bmatrix} \xi^1 & 2 \\ 2 & \xi^2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \succeq_{(2)} -\begin{bmatrix} \xi^3 \\ 160 \\ \xi^4 \end{bmatrix}, \end{aligned} \tag{4.4.37}$$

where, ξ^i, $i = 1, \ldots, 4$ denotes a random variable. Let $\xi^1 := 4 \pm \alpha$, $\xi^2 := 2 \pm \alpha$, $\xi^3 := 200 \pm 10\beta$, and $\xi^4 := 40 \pm 5\beta$ where α and β are equal to 1. Where $\xi^3 := 200 \pm 10\beta$ suggests that the ξ^1 is a binary between 3 and 5.

To solve the optimization problem (4.4.37), Homem-de-Mello et al. [70] linearized the program and eliminated the redundant constraints. They obtained the optimal solution of problem (4.4.37) (with $\alpha = \beta = 1$) to be $x = (28.18, 34.55)$ and the corresponding objective value to be $f = 153.44$. Here we reformulate the optimization problem (4.4.37) as discussed in Section 4.2.3 and solved the reformulated problem by the proposed algorithms. We set the penalty parameter $\rho = 50$ and double it at each iteration. The optimal solution obtained by the proposed algorithms after 4 iteration is $x = (27.99, 34.66)$ and the corresponding objective value is $f = 153.29$.

4.4.2 A budget allocation model

The purpose of this example is to compare the efficiency of the level function method, and a modified cutting-plane method based on the exact penalization scheme with the linearized method proposed by Armbruster and Luedtke [4].

In what follows we present the budget allocation problem and study the behavior of the proposed model and methods to solve a simple budget allocation problem. This example is inspired by the budget allocation problem of Armbruster and Luedtke [4] and the example in [71]. Here we restate the problem:

Example 4.2. *Given a fixed budget, the problem is to determine what fraction of the budget to allocate to a set of candidate projects, $t \in \mathcal{T}$ with $|\mathcal{T}| = T$. The quality of a budget allocation is characterized by d distinct objectives, for which larger values are preferred. Each project $t \in \mathcal{T}$ is characterized by a d-dimensional random vector of reward rates R_t for these objectives. Thus, given a feasible budget allocation $x \in \mathcal{X} := \{x \in \mathbb{R}^T_+ : x \cdot 1 = 1\}$, the values of the d objectives are $\sum_{t\in\mathcal{T}} R_t x_t$. We assume that we are given a d-dimensional random vector Y that indicates a minimal acceptable joint performance of these objectives, and we require the performance of the chosen budget allocation to stochastically dominate Y. Subject to this condition, the goal is to maximize a weighted combination of the expected values of the measures:*

$$\begin{array}{ll} \max\limits_{x\in\mathcal{X}} & \sum\limits_{t\in\mathcal{T}} w^T \mathbb{E}[R_t] x_t \\ \text{s.t.} & \sum\limits_{t\in\mathcal{T}} R_t x_t \succeq^{Plin}_{(2)} Y, \end{array} \tag{4.4.38}$$

where $w \in \mathbb{R}^d_+$ is a given weight vector.

For the test instances, we assumed that the reward rate $R := [R_1, R_2, \ldots, R_T]$ are one of N equally likely scenarios $\{R^j : j \in \mathcal{N}\}$ sampled from a joint normal distribution with mean μ and covariance matrix Σ. The components of μ are chosen randomly from $U[10, 20]$ and the covariance matrix Σ was calculated as follows. The coefficient of variations were chosen from $U[0.2, 1.1]$. The correlation of any two distinct elements

(t,k) and (t',k') were chosen from $U[-0.2,0.4]$ if they share a project $(t=t')$ and from $U[-0.1,0.1]$ if they share an objective $(k=k')$ and were 0 otherwise. The benchmark random vector Y was determined from an allocation in which all projects are allocated an equal fraction of the budget, but to avoid being overly conservative, was then reduced by a fixed fraction δ of its mean. Specifically, a given realizations $R_t^j \in \mathbb{R}^d$, for each scenario j and project t, realization j of Y has a probability $q_y(j) = 1/N$ and is given by $Y^j = B^j - \delta(\frac{1}{N}\sum_{k=1}^{N} B^k)$ where $B^j = \frac{1}{T}\sum_{t\in\mathcal{T}} R_t^j$. In this example, we set the $\delta = 0.1$ and weight all objectives equally in the objective, $w = (1,\ldots,1)$.

Table 4.4.1 shows the computation times to solve these instances using the exact penalization scheme and solved by the projected level function algorithm and the modified cutting plane algorithm. Since the results of Algorithm 4.1 are identical to Algorithm 4.2, we only present the results for Algorithm 4.2. For these experiments, we varied the number of objectives $d \in \{3,5\}$, the number of projects $T \in \{50,100\}$, and the number of scenarios $N = M \in \{100,300,500\}$. For each combination of these parameters we display the average computation time in seconds over five instances at that size.

(d,T)	Algorithms	$N=100$	$N=300$	$N=500$
(3,50)	Algorithm 4.2	2.36	13.05	46.86
	Algorithm 4.4	24.13	219.67	739.27
(3,100)	Algorithm 4.2	2.69	14.78	53.48
	Algorithm 4.4	104.80	-	-
(5,50)	Algorithm 4.2	5.38	14.04	90.18
	Algorithm 4.4	24.10	236.28	527.14
(5,100)	Algorithm 4.2	10.20	26.69	98.47
	Algorithm 4.4	117.94	-	-

TABLE 4.4.1: Average solution times in seconds of five instances solved by projected level function algorithm and the modified cutting plane algorithm. The "'-"' indicate that the algorithms could not solve the problem within 30 minutes limit.

(d,T)	$N=100$	$N=300$	$N=500$
(3,50)	0.3	12.3	86.2
(3,100)	0.3	8.9	61.6
(5,50)	0.6	37.8	181.8
(5,100)	0.7	23.0	105.6

TABLE 4.4.2: Average solution times in seconds of five instances solved using linear SSD model proposed in [4].

These results indicate that with the exact penalization scheme and Algorithm 4.2 it is possible to solve instances with a relatively large number of scenarios with lower computation time compared to the linear SSD formulation model's results shown in Table 4.4.2. Although, the opposite is true for lower number of scenario, but one advantage of the proposed exact penalized model and the solution methods is that they can deal with both linear and nonlinear underlying functions. Furthermore, Algorithm 4.4 proved to be less efficient. This is because as the sample size increases, the construction of set $\mathcal{J}_t$ in Step 3 of the algorithm takes longer time.

4.4.3 Portfolio performance

Suppose that we have a fixed capital to be invested in n assets. Let R_i, $i = 1, \ldots, n$, denote the return of asset i. In practice, the return is often uncertain and we use a random variable ξ to describe the uncertainty. Specifically, we write R_i as $R_i(\xi)$ and in doing so we are assuming that all n assets have an identical random factor depending on ξ.

To simplify the discussion, we normalize the capital to 1 and use x_i, $i = 1, \ldots, n$, to denote the fraction of capital to be invested in asset i. The portfolio return can then be formulated as:

$$f(x, \xi) := R_1(\xi)x_1 + R_2(\xi)x_2 + \cdots + R_n(\xi)x_n. \tag{4.4.39}$$

We use the optimization problem (4.2.5) to optimize our investment strategy. To ease the presentation, we repeat the model:

$$\begin{array}{ll} \min\limits_{x \in X} & -\mathbb{E}[f(x, \xi)] \\ \text{s.t} & \nu^T g(x, \xi) \succeq^{Plin}_{(2)} \nu^T Y(\xi), \ \forall \nu \in S, \end{array} \tag{4.4.40}$$

where f is defined by (4.4.39). We need to specify $g(x, \xi)$ and X. The random variable $Y(\xi)$ plays the role of a benchmark outcome. For example, one may consider $Y(\xi) = g(\bar{x}, \xi)$, where $\bar{x} \in \mathcal{X}$ is some reasonable value of the decision vector, which is currently employed in the system. Note that $g(x, \xi)$ and $Y(\xi)$ are m-dimensional random vectors, rather than scalar variables. Additionally, we use set of linear constraints to define the set S, see (4.2.4).

To further examine the efficiency of the multivariate SSD model, we calculate the Conditional Value at Risk (CVaR) for random variable $f(x^*, \xi)$ where x^* is an approximate optimal solution obtained from solving (4.2.23). By definition for a specified probability level α, the Value at Risk (VaR) of a portfolio is the lowest amount C such that, with probability α, the profit does not fall below C. The CVaR_α is the conditional expectation of profit below C. In our context,

$$\text{CVaR}_\alpha(f(x^*, \xi)) = \sup_{\text{C}} \left\{ \text{C} - \frac{1}{\alpha}\mathbb{E}[(\text{C} - f(x^*, \xi))_+] \right\}, \tag{4.4.41}$$

where $\alpha \in (0, 1)$ is a pre-specified constant. Three values of α are commonly considered: 0.90, 0.95, 0.99. However, in our analysis we focus on the case of $\alpha = 0.95$.

Let us now estimate the penalty parameter ρ through Theorem 4.6. Referring back to Theorem 4.6, we need to calculate κ, δ, and D. Let $x_0 \in \mathcal{X}$ be the weights for an equally

weighted portfolio, the $\sum_{i=1}^{N} p_i \kappa(\xi^i) = 0.0084$. The δ can be calculated as follows:

$$\delta := -\max_{j\in 1,\dots,N}\left\{\max_{\nu\in S}\left(\max_{\mathcal{J}\in\mathcal{N}}\sum_{i\in\mathcal{J}} p_i(\nu^T\eta_j - \nu^T G(x_0,\xi^i)) - \sum_{i=1}^{N} p_i(\nu^T\eta_j - \nu^T Y(\xi^i))_+\right)\right\},$$

and we found $\delta \approx 4.865\text{E-}005$. We choose $D = 1$ and estimate the penalty parameter as follows:

$$\rho \geq \sum_{i=1}^{N} p_i \kappa(\xi^i)\delta^{-1} D = 192.39.$$

Example 4.3. *We consider m history of percentage returns, for three different group of n assets. Each of these groups could belong to a different Index. Our aim is to find an optimal investment strategy for a fixed capital in the n assets which maximized the expected profit subject to certain risk averse measures. Particularly we consider the following model:*

$$\begin{array}{ll}\min\limits_{x\in X} & -\mathbb{E}[f(x,\xi)],\\ \text{s.t} & \nu^T g(x,\xi) \succeq_{(2)} \nu^T Y(\xi),\end{array}$$

where $g(x,\xi) = [g_1(x,\xi)\ g_2(x,\xi)\ g_3(x,\xi)]$ and $Y(\xi) = [Y_1(\xi)\ Y_2(\xi)\ Y_3(\xi)]$. We apply the exact penalization as explained in Section 4.2.3 ans solve the reformulated problem by the level function algorithms (Algorithm 4.1 and 4.2), the cutting plane method (Algorithm 4.4), and the cutting-surface method (Algorithm 4.5). For the purpose of this example we set the upper bound and lower bound on the capital invested equal to 0.2 and 0, respectively.

We collected 300 daily historical returns of 53 FTSE 100, 53 Nasdaq100 and 30 Dow Jones assets prior to March 2011. We use the first 100 observations to construct the portfolio strategy. We solve the optimization problem using level function algorithms, modified cutting-plane method, and the cutting surface method. Table 4.4.3 shows the result of this example. In this example each component of the vector $g(x,\xi)$ corresponds to the sum of return of the assets belonging to each of the three indices computed as described in (4.4.39).

Algorithm	Iter.	Time	No. Assets	Return	CVaR
4.1	10	0.0188	6	0.034	0.015
4.2	9	0.0174	6	0.034	0.015
4.4	4	0.0166	6	0.034	0.014
4.5	6	0.653	6	0.034	0.015

TABLE 4.4.3: Time is in minutes. No. Assets refers to the number of assets in the optimal portfolio. The expected return of the benchmark portfolio $Y = [0.0051\ 0.0085\ 0.0069]$.

As it can be seen all four algorithms result in very similar portfolios with identical expected return and number of assets in the portfolio.

We set up a backtest and use the remaining 200 observations to construct an out-of-sample test in order to investigate the performance of the selected portfolio. Figures 4.4.1 and 4.4.2 show the difference of return on selected portfolio and benchmark portfolio. The benchmark portfolio represent the average return of the three indices. The comparison of the return to each index individually is presented in Figures A.1.1-A.1.6 in the Appendix A.

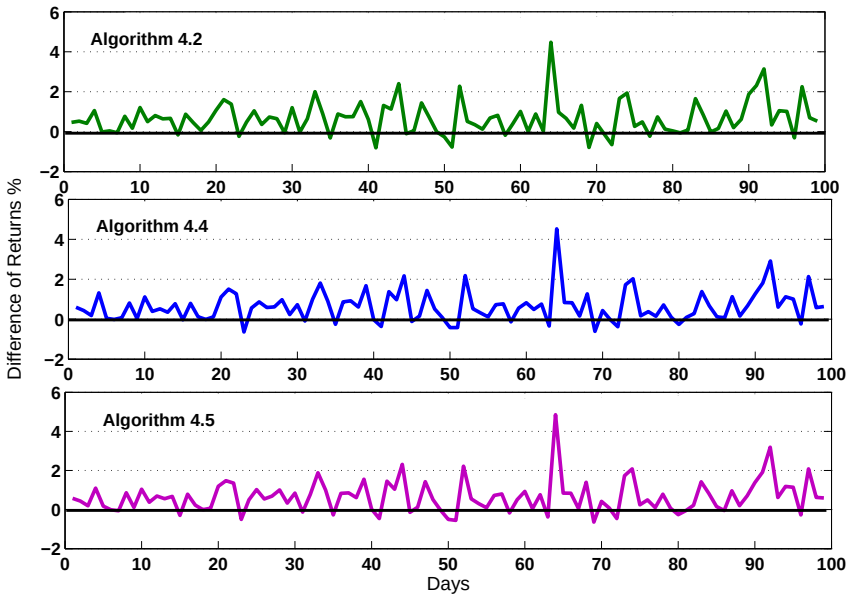

FIGURE 4.4.1: Backtesting of the difference of return on selected portfolios and indices. The benchmark portfolio is the average return of the indices.

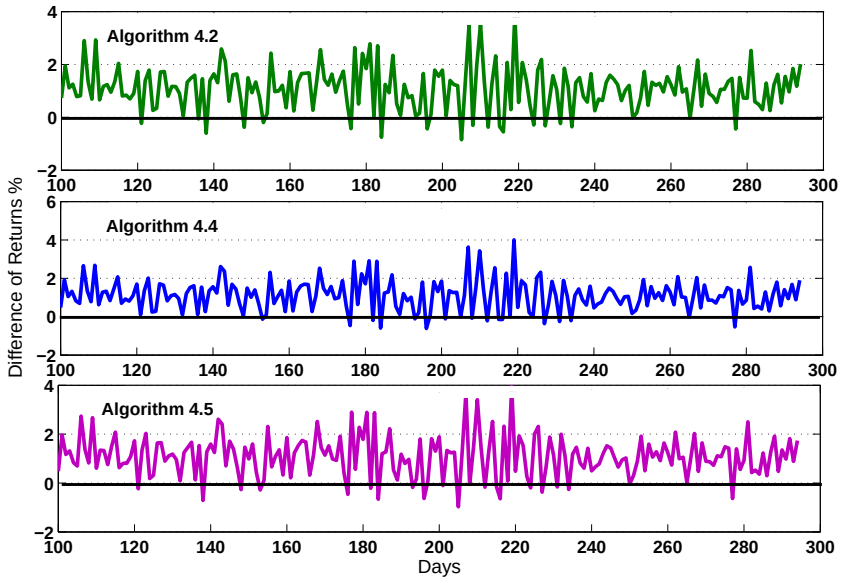

FIGURE 4.4.2: Out of sample test of the difference of return on selected portfolios and indices. The benchmark portfolio is the average return of the indices.

It can be seen that in both Figures 4.4.1 and 4.4.2, the line lies mostly above the zero line which means that the generated portfolio return is higher than the benchmark portfolio.

Furthermore, to investigate the dominance relationship we present the graph of cumulative distribution functions of portfolio return generated by the multi-SSD model using the Algorithms 4.2-4.5, Markowtitz model and the benchmark portfolio in Figure 4.4.3. It is clear that the generated portfolio strategy dominates the benchmark portfolio. Moreover, to see the performance of the generated strategy out-of-sample we present

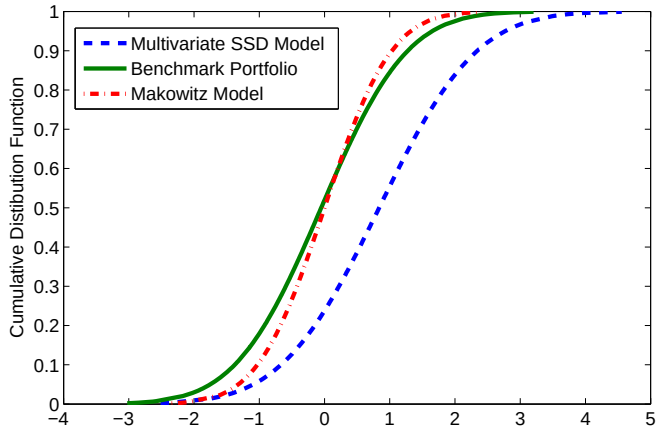

FIGURE 4.4.3: Cumulative distribution functions for the portfolio strategy generated by the Multi-SSD models, Markowitz model and the benchmark portfolio.

graph of cumulative return of the of portfolio return generated by the multi-SSD model using the Algorithms 4.2-4.5, Markowtitz model and the benchmark portfolio in Figure 4.4.4. It can be seen that the return generated by the portfolio strategy based on the Multivariate SSD model is much higher compared to the Markowitz model and the benchmark portfolio.

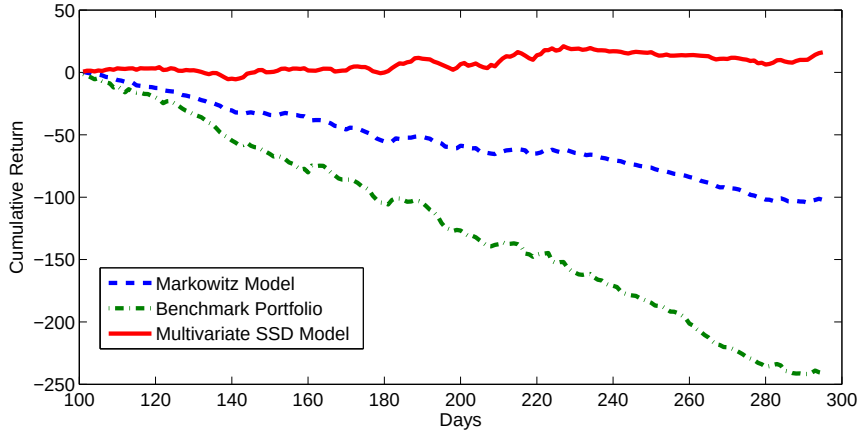

FIGURE 4.4.4: Out-of-sample cumulative return for the generated portfolio strategy based on the Multivariate SSD models, Markowitz model and the benchmark portfolio).

To illustrate the benefit of using multivariate stochastic dominance constraints, we compare the portfolio strategy constructed by the optimization problem (4.2.22) with an

investment strategy generated by Markowitz model as described below:

$$\begin{aligned} \max_{x\in\mathcal{X}} \quad & \mathbb{E}[f(x,\xi)] - \lambda\mathbb{E}[R(x,\xi)] \\ \text{s.t.} \quad & \mathbb{E}[g_i(x,\xi)] \geq R_i^b,\ i=1,\dots,m, \\ & \sum_{i=1}^{n} x_i = 1,\ x \geq 0,\ x \in \mathcal{X}, \end{aligned} \tag{4.4.42}$$

where $\lambda = 1$ is a fixed nonnegative number, $\mathbb{E}[R(x,\xi)]$ is the portfolio variance, R_i^b is the benchmark return set equal to the index i, $\mathbb{E}[g_i(x,\xi)]$ is the return of the asset belonging to index i, and $\mathbb{E}[f(x,\xi)]$ is the return defined as in (4.4.39).

Table 4.4.4 compares the portfolio generated by Markowitz model to the generated portfolio by the multivariate SSD model. As it can be seen, although the number of assets in the optimal portfolio are the same but the portfolio generated by the Markowitz model has a lower return and a higher CVaR.

Model	No. Assets	Return	CVaR
Multivariate SSD	6	0.034	0.014
Markowitz	6	0.032	0.018

TABLE 4.4.4: Time is in minutes. No. Assets refers to the number of assets in the optimal portfolio.

Figures 4.4.5 and 4.4.6 present the result of the backtest and out-of-sample test as described earlier. As it can be seen the portfolio generated by the optimization problem (4.2.23) outperforms the strategy generated by the Markowitz model (4.4.42) by having relatively higher returns both in-sample and out-of-sample.

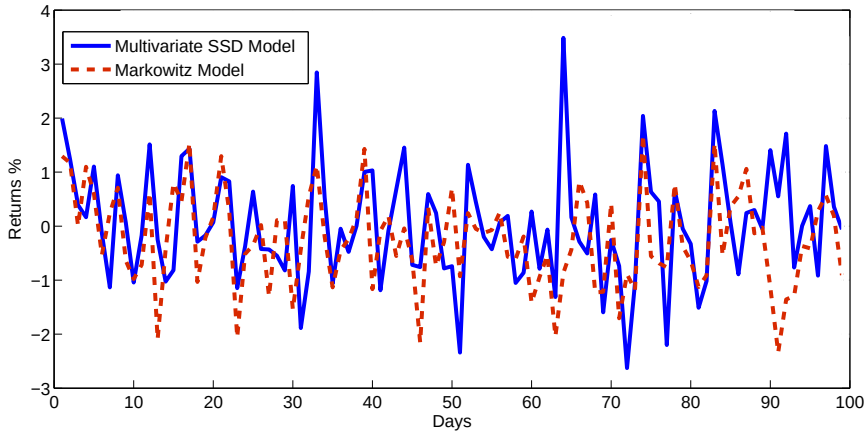

FIGURE 4.4.5: Comparing the backtest of the portfolio return of the optimization problem with multivariate SSD constraint and the Markowitz model.

To further compare the performance of the the two portfolio we use the Sortino ratio. The Sortino ratio measures the risk-adjusted return of an investment asset, portfolio or strategy. We used risk free rate (0.5%) and the benchmark portfolio as the required

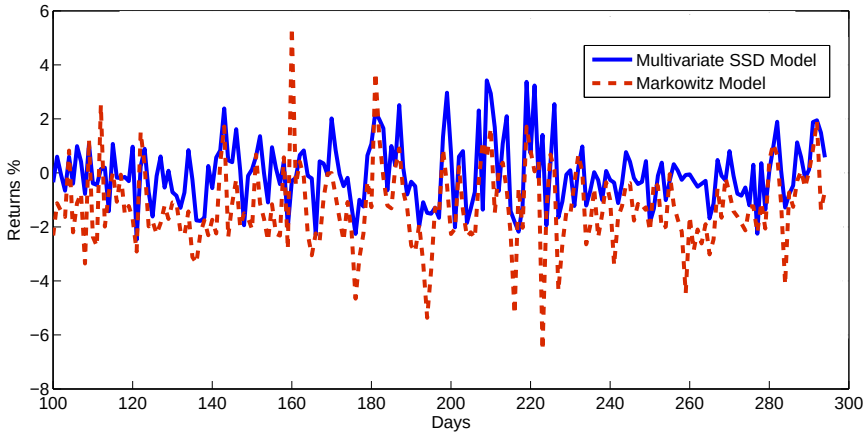

FIGURE 4.4.6: Comparing out-of-sample test of the portfolio return of the optimization problem with multivariate SSD constraint and the Markowitz model.

rate of return. We calculated the Sortino ratio both at the 100^{th} day and 300^{th} day. The results are shown in Table 4.4.5. As it can be seen the portfolio generated by the multivariate SSD model out perform the portfolio generated by the Markowitz model by having higher risk-adjusted return.

Model	Required return	100^{th} day Sortino ratio	300^{th} day Sortino ratio
Multivariate SSD Model	Benchmark	0.3969	0.3903
	Risk-free	0.2643	0.0749
Markowitz Model	Benchmark	0.1716	0.1308
	Risk-free	0.1795	0.0637

TABLE 4.4.5: Sortino ratio of the portfolio generated by optimization problem with multivariate SSD constraints and the Markowitz model.

Furthermore, we test the algorithms for various number of assets and record the CPU time. Figure 4.4.7 presents the result for this test. As it can be seen, all algorithms solve relatively large problems within a reasonable time. Additionally, we investigate the performance of the Algorithm 4.2, 4.4, and 4.5 as the number of observations increases. This is illustrated in Figure 4.4.8. Although the Figure 4.4.8 shows that the Algorithm 4.5 becomes inefficient, in our numerical tests increasing the number of observations did not result in a better portfolio.

4.5 Conclusion

In this chapter we studied stochastic programming with multivariate second order stochastic dominance constraints. Specifically, we proposed an exact penalty method for second order multivariate stochastic dominance constraints. Furthermore, we solved the penalized problem (4.2.22) using the level function method as well as a modified cutting-plane method inspired by the methods proposed in [120, 52]. These method were compared

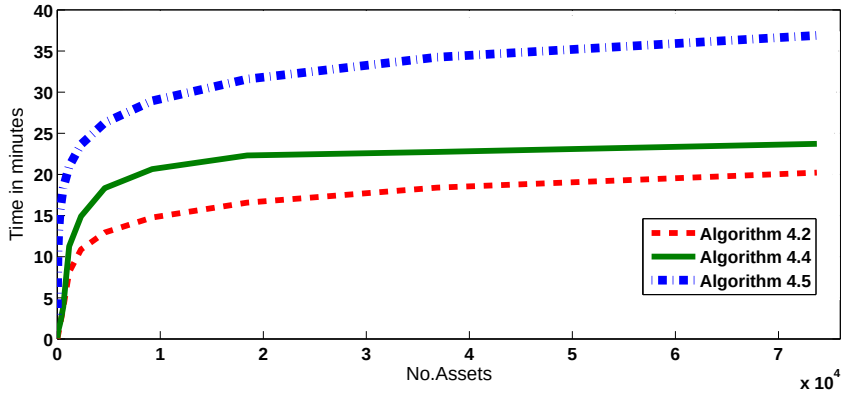

FIGURE 4.4.7: Graph of CPU time for various number of instruments for each algorithm.

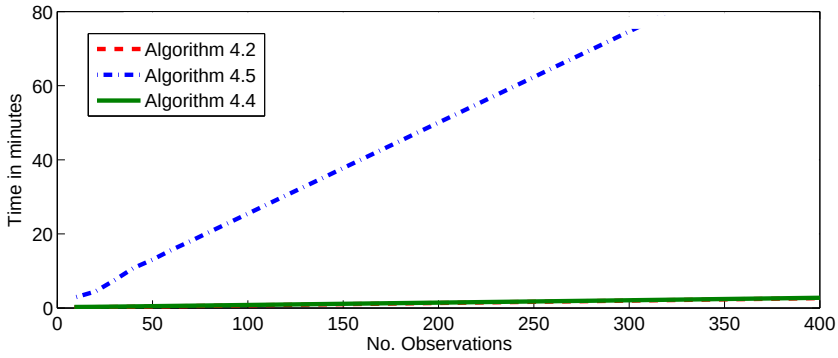

FIGURE 4.4.8: Graph of CPU time for various number of instruments for each algorithm.

to the cutting surface method proposed in [70], and the linearized method proposed in [4].

We applied the penalization scheme and the numerical methods to an academic test problem, a budget allocation problem, and a portfolio optimization problem. The academic test results showed that the penalization approach and the numerical methods results in similar optimal solution as the solution generated in [70, Section 2.2]. The budget allocation problem showed that the proposed method solved with Algorithm 4.2 is more efficient compared to the linearized method when the sample size is large. However, this is not the case when sample size is relatively small. The main advantage of our proposed method to the linearized method is that it can deal with nonlinear underlying functions. In the portfolio optimization problem, we used data of 136 assets from three different indices (FTSE 100, Nasdaq100, and Daw Jones). To investigate the performance of generated portfolio strategy, we set up a backtest and an out-of-sample test and compared the performance of the selected portfolio to the corresponding indices. We concluded that the generated portfolio performs better than the indices in sense of higher return both in-sample and out-of-sample.

Furthermore, to illustrate the benefit of considering multivariate stochastic dominance, we introduced the Markowitz model (4.4.42) and compared the performance of the Multivariate SSD model to Markowitz model both in-sample and out-of-sample as well as based on the Sortino ratio. It was concluded that the portfolio optimization problem with multivariate SSD constraints outperforms the portfolio optimization problem based on Markowitz model by having higher risk-adjusted return.

Moreover, we performed a test to investigate the effect of number of instrument on the computation time for each algorithm. These test suggested as anticipated that the projected level function algorithm and the cutting plane method can solve a large problems within reasonable time.

Chapter 5

Robust Reward-Risk Ratio Optimization

5.1 Overview

Stochastic programming has established itself as a powerful modeling tool when an accurate probabilistic description i.e. accurate values for the system parameters and specific probability distributions for the random variables are available. However, such information is rarely available in practice. In such situations, there are two major ways to deal with the uncertainties. One is through sample average approximations (SAA) also known as Monte Carlo method, where SAA of the expected value of the underlying function is constructed using empirical data, for a detailed discussion see [127]. The other is to use partial information such as moment to identify a set of possible probability distributions within which the true distribution lies. A robust optimization approach to this problem is based on making the decision that would be appropriate given the worst probability distribution amongst the set of possible distributions.

The robust optimization for stochastic programming can be traced back to the work by Scarf [125]. There has been extensive improvements in this field primarily driven by applications in risk management, finance and engineering [61, 65, 102, 154]. Given historical data, it is easier to estimate moment information of random parameters than to derive their probability distributions. This motivates the use of moment information in developing uncertainty models for random parameters.

The problem of moment has been studied by Stieltjes [136] in the ninetheenth century. The problem is related to the characterization of a feasible sequence of moments. Schmudgen [126], Putinar [110], and Curto and Fialkow [30] derived necessary and sufficient conditions sequences of moments with different settings. The problem of moments is also related to optimization over polynomials (the dual theory of moment). Lasserre

[87] and Parrilo [107] among others proposed relaxation hierarchies for optimization over polynomials using moment results. Bertsimas and Popescu [19] studied further the optimal inequalities given moment information. Moment problems in finance such as option pricing problems have been investigated in the literature (see [18, 94, 23]).

Inspired with the notion of reward-risk ratio optimization, we study robust stochastic reward-risk ratio optimization. We propose two robust formulations, one based on mixture distribution, and the other based on the first moment approach. We propose a sample average approximation formulation as well as a penalty scheme for the two robust formulations, respectively and solve the latter with the level function method proposed by Lemarechal et al. [89] and extended by Xu [146].

The main contribution of this chapter can be summarized as follows:

- We propose a robust optimization problem for a reward-risk ratio optimization based on mixture distribution and first order moment approach. For the case of mixture distribution a sample average approximation formulation is also presented. Moreover, an exact penalization scheme is proposed for the first order moment approach to handle the semi-infinite constraints in the dual problem.
- The proposed methods are applied to the Sortino performance ratio and the robust formulations based on both mixture distribution and the first order moment approach are derived.
- We investigate the numerical efficiency and accuracy of the proposed methods by presenting a portfolio optimization problem and a fund of funds problem based on real world data. We further, set up backtest and out-of-sample test to inspect the performance of the generated portfolios and compare them to the benchmark portfolio.

In the rest of this chapter we focus on numerical aspect of robust reward-risk ratio optimization. In Section 5.2 we introduce a reward-risk ratio optimization with one sided risk measure. In Section 5.2.2, we present robust formulation of reward-risk ratio optimization problem with one sided risk measure based on both mixture distribution and first moment approach followed with an exact penalization scheme. In Section 5.4, we apply the proposed methods to Sortino performance ratio. In Section 5.5, we present some numerical test results. Finally, in Section 5.6 we present the conclusions.

5.2 Robust Reward-Risk Ratio Optimization

5.2.1 Introduction

There are two basic approaches to the problem of portfolio selection under uncertainty. One is the stochastic dominance approach, based on the axiomatic model of risk-averse preferences. The optimization problems that arise are not easy to solve in practice as it was discussed in previous chapters. The other is the reward-risk analysis. According to the reward-risk analysis, the portfolio choice is made with respect to two criteria, the expected portfolio return and the portfolio risk. A portfolio is preferred to another one if it has higher expected return and lower risk.

Related to the reward-risk analysis is the reward-risk ratio optimization. Since the publication of the Sharpe ratio, see [132], which is based on the mean-variance analysis, some new performance measures like the STARR ratio, the Minimax measure, Sortino ratio, Farinelli-Tibiletti ratio and most recently the Rachev ratio and the Generalized Rachev ratio have been proposed (for an empirical comparison, see Biglova et al. [22], Rachev et al. [124] and the references therein). The new ratios take into account empirically observed phenomena, that the distributions of asset's returns are fat-tailed and skewed, by incorporating proper reward and risk measures.

In this chapter, we focus on general performance measure optimization with one-sided risk measure, where only downward variations are penalized. We will discuss this in details in Section 5.4. Specifically, we consider following optimization problem:

$$\max_{x \in X} \quad \frac{\mathbb{E}[\mu(x,\xi) - Y(\xi)]}{\mathbb{E}[(Y(\xi) - \mu(x,\xi))_+]}, \tag{5.2.1}$$

where $Y(\xi)$ is a benchmark and $(a)_+ = \max(a, 0)$. Note that, the numerator is concave and equivalent to the expected excess return over the benchmark, while the denominator is a convex one sided risk measure. In what follows, we explore the possible simplification of the problem that would facilitate the numerical solution.

Pinar et al. [108] showed that under some assumptions, the problem (5.2.1) can be reformulated as:

$$\begin{array}{ll} \max\limits_{(x,\tau) \in X \times \mathbb{R}} & \tau \\ \text{s.t.} & \mathbb{E}[\mu(x,\xi) - Y(\xi) - \tau(Y(\xi) - \mu(x,\xi))_+] \geq 0, \\ & x \in X. \end{array} \tag{5.2.2}$$

In what follows, we will present the robust reformulation of the problem (5.2.2) based on the mixture distribution approach and the first order moment approach.

5.2.2 Robust formulations

In mathematical optimization models, we commonly assume that the data inputs are precisely known and ignore the influence of parameter uncertainties on the optimality and feasibility of the models. Robust optimization addresses the issue of data uncertainties from the perspective of computational tractability. Since its introduction by H. Scarf in 50s [125], the robust optimization model also known as the minimax stochastic program, has granted a lot of interest in both academic and practitioner's communities. In what follows, we present a robust formulation for the optimization problem (5.2.1).

Let $\mathcal{P}$ denote a set of probability distributions which contains the true probability. We consider a robust scheme for problem (5.2.1) as follows:

$$\max_{x\in X}\min_{P\in\mathcal{P}} \quad \frac{\mathbb{E}_P[\mu(x,\xi)-Y(\xi)]}{\mathbb{E}_P[(Y(\xi)-\mu(x,\xi))_+]}. \tag{5.2.3}$$

Moreover, the robust counterpart of the problem (5.2.2) can be formulated as:

$$\begin{array}{ll}\max\limits_{(x,\tau)\in X\times\mathbb{R}} & \tau \\ \text{s.t.} & \min\limits_{P\in\mathcal{P}} \mathbb{E}_P[\mu(x,\xi)-Y(\xi)-\tau(Y(\xi)-\mu(x,\xi))_+]\geq 0.\end{array} \tag{5.2.4}$$

In the above two formulations, the robustness is in the sense that given the set of probability measures $\mathcal{P}$, an optimal solution is sought against the worst probability measure which is used to compute the expected value of the objective function. Note that, the robust problem (5.2.4) depends on the choice of $\mathcal{P}$, and an optimal solution of problems (5.2.3) and (5.2.4) provide a lower bound of the true optimal values of the true problems (5.2.1) and (5.2.2), respectively.

Theorem 5.1. *Consider the two robust optimization problems (5.2.4) and (5.2.3). The solution of problem (5.2.4) is equivalent to the solution of problem (5.2.3).*

Proof. Let (x^*,τ^*) be the optimal solution to problems (5.2.4). Let $\hat{x}$ be the optimal solution of the problem (5.2.3) and $\hat{\tau}$ be defined as

$$\min_{P\in\mathcal{P}} \frac{\mathbb{E}_P[\mu(\hat{x},\xi)-Y(\xi)]}{\mathbb{E}_P[(Y(\xi)-\mu(\hat{x},\xi))_+]} = \hat{\tau}.$$

Then

$$\mathbb{E}_P[(\mu(x^*,\xi)-Y(\xi))-\tau^*(Y(\xi)-\mu(x^*,\xi))_+]\geq 0,\ \forall P\in\mathcal{P}.$$

Consequently

$$\frac{\mathbb{E}_P[\mu(x^*,\xi)-Y(\xi)]}{\mathbb{E}_P[(Y(\xi)-\mu(x^*,\xi))_+]}\geq \tau^*,\ \forall P\in\mathcal{P},$$

then

$$\min_{P\in\mathcal{P}} \frac{\mathbb{E}_P[\mu(x^*,\xi)-Y(\xi)]}{\mathbb{E}_P[(Y(\xi)-\mu(x^*,\xi))_+]}\geq \tau^*.$$

Therefore, we can conclude that (x^*, τ^*) is a feasible solution of problem (5.2.4), and $\hat{\tau} \geq \tau^*$.

Similarly, for the problem (5.2.3) we can say that

$$\frac{\mathbb{E}_P[\mu(\hat{x}, \xi) - Y(\xi)]}{\mathbb{E}_P[(Y(\xi) - \mu(\hat{x}, \xi))_+]} \geq \hat{\tau}, \ \forall P \in \mathcal{P}.$$

Then

$$\mathbb{E}_P[(\mu(\hat{x}, \xi) - Y(\xi)) - \hat{\tau}(Y(\xi) - \mu(\hat{x}, \xi))_+] \geq 0, \ \forall P \in \mathcal{P}.$$

Therefore

$$\min_{P \in \mathcal{P}} \mathbb{E}_P[(\mu(\hat{x}, \xi) - Y(\xi)) - \hat{\tau}(Y(\xi) - \mu(\hat{x}, \xi))_+] \geq 0.$$

Consequently, we can conclude that $(\hat{x}, \hat{\tau})$ is feasible solution of problem (5.2.4), and $\hat{\tau} \leq \tau^*$. □

In what follows, we focus on solving problem (5.2.4), we can reformulate the optimization problem (5.2.4) as follows:

$$\begin{array}{ll} \min\limits_{(x,\tau) \in X \times \mathbb{R}} & -\tau \\ \text{s.t.} & -[\min\limits_{P \in \mathcal{P}} \mathbb{E}_P[(\mu(x, \xi) - Y(\xi)) - \tau(Y(\xi) - \mu(x, \xi))_+]] \leq 0. \end{array} \tag{5.2.5}$$

Note that, the constraint is the optimization problem (5.2.5) can equivalently be written as

$$\max_{P \in \mathcal{P}} \ -\mathbb{E}_P[(\mu(x, \xi) - Y(\xi)) - \tau(Y(\xi) - \mu(x, \xi))_+] \leq 0. \tag{5.2.6}$$

Consequently, the optimization problem (5.2.5) can be reformulated as follows:

$$\begin{array}{ll} \min\limits_{(x,\tau) \in X \times \mathbb{R}} & -\tau \\ \text{s.t.} & \max\limits_{P \in \mathcal{P}} \ -\mathbb{E}_P[(\mu(x, \xi) - Y(\xi)) - \tau(Y(\xi) - \mu(x, \xi))_+] \leq 0. \end{array} \tag{5.2.7}$$

In what follows, we focus on the constraint of optimization problem (5.2.7).

The robust optimization problem is also referred to as minimax, maximin or distributional robust optimization in the literature of stochastic programming. Research on minimax robust optimization dates back to the pioneering work by Zackova [150] and Dupačová [41]. More recently, substantial extensions have been done by Shapiro and Kleywegt [130], Shapiro and Ahmed [128], and Bertsimas et al. [16]. Relationship between robust optimziation and minimax/distributional robust optimization have been discussed by Goh and Sim [60] and Xu et al. [147].

There are various ways to define the set $\mathcal{P}$ depending on the availability of information on the distribution of the unknown parameter. In this chapter, we consider two specific

cases, one is that $\mathcal{P}$ is defined as a mixture of a set of known probability distributions; the other is that $\mathcal{P}$ is defined through first order moments.

5.2.3 Mixture distribution

Assume that ξ is only known to belong to a set of distributions which consists of all the mixutres of some predetermined likelihood distributions, i.e.,

$$P \in \mathcal{P} \triangleq \left\{ \sum_{l=1}^{L} \lambda_l P_l : \ \sum_{l=1}^{L} \lambda_l = 1, \ \lambda_l \geq 0, \ l = 1, \ldots, L \right\}, \tag{5.2.8}$$

where $P_1, \ldots, P_L$ is the set of probability measures for $l = 1, \ldots, L$ and L denotes the number of the likelihood distributions.

In this setup, we assume that probability distributions P_l, $l = 1, \ldots, L$, are known and the true probability distribution is in the convex hull of them. Mixture distribution has already been studied in robust statistics and used in modeling. More recently, Zhu and Fukushima [153] studied robust optimization of CVaR of a random function under mixture probability distributions. Here we apply the approach to a fractional optimization problem.

With the $\mathcal{P}$ defined as in (5.2.8), we can write problem (5.2.6) as follows:

$$\begin{array}{ll} \max & -\sum\limits_{l=1}^{L} \lambda_l \mathbb{E}_{P_l}[(\mu(x,\xi) - Y(\xi)) - \tau(Y(\xi) - \mu(x,\xi))_+] \\ \text{s.t.} & \sum\limits_{l=1}^{L} \lambda_l = 1, \\ & \lambda_l \geq 0, \ l = 1, \ldots, L. \end{array} \tag{5.2.9}$$

Due to linearity with respect to variable λ_l, $l = 1, \ldots, L$, the problem (5.2.9) is equivalent to

$$\max_{l=1,\ldots,L} \ -\mathbb{E}_{P_l}[(\mu(x,\xi) - Y(\xi)) - \tau(Y(\xi) - \mu(x,\xi))_+].$$

Consequently, the problem (5.2.7) can be reformulated as follows:

$$\begin{array}{ll} \min\limits_{(x,\tau)\in X\times\mathbb{R}} & -\tau \\ \text{s.t.} & -\mathbb{E}_{P_l}[(\mu(x,\xi) - Y(\xi)) - \tau(Y(\xi) - \mu(x,\xi))_+] \leq 0, \ \text{for } l = 1, \ldots, L. \end{array} \tag{5.2.10}$$

In the case when ξ has a finite discrete distribution, that is, ξ takes finite number of values $\xi^1, \ldots, \xi^N$, the formulation (5.2.10) can be simplified as follows:

$$\begin{aligned} \min_{(x,\tau)\in X\times\mathbb{R}} \quad & -\tau \\ \text{s.t.} \quad & -\sum_{i=1}^{N} p_l^i[(\mu(x,\xi_l^i) - Y(\xi_l^i)) - \tau(Y(\xi_l^i) - \mu(x,\xi_l^i))_+] \leq 0, \text{ for } l = 1, \ldots, L. \end{aligned} \tag{5.2.11}$$

where p_l^i corresponds to the probability measure of P_l in scenario i.

In the case when P_l satisfies continuous distribution, it might be undesirable to compute expected values of the underlying random functions with respect to the distribution of the probability measure. One way to tackle this issue is through sample average approximation. For a fixed l, let $\xi_l^1, \ldots, \xi_l^{N_l}$ denote independent and identically distributed random variable to ξ with distribution P_l. Then $\mathbb{E}_{P_l}[f(x,\xi)]$ can be approximated by

$$\frac{1}{N_l}\sum_{i=1}^{N_l} f(x,\xi_l^i).$$

Consequently, The sample average approximation of the optimization problem (5.2.10) can be formulated as follows:

$$\begin{aligned} \min_{(x,\tau)\in X\times\mathbb{R}} \quad & -\tau \\ \text{s.t.} \quad & -\sum_{i=1}^{N} p_l^i[(\mu(x,\xi_l^i) - Y(\xi_l^i)) - \tau(Y(\xi_l^i) - \mu(x,\xi_l^i))_+] \leq 0, \text{ for } l = 1, \ldots, L. \end{aligned} \tag{5.2.12}$$

The ξ_l^i can be generated by computer simulation under probability distribution P_l.

5.2.4 First order moment approach

Recall, the robust optimization problem (5.2.7). Let set $\mathcal{P}$ denote the possible probability distributions that is assumed to include the true P. In what follows we discuss the approach used to solve the optimization problem (5.2.7). Recall problem (5.2.5) and let

$$\min_{P\in\mathcal{P}} \mathbb{E}_P[G(x,\xi,\tau)] := \mathbb{E}_P[(\mu(x,\xi) - Y(\xi)) - \tau(Y(\xi) - \mu(x,\xi))_+]. \tag{5.2.13}$$

We denote by $\mathcal{B}$ the set of probability measures on $(\Xi, \mathcal{F})$ and $\mathbb{E}_P[G(x,\xi,\tau)]$ is given by the integral

$$\mathbb{E}_P[G(x,\xi,\tau)] = \int_{\Xi} G(x,\xi,\tau)Pd(\xi).$$

It is reasonable to assume that we have some knowledge about certain moments of corresponding probability distribution. That is, the set $\mathcal{P}$ is defined by moment constraints

as follows

$$\mathcal{P} := \left\{ P \in \mathcal{B} : \begin{array}{l} \mathbb{E}_P[\psi_i(\xi)] = b_i,\ i = 1, \ldots, p, \\ \mathbb{E}_P[\psi_i(\xi)] \leq b_i,\ i = p+1, \ldots, q, \end{array} \right\},$$

where $\psi_i : \Xi \to \mathbb{R}$, $i = 1, \ldots, q$, are measurable functions. In practice, b_i is the mean value of a random variable $\psi_i(\xi)$, which can often be estimated. Moment of distributions are well discussed in the literature (see [142, 143] and references therein).

In what follows, we use the first order moment condition to derive the dual of problem (5.2.13). For a given $x \in X$, the Lagrangian dual of the problem (5.2.13) is as follows:

$$\max_{\lambda \in \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^{q-p}_+} \min_{P \succeq 0} L_x(P, \lambda), \tag{5.2.14}$$

where

$$L_x(P, \lambda) := \mathbb{E}_P[G(x, \xi, \tau)] + \lambda_0(1 - \mathbb{E}_P[1]) + \sum_{j=1}^{q} \lambda_j (b_i - \mathbb{E}_P[\psi_i(\xi)]).$$

Consequently, we can write the dual problem (5.2.14) in the form

$$\begin{array}{ll} \max\limits_{\lambda \in \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^{q-p}_+} & \lambda_0 + \sum\limits_{i=1}^{q} b_i \lambda_i \\ \text{s.t.} & \lambda_0 + \sum\limits_{i=1}^{q} \lambda_i \psi_i(\xi) \geq G(x, \xi, \tau),\ \xi \in \Xi. \end{array} \tag{5.2.15}$$

Note that, if the set Ξ is finite, the problem (5.2.13) and its dual (5.2.15) are linear programming problems. In that case there is no duality gap between these problems unless both are feasible. If the set Ξ is infinite, then the dual problem (5.2.15) becomes a linear semi-infinite programming problem. In that case one needs to verify some regularity conditions in order to ensure "no duality gap" property. For further detail see [129, Page 310].

We can reformulate the problem (5.2.15) as

$$\begin{array}{ll} \min\limits_{\lambda \in \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^{q-p}_+} & -\lambda_0 - \sum\limits_{i=1}^{q} b_i \lambda_i \\ \text{s.t.} & \lambda_0 + \sum\limits_{i=1}^{q} \lambda_i \psi_i(\xi) \geq G(x, \xi, \tau),\ \xi \in \Xi. \end{array} \tag{5.2.16}$$

Applying the dual formulation (5.2.16) to the problem (5.2.6), we can write problem (5.2.7) as the following semi-infinite problem

$$\begin{aligned} \min_{x\in X,\tau\in\mathbb{R},\lambda\in\mathbb{R}\times\mathbb{R}^p\times\mathbb{R}_+^{q-p}} \quad & -\tau \\ \text{s.t.} \quad & \lambda_0 + \sum_{i=1}^{q} b_i\lambda_i \le 0, \\ & \lambda_i \ge 0, \text{ for } i = p+1,\ldots,q, \\ & \lambda_0 + \sum_{i=1}^{q} \lambda_i\psi_i(\xi) \ge G(x,\xi,\tau),\ \xi\in\Xi, \end{aligned} \tag{5.2.17}$$

In what follows, we will discuss an exact penalization to deal with the infinite number of constraint in optimization problem (5.2.17).

5.2.4.1 An exact penalization

Bertocchi et al. [14] proposed an exact penalization approach for handling general semi-infinite constraints where underlying functions are nonlinear in ξ. Recall that

$$G(x,\xi,\tau) = (\mu(x,\xi) - Y(\xi)) - \tau(Y(\xi) - \mu(x,\xi))_+.$$

Assume that the support set of Ξ is compact and for each $\xi\in\Xi$, $G(x,\xi,\tau)$ is convex w.r.t. x. Let

$$R(x,\tau,\lambda_0,\lambda,\xi) := G(x,\xi,\tau) - \lambda_0 - \lambda^T\psi(\xi),$$

and $w := (x,\tau,\lambda_0,\lambda)$. The semi-infinite constraint of (5.2.17) can be written as

$$\max_{\xi\in\Xi} R(w,\xi) \le 0. \tag{5.2.18}$$

Definition 5.2. Let $\mathcal{F}$ denote the set of solutions to (5.2.18). Problem (5.2.18) is said to satisfy Slater condition if there exist a positive number $\bar{\delta}$ and a point $\bar{w}\in\mathcal{F}$ such that

$$\max_{\xi\in\Xi} R(\bar{w},\xi) \le -\bar{\delta}.$$

The problem (5.2.18) is said to satisfy strong Slater condition if there exists a positive number γ such that for any $w\in\mathcal{F}$ with $R(w,\xi) = 0$ for some $\xi\in\Xi$, there exists a point $\hat{w}$ with $R(\hat{w},\xi) < 0$ for all $\xi\in\Xi$ and

$$\|w - \hat{w}\| \le \gamma \min_{\xi\in\Xi} -R(\hat{w},\xi)_+.$$

Note that, the strong Slater condition requires the Slater condition. The notion of strong Slater condition is introduced by Gugat [65] for deriving error bound of a semi-infinite convex system of inequalities.

Lemma 5.3. *Assume: a) the inequality system (5.2.18) satisfies the strong Slater condition, b) function $R(\cdot,\xi)$ is convex on W for every $\xi \in \Xi$. Then there exists a positive number $\gamma > 0$ such that*

$$d(w,\mathcal{F}) \leq \gamma \max_{\xi\in\Xi}(R(w,\xi))_+,$$

where $d(w,\mathcal{F})$ denotes the distance between point w and set $\mathcal{F}$ and $a_+ = max(a,0)$ for $a \in \mathbb{R}$.

For the proof refer to [14, Lemma 2.1].

Theorem 5.4. *[14, Theorem 2.1] Consider the following convex program:*

$$\begin{array}{ll} \min & Q(w), \\ \text{s.t.} & H(x,\xi) \leq 0, \\ & w \in W, \end{array} \tag{5.2.19}$$

where $W = X \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}_+^{q-p}$. Assume: a) the inequality system (5.2.18) satisfies the strong Slater condition, b) function $H(\cdot,\xi)$ is convex on W for every $\xi \in \Xi$, c) function Q is convex and Lipschitz continuous with modulus C. Then there exists a positive number $\bar{\rho}$ such that the set of optimal solutions of (5.2.19) coincides with the set of th optimal solutions of the following penalized problem:

$$\begin{array}{ll} \min & Q(w) + \rho \max_{\xi\in\Xi}(H(w,\xi))_+, \\ \text{s.t.} & w \in W, \end{array} \tag{5.2.20}$$

where $\rho \geq \bar{\rho}$.

Proof. Under the strong Slater condition, it follows by Lemma 5.3 that there exists a constant $\gamma > 0$ such that

$$d(w,\mathcal{F}) \leq \gamma \max_{\xi\in\Xi}(H(w,\xi))_+.$$

Let ρ be a positive constant such that $\rho > \gamma C$. By [29, Proposition 2.4.3] we can show that the two optimal solution sets of problems (5.2.20) and (5.2.19) coincide. This prove the existence of a positive constant $\bar{\rho} := \gamma C$. The proof is complete. □

We can reformulate problem (5.2.17) by Theorem 5.4 as the following penalized minimization problem:

$$\begin{array}{ll} \min\limits_{w} & -\tau + \rho\varphi(w), \\ \text{s.t.} & \lambda_0 + \sum\limits_{i=1}^{q} b_i\lambda_i \leq 0, \\ \text{s.t.} & \lambda_i \geq 0, \text{ for } i = p+1,\ldots,q. \end{array} \tag{5.2.21}$$

where $w := (x, \tau, \lambda_0, \lambda)$, and

$$\varphi(w) := \max_{\xi \in \Xi}(R(w, \xi))_+.$$

This is a deterministic nonsmooth convex program with simple constraints. Well known method such as cutting plane method, bundle methods and subgradient methods can be exploited to solve it. In what follows, we outline the level function method due to Lemarechal et al. [89] and extended by Xu [146] for solving optimization problem (5.2.21).

5.3 Solution Methods

In this section we will discuss the solution methods for solving optimization problem (5.2.10) and (5.2.21).

The optimization problem (5.2.10) can be solved by MATLAB built-in solver "fmincon" which is suitable for nonlinear constraint optimization problems. However, the problem (5.2.21) is a deterministic nonsmooth convex program with simple constraint which is best solved by algorithms such as the level function method as mentioned earlier.

5.3.1 Level function method

In this section, we focus on solving optimization problem (5.2.21) with the projected level function algorithm as discussed in Chapter 3 (Algorithm 3.2). In what follows, we refer to Algorithm 3.2 as Algorithm 5.1.

let $\vartheta(w, \rho)$ be defined as follows:

$$\vartheta(w, \rho) := -\tau + \rho \max_{\xi \in \Xi}(R(w, \xi))_+,$$

where $w := (x, \tau, \lambda_0, \lambda)$. Let $\zeta_k \in \partial_w \vartheta(w, \rho)$, then

$$\sigma_{w_k}(w) = \zeta_k^T (w - w_k) / \|\zeta_k\|,$$

is a level function of $\vartheta(w, \rho)$ at w_k.

Theorem 5.5. *Let $\{w_k\}$ be generated by Algorithm 5.1. Assume the conditions of Theorem 5.4 are satisfied. Then*

$$\Delta(k) \leq \epsilon, \text{ for } k > M^2 \Upsilon^2 \epsilon^{-2} \lambda^{-2} (1 - \lambda^2)^{-1},$$

where Υ represents the diameter of the solution set W, ϵ and λ are given in the projected level function algorithm (Algorithm 3.2).

For the proof refer back to the proof of Theorem 4.7 in Chapter 4 and [146, Theorem 3.3].

In the projected level function algorithm (Algorithm 5.1) the penalty parameter in $\vartheta(w, \rho)$ is fixed. In some cases, it might be difficult to estimate the penalty parameter. One way to tackle this issue as discussed in Chapter 4 is to start with an estimate of penalty parameter and solve the resulting penalized problem with the level function algorithms. The feasibility of the obtained solution is checked: if it is feasible the optimal solution is obtained, otherwise, the penalty parameter is increased the process is repeated. This kind of procedure in known as *Simple Penalty Function Method* in the literature of optimization, see for instance [138, Algorithm 10.2.3]. We describe the aforementioned procedure formally in the following algorithm for the penalized problem (5.2.21).

Algorithm 5.2 (Simple Penalty Function Method for penalized problem (5.2.21)).

Step 1. Let $\bar{\epsilon}$ be a positive number. Let ρ_0 be an intial estimate of the penalty parameter. Set $t := 0$.

Step 2. For $\bar{\rho} := \rho_t$, apply Algorithm 5.1 to solve problem (5.2.21). Let x_t denote the solution obtained from solving the problem.

Step 3. If $\max_{\xi \in \Xi}(R(w, \xi))_+ \leq \bar{\epsilon}$, stop; otherwise, set $w_{t+1} := w_t$, $\rho_{t+1} := 10\rho_t$ and $t := t + 1$, go to step 2.

Algorithm 5.2 terminates in a finite number of iterations in that the exact penalty parameters for problem (5.2.21) is finite, see Theorems 5.4.

In what follows, we first discuss the financial performance ratio optimization and present some examples. Then we present the robust reformulation of the performance ratio followed by some numerical results.

5.4 Financial Performance Ratios

A major topic of debate in modern asset allocation modeling and managing techniques is how to choose the best performance ratio. Decades ago, Sharpe [132] introduced the well-known *Sharpe Ratio* for managing mutual funds. Subsequently, Zenios [152], Zenios and Kang [151] and Sharpe [133] improved the ratio suggesting to refer the performance to a benchmark. Although this ratio is fully compatible with normally distributed

returns, it loses reliability as soon this property is relaxed (see [44, 88]). Consequently, a number of alternative performance measures such as Sortino ratio [134], MiniMax ratio [149] and Stable ratio [22, 74] have been proposed in the literature.

The well-known Sharpe ratio [132] of a portfolio with return $\mu(x,\xi)$ and a benchmark $Y(\xi)$ can be calculated as:

$$\Phi_{Sharpe}(\mu(x,\xi), Y(\xi) = \frac{\mathbb{E}[\mu(x,\xi) - Y(\xi)]}{\sigma(\mu(x,\xi) - Y(\xi))},$$

where σ denotes the standard deviation. Using the standard deviations as a measure of risk results in equal penalization of both upside and downside deviations to the benchmark. Therefore, this type of ratio is suitable for investment where the main concern is to control the stability of return around the benchmark. On the other hand, if the investment is more concerned with the trade off between large favorable/unfavorable deviations from the benchmark, the Sortino ratio is more appropriate [134].

Definition 5.6. The Sortino performance ratio for a portfolio with return $\mu(x,\xi)$ and benchmark $Y(\xi)$ is defined as:

$$\Phi_{SS}(\mu(x,\xi), Y(\xi) = \frac{\mathbb{E}[\mu(x,\xi) - Y(\xi)]}{\mathbb{E}[((Y(\xi) - \mu(x,\xi))_+)^d]^{1/d}},$$

where $d > 0$, and denote the left orders of the performance ratio.

Sortino ratio substitutes the standard deviation as a measure of risk with left partial moment of order d. Therefore, the only penalizing volatility is the undesirable one below the benchmark. The original Sortino ratio [134] is defined for $d = 2$, it has been extended to $d \geq 1$ in [22, 123]. More recently, the case when $d > 0$ has been considered by Farinelli and Tibeletti [54, 55].

In the remaining of this chapter, we concentrate on the Sortino performance ratio [134] and propose a robust financial performance optimization model based on this ratio. Recall, robust optimization problem (5.2.7), we can equivalently define the robust formulation of the Sortino optimization problem as:

$$\begin{array}{ll} \min\limits_{(x,\tau)\in X\in\mathbb{R}} & -\tau \\ \text{s.t.} & -\max\limits_{P\in\mathcal{P}} \mathbb{E}_P[(\mu(x,\xi) - Y(\xi)) - \tau((Y(\xi) - \mu(x,\xi))_+^d)^{1/d}] \leq 0. \end{array} \tag{5.4.22}$$

In what follows, we present the robust formulation of the (5.4.22) based on the mixture distribution uncertainty as discussed in Section 5.2.3 and first order moment approach discussed in Section 5.2.4.

Recall the discussion in Section 5.2.3 on robust optimization based on mixture distribution, specifically problem (5.2.10). We can reformulate problem (5.4.22) as a robust

optimization based on mixture distribution as follows:

$$\begin{aligned} \min_{(x,\tau)\in X\times\mathbb{R}} \quad & -\tau \\ \text{s.t.} \quad & -\mathbb{E}_{P_l}[(\mu(x,\xi)-Y(\xi))-\tau((Y(\xi)-\mu(x,\xi))_+^d)^{1/d}]\le 0\ l=1,\dots,L. \end{aligned} \tag{5.4.23}$$

Moreover, in the case when ξ has a finite distribution, that is, ξ takes finite number of values $\xi^i,\dots,\xi^N$, the formulation (5.4.23) can be simplified as follows:

$$\begin{aligned} \min_{(x,\tau)\in X\times\mathbb{R}} \quad & -\tau \\ \text{s.t.} \quad & -\sum_{i=1}^{N} p_l^i[(\mu(x,\xi_l^i)-Y(\xi_l^i))-\tau((Y(\xi_l^i)-\mu(x,\xi_l^i))_+^d)^{1/d}]\le 0\ l=1,\dots,L. \end{aligned} \tag{5.4.24}$$

where p_l^i corresponds to the probability measure of P_l in scenario i.

Consequently, The sample average approximation of the optimization problem (5.4.23) can also be formulated as follows:

$$\begin{aligned} \min_{(x,\tau)\in X\times\mathbb{R}} \quad & -\tau \\ \text{s.t.} \quad & -\sum_{i=1}^{N} p_l^i[(\mu(x,\xi_l^i)-Y(\xi_l^i))-\tau((Y(\xi_l^i)-\mu(x,\xi_l^i))_+^d)^{1/d}]\le 0\ l=1,\dots,L. \end{aligned} \tag{5.4.25}$$

As discussed earlier, the ξ_l^i can be generated by computer simulation under probability distribution P_l.

Let us now present the reformulation of problem (5.4.22) based on first order moment approach as discussed in Section 5.2.4. The problem (5.4.22) can be reformulated based on first order moment approach as follows:

$$\begin{aligned} \min_{x\in X,\tau\in\mathbb{R},\lambda\in\mathbb{R}\times\mathbb{R}^p\times\mathbb{R}_+^{q-p}} \quad & -\tau \\ \text{s.t.} \quad & \lambda_0+\sum_{i=1}^{q} b_i\lambda_i\le 0, \\ & \lambda_i\ge 0,\ \text{for } i=p+1,\dots,q, \\ & \lambda_0+\sum_{i=1}^{q}\lambda_i\psi_i(\xi)\ge G(x,\xi,d,\tau),\ \xi\in\Xi, \end{aligned} \tag{5.4.26}$$

where

$$G(x,\xi,d,\tau) := (\mu(x,\xi_l^i)-Y(\xi_l^i))-\tau((Y(\xi_l^i)-\mu(x,\xi_l^i))_+^d)^{1/d}.$$

Subsequently, the problem (5.4.26) can be reformulated based on the penalization scheme discussed in Section 5.2.4.1 as:

$$\begin{aligned} \min_{w} \quad & \vartheta(w,\rho) := -\tau + \rho \max_{\xi\in\Xi}(R(w,\xi))_+ \\ \text{s.t.} \quad & \lambda_0 + \sum_{i=1}^{q} b_i\lambda_i \leq 0, \\ & \lambda_i \geq 0, \text{ for } i = p+1,\ldots,q, \\ & w \in W, \end{aligned} \tag{5.4.27}$$

where

$$R(w,\xi) := G(x,\xi,d,\tau) - \lambda_0 - \lambda^T\psi(\xi).$$

As mentioned earlier, this is a deterministic nonsmooth convex program with simple constraints which can be solved by well known method such as level function method.

5.5 Numerical Tests

We have carried out a portfolio performance ratio optimization and a fund of funds investment problem on the proposed models and algorithm by using MATLAB 7.10 and IBM ILOG CPLEX 12.4 installed on a HP Notebook PC with Windows 7 operating system, and Intel Core i7 processor. We have integrated the Algorithm 5.2 in Algorithm 5.1 and set the initial penalty parameter equal to 500. Further, we set the $\lambda = 0.5$ and $\epsilon = 0.0001$.

Recall Sortino robust optimization problem (5.4.23), the nominal counter part of this problem can be formulated as follows:

$$\begin{aligned} \min_{(x,\tau)\in X\in\mathbb{R}} \quad & -\tau \\ \text{s.t.} \quad & -\mathbb{E}[\mu(x,\xi) - Y(\xi) - \tau((Y(\xi) - \mu(x,\xi))_+^d)^{1/d}] \leq 0, \end{aligned} \tag{5.5.28}$$

where for the purpose of our numerical test, we set $d = 2$ for both the mixture distribution problem as well as the nominal problem.

Particularly, we consider a portfolio optimization problem with real world test data to investigate the efficiency of the proposed robust optimization models (5.4.23) and (5.4.27), and compare the results to portfolio strategy generated by the nominal problem (5.5.28). Moreover, we consider a fund of funds problem, which is an investment strategy of holding a portfolio of other investment funds rather than investing directly in shares, bonds or other securities, and compare the results to the strategy generated by the nominal problem (5.5.28).

5.5.1 Portfolio performance ratio

Suppose that we have a fixed capital to be invested in n assets. Let R_i, $i = 1, \ldots, n$, denotes the return of asset i. In practice, the return is often uncertain and we use a random variable ξ to describe the uncertainty. Specifically, we write R_i as $R_i(\xi)$ and in doing so we are assuming that all n assets have identical random factors.

To simplify the discussion, we normalize the capital to 1 and use x_i, $i = 1, \ldots, n$, to denote the fraction of capital to be invested in asset i. The portfolio return can then be formulated as:

$$\mu(x, \xi) := R_1(\xi)x_1 + R_2(\xi)x_2 + \cdots + R_n(\xi)x_n. \tag{5.5.29}$$

Example 5.1. *We consider m history of rate of return, for a group of n assets. Our aim is to find a robust optimal investment strategy for a fixed capital in the n assets which minimized the ratio of the risk and expected excess return. Particularly we consider the following model:*

$$\begin{array}{ll} \min\limits_{(x,\tau)\in X\in\mathbb{R}} & -\tau \\ \text{s.t.} & -\max\limits_{P\in\mathcal{P}} \mathbb{E}_P[\mu(x,\xi) - Y(\xi) - \tau((Y(\xi) - \mu(x,\xi))_+^d)^{1/d}] \leq 0, \end{array} \tag{5.5.30}$$

we apply the the reformulations based on the mixture distribution uncertainty and first order moment problem discussed in Section 5.4, respectively.

We collected 2722 daily stock returns of 34 FTSE 100 assets from Jan 2005 to June 2012. We use the first 1700 observations to generate portfolio strategies and the remaining observations in constructing an out-of-sample test. We solve the mixture distribution problem with the built-in MATLAB function "fmincon", while the first order moment problem is solved through the level function algorithm (Algorithm 5.1).

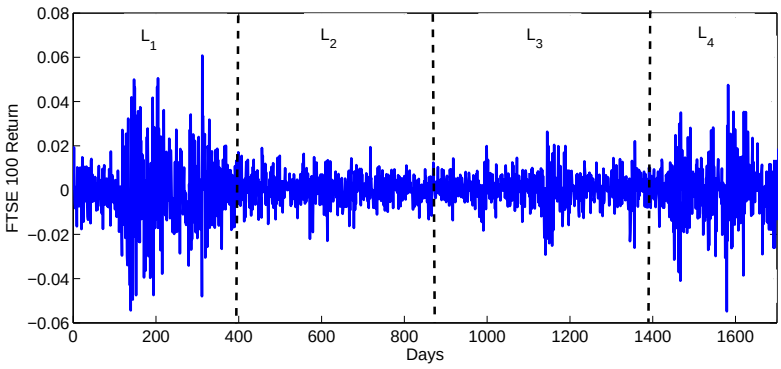

FIGURE 5.5.1: Time series analysis of the FTSE 100 Index used for setting the number of L and the corresponding probabilities.

Periods	Mean (10^{-4})	Variance (10^{-4})
L_1	-4.6674	2.6880
L_2	4.3397	0.3923
L_3	5.5324	0.5199
L_4	-5.3644	1.8330

TABLE 5.5.1: Expected value and variance of returns of the FTSE 100 index in different time periods (L).

Figure 5.5.1 shows the return of FTSE 100 index over the first 1700 observations. It can be seen that the data set can be divided into four subsection. Furthermore, the expected value and variances of returns of the FTSE 100 index corresponding to these different time periods are listed in Table 5.5.1. In this example, according to our observation, we assume that the samples are generated by the mixture distribution of four likelihood distributions. Specifically, we assume that samples within each time period are generated by normal distributions with means and variances shown in Table 5.5.1.

In the computation of the nominal portfolio optimization problem, we set $L = 1$ and $N_l = 1700$, i.e., all the samples are used in the model by assuming that they are generated by one nominal probability distribution. In the computation of the mixture distribution model, we set $L = 4$ and $N_1 = 400$, $N_2 = 490$, $N_3 = 510$, and $N_4 = 300$, where we assume the samples within each time period are generated by the corresponding likelihood distribution. In the computation of the first order moment problem, we let the $q = 1$ and define $\psi_1(\xi)$ as the return function of the FTSE 100 index. Moreover, in these tests we set the upper bound and lower bound on weights equal to 0.6 and 0, respectively.

The results for the nominal problem, mixture distribution problem and first order moment problem are presented in Table 5.5.2.

Model	Time(min)	Iter	No.Assets	Return	Risk
Nominal	0.0156	6	3	-0.0076	0.7542
Mixture distribution	0.0316	15	8	0.0016	0.2798
First order moment	0.9444	10	14	0.0033	0.1765

TABLE 5.5.2: The results of the three models for 34 stocks of FTSE 100 index, where the benchmark is considered to be equal to the return of FTSE 100 index.

As it can be seen, the two robust portfolio models outperform the nominal portfolio in sense of having both higher return and lower risk, while the first moment model having the best performance.

Figures 5.5.2 and 5.5.3 show the backtest and out-of-sample test of the generated portfolios for the three models. As it can be seen, the return of the nominal portfolio performs poorly compared to the return of the portfolios generated by the mixture distribution

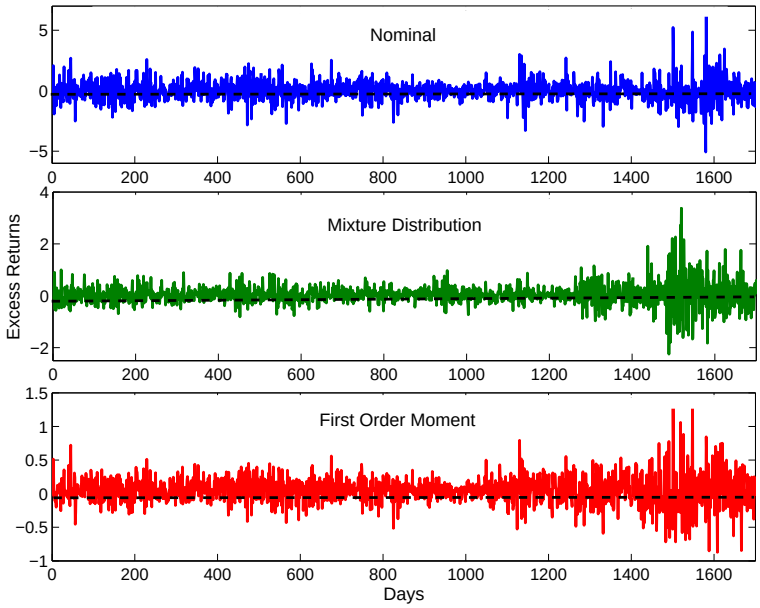

FIGURE 5.5.2: Backtest of the of the excess return of the three generated portfolios.

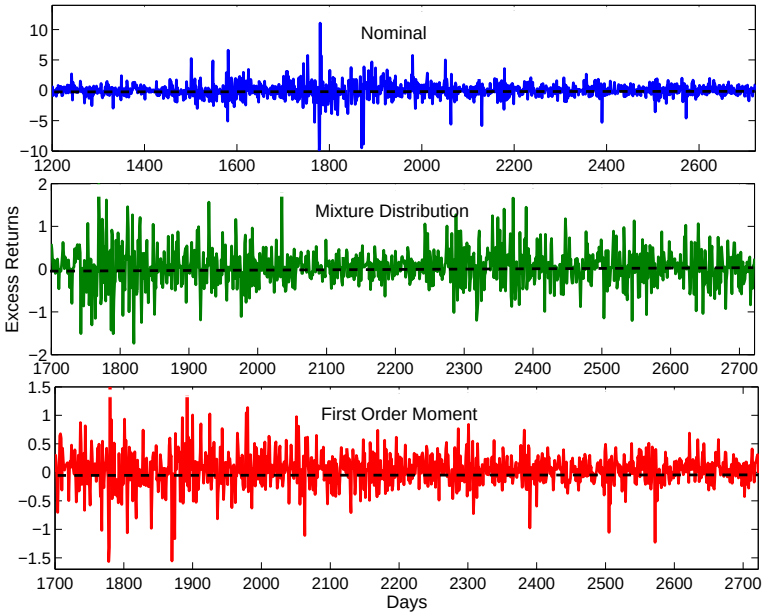

FIGURE 5.5.3: Out of sample test of the excess return of three generated portfolios.

and first order moment reformulations. Moreover, it can be seen that the portfolio constructed by the mixture distribution formulation is less conservative compared to the first order moment's problem both in-sample and out-of-sample.

Moreover, to see the performance of the generated strategies out-of-sample we present graph of cumulative return of the of portfolio return generated by the nominal, mixture distribution and the first order models in Figure 5.5.4. It can be seen that the return

generated by the two robust formulations is much higher compared to that of the nominal portfolio.

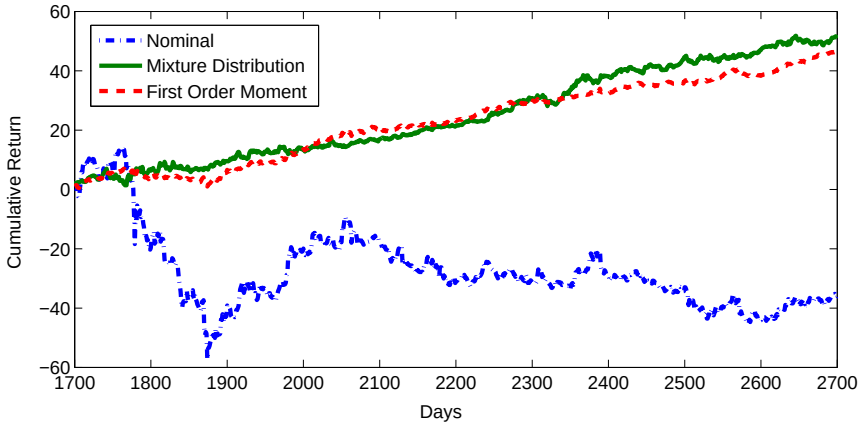

FIGURE 5.5.4: Out-of-sample cumulative return for the generated portfolio strategy based on the nominal, mixture distribution and first order moment models.

Figure 5.5.5 and 5.5.6 presents the risk profile of the generated portfolios both in-sample and out-of-sample. It can be seen that the portfolio generated based on the first order moment formulation has lower risk exposure compared to the other two models.

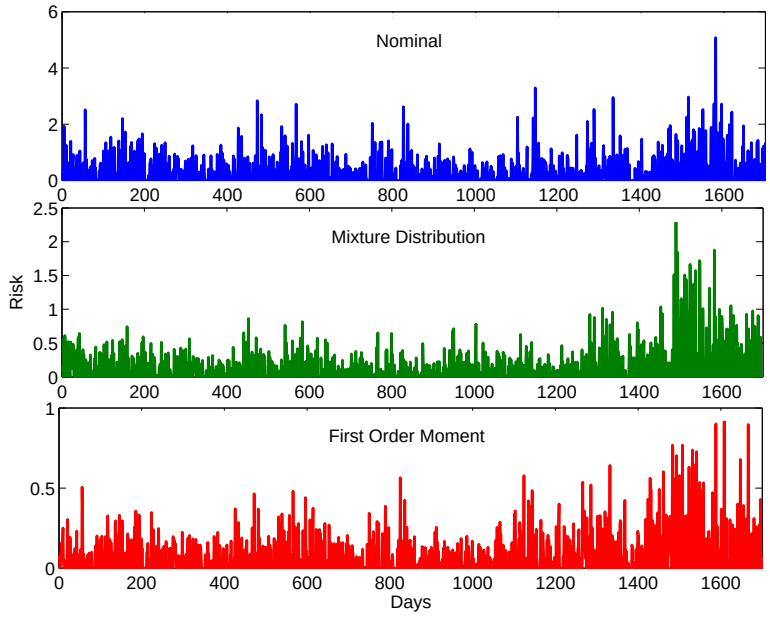

FIGURE 5.5.5: Comparison of the risk measure in-sample.

It should be noted that the 2008 financial crisis occurred in days 1400 to 2000 and European sovereign-debt crisis occurred around days 2370 to present. As it can be seen, the risk of the first order moment is the least during both of these periods compared to the nominal strategy and the mixture distribution model. Additionally, the negative excess return of the robust portfolios is much lower compared to the nominal portfolio in

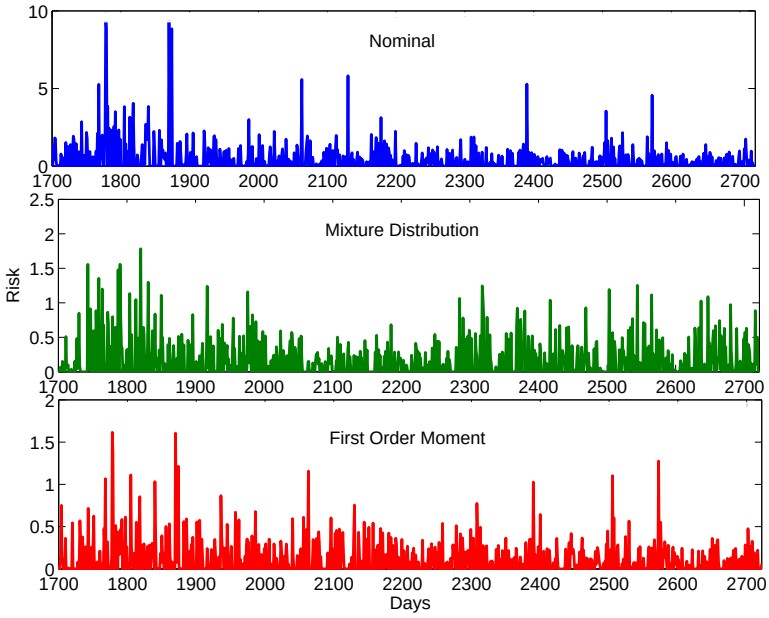

FIGURE 5.5.6: Comparison of the risk measure out-of-sample.

2008 financial crisis period, while the excess return of these portfolios are mostly positive in period coinciding with the European sovereign-debt crisis.

It should be mentioned that although the excess return of the robust portfolios are lower than of that of nominal portfolio but in many financial and non-financial institutions such as pension funds and national insurance systems having consistently positive low risk return is more attractive than having high but very risky returns.

5.5.2 Fund of funds investment

A "fund of funds" (FOF) is an investment strategy of holding a portfolio of other investment funds rather than investing directly in shares, bonds or other securities. In this section we consider four funds representing FTSE 100 Index (Fund 1), S&P 500 (Fund 2), Nasdaq100 (Fund 3), and Hang Seng (Fund 4).

We have collected 2722 historical rate of return and let the benchmark be average return of the four funds. As before, we have used the first 1700 observations to generate the portfolio strategy and remaining observations are used to set up an out-of-sample test to investigate the performance of the portfolio.

Figure 5.5.7 shows the return of the four funds over the first 1700 observations. It can be seen that the data set can be divided into three subsection. Furthermore, the expected value and variances of returns of the four funds corresponding to different time periods are listed in Table 5.5.3. In this example, according to our observation, we assume that the samples are generated by the mixture distribution of three likelihood distributions.

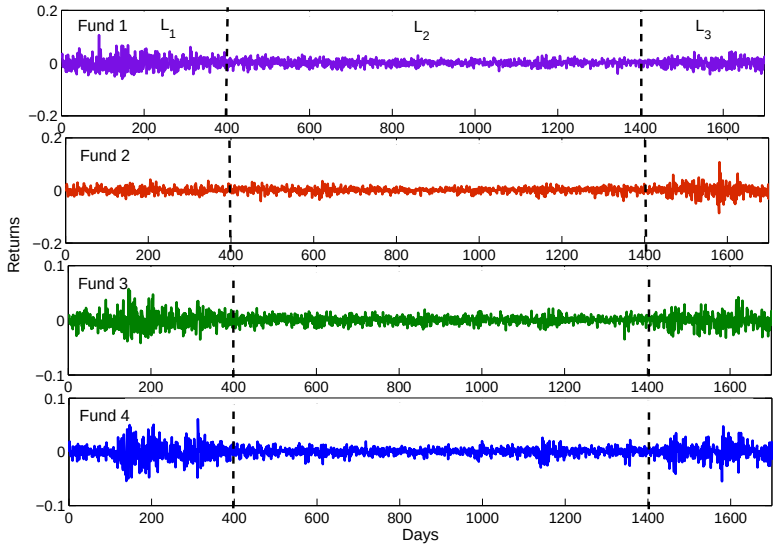

FIGURE 5.5.7: Time series analysis of the FTSE 100 Index used for setting the number of L and the corresponding probabilities.

Period	Mean (10^{-3})				Variance (10^{-3})			
	Fund 1	Fund 2	Fund 3	Fund 4	Fund 1	Fund 2	Fund 3	Fund 4
L_1	-0.4667	-0.2322	-0.2055	-0.2288	0.2688	0.2251	0.5809	0.1296
L_2	0.4870	0.4242	0.4211	0.7660	0.0456	0.0443	0.1073	0.0860
L_3	-0.5280	-0.4772	0.0752	0.2589	0.1827	0.1463	0.2103	0.4437

TABLE 5.5.3: Expected value and variance of returns of the four funds in different time periods (L).

Specifically, we assume that samples within each time period are generated by normal distributions with means and variances shown in Table 5.5.3.

In the computation of the nominal portfolio optimization problem, we set $L = 1$ and $N_l = 1700$, i.e., all the samples are used in the model by assuming that they are generated by one nominal probability distribution. In the computation of the mixture distribution model, we set $L = 3$ and $N_1 = 400$, $N_2 = 1000$, and $N_3 = 300$, where we assume the samples within each time period are generated by the corresponding likelihood distribution. In the computation of the first order moment problem, we let the $q = 1$ and define $\psi_1(\xi)$ as the return function of the equally weighted portfolio of the four funds. In what follows, we set the upper bound and lower bound on the capital invested equal to 0.6 and 0, respectively.

Table 5.5.4 presents the results for this example. As can be seen, the first order moment portfolio has the highest expected return with the least risk. Furthermore, the expected return and the associated risk of the mixture distribution are slightly lower and higher than of that of the first order moment problem, respectively. Moreover, it can be seen that the portfolio strategies generated by the two robust models are more diversified compared to the nominal strategy.

Model	Time(min)	Iter	Funds	Return	Risk
Nominal	0.0176	8	0.0000, 0.6000, 0.4000, 0.0000	0.0206	0.0048
Mixture distribution	0.0182	8	0.2387, 0.6000, 0.1613, 0.0000	0.0157	0.0033
First order moment	0.0863	7	0.1493, 0.4793, 0.3161, 0.0554	0.0212	0.0024

TABLE 5.5.4: The results of the three models for four funds, where the benchmark is considered to be an equally weighted portfolio of these funds.

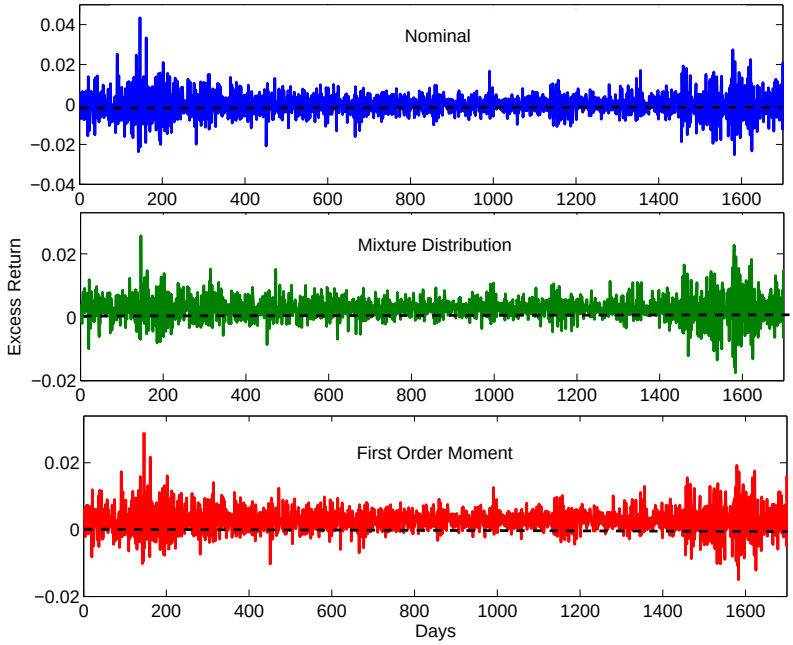

FIGURE 5.5.8: Comparison of the excess return of portfolio in-sample.

Figure 5.5.8 and 5.5.9 demonstrate the performance of the generated portfolio in-sample and out-of-sample, respectively. As can be seen, the two robust optimizations have positive excess returns both in-sample and out-of-sample whereas the excess return of the nominal portfolio varies a lot around the zero line. Furthermore, it can be seen that the excess return of the mixture distribution portfolio and that of first order moment portfolio, are almost always positive in-sample. Moreover, it can be seen that the excess return of the robust formulations outperform that of the nominal portfolio in the sense of having lower negative excess return and equivalent positive excess return.

Moreover, to see the performance of the generated strategies out-of-sample we present graph of cumulative return of the of portfolio return generated by the nominal, mixture distribution and the first order models in Figure 5.5.10. It can be seen that the return generated by the two robust formulations is much higher compared to that of the nominal portfolio.

Figure 5.5.11 and 5.5.12 illustrate the risk associated with each portfolio in-sample and out-of-sample, respectively. As anticipated, the risk of the first order moment model

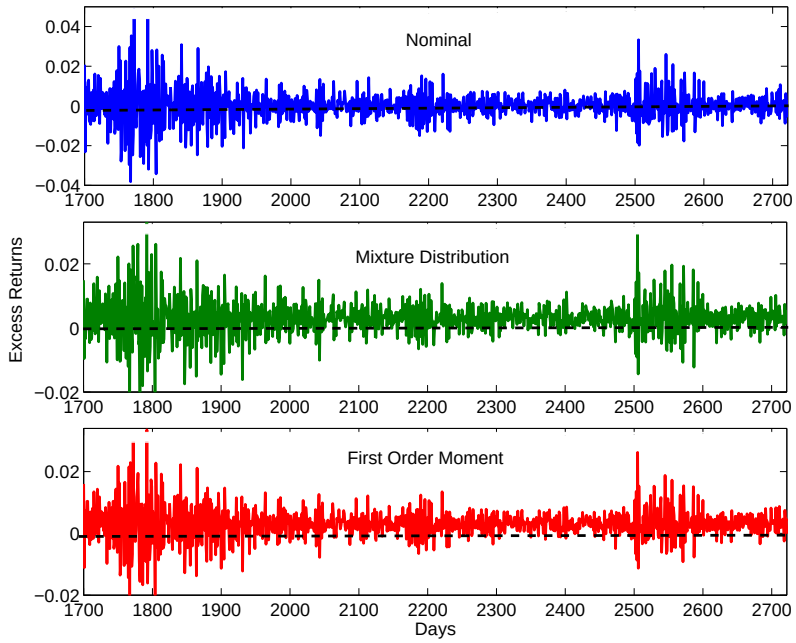

FIGURE 5.5.9: Comparison of the excess return of portfolio out-of-sample.

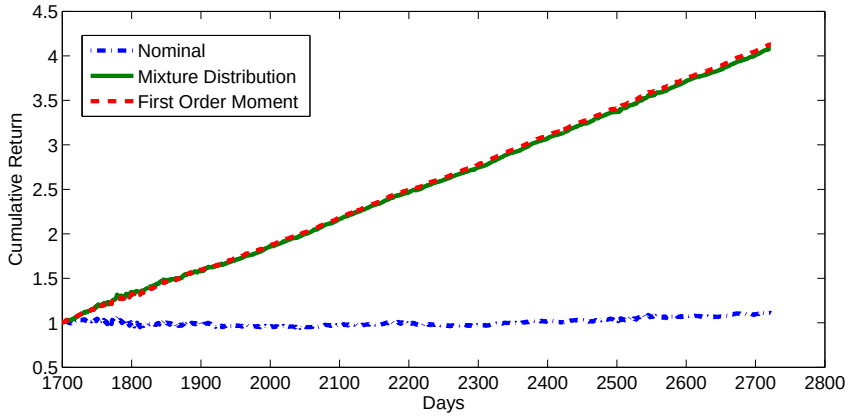

FIGURE 5.5.10: Out-of-sample cumulative return for the generated portfolio strategy based on the nominal, mixture distribution and first order moment models.

is least compared to the other two models both in-sample and out-of-sample, followed closely by the mixture distribution model.

As discussed in the previous example, the 2008 financial crisis occurred in days 1400 to 2000 and European sovereign-debt crisis occurred around days 2370 to present. As it can be seen, the risk of the first order moment is the least during both of these periods compared to the nominal strategy and the mixture distribution model. Additionally, the negative excess return of the robust portfolios is much lower compared to the nominal portfolio in 2008 financial crisis period, while the excess return of these portfolios are mostly positive in period coinciding with the European sovereign-debt crisis.

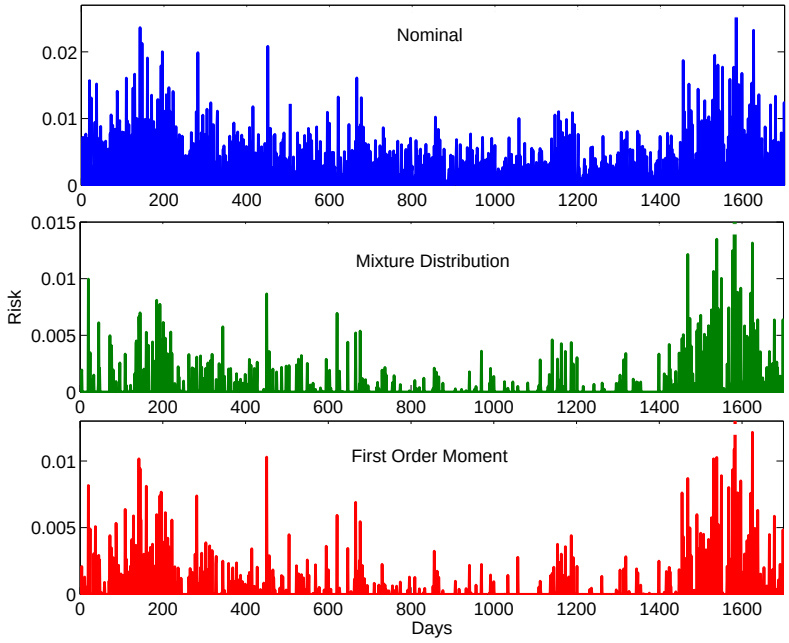

FIGURE 5.5.11: Comparison of the risk measure in-sample.

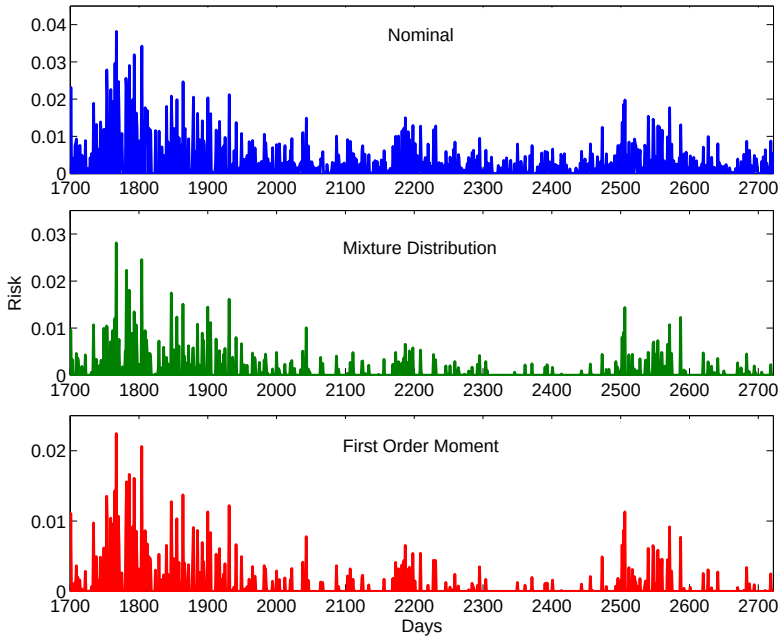

FIGURE 5.5.12: Comparison of the risk measure out-of-sample.

In order to investigate the efficiency of the models and the solution methods with respect to CPU time, we have solved the three model with the proposed algorithms first for an increasing number of assets with a fixed sample size and second for an increasing number of samples with a fixed number of assets. As it can be seen in Figures 5.5.13 and 5.5.14 the mixture distribution model and the nominal model solved by the built-in MATLAB function "fmincon" are more efficient compared to the first order moment model solved

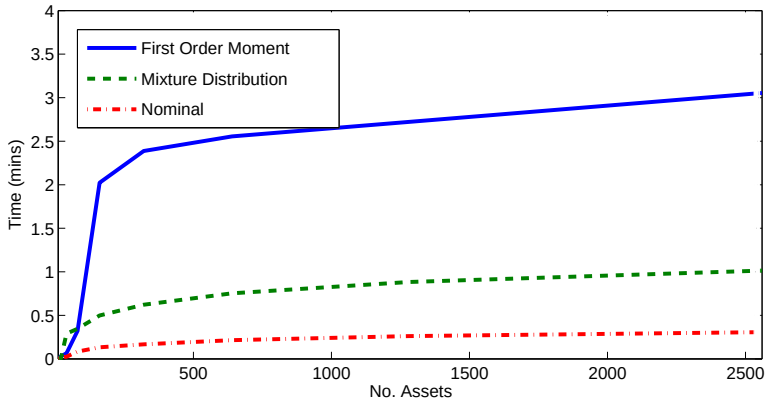

FIGURE 5.5.13: Comparison of the CPU time versus the number of assets for a sample of 100.

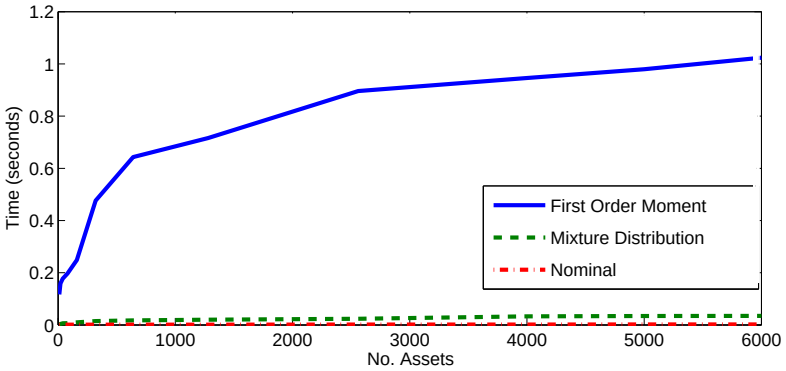

FIGURE 5.5.14: Comparison of CPU time versus the sample size when the asset number is fixed at 100.

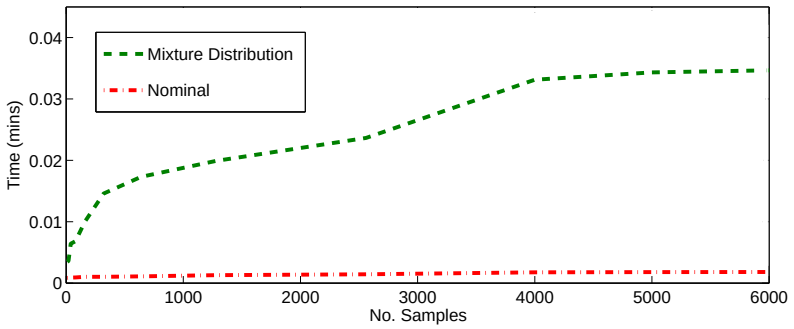

FIGURE 5.5.15: Zoom in of the comparison of CPU time versus the sample size for the mixture distribution and the nominal models.

by the level function algorithm. However, the first order moment is still efficient with a CPU time of less than four minutes in both cases.

5.6 Conclusion

In this chapter we focused on a robust formulation for a performance ratio optimization based on one sided variability measure. We propose a robust optimization problem for a reward-risk ratio optimization based on mixture distribution and first order moment approach. For the case of mixture distribution, a sample average approximation formulation was presented. Moreover, an exact penalization scheme was proposed for the first order moment approach to handle the semi-infinite constraints in the dual problem. The proposed methods were applied to the Sortino performance ratio and the robust formulations based on both mixture distribution and the first moment approach are derived.

We investigated the numerical efficiency and accuracy of the proposed methods by presenting a portfolio optimization problem and a fund of funds problem based on real world data. We further, set up backtest and out-of-sample test to inspect the performance of the generated portfolios and compare them to the benchmark portfolio.

Based on the numerical tests results we can conclude that both of the robust formulations result in more conservative but better portfolios compared to the nominal strategy in a sense of both expected return and the associated risk. Moreover, the first order moment problem results in the most conservative solution, where the excess return is lower but mostly positive compared to that of the mixture distribution model and the nominal model. Furthermore, the investigation of the efficiency of the models and methods with respect to the CPU time revealed that the mixture distribution model and the nominal model solved with the MATLAB built-in function “fmincon” are more efficient compared to the first order moment model solved by the level function algorithm. However, the largest first order moment problem with respect to both sample size and the number of assets were solved in less than four minutes.

Chapter 6

Concluding Remarks

6.1 Research Outcomes

There are three basic approaches to the problem of portfolio selection under uncertainty; stochastic dominance, expected utility maximization and the reward-risk analysis. In this thesis we focused on stochastic optimization problems with stochastic dominance constraints as well as reward-risk ratio optimization which is related to reward-risk analysis.

The stochastic dominance notion has been employed in many areas including, medicine and health (Madden [96]), poverty and inequality studies (Jeffrey and Eidman [76], Anderson [1]), agriculture (Davidson, Duclos [33]) and financial decision making (see Annaert et al [3], Levy [91]. Eeckhoudt [43] and references therein). We focused on application of stochastic dominance in portfolio optimization. The advantages of using stochastic dominance model was discussed in Chapter 2. Application of second order stochastic dominance as a criteria of choice proves to be difficult. Generally, in solving stochastic programming problems with second order stochastic dominance, one needs to deal with three main issues; a) the expectation of random functions in the objective and the constraint, b) infinite number of constraints, c) the non-smoothness arising from the plus function in the constraints.

The main contribution of this thesis can be summarized as follows:

- To overcome the difficulties associated with scalar second order stochastic dominance a recently developed exact penalization scheme for such problems is exploited. Moreover, a penalization scheme is developed for the multivariate second order stochastic dominance by exploiting Clark's exact penalty function [29, Proposition 2.4.3] and Robinson's error bound [113]. The multivariate stochastic

dominance constraints are reformulated and it is shown that the reformulated problem satisfies the Slater Constraint Qualification under some moderate conditions. Furthermore, an exact penalization scheme based on L_∞-norm is derived.

- The resulting penalized problems are non-smooth convex optimization problems which can efficiently be solved with numerical methods such as stochastic approximation methods and level function methods. The convergence analysis regarding the solution methods for each specific problem is presented. Moreover, a modified cutting plane method is proposed for the multivariate stochastic dominance model. This cutting plane method differs from those in the literature [120] in that it applies to the maximum of the constraint functions rather than each constraint function. Moreover, this modified cutting plane method uses the cutting plane representation proposed in [81], so it differ from the methods proposed in [70, 73]. The idea of applying the cutting-plane method to the maximum of the constraint functions is similar to the idea in algorithm proposed by Fábián et al. [52]. However, their method is applied to linear models while this modified cutting plane method is also applicable to nonlinear case. Furthermore, the proposed numerical methods provides an alternative approach to the existing cutting surface method for multivariate stochastic dominance introduced by Homem-de-Mello and Mehrota [70] and the linearized method proposed by Armbruster and Luedtke [4].

- Moreover, we focused on robust reward-risk ratio optimization to address the issue of data uncertainties from the perspective of computational tractability. We considered robust formulations based on mixture distribution approach and first order moment approach. The problems arising from the mixture distribution approach can be solved with nonlinear solvers such as MATLAB built-in solver "fmincon", while the first order moment approach results in a semi-infinite programming problem. To overcome difficulties associated with this type of problem, we proposed an exact penalization method to deal with the infinite number of constraints in optimization problem. This resulted in reformulation of the optimization problem as a deterministic non-smooth convex program with simple constraints which can be solved with well-known methods such as level function methods where the convergence analysis were presented.

The numerical methods discussed in this thesis have some advantages and limitations. These methods can efficiently solve non-smooth, nonlinear, convex optimization problems within reasonable computation times. The stochastic approximation method requires calculation of only one approximate subgradient per iteration and can be applied to the case when the underlying functions are highly nonlinear and/or non-smooth, and the distribution of the random variable may be unknown. The level function methods require calculation of a subgradient instead of an approximate subgradient of the objective function at each iterate and therefore it applies to the problem with known distribution

of the random variable or the sample average approximated problems. A clear advantage of the method is that we can estimate the number of iterations needed for a specified precision. The modified cutting plane method discussed differs from those in literature [120] in that it applies to the maximum of the constraint functions rather than each constraint function. This saves considerable computation time because at each iteration, our cutting-plane method requires the addiction of a couple of linear constraints only. The approach also differs from that in [70, 73] because our modified cutting-plane method uses the cutting-plane representation proposed in [81]. The idea of applying the cutting-plane method to the maximum of the constraint functions is similar to the idea in algorithm proposed by Fábián et al. [52]. Note that Fábián's algorithm is applied to linear models while our modified cutting plane method is applicable to nonlinear case. Furthermore, the level function methods and the modified cutting plane method are not sensitive to an increase in the size of the problem, as an increase in either the sample size or the number of instruments does not have a significant impact on the performance of these methods. It should be noted that, if the problem considered is linear then, some off the shelf linear programming softwares would outperform these methods in sense of computation time.

6.2 Future Research

Each of the chapters in this thesis contributed to the existing literature. So far we have concentrated on second order stochastic dominance. This research can be extended by considering stochastic optimization problems with first order stochastic dominance, which is closely related to the Value at Risk (VaR) measure. VaR is a widely used risk measure of the risk of loss on a specific portfolio of financial assets. It is commonly used in risk management, risk measurement, financial reporting and computing regulatory capital (Basel II, and III).

Another interesting possibility for continuing this research could lie in investigating the properties of multivariate stochastically weighted dominance [72], in which the vector of weights ν is treated as a random vector. Such an approach is much less restrictive than the deterministic weighted approach considered in this research.

Moreover, we have only concentrated on problems where the random variable is discrete. We could further extend this research and consider problems with continuous random variables. This type of problem has extensively been discussed in theoretical literature. However, there has not been extensive research done on numerical analysis and performance.

Furthermore, as it was discussed in this thesis, there are two basic approaches to the problem of portfolio selection under uncertainty. One of them is the stochastic dominance approach, and the other is the reward-risk analysis which is also related to the

reward-risk ratio optimization. In this thesis we considered robust optimization of this type of problem based on mixture distribution and first order moment approach. However, this research can be extended by considering distributionally robust optimization under moment uncertainty where uncertainty is described in both the distribution form (discrete, Gaussian, exponential, etc.) and moments (mean and covariance matrix).

Appendix A

Appendix

A.1 Figures for Chapter 4

The backtest and out-of-sample comparison of the generated portfolio with the multivariate SSD model to the indices. We present the figures related to the backtests followed by the out-of-sample tests.

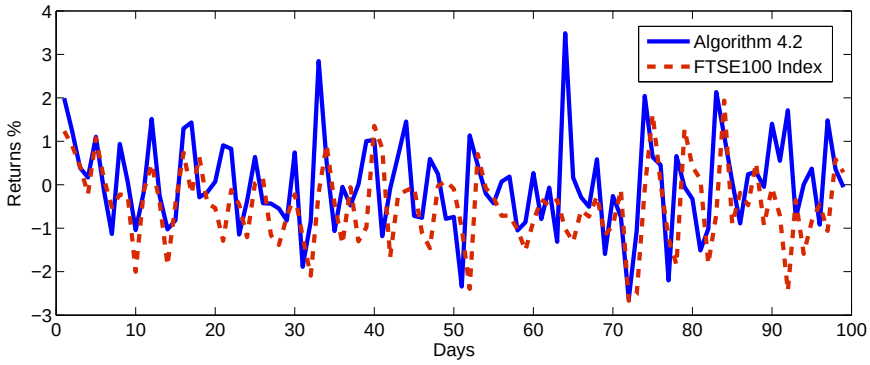

FIGURE A.1.1: Backtest comparison of the Multivariate SSD model and the FTSE 100 Index.

As it can be seen the return of the portfolio strategy generated by the proposed model and algorithms out perform the return of each individual index both in-sample and out-of-sample.

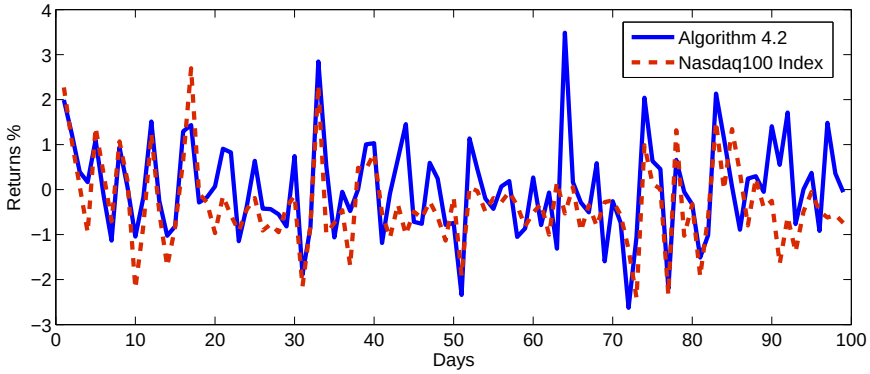

FIGURE A.1.2: Backtest comparison of the Multivariate SSD model and the Nasdaq 100 Index.

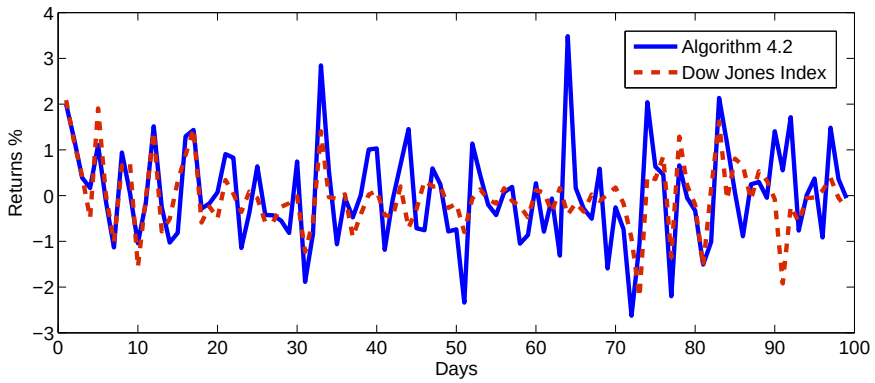

FIGURE A.1.3: Backtest comparison of the Multivariate SSD model and the Daw Jones Index.

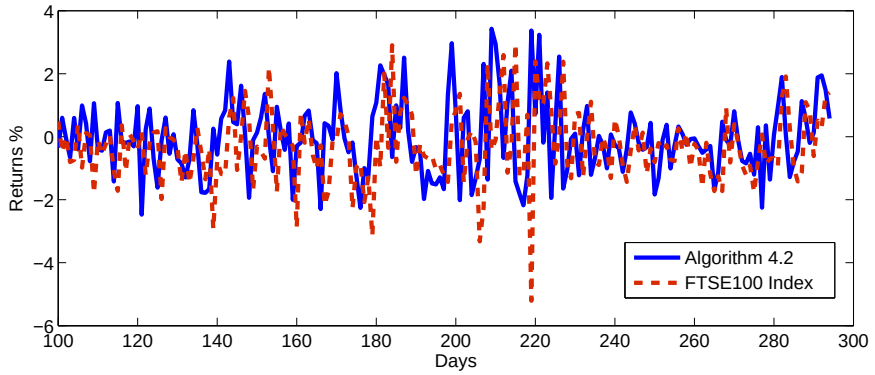

FIGURE A.1.4: Out-of-sample comparison of the Multivariate SSD model and the FTSE 100 Index.

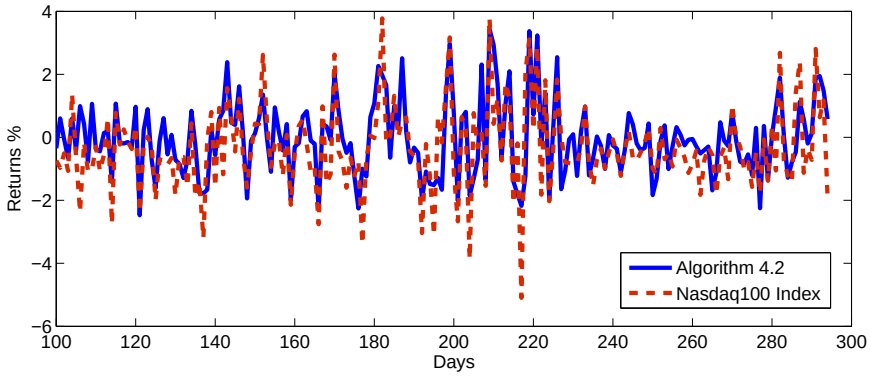

FIGURE A.1.5: Out-of-sample comparison of the Multivariate SSD model and the Nasdaq 100 Index.

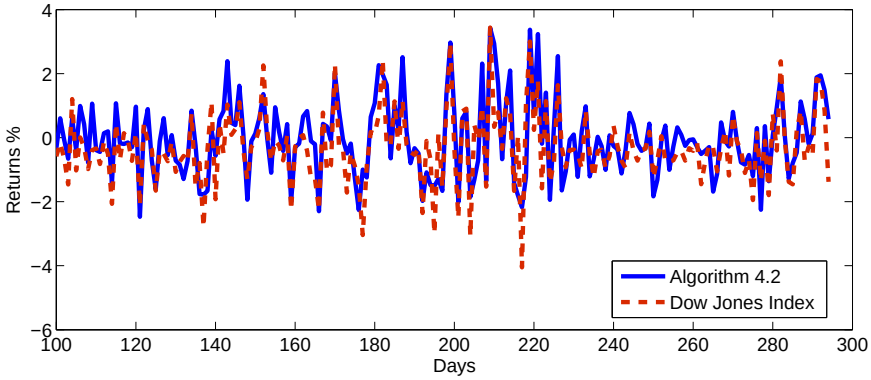

FIGURE A.1.6: Out-of-sample comparison of the Multivariate SSD model and the Daw Jones Index.

Bibliography

[1] G. Anderson. Nonparametric tests of stochastic dominance in income distributions. *Economic Review*, 64:1183–1193, 1996.

[2] F. Andersson, H. Mausser, D. Rosen, and S. Uryasev. Credit risk optimization with conditional value at risk criterion. *Mathematical Programming*, 89:273–291, 2001.

[3] J. Annaert, S. Osselaer, and B. Verstraete. Performance evaluation of portfolio insurance strategies using stochastic dominance criteria. *Journal of Baking and Finance*, 33:272–280, 2009.

[4] B. Armbruster and J. Luedtke. Models and formulations for multivariate dominance constrained stochastic programs. *Northwestern University*, 2011.

[5] P. Artzner, F. Delbaen, J. M. Eber, and D. Heath. Thinking coherently. *Risk*, 10:68–91, 1997.

[6] P. Artzner, F. Delbaen, J. M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–227, 1999.

[7] B. M. Baker and J. Sheasby. Accelerating the convergence of subgradient optimization. *European Journal of Operational Research*, 117:136–144, 1998.

[8] V. Bawa. Optimal rules for ordering uncertain prospects. *Journal of Financial Economics*, 2:95–121, 1975.

[9] E. Beale. On minimizing a convex function subject to linear inequalities. *Journal of the Royal Statistical Society B*, 17:173–184, 1955.

[10] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23:769–805, 1998.

[11] A. Ben-Tal and A. Nemirovski. Robust solutions to uncertain programs. *Mathematics of Operations Research*, 25:1–13, 1999.

[12] A. Ben-Tal and A. Nemirovski. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming*, 88:411–424, 2000.

[13] D. Bernoulli. Exposition of a new theory on the measurement of risk. *Econometrica*, 22:23–36, 1954. Translation of a paper originally published in Latin in St. Petersburg in 1738.

[14] M. Bertocchi, M. T. Vespucci, and H. Xu. Robust approaches for two stage stochastic programs with applications in capacity expansion planning in energy industry.

[15] Bertsimas and D. D. Brown. Robust linear optimization and coherent risk measures. *Technical Report, Massachusetts Institute of Technology*, 2005.

[16] D. Bertsimas, X. V. Doan, K. Natarajan, and C. P. Teo. Models for minimax stochastic linear optimization ptoblems with risk aversion. *Mathematics of Operations Research*, 35:580–602, 2010.

[17] D. Bertsimas, D. Pachamanova, and M. Sim. Robust linear optimization under general norms. *Operations Research Letters*, 32:510–516, 2004.

[18] D. Bertsimas and I. Popescu. On the relation between option and stock prices: a convex optimization. *Operations Research*, 50:358–374, 2002.

[19] D. Bertsimas and I. Popescu. Optimal inequalities in probability theory: a convex optimization approach. *SIAM Journal of Optimization*, 15:780–804, 2005.

[20] D. Bertsimas and M. Sim. Robust discrete optimization and network flows. *Mathematical Programming*, 98:49–71, 2003.

[21] D. Bertsimas and M. Sim. The price of robustness. *Operations Research*, 52:35–53, 2004.

[22] Ortobelli S. Rachev S. Biglova, A. and S. Stoyanov. Different approaches to risk estimation in portfolio theory. *The Journal of Portfolio Management*, 31:103–112, 2004.

[23] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[24] M Branda. Reformulation of general chance constrained problems using the penalty functions. *Stochastic Programming E-print Series*, 2, 2010.

[25] W. Breen and J. Savage. Portfolio distributions and tests of security selection models. *The Journal of Finance*, 23:805–819, 1968.

[26] J. R. Brige and F. Louveaux. *Introduction to Stochastic Programming*. Springer-Verlag, 1997.

[27] A. Charnes and W. W. Cooper. Chance-constrained programming. *Management Science*, 5:73–79, 1959.

[28] P. Cheridito and M. Delbaen, F. Kupper. Coherent and convex risk measures for bounded cádlág processes. *Stochastic Processes Their Application*, 112:1–22, 2004.

[29] F. H. Clarke. *Optimization and non-smooth analysis*. Wiley, New York, 1983.

[30] R. E. Curto and L. A. Failow. *Solution of the truncated complex moment problem for flat data*. Memoirs of the American Mathematical Society, 1996.

[31] G. B. Dantzing. *Linear programming and extensions*. Princeton University Press, Princeton, NJ, 1955.

[32] G. B. Dantzing. Linear programming under uncertainty. *Management Science*, 1:197–206, 1955.

[33] R. Davidson and J. Y. Duclos. Statistical inference for stochastic dominance and for the measurement of poverty and inequality. *Econometrics*, 68:1435–1465, 2000.

[34] F. Delbaen. *Coherent risk measures on general probability spaces. Essays in Honour of Dieter Sondermann*. Springer-Verlag, Berlin, Germnay, 2002.

[35] D. Dentcheva and G Martinez. Two-stage stochastic optimization problems with stochastic ordering constraints on the recourse. *European Journal of Operational Research*, 219:1–8, 2012.

[36] D. Dentcheva and A. Ruszczyński. Optimization with stochastic dominance constraints. *SIAM Journal of Optimization*, 13:548–566, 2003.

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

[38] D. Dentcheva and A. Ruszczyński. Portfolio optimization with stochastic dominance constraints. *journal of Banking and Finance*, 30:433–451, 2006.

[39] D. Dentcheva and A. Ruszczyński. composite semi-infinite optimization. *Control and Cybernetics*, 36:634–646, 2007.

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

[41] J. Dupačová. The minimax approach to stochastic programming and an illustrative application. *Stochastics*, 20:73–88, 1987.

[42] J. Dupačová, A. Gaivoronski, and Szantai T. Kos, Z. Stochastic programming in water management: A case study and a comparison of solution techniques. *European Journal of Operational Research*, 52:28–44, 1991.

[43] L. Eeckhoudt, H. Schlesinger, and I. Tsetlin. Approximating of risks via stochastic dominance. *Journal of Economic Theory*, 144:994–1003, 2009.

[44] A. Ekholm and D. Pasternack. The negative news threshold- an explanation of negative skewness in stock returns. *The European Journal of Finance*, 11:511–529, 2005.

[45] L. El-Ghaoui and H Lebret. Robust solutions to least-square problems to uncertain data matrices. *SIAM Journal on Matrix Analysis and Applications*, 18:1035–1064, 1997.

[46] L. El-Ghaoui, F. Oustry, and H Lebret. Robust solutions to uncertain semidenite programs. *SIAM Journal on Optimization*, 9:33–52, 1998.

[47] Y. Ermoliev. Random optimization and stochastic programming. *Lecture Notes in Mathematics*, 112:104–115, 1969.

[48] Y. Ermoliev. *Stochastic Programming Methods*. Nauka, Moscow, 1976.

[49] Y. Ermoliev. Stochastic quasi-gradient methods and their application to system optimization. *Stochastics*, 9:1–36, 1983.

[50] Y. Ermoliev and V. I. Norkin. On nonsmooth and discontinuous probles of stochastic systems optimization. *European Journal of Operational Research*, 101:230–244, 1997.

[51] Y. Ermoliev and J. B. Wets. *Numerical Techniques for Stochastic Optimization*. Springer-Verlag, Berlin,, 1988.

[52] C. I. Fábián, G. Mitra, and D. Roman. Processing second order stochastic dominance models using cutting-plane representations. *Mathematical Programming Series A*, 2009.

[53] E. F. Fama and J. Macbeth. Risk, return and equilibrium: Empirical tests. *Journal of Political Economy*, 81:607–636, 1973.

[54] S. Farinelli, M. Ferreira, D. Rossello, M. Thoeny, and L. Tibiletti. Beyond sharpe ratio: Optimal asset allocation using different performance ratios. *Journal of Banking & Finance*, 32:2057–2063, 2008.

[55] S. Farinelli, M. Ferreira, D. Rossello, M. Thoeny, and L. Tibiletti. Optimal asset allocation aid system: From "one-size" vs "tailor-made" performance ratio. *European Journal of Operational Research*, 192:209–215, 2009.

[56] P. C. Fishburn. *Decision and Value Theory*. John Wiley & Sons, New York, 1964.

[57] P. C. Fishburn. *Utility Theory for Decision Making*. John Wiley & Sons, New York, 1970.

[58] P. C. Fishburn. *Stochastic Dominance: An Approach to Decision Making Under Risk*. D. C. Heath, Lexington, MA, 1978.

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

[60] J. Goh and M. Sim. Distributionally robust optimization and its tractable approximations. *Operations Research*, 58:902–917, 2010.

[61] D. Goldfarb and G. Iyengar. Robust portfolio selection problems. *Mathematics of Operations Research*, 28:1–38, 2003.

[62] Gollmer, U., F. Neise, and R. Schultz. Stochastic programs with first-order dominance constraints induced by mixed-integer linear recourse. *SIAM Journal of Optimization*, 19:552–571, 2008.

[63] R. Gollmer, U. Gotzes, F. Neise, and R. Schultz. Risk modeling via stochastic dominance in power systems with dispersed generation. *Technical Report Department of Mathematics, University of Duisburg-Essen*, 2007.

[64] R. Gollmer, U. Gotzes, F. Neise, and R. Schultz. Second-order stochastic dominance constraints induced by mixed-integer linear recourse. *Technical Report Department of Mathematics, University of Duisburg-Essen*, 2007.

[65] M. Gugat. Error bounds for infinite systems of convex inequalities without slater's condition. *Mathematical Programming*, 88:255–275, 2000.

[66] S. Gustafon. Investigating semi-infinite programs using penalty function and lagrangian methods. *Australian Mathematical Society, Series B*, 28:158–169, 1986.

[67] G. Hanoch and H. Levy. The efficiency analysis of choices involving risk. *The Review of Economics Studies*, 36:335–346, 1969.

[68] J. Hardar and W. Russel. Rules for ordering uncertain prospects. *The American Economics Review*, 59:25–34, 1969.

[69] G. H. Hardy, J. E. Littlewood, and G. Polya. Inequalities. 1934.

[70] T. Homem-de Mello and S. Mehrota. A cutting surface method for uncertain linear programs with polyhedral stochastic dominance constraints. *SIAM Journal of Optimization*, 20:1250–1273, 2009.

[71] J. Hu, T. Homem-de Mello, and S. Mehrota. Risk adjusted budget allocation models with application in homeland security. *To appear in IIE Transactions*, 43:819–839, 2011.

[72] J. Hu, T. Homem-de Mello, and S. Mehrota. Concepts and application of stochastically weighted stochastic dominance. April, 2011.

[73] J. Hu, T. Homem-de Mello, and S. Mehrota. Sample average approximation of stochastic dominance constrained programs. *Mathematical Programming, Series A*, October 2010.

[74] I. Huber, S. Ortobelli, S. T. Rachev, and E. Schwartz. Portfolio choice theory with non-gaussian distributed returns. *In: Rachev, S. T. (Ed.), Handbook of Heavy Tailed Distribution in Finance, Elsevier, Amsterdam*, pages 547–594, 2003.

[75] G. Infanger. *Stochastic Programming, International Series in Operations Research and Management Science,*. Springer Science, 2011.

[76] S. R. Jeffrey and V. R. Eidman. The application of multivariate stochastic dominance criteria to agricultural economic problems. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie*, 39:193–209, 1991.

[77] P. Kall. *Stochastic Programming.* John Wiley & Sons, 1994.

[78] P. Kall and J. Mayer. Stochastic linear programming: Models, theory, and computations. *Springer, International Series in Operations Research and Management Science*, 2005.

[79] J. Keifer and Wolfowitz. Stochastic estimation of the maximum of a regression function. *Annual Mathematical Statistics*, 23:462–466, 1952.

[80] J. E. Kelley. A cutting plane method for solving convex programs. *SIAM Journal of Applied Mathematics*, 8:703–712, 1960.

[81] W. K. Klein Haneveld and M. H. Ven Der Vlerk. Integrated chance constraints: reduced forms and an algorithm. *Computational Management Science*, 3:254–269, 2006.

[82] A. J. Kleywegt, A. Shapiro, and T. Homem-de Mello. The sample average approximation method for stochastic discrete optimization. *SIAM Journal of Optimization*, 12:479–502, 2002.

[83] P. Krokhmal, J. Palmquist, and S. Uryasev. Portfolio optimization with conditional value-at-risk objective and constraints. *Journal of Risk*, 4:43–68, 2001.

[84] Y. Kroll and H. Levy. Stochastic dominance: a review and some new evidence. *Research in Finance*, 2:163–227, 1980.

[85] A. Künzi-Bay and J. Mayer. Computational aspects of minimizing conditional value-at-risk. *computational Management Science*, 3:3–27, 2006.

[86] T. Larsson, M. Patriksson, and A. Strömberg. Conditional subgradient optimization - theory and application. *European Journal of Operational Research*, 88:382–403, 1996.

[87] J. B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM journal on Optimization*, 11:796–817, 2001.

[88] H. E. Leland. Beyond mean-variance: performance measurement in a nonsymmetrical world. *Financial Analyst Journal*, 55:27–35, 1999.

[89] C. Lemarechal, A. Nemirovskii, and Y. Nestrov. New variants of bundle methods. *Mathematical Programming*, 69:111–147, 1995.

[90] H. Levy and M. Sarnat. Stochastic dominance and expected utility: survey and analysis. *Management Science*, 38:555–593, 1992.

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

[92] J. Linderoth, A. Shapiro, and S. Wright. The empirical behavior of sampling methods for stochastic programming. *Annuals of Operations Research*, 142:215–241, 2006.

[93] Y. Liu and H. Xu. Stability and sensitivity analysis of stochastic programs with second order dominance constraints. *Stochastic Programming E-Print Series*, 2010.

[94] A. W. Lo. Semi-parametric upper bounds for option prices and expected payoffs. *Journal of Financial Economics*, 19:373–387, 1987.

[95] J. Luedtke. New formulation for optimization under stochastic dominance constraints. *SIAM Journal on Optimization*, 19:1433–1450, 2008.

[96] D. Madden. Mental stress in ireland, 1994–2000: a stochastic dominance approach. *Health Economics*, 18:1202–1217, 2009.

[97] W. K. Mak, D. P. Morton, and R. K. Wood. Monte carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters*, 24:47–56, 1999.

[98] H. M. Markowitz. Portfolio selection. *Journal of Finance*, 7:77–92, 1952.

[99] H. M. Markowitz. *Portfolio Selection.* John Wiley & Sons, New York, 1959.

[100] H. M. Markowitz. *Mean-Variance Analysis in Portfolio Choice and Capital Markets.* Blackwell, Oxford, 1987.

[101] A. Müller and D. Stoyan. *Comparison Methods for Stochastic Models and Risks.* Wiley, Chichester, 2002.

[102] K. Natarajan, D. Pachamanova, and M. Sim. Incorporationg asymmetric distributional information in robust value-at-risk optimization. *Management Science*, 54:573–585, 2008.

[103] Y. Nie, X. Wu, and T. Homem-de Mello. Optimal path problems with second-order stochastic dominance constraints. *Northwestern University*, 2009.

[104] W. Ogryczak and A. Ruszczyński. From stochastic dominance to mean-risk models: semideviations as risk measures. *European Journal of Operational Research*, 116:33–50, 1999.

[105] W. Ogryczak and A. Ruszczyński. On consistency of stochastic dominance and mean-semideviations models. *Mathematical Programming*, 89:217–232, 2001.

[106] W. Ogryczak and A. Ruszczyński. Dual stochastic dominance and related mean-risk models. *SIAM Journal of Optimization*, 13:60–78, 2002.

[107] P. Parrilo. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*. PhD Thesis, California Institute of Technology, 2000.

[108] M. Pinar and R. H. Tütüncü. Robust profit opportunities in risky financial portfolios. *Operations Research Letters*, 33:331–340, 2005.

[109] B. T. Poljak. Nonlinear programming methods in the presence of noise. *Mathematical programming*, 14:87–97, 1978.

[110] M. Putinar. Positive polynomials on compact semialgebraic sets. *Indian university Mathematics Journal*, 42:969–984, 1993.

[111] J. Quirk and R. Saposnik. Admissibility and measurable utility functions. *The Review of Economi Studies*, 29:140–146, 1962.

[112] H. Robbins and S. Monro. A stochastic approximation method. *Annual Mathematical Statistics*, 22:400–407, 1951.

[113] S. M. Robinson. An application of error bounds for convex programming in a linear space. *SIAM Journal of Control and Optimization*, 13:271–273, 1975.

[114] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, N. J., 1970.

[115] R. T. Rockafellar and S. Uryasev. Optimization of conditional value at risk. *The Journal of Risk*, 2, 2000.

[116] R. T. Rockafellar, S. Uryasev, and M Zabarankin. Deviation measures in risk analysis and optimization. *Finance Stochastics*, 10:51–74, 2006.

[117] D. Roman, K. Darby-Dowman, and G. Mitra. Portfolio construction based on stochastic dominance and target return distributions. *Mathematical Programming*, 108:541–569, 2000.

[118] D. Roman and G. Mitra. Portfolio selection models: a review and new directions. *Wilmott Journal*, 1:1–17, 2009.

[119] M. Rothschild and J. F. Stiglitz. Increasing risk: I. a definition. *Journal of Economics Theory*, 2:225–243, 1970.

[120] G. Rudolf and A. Ruszczyński. Optimization problems with second order stochastic dominance constraints: Duality, compact formulations, and cut generation methods. *SIAM Journal on Optimization*, 19:1326–1343, 2008.

[121] A Ruszczyński and A. Shapiro. *Stochastic Programming.* Elsevier, Amsterdam, 2003.

[122] A Ruszczyński and A. Shapiro. Optimization of convex risk functions. *Mathematics of Operation Research*, 31:433–452, 2006.

[123] Rachev S., S. T. Ortobelli, S. Stoyanov, Fabozzi F., and Biglova A. Desirable properties of an ideal risk measure in portfolio theory. *International Journal of Theoretical and applied Finance*, 11:19–54, 2008.

[124] Rachev S., Jašik T., Biglova A., and Fabozzi F. Risk and return in momentum strategies: Profitability from portfolios based on risk-adjusted stock ranking criteria. *Technical report, Department of probability and applied statistics, University of California, Santa Barbara*, 2005.

[125] H. Scarf. A minmax solution of an inventory problem. *K. S. Arrow, s. Karlin, H. E. Scarf. Studies in the Mathematical Theory of Inventory and Production, Stanford University Press*, 1958.

[126] K. Schm udgen. The k-moment problem for compact semialgebraic sets. *Mathematische Annalen*, 289:203–206, 1991.

[127] A. Shapiro. Monte carlo sampling methods, in stochastic programming. *Handbook in Operations Research and Management Science, Stochastic Programming*, 10:353–425, 2003.

[128] A. Shapiro and S. Ahmed. On a class of minimax stochastic programs. *SIAM Journal on Optimization*, 14:1237–1249, 2004.

[129] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory.* SIAM, Philadelphia, PA, 2009.

[130] A. Shapiro and A. J. Kleywegt. Minimax analysis of stochastic problems. *Optimization methods and software*, 17:523–542, 2002.

[131] A. Shapiro and A. Nemirovski. On complexity of stochastic programming problems, in:. *Continuous Optimization: Current Trends and Applications*, 2005.

[132] W. F. Sharpe. Mutual funds performance. *Journal of Business*, pages 119–138, 1966.

[133] W. F. Sharpe. The sharpe ratio. *Journal of Portfolio Management*, Fall:49–58, 1994.

[134] F. A. Sortino and S. Satchell. *Managing Downside Risk in Financial Markets*. Butterworth Heinemann, 2001.

[135] A. Soyster. Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research*, 21:1154–1157, 1973.

[136] T. J. Stieltjes. Recherches sur les fractions continues. *Annales de la Facculte des Sciences de Toulouse*, 8:1–122, 1894.

[137] S. Sun, H. Xu, R. Meskarian, and Y. Wang. Exact penalization, level function method and modified cutting-plane method for stochastic programs with second order stochastic dominance constraints. October,2011.

[138] W. Sun and Y. Yuan. *Optimization Theory and Methods*. Springer, New York, 2006.

[139] S. Tesfatsion. Stochastic dominance and the maximization of expected utility. *The Review of Economics Studies*, 43:301–315, 1976.

[140] S. Uryasev. Conditional value-at-risk: Optimization algorithms and applications. *Financial Engineering News*, 14:1–5, 2000.

[141] J. Von Neumann and O. Morgenstern. Theory of games and economic behavior. *The Review of Economics Studies*, 29:140–146, 1962.

[142] S. W. Wallace and H. Kjetil. Generating scenario trees for multi-stage decision problems. *Management Science*, 47:295–307, 2001.

[143] S. W. Wallace, H. Kjetil, and K. Michal. A heuristic for moment-matching scenario generation. *Computational Optimization and Application: An International Journal*, 24:169–185, 2003.

[144] G. A. Whitmore. Third-degree stochastic dominance. *The American Economics Review*, 60:457–459, 1970.

[145] G. A. Whitmore and M. C. Findlay. *Stochastic Dominance: An Approach to Decison-Making Under Risk*. D. C. Health, Lexington, MA, 1978.

[146] H. Xu. Level function method for quasiconvex programming. *Journal of Optimization Theory and Applications*, 108:407–437, 2001.

[147] H. Xu, C. Caramanis, and S. Mannor. A distributional interpretation of robust optimization. *Mathematics of Operations Research*, 37:95–110, 2012.

[148] Y. Yamai and T. Yoshiba. Comparative analysis of expected shortfall and value at risk: expected utility maximization and tail risk. *Monetary and Economics Studies*, 20:57–86, 2002.

[149] M. R. Young. A minimax portfolio selection rule with linear programming solution. *Management Science*, 44:673–683, 1998.

[150] J. Zackova. On minimax solution of stochastic linear programming problems. *Casopis pro Pestovani Matematiky*, 91:423–430, 1966.

[151] S. A. Zenios. Mean-absolute deviation portfolio optimization for mortgage backed securities. *Annals of Operations Research*, 45:433–450, 1993.

[152] S. A. Zenios. A model for portfolio management with mortgage backed securities. *Annals of Operations Research*, 43:337–356, 1993.

[153] S. Zhu and M. Fukushima. Worst-case conditional value-at-risk with application to robust portfolio management. *Operations Research*, 57:1155–1168, 2009.

[154] S. Zymler, D. Kuhn, and B. Rustem. Worst-case conditional value-at-risk of non-linear portfolios. *Optimization Online*, 2009.