Coherent Distortion Risk Measures in Portfolio Selection

by

Ming Bin Feng

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Actuarial Science

Waterloo, Ontario, Canada, 2011

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

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

Abstract

The theme of this thesis relates to solving the optimal portfolio selection problems using linear programming. There are two key contributions in this thesis. The first contribution is to generalize the well-known linear optimization framework of Conditional Value-at-Risk (CVaR)-based portfolio selection problems (see Rockafellar and Uryasev [50, 51]) to more general risk measure portfolio selection problems. In particular, the class of risk measure under consideration is called the Coherent Distortion Risk Measure (CDRM) and is the intersection of two well-known classes of risk measures in the literature: the Coherent Risk Measure (CRM) and the Distortion Risk Measure (DRM). In addition to CVaR, other risk measures which belong to CDRM include the Wang Transform (WT) measure, Proportional Hazard (PH) transform measure, and lookback (LB) distortion measure. Our generalization implies that the portfolio selection problems can be solved very efficiently using the linear programming approach and over a much wider class of risk measures.

The second contribution of the thesis is to establish the equivalences among four formulations of CDRM optimization problems: the return maximization subject to CDRM constraint, the CDRM minimization subject to return constraint, the return-CDRM utility maximization, the CDRM-based Sharpe Ratio maximization. Equivalences among these four formulations are established in a sense that they produce the same efficient frontier when varying the parameters in their corresponding problems. We point out that the first three formulations have already been investigated in Krokhmal et al. [36] with milder assumptions on risk measures (convex functional of portfolio weights). Here we apply their results to CDRM and establish the fourth equivalence. For every one of these formulations, the relationship between its given parameter and the implied parameters for the other three formulations is explored. Such equivalences and relationships can help verifying consistencies (or inconsistencies) for risk management with different objectives and constraints. They are also helpful for uncovering the implied information of a decision making process or of a given investment market.

We conclude the thesis by conducting two case studies to illustrate the methodologies and implementations of our linear optimization approach, to verify the equivalences among four different problem formulations, and to investigate the properties of different members of CDRM. In addition, the efficiency (or inefficiency) of the so-called $\frac{1}{n}$ portfolio strategy in terms of the trade off between portfolio return and portfolio CDRM. The properties of optimal portfolios and their returns with respect to different CDRM minimization problems are compared through their numerical results.

Acknowledgments

I offer my sincerest gratitude to my advisor and dear friend,Dr Ken Seng Tan, for his supervision, advice, and guidance from the very beginning of this research project. Dr Tan has been giving me extraordinary experiences throughout the research process. Above all and the most needed, he provided me the utmost encouragement and support when I experience difficulties and depressions. I am also extremely grateful to Dr Phelim Boyle for his generosity to share with me his research ideas and to have me involved in his research projects.

Thanks to the Department of Statistics and Actuarial Science and Dr Phelim Boyle for providing me with the financial means to complete this thesis.

Special thanks to Mary Lou Dufton, Joan Hatton and other administrative staffs for their continuous supports and assistances in the development and preparation of this thesis.

I would like to thank all the friends who have been of invaluable assistance when I hit a bottleneck during the period of completing this thesis. My fellow classmates have always been the source of intellectual stimulation.

Finally, my deepest gratitude is sent to my family for their warm support, which brings belief and hope into my life. To may parents for their understanding and endless love. I give my heartfelt gratitude to my beloved fiancée, Jiayi Li, her never-ending encouragement and unfailing love.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

This thesis studies the mathematical and computational properties of *Coherent Distortion Risk Measures (CDRMs)*, incorporates CDRM in various portfolio selection models and shows the equivalences among these models.

Coherent Distortion Risk Measure, as its name reveals, is the intersection of two well-known classes of risk measures: *coherent risk measures (CRMs)* and *distortion risk measures (DRMs)*. CRM was first proposed by Artzner et al. [3] via an axiomatic approach, in which mathematical properties of risk measures were derived from a set of intuitive principles. DRM was first proposed and studied as an insurance premium principle in series of papers by Wang [64, 65] and Wang et al. [69]. DRM was studied as a risk measure by Wirch and Hardy [71] and has become a popular class of risk measure since then. These two classes of risk measures have certain desirable conceptual, mathematical and computational properties, some of which are common for both classes while the others are different. Further discussions of CRM and DRM will be made in Section 3.2 and Section 3.3 respectively. CDRM enjoys properties of both CRM and DRM hence can be applied to portfolio selection problems in a unique way that neither CRM or DRM can in general. In particular, any CDRM can be represented as a convex combination of Conditional Value-at-Risk (CVaR) at different confidence levels but neither CRM nor DRM has such an amenable representation.

The seminal work of Markowitz [44] gave rise to long lasting research efforts in the field of portfolio selection. Portfolio selection problems are solved in order to exploit the so-call risk-reward trade off. Given a universe of assets, any rational investor wants to select a portfolio consisting these assets such that it both suits the investor's risk appetite and generates the highest possible portfolio return. With selected risk and return measures, an investor can have different ways to describe his/her risk-reward trade off, which in turn induce different formulations of portfolio selection problems.

This thesis studies four formulations of portfolio selection problems and establishes equivalences among these four problems via *Karush-Kuhn-Tucker (KKT) conditions*. Unless stated otherwise, the expected return and CDRM of a portfolio are the selected return and risk measures in our portfolio selection problems. In our portfolio section problem formulations, the n-vector $\boldsymbol{x}$[1] can be interpreted as a portfolio consisting of n instruments. Depending on the particular application x_i may represent the number of units of instrument i in the portfolio or the proportion of initial wealth invested in instrument i. $\boldsymbol{D} \in \mathbb{R}^n$ denotes the set of feasible portfolios that satisfies given constraints such as budget constraint, non-negativity constraints, etc besides the risk and/or return constraints that are given separately. Finally $R(\boldsymbol{x})$ and $\rho(\boldsymbol{x})$ denote the expected return and the risk measures for portfolio $\boldsymbol{x}$ respectively, unless specified otherwise. Detailed formulations, notations and discussions of these formulations will be presented in Chapter 4. As a preview, the following four portfolio selection problems are studied in this thesis:

1. **Return maximization.** Select a feasible portfolio that has the highest expected portfolio return given that its portfolio risk is at most η

$$\max_{\boldsymbol{x}} \{R(\boldsymbol{x}) | \rho(\boldsymbol{x}) \leq \eta, \boldsymbol{x} \in \boldsymbol{D}\} \tag{1.1}$$

2. **Risk minimization.** Select a feasible portfolio that has the lowest portfolio risk given that its expected portfolio return is at least μ

$$\min_{\boldsymbol{x}} \{\rho(\boldsymbol{x}) | R(\boldsymbol{x}) \geq \mu, \boldsymbol{x} \in \boldsymbol{D}\} \tag{1.2}$$

3. **Utility maximization.** Select a feasible portfolio that has the highest utility that is expressed as the expected portfolio return minus the product of a risk aversion parameter $\tau > 0$ and portfolio risk[2]

$$\max_{\boldsymbol{x}} \{R(\boldsymbol{x}) - \tau\rho(\boldsymbol{x}) | \boldsymbol{x} \in \boldsymbol{D}\} \tag{1.3}$$

[1] Throughout this thesis, all vectors and matrices are specified in italic boldface while all scalers are specified as italic.

[2] $R(\boldsymbol{x}) - \tau\rho(\boldsymbol{x})$ is referred to as a utility function in the literature (see [51] for example) but it is not the same as the classical Von Neumann-Morgenstern utility.

4. **Sharpe Ratio maximization.** Select a feasible portfolio that has the highest Sharpe Ratio, expressed as ratio of the difference of portfolio return and a benchmark return level ν and the portfolio risk

$$\max_{\boldsymbol{x}}\{\frac{R(\boldsymbol{x})-\nu}{\rho(\boldsymbol{x})}|\boldsymbol{x}\in\boldsymbol{D}\} \tag{1.4}$$

This thesis is structured as follows: the remainder of this chapter motivates the study of CDRM, its different formulations in portfolio selection problems and provides a comprehensive literature review on developments in risk measures, portfolio selection, and the conjunction of the two fields. Chapter 2 presents a number of preliminaries such as linear programming (LP), fractional linear programming (LFP), Karush-Kuhn-Tucker (KKT) conditions, and CVaR optimization shortcut for ease of later discussions. Chapter 3 studies various properties of CRM, DRM and CDRM. Chapter 4 studies four formulations of CDRM portfolio selection problem and establishes equivalences among those formulations. Chapter 5 conducts case studies and presents numerical results to verify the equivalences shown in Chapter 4 and to study some interesting observations when applying our methodologies to real data. Chapter 6 gives concluding remarks for this thesis and proposes plausible extensions for future research.

1.1 Motivation

Although a rigid definition of the risk-reward trade off, or sometimes referred to as the risk-return trade off, varies across different applications, it is commonly known as the principle that potential return rises with an increase in risk. A quantitative measure of risk-reward trade off in literature, for example, is the expected excess return on a broad stock market index divided by its standard deviation, commonly known as the *Sharpe ratio*[3]. Other measures of return and risk are used in other applications. One challenge to portfolio managers is to construct portfolios that fully utilize a preselected measure of risk-reward trade off. For example, an investor may want to construct a portfolio with the highest Sharpe ratio so that the highest excess return can be obtained for each unit of risk, from a statistical point of view. Consider yet another example, when reinsurance companies take layers of risks from insurance companies in exchange for premiums, they may want to collect as much premiums as possible given that their existing level of reserve is sufficient

[3] Also known as the *price of risk* and the *Return on Risk Adjusted Capitals (RORAC)*, depending on particular application.

to cover losses from the risks they have taken 95% of the time. In the aforementioned examples and many others, it is clear that utilization of risk-reward trade off naturally induces various optimization problems.

Although the risk-reward trade off is an intuitive principle and is widely accepted in practice, formulating the induced optimization problems into mathematical programming problems that are computationally amenable computationally is a challenging task. The source of difficulty is two fold. Firstly there are many risk and return measures and it is hard to make appropriate choice in practice. In particular the choice of risk measure is in itself an ongoing debate and a universally accepted risk measure is yet to be decided. Secondly for predetermined choices of risk and return measures, understanding their mathematical and computational properties is of critical importance in formulating them into programming problems. Otherwise the induced portfolio selection problem could be ill-structured in a sense that it cannot be solved sufficiently and hence is of little practical value.

Searching for useful risk measures can help addressing the first difficulty. Prior to the work of Artzner et al. [3], the search of useful risk measure remained in a "passive" phase in a sense that new risk measures are proposed by observing and tackling disadvantages of old ones or extensions of old ones. For example, variance as a risk measure was criticized since it penalizes both unexpected portfolio increases and unexpected portfolio declines. Semi-variance was then proposed to tackle such shortcoming, which was later extended to partial moments. In such phase of search, efforts were paid mostly to find out what properties that a useful risk measure should not have. Artzner et al. [3] not only proposed a new class of risk measures, but also initiated an "active" way of searching for desirable risk measures. In such active search, a set of intuitive principles, such as the principle of risk diversification and impossibility of hedging similar risks, are proposed and explored in an intuitive way and are termed "axioms". Such axioms are then abstracted in mathematical expressions. The resulting properties, such as subadditivity and comonotonic additivity, are exploited from mathematical and computational perspectives. Simply put, an active search pays efforts to find out what properties that a useful risk measure should have. Both CRM and DRM are products of such active search but with different sets of axioms. Detailed discussions of the long lasting search of risk measures in literature are given in Section 1.2.

By construction, CRM and DRM are useful risk measures because they both have some intuitive and desirable properties as risk measures, namely their underlying set of axioms. Moreover, the mathematical properties of CRM and DRM derived from their underlying set of axioms ease their formulations in mathematical programming in different ways. CDRM, as the intersection of two classes of risk measures, enjoys conceptual, mathematical, and computational properties of both CRM and DRM. As a result, it is possible to derive further and stronger results for CDRM in the context of portfolio selection. Such possibility motivates our study of CDRM and its applications in portfolio selection problems.

However, even with well chosen risk and return measures which are conceptually intuitive and computationally amenable, there are various ways to formulate the induced portfolio selection problem. Within a given feasible set of portfolios[4], common formulations of portfolio selection problem include minimization of risk given expected return constraints, maximization of expected return given risk constraints, maximization of a special utility function that is a linear combination of return and risk, and maximization of Sharpe Ratio. It is worth noting that the last formulation is generally more complicated than the others from mathematical programming perspective. This thesis establishes equivalences among these four formulations in a sense that these formulations produce the same optimal solution if their parameters satisfy certain equivalent conditions.

The equivalences established in this thesis can help reconciling different investment objectives and help detecting inconsistencies for different investment criteria. For instance, two investors with distinct objectives (one wants return maximization while the other wants risk minimization, for example) may end up with the same optimal portfolio because a certain conditions are satisfied. Therefore under these conditions the two different investment objectives are indeed equivalent. In practice, investors sometimes have multiple investment decision criteria at the same time and it is of interest to tell whether these multiple criteria are consistent with each other, i.e., whether the optimal portfolio for one criterion is the same as that for the other criteria. For example, different departments of an investment bank may use different investment criteria in investment decisions, it is helpful to verify the consistencies (or inconsistencies) among these criteria to ensure the portfolio performance of the bank's overall portfolio.

With such equivalences, the four aforementioned types of formulations can be used interchangeably. Such equivalences provide portfolio managers with extra freedom in choosing decision criteria base on their own needs and conventions. For example, management may have a risk minimization problem in mind since it has decided on a minimum acceptable level of return but the company's internal software is specialized in solving return maximization problems. Our results can help recasting the risk minimization problem as a return maximization problem so that management needs can be satisfied and computational advantages can be taken in the mean time. Moreover, the equivalences reveal linkages among these formulations. For example, if management preselects‘ a minimum acceptable level of return, the equivalences can infer its equivalent maximum acceptable risk level, risk aversion parameter value, and benchmark return level. Such additional informa-

[4]The set of portfolios that satisfies portfolios constraints other than risk and return constraints.

tion can help companies in checking consistencies in decision making process with different decision criteria, in inferring competitors' investing strategies from limited available information, etc.The growing complexity of financial risk management problems demand uses of sophisticated risk measures which in turn induce complicated mathematical programming problems. The benefits of equivalences among different formulations will become more and more significant by connecting a seemingly overwhelming problem with a well studied problem that is readily applicable and can be solved efficiently.

1.2 Literature Review

The pioneering work of Markowitz [44] addressed and solved portfolio selection problems via mathematical programming models. The mean-variance optimization framework and the related works of Sharpe [54], Lintner [43], Merton [46] and others' contributed to the revolutionary development of a new subject of study: the *Mordern Portfolio Theory (MPT)*. Despite the critiques on its unrealistic assumptions about investors and markets, MPT has been one of the most important models in portfolio selection and risk management. Furthermore, MPT is appreciated in many corporate finance applications. Such concepts as the *Capital Asset Pricing Model (CAPM)*, the *Efficient Frontier*, the *Capital Allocation Line (CAL)*, the *Capital Market Line (CML)*, the *Security Market Line (SML)* have gained critical importances in both academic research and industrial applications. In short, MPT's conceptual simplicity in interpretation and structural amenability in computation lend itself the practical and theoretical importance.

One of the inspirations that MPT offers is the linkage between risk measures and mathematical programming. Developments in the subject of portfolio selection can be made through either search of new risk measures or improvements in modeling and solving mathematical programming. On one hand, proposals of new risk measures and exploitations of their structural properties ease the formulation of such risk measures in programming problems and provide meaningful interpretations to programming problems' inputs/outputs. On the other hand, advances in mathematical programming usually enlarges the applicability of existing portfolio selection problems by exploring more efficient solution methods, relaxing previous model assumptions, and developing more advanced models for more sophisticated risk measures. Note that other aspects such as economic factors in the model, modeling errors, and parameter estimation errors involved are also important in this subject but they are outside the scope of this thesis.

1.2.1 Advances in Risk Measures

Besides academic interests, the long lasting search for useful risk measures is in part stimulated by practical needs in business and investment risk management. The definition of risk varies across different industries and applications. Moreover, such definition can be controversial even within one particular discipline. Nevertheless, there are plenty of literatures dedicated to exploring sensible risk measures in different applications. This section aims to provide a comprehensive literature review on the advances in risk measures without overwhelming mathematical details and notational burdens.

Variation from expected return, or variance of returns in statistical terminology, seems to be a logical representation for risk and hence became one of the first risk measures studied in literature. Suppose there are n risky assets in the market and the return on asset i is denoted by a random variable $\boldsymbol{R}_i$. Denote the variance-covariance matrix of asset returns by $\boldsymbol{\Sigma}$, which usually is either given or estimated from data. Then for any portfolio $\boldsymbol{x}$, whose i^{th} entry x_i for $i = 1, \cdots, n$ denotes the proportion of initial wealth invested in asset i, the variance of portfolio return is given by[5]

$$\sigma_p^2 = \boldsymbol{x}^T \boldsymbol{\Sigma} \boldsymbol{x}$$

The variance of portfolio return as above was employed as the risk measure in the seminal work of Markowitz's [44] yet it has one obvious shortcoming as a risk measure. Variance holds a symmetric view towards unexpected portfolio growths and unexpected portfolio declines thus penalizes them in the same way. However, empirical findings show that investors view profits and losses differently, which contradicts with the symmetric measures that variance provides.

Semivariance, proposed by Markowitz [45] (first edition published in 1959), is alternative risk measure that tackles such symmetry issue. In line with the previous notations, the semivariance of portfolio $\boldsymbol{x}$ is given by:

$$\sigma_{p,min}^2 = E[((\sum_{i=1}^{n} x_i \overline{\boldsymbol{R}}_i - \sum_{i=1}^{n} x_i \boldsymbol{R}_i)^+)^2]$$

where $E[\cdot]$ denotes the expectation operator, $\overline{\boldsymbol{R}}_i = E[\boldsymbol{R}_i]$ denotes the expected return on asset i for $i = 1, \cdots, n$, and $(a)^+ = \max\{a, 0\}$ for $a \in \mathbb{R}$.

[5]Throughout this thesis, superscript capital T denotes matrix transposition and any vector without transposition operation is a column vector.

Bawa and Lindenberg [6] further extended the mean-semivariance optimization framework to a mean-lower partial moment optimization framework which considered a whole class of risk measures that hold asymmetric view towards profit and loss.

The mean-variance optimization framework has yet another shortcoming in failing to consider higher moments such as skewness and kurtosis for general portfolio return distributions. Such measures are found important in investment decisions empirically. Research efforts on incorporating higher moments of portfolio return distributions are evidenced by Lai [41], Chunhachinda [15], and Harvey et al. [27].

Other investment considerations are also important in a portfolio selection model. Roy [52] pointed out the *safety first principle* which states that investor want to ensure the safety of the investment principal before considering the risk and reward trade-off. This principle inspired the developments for numerous downside risk measures such as the *Expected Gain-Confidence Limit Criterion* discussed in Baumol [5] and was later polished as a more well-known risk measure called the *Value-at-Risk (VaR)*.

VaR measures the potential loss in value of a risky asset or portfolio (the negate of its return) over a defined period for a given confidence level. In many financial applications VaR is set as a capital requirement. For example, VaR at 95% is defined as the minimal amount of capital that is required in order to cover portfolio losses in 95% of cases. More precisely, given a confidence level $\alpha \in (0, 1)$, the Value-at-Risk at confidence level α of a loss random variable $R^* = -R$, $VaR_\alpha(R^*)$ is defined as the α-quantile of R^*, i.e.,

$$VaR_\alpha(R^*) = \inf\{r | Pr(R^* \le r) \ge \alpha\} \tag{1.5}$$

VaR has been widely used in practice since its introduction. Moreover, it was adopted as the "1st pillar" in Basel II, which are recommendations on banking laws and regulations issued by Basel Committee on Banking Supervision. Jorion [30] and Pritsker [48] provided discussions on VaR's high status in industry regulations. Despite its conceptual simplicity and practical popularity, VaR suffers from being unstable computationally and difficult to work with numerically. Moreover, VaR has been criticized among academics. Firstly, VaR does not provide information on the severity of losses beyond the threshold amount indicated by itself. Secondly, VaR fails to be coherent in the sense of Artzner et al. [3].

The seminal papers of Artzner et al. [3] constructed CRM based on a particular set of axioms and showed that VaR violates the subadditivity axiom in general. Besides pointing out the lack of coherence of VaR, Artzner et al. [3] also proposed a member of CRM as an alternative risk measure: the *Tail Conditional Expectation.* Such risk measure is studied intensively by different authors in various contexts, hence is sometimes defined and termed differently. Wirch and Hardy [71] defined the same risk measure as *Conditional Tail Expectation (CTE)*; Bertsimas et al. [12] termed a similar risk measure as the *Mean Shortfall*and a similar risk measure with seemingly identical term *Expected Shortfall* is studied by Acerbi et al. [1]. Rockafellar and Uryasev [50] termed their risk measure the *Conditional Value-at-Risk (CVaR)* and we will follow their terminology in this thesis. Although these risk measures are defined similarly and coincide with each other in some cases, their differences should not be ignored. Readers may refer to Rockafellar and Uryasev [51] for comprehensive discussions on CVaR to its full generality. In line with previous notations, the Conditional Value-at-Risk at level α, $CVaR_\alpha(R^*)$ is defined as the mean of the α-tail distribution of R^*, where the distribution in question is the one with cumulative distribution function defined by

$$F_\alpha(r) = \begin{cases} 0 & \text{for } Pr(R^* \leq r) < VaR_\alpha(R^*) \\ \frac{Pr(R^* \leq r) - \alpha}{1-\alpha} & \text{for } Pr(R^* \leq r) \geq VaR_\alpha(R^*) \end{cases} \tag{1.6}$$

Note that $E[R^*|R^* > VaR_\alpha(R^*)]$ and $E[R^*|R^* \geq VaR_\alpha(R^*)]$ are referred to as the upper α-CVaR($CVaR^+$) and the lower α-CVaR($CVaR^-$) respectively but neither of them is a proper definition of CVaR. These three quantities have the same value in some cases but are different for general loss distributions. Detailed discussions can be found in Rockafellar and Uryasev [51].

Distortion risk measures (DRM) is another well known class of risk measures that is constructed via similar axiomatic approach. The set of axioms underlies DRM was originally developed for and applied in insurance premium principles, see Goovaerts et al. [25] and Wang et al. [69]. Such premium principles are studied as risk measures in Wang [66, 67] and are known as DRM since then. Campana [13] considered the applications of DRM for discrete loss distributions. Gourieroux [26] provided a statistical framework for analyzing the sensitivities of DRMs with respect to various risk aversion parameters. Generally speaking, a DRM is the expectation of portfolio loss random variable under a distorted probability measure. Different distortions reflect different risk appetites of decision makers. From a mathematical point of view, DRM is a Choquet integral and all the standard results about Choquet integrals, such as those discussed in Denneberg [19], are applicable to DRM. Detailed discussions of DRM is given in Section 3.3.

Although there are commonalities in the underlying axioms for CRM and DRM and there are risk measures, such as CVaR, that belong to both CRM and DRM, these two classes of risk measures are not subclasses of each other. Kusuoka [40] studied subclasses of CRM and proved a representation theorem for comonotonic law-invariant coherent risk measures. For continuous loss distributions, any comonotonic law-invariant coherent risk measure can be represented as a convex combination of CVaR's at different confidence levels. Bertsimas and Brown [11] proved similar results for discrete loss distributions and strengthen their claim by proving only finite number of CVaR's are needed in the representation. To the best of the author's knowledge, Bellini and Caperdoni [8] was the first paper that synchronized CRM and DRM and studied the intersection of both classes and named it coherent distortion risk measures (CDRM). They showed that CDRM coincides with the class of comonotonic law-invariant coherent risk measures. Acerbi et al. [2] studied spectral measures of risk and applied them in portfolio selection problems. A spectral risk measure with admissible spectrum can be seen as a CDRM yet it overlooks its connection the underlying distortion function. Acerbi et al. [2] also provided an optimization scheme for portfolio selection problems with spectral risk measures. Inspired by the linear optimization scheme for CVaR and spectral risk measures as well as the synchronization of CRM and DRM, we consider applying the convex combination representation of CVaR's for CDRM in portfolio selection problems.

1.2.2 Advances in Mathematical Programming

Mathematical programming concerns the optimal allocation of resources. Its origin can be traced back to the World War II when the armies looked for ways to minimize their own costs and maximize losses to their enemies. The simplest model, linear programming, was developed as an academic discipline in 1939 by Russian mathematician Kantorovich and Dutch mathematician Koopmans, recognized by Kantorovich [33][6]. The founders of the subject, however, are generally regarded as George B. Dantzig, who devised the simplex method in 1947 and published in Dantzig [16], and John von Neumann, who established the theory of duality in the same year based on Neumann [63]. Important generalizations of linear programming include:

1. Integer programming (IP), whose decision variables are restricted to integers, sometimes referred to as the integrality constraints.
2. Quadratic programming (QP), whose objective function is a quadratic function of decision variables subject to linear constraints on these variables.
3. Nonlinear programming (NLP), whose objective function and/or constraints are nonlinear.

[6] The original paper was published in Russian in 1939, the reference is its translated republication in 1960.

Fractional programming is a subclass of nonlinear programming where the objective function is a fraction of two functions. Such optimization problems arise naturally in many applications such as maximizing business efficiency ratio in corporate finance and minimizing pressure (ratio of force and area) on a surface in physical applications. References on nonlinear programming and fractional programming include Dinkelbach [22], Schaible and Ibaraki [53] and Stancu-Minasian [58]. Singh [55] studied the optimality conditions in fractional programming with and without differentiability requirements for objective and constraints.

Linear fractional programming (LFP) is a special case of fraction programming and is employed in this thesis to solve CDRM-based Sharpe Ratio maximization problems. Charnes and Cooper [14] discusses how to solve LFP by solving at most two related LPs via a variable transformation method. Tantawy [60] and Tantawy [61] provided two more solution methods and performed sensitivity analysis for LFPs.

Robust optimization and stochastic programming can also be viewed as generalizations of linear programming but they deserve a separate discussion. The aforementioned generalizations generalize the form of objective function and/or constraints but nevertheless are deterministic models which known and fixed problem parameters are assumed. However, any real world problem inevitably includes some degree of uncertainties in its parameter values. Robust optimization and stochastic programming aim to tackle such difficulty and take parameter uncertainties into consideration. Robust optimization considers parameters that are known only within certain bounds and aims to find a solution that is feasible for all parameter ranges and optimal in some sense. Its first introduction dates back to 1973 in Soyster [57]. Stochastic programming assumes parameters are random variables and their probability distributions are known or can be estimated. The goal then is to find a feasible solution for all possible data instances and maximizes the expectation of some function of the decisions variables[7].

1.2.3 Advances in Portfolio Selection

Advances in portfolio selection depend largely in advances in risk measures as well as in mathematical programming. Understandings of risk measures are the theoretical bases for formulations of portfolio selection problems. Mathematical programming determines the breadth and depth of a risk measure's applicability in practice. As discussed above, the origin of mathematical programming predates that of risk measures hence the latter

[7]Extensive description of various methodologies for robust optimization and stochastic programming can be found, along with other resources, at URL http://stoprog.org/

has been a major determinant in the advances in portfolio selection. Many mathematical models along with their solution methods are usually readily in place once the structural properties of a new risk measure is explored and can formulated into a programming problem.

Markowitz's mean-variance portfolio selection framework used quadratic programming since variance is by nature a quadratic function of portfolio weights. Skewness coefficient and kurtosis coefficient of portfolio loss are the 3^{rd} and 4^{th} order polynomials of portfolio weights hence can be naturally formulated as corresponding 3^{rd} and $4^{t}h$ order programming problems as discussed in Jurczenko et al. [32]). Portfolio selection problems with semivariance as risk measure can be formulated as a second-order conic programming (SOCP). By the same token, p-order lower partial moment risk measures can be formulated as p-order conic programming (pOCP). References can be found in Tomoyuki [62] and Soberanis et al. [56].

VaR's widespread popularity has stimulated extensive research efforts in formulating a mean-VaR portfolio selection framework. Basak and Shapiro [4] analyzed optimal portfolio and wealth/consumption policies with VaR as risk measure. Puelz [49] presented and compared four models for VaR portfolio selection problems. Gaivoronski and Pflug [24] studied mean-VaR optimization problems with emphasis on its structural properties and computational aspects. Benati and Rizzi [10] formulates mean-VaR portfolio selection problem as mixed integer linear programs and showed that such problem is NP-hard. A comment on this paper by Lin [42] showed that one claim in Benati and Rizzi [10] was only partly true and proposed an alternative model for mean-VaR portfolio problem.

In actuarial and financial literature, CVaR is usually regarded as a superior risk measure than VaR. Besides coherence, CVaR is also mathematically and computationally more stable with respect to portfolio weights than VaR is. Remarkably, Rockafellar and Uryasev [50] developed a CVaR minimization scheme under which portfolio selection problem with CVaR objective can be formulated as a linear programming. Krokhmal et al. [36] made use of such formulation and extended it to problems with CVaR constraints. Such linear formulation for CVaR was also developed independently by Bertsimas et al. [12]. Bertsimas et al. [12] also studied some properties of the mean-CVaR efficient frontier. Fábián [23] exploited the linear optimization scheme of CVaR and considered CVaR in stochastic models. Krokhmal and Soberanis [37] investigated a subclass of coherent risk measures, termed the higher moment coherent risk measures (HMCR), formulated HMCR into p-order conic programs, and employed a decomposition scheme to solve HMCR portfolio selection problems efficiently.

Chapter 2

Preliminaries

This thesis studies structural properties of CDRM, considers four formulations of mean-CDRM portfolio selection problems, and establishes equivalences among these formulations via Karush-Kuhn-Tucker (KKT) optimality conditions. Reviews on linear programming, linear fractional programming, KKT optimality conditions, and CVaR minimization short-cut are important preliminaries to ease later discussions in this thesis. Full coverage of these preliminaries is outside the scope of this thesis yet some important results are presented in the current chapter.

2.1 Linear Programming and Fractional Programming

As the foundation of mathematical programming, linear programming(LP) is a technique for the optimization of a linear objective function, subject to linear equality and/or linear inequality constraints. When put into specific context, LP can be a powerful modeling method for determining a way to achieve the best outcome when some requirements are specified. For example, the problem of profit maximization subject to limited resources can be formulated as an LP. Mathematically, the *standard inequality form* of a minimization LP problem is given by

$$\begin{array}{rl} \text{minimize} & \boldsymbol{c}^T\boldsymbol{x} \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} \;\leq\; \boldsymbol{b} \end{array} \tag{2.1}$$

where $\boldsymbol{x} \in \mathbb{R}^n$ denotes the vector of decision variables, $\boldsymbol{A}$ is an m-by-n matrix of known coefficients, $\boldsymbol{c} \in \mathbb{R}^n$ and $\boldsymbol{b} \in \mathbb{R}^m$ are both vectors of known coefficients.

$\boldsymbol{c}^T\boldsymbol{x}$, the linear function of $\boldsymbol{x}$ to be minimized is called the objective function. The set of linear equations $\boldsymbol{Ax} \leq \boldsymbol{b}$ are the constraints which specify a convex polytope over which the objective function is to be minimized. Let $\boldsymbol{D}$ be the set of feasible solutions for (2.1), i.e., $\boldsymbol{D} = \{\boldsymbol{x} \in \mathbb{R}^n | \boldsymbol{Ax} \leq \boldsymbol{b}\}$. It is sometimes convenient write $\boldsymbol{D}$ in terms of the individual constraint as[1] $\boldsymbol{D} = \{\boldsymbol{x} \in \mathbb{R}^n | \boldsymbol{A}_{i\cdot}^T\boldsymbol{x} \leq b_i, i = 1, \cdots, m\}$, so that $\boldsymbol{A}_{i\cdot}\boldsymbol{x}$ is the gradient of the i^{th} constraint and b_i is its right hand side. Unless stated otherwise, we assume that $\boldsymbol{D}$ is a bounded convex set that satisfies some regularity conditions specified in Section 2.2.

Admittedly the linearity requirements on objective and constraints may hinder the applicability of LP in practice. Yet there are plenty of modeling techniques and solution methods which allows LP to outreach a wide range of real life problems. Although it is not the purpose of this thesis to provide a thorough coverage on modeling and solving LP, some relevant LP modeling techniques for discussions in this thesis are shown below:

1. For LP maximization problems we can easily minimize the negative of its original objective function in order to convert it in the form of (2.1)

2. For any linear equality constraint

$$\boldsymbol{A}_{i\cdot}^T\boldsymbol{x} = b_i$$

   we can replace it by two linear inequality constraints:

$$\begin{aligned} \boldsymbol{A}_{i\cdot}^T\boldsymbol{x} &\leq b_i \\ \boldsymbol{A}_{i\cdot}^T\boldsymbol{x} &\geq b_i \end{aligned}$$

3. For optimization problems of the form (known as *minimax* and *maximin* problems)

$$\begin{aligned} &\min_{\boldsymbol{x}\in\boldsymbol{D}} \quad \{\max\{\boldsymbol{c}_1^T\boldsymbol{x}, \boldsymbol{c}_2^T\boldsymbol{x}, \cdots, \boldsymbol{c}_k^T\boldsymbol{x}\}\} \quad \text{, or} \\ &\max_{\boldsymbol{x}\in\boldsymbol{D}} \quad \{\min\{\boldsymbol{c}_1^T\boldsymbol{x}, \boldsymbol{c}_2^T\boldsymbol{x}, \cdots, \boldsymbol{c}_k^T\boldsymbol{x}\}\} \end{aligned}$$

   where $\boldsymbol{c}_i \in \mathbb{R}^n, \forall i = 1, 2, \cdots, k$. We can reformulate the problem by adding one auxiliary variable z and k linear constraints:

$$\begin{aligned} &\min_{\boldsymbol{x}\in\boldsymbol{X}, z\in\mathbb{R}} \quad \{z | z \geq \boldsymbol{c}_i^T\boldsymbol{x}, i = 1, \cdots, k\} \quad \text{, or} \\ &\max_{\boldsymbol{x}\in\boldsymbol{X}, z\in\mathbb{R}} \quad \{z | z \leq \boldsymbol{c}_i^T\boldsymbol{x}, i = 1, \cdots, k\} \end{aligned}$$

[1] For any matrix $\boldsymbol{A}$, $\boldsymbol{A}_{\cdot i}$ and $\boldsymbol{A}_{i\cdot}$ denotes the i^{th} column and the i^{th} row of $\boldsymbol{A}$ respectively. Both notations should be understood as column vectors if no transposition operation is specified.

Remark 2.1.1. *The third modeling technique discussed above can be applied in a variety of cases*

- *Minimization of absolute values,* $\min\{|\boldsymbol{c}^T\boldsymbol{x}|\} = \min\{\max\{\boldsymbol{c}^T\boldsymbol{x}, 0\}\}$.
- *Minimization a sum of maximums or maximization of a sum of minimums.*
- *We can even apply it to a usual objective function by realizing that* $\min\{\boldsymbol{c}^T\boldsymbol{x}\} = \min\max\{\boldsymbol{c}^T\boldsymbol{x}, -\infty\}$ *and* $\max\{c^T\boldsymbol{x}\}\max\min\{\boldsymbol{c}^T\boldsymbol{x}, \infty\}$. *In this way we effectively replace the original linear objective function by a linear constraint.*

Popular solution methods for LP include the Simplex algorithm and the various interior point algorithms. Although the theoretical bound of computational complexity is exponential in problem size for the Simplex algorithm and is polynomial in problem size for interior point algorithms, the Simplex algorithm is found to be more efficient than interior point algorithms in practice. Therefore Simplex algorithm is employed in this thesis to provide numerical demonstrations.

Despite the popularity and powerful modeling techniques of LP, in many practical applications we may want to optimize a ratio of two functions. For example, business efficiency ratio is defined as the ratio of expenses and revenue (with a few variations) and it is natural for management to minimize such ratio. As mentioned in Chapter 1, Sharpe ratio is an important measure of risk-reward trade off and maximization of Sharpe Ratio is a canonical portfolio selection problem. Fractional programming is then a natural formulation for such ratio optimization problems. In line with previous notations, the *standard inequality form* of a maximization fractional programming problem is given by

$$\begin{array}{lrcl} \text{maximize} & \frac{f(\boldsymbol{x})}{g(\boldsymbol{x})} & & \\ \text{subject to} & h_i(\boldsymbol{x}) & \leq & 0, \qquad i = 1, \cdots, m \end{array} \tag{2.2}$$

where $f(\cdot) : \mathbb{R}^m \mapsto \mathbb{R}$, $g(\cdot) : \mathbb{R}^m \mapsto \mathbb{R}$, and $h_i(\cdot) : \mathbb{R}^m \mapsto \mathbb{R}$ for $i = 1, \cdots, m$ are all functions of the vector of decision variables $\boldsymbol{x}$.

Without knowing the particular structural properties for $f(\cdot)$, $g(\cdot)$, and $h_i(\cdot), i = 1, \cdots, m$, it is difficult to derive optimality conditions and solution methods for Problem (2.2).

Definition 2.1.1. *A function $f(\cdot)$ is called convex on $\boldsymbol{D}$ if $\forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in \boldsymbol{D}$ and $\forall \lambda \in [0, 1]$, we have*

$$f(\lambda\boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2) \leq \lambda f(\boldsymbol{x}_1) + (1-\lambda)f(\boldsymbol{x}_2) \tag{2.3}$$

Furthermore, the function $f(\cdot)$ is strictly convex on D if $\forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in \boldsymbol{D}$, $\boldsymbol{x}_1 \neq \boldsymbol{x}_2$ and $\forall \lambda \in (0,1)$, the inequality (2.3) is strict.

Definition 2.1.2. *A function $f(\cdot)$ is called quasi-convex on $\boldsymbol{D}$ if $\forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in \boldsymbol{D}$ and $\forall t \in [0,1]$:*

$$f[\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2] \leq \max\{f(\boldsymbol{x}_1), f(\boldsymbol{x}_2)\} \tag{2.4}$$

Furthermore, the function $f(\cdot)$ is called explicit quasiconvex on $\boldsymbol{D}$ if $\forall x_1, x_2 \in \boldsymbol{D}$, $f(\boldsymbol{x}_1) \neq f(\boldsymbol{x}_2)$ and $\forall t \in (0,1)$, the inequality (2.4) is strict.

Corollary 2.1.1. *(Special case of Corollary 2.4.1 in Stancu-Minasian (1997))*

For n-vectors coefficients $\boldsymbol{c}$ and $\boldsymbol{d}$, n-vector variable $\boldsymbol{x}$ and constants c_0 and d_0. Let $\boldsymbol{D} \subseteq \mathbb{R}^n$ be a bounded convex set, the function $F : \boldsymbol{D} \mapsto \mathbb{R}$ defined as

$$F(\boldsymbol{x}) = \frac{\boldsymbol{c}^T \boldsymbol{x} + c_0}{\boldsymbol{d}^T \boldsymbol{x} + d_0}$$

is an explicit quasiconcave function provided that $\boldsymbol{d}'x + d_0 \neq 0$ and the sign of $\boldsymbol{d}'x + d_0$ remains unchanged $\forall \boldsymbol{x} \in \boldsymbol{D}$.

Theorem 2.1.1. *(Variation of Theorem 2.3.5 in Stancu-Minasian [58])*

If $F : \boldsymbol{D} \mapsto \mathbb{R}$ is explicit quasiconcave on the convex set $\boldsymbol{D}$, then any local maximum of function $F(\cdot)$ is a global maximum of $F(\cdot)$ on $\boldsymbol{D}$.

If $f(\cdot)$, $g(\cdot)$, and $h_i(\cdot), i = 1, \cdots, m$ are all linear functions of $\boldsymbol{x}$ in Problem (2.2), then Problem (2.2) reduces to a well-known class of programming problems called the *linear fractional programming (LFP)*. In line with previous notations, the *standard inequality form* of a maximization LFP problem is given by:

$$\begin{array}{rccc} \text{maximize} & \frac{f(x)}{g(x)} & = & \frac{\boldsymbol{c}^T \boldsymbol{x} + c_0}{\boldsymbol{d}^T \boldsymbol{x} + d_0} \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} & \leq & \boldsymbol{b} \end{array} \tag{2.5}$$

where $\boldsymbol{c} \in \mathbb{R}^n$ and $\boldsymbol{d} \in \mathbb{R}^n$ are vectors of known coefficients, c_0 and d_0 are scalars of known coefficients.

There are Simplex method and interior point algorithms for LFP similar to those for LP. However such algorithms are not usually available in commercial software because any LFP can be solved by solving at most two LP's through a simple variable transformation method, studied by Charnes and Cooper[14].

Consider the following variable transformation:

$$\boldsymbol{y} = t\boldsymbol{x} \tag{2.6}$$

where $t \geq 0$ is a scalar such that:

$$t = \frac{1}{d^T x + d_0} \Leftrightarrow d^T y + d_0 t = 1 \tag{2.7}$$

Multiplying the numerator and the denominator of the objective as well as the constraints of (2.5) by t, and using the relation (2.7), we have:

$$\begin{array}{rrcrcl} \text{maximize} & \boldsymbol{c}^T\boldsymbol{y} & + & c_0 t & & \\ \text{subject to} & \boldsymbol{A}\boldsymbol{y} & - & \boldsymbol{b}t & \leq & \boldsymbol{0} \\ & \boldsymbol{d}^T\boldsymbol{y} & + & d_0 t & = & 1 \\ & & & t & \geq & 0 \end{array} \tag{2.8}$$

Denote the feasible set for Problem (2.8) by $\boldsymbol{D}^* = \{(\boldsymbol{y}^T, t)^T \in \mathbb{R}^n \times \mathbb{R} | \boldsymbol{A}\boldsymbol{y} - \boldsymbol{b}t \leq 0, \boldsymbol{d}^T\boldsymbol{y} + d_0 t = 1, t \geq 0\}$. Recall that $\boldsymbol{D} = \{\boldsymbol{x} \in \mathbb{R}^n | \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}\}$ is the feasible set for Problem 2.5 which is assumed to be a bounded set.

Lemma 2.1.1. *For any feasible solution $(\boldsymbol{y}^T, t)^T \in \boldsymbol{D}^*$ of the problem (2.8), we have $t > 0$*

Proof. Assume that $(\hat{\boldsymbol{y}}^T, 0)^T \in \boldsymbol{D}^*$ and let $\hat{\boldsymbol{x}} \in \boldsymbol{D}$. Since $\boldsymbol{A}\hat{\boldsymbol{y}} \leq \boldsymbol{0}$ it follows that $\boldsymbol{x}_r = \hat{\boldsymbol{x}} + r\hat{\boldsymbol{y}} \in D$. Yet r can be as large as required, which implies that $\boldsymbol{D}$ is unbounded and leads to a contradiction. □

Theorem 2.1.2. *(Variation of Theorem 3.4.1 in Stancu-Minasian [58])*

If $\boldsymbol{d}^T\boldsymbol{x}^* + d_0 > 0$ where $\boldsymbol{x}^*$ is the optimum solution for (2.5) and $(\boldsymbol{y}^{*T}, t^*)^T$ is an optimum solution for (2.8), then $x = \frac{\boldsymbol{y}^*}{t^*}$ is an optimum solution for (2.5).

Proof. Since $(\boldsymbol{y}^{*T}, t^*)^T$ is an optimum solution for (2.8), we have:

$$\boldsymbol{c}^T\boldsymbol{y}^* + c_0 t^* \geq \boldsymbol{c}^T\boldsymbol{y} + c_0 t, \qquad \forall (\boldsymbol{y}^T, t)^T \in D^* \tag{2.9}$$

Assume that $\boldsymbol{x} = \frac{\boldsymbol{y}^*}{t^*}$ were not an optimum solution for (2.5). Hence there is another point $x^0 \in D$ which is optimum, i.e.,

$$\frac{\boldsymbol{c}^T\boldsymbol{x}^0 + c_0}{\boldsymbol{d}^T\boldsymbol{x}^0 + d_0} > \frac{\boldsymbol{c}^T(\frac{\boldsymbol{y}^*}{t^*}) + c_0}{\boldsymbol{d}^T(\frac{\boldsymbol{y}^*}{t^*}) + d_0}$$

But

$$\frac{\boldsymbol{c}^T(\frac{\boldsymbol{y}^*}{t^*}) + c_0}{\boldsymbol{d}^T(\frac{\boldsymbol{y}^*}{t^*}) + d_0} = \frac{\boldsymbol{c}^T\boldsymbol{y}^* + c_0 t^*}{\boldsymbol{d}^T\boldsymbol{y}^* + d_0 t^*} = \boldsymbol{c}^T\boldsymbol{y}^* + c_0 t^*$$

since

$$\boldsymbol{d}^T\boldsymbol{y}^* + d_0 t^* = 1, \qquad (\boldsymbol{y}^{*T}, t^*)^T \in D^*$$

Hence

$$\frac{\boldsymbol{c}^T\boldsymbol{x}^0 + c_0}{\boldsymbol{d}^T\boldsymbol{x}^0 + d_0} > \boldsymbol{c}^T\boldsymbol{y}^* + c_0 t^* \tag{2.10}$$

Since $\boldsymbol{x}^0$ is an optimum solution of (2.5) and by assumption $\boldsymbol{d}^T\boldsymbol{x}^0 + d_0 > 0$. Let $\boldsymbol{d}^T\boldsymbol{x}^0 + d_0 = \epsilon$ and let $\hat{\boldsymbol{y}} = \frac{\boldsymbol{x}^0}{\epsilon}$, $\hat{t} = \frac{1}{\epsilon}$. The point $(\hat{\boldsymbol{y}}^T, \hat{t})^T$ is a feasible solution for (2.8) since

$$\begin{aligned} \boldsymbol{A}\hat{\boldsymbol{y}} = \frac{\boldsymbol{A}\boldsymbol{x}^0}{\epsilon} &\leq \frac{b}{\epsilon} = b\hat{t} \\ \boldsymbol{d}^T\hat{\boldsymbol{y}} + d_0\hat{t} = \boldsymbol{d}^T\frac{\boldsymbol{x}^0}{\epsilon} + d_0\frac{1}{\epsilon} &= \frac{\boldsymbol{d}^T\boldsymbol{x}^0 + d_0}{\epsilon} = 1 \\ \hat{t} &\geq 0 \end{aligned}$$

We have

$$\frac{\boldsymbol{c}^T\boldsymbol{x}^0 + c_0}{\boldsymbol{d}^T\boldsymbol{x}^0 + d_0} = \frac{\frac{\boldsymbol{c}^T\boldsymbol{x}^0}{\epsilon} + \frac{c_0}{\epsilon}}{\frac{\boldsymbol{d}^T\boldsymbol{x}^0}{\epsilon} + \frac{d_0}{\epsilon}} = \frac{\boldsymbol{c}^T\hat{\boldsymbol{y}} + c_0\hat{t}}{\boldsymbol{d}^T\hat{\boldsymbol{y}} + d_0\hat{t}} = \boldsymbol{c}^T\hat{\boldsymbol{y}} + c_0\hat{t} \tag{2.11}$$

(2.10) and (2.11) imply

$$\boldsymbol{c}^T\hat{\boldsymbol{y}} + c_0\hat{t} > \boldsymbol{c}^T\boldsymbol{y}^* + c_0 t^* \tag{2.12}$$

which contradicts (2.9). The proof is complete. □

Note that Theorem 2.1.2 was proved under the assumption that $\boldsymbol{d}^T\boldsymbol{x}^* + d_0 > 0$ where $\boldsymbol{x}^*$ is the optimum solution for (2.5). If $\boldsymbol{d}^T\boldsymbol{x}^* + d_0 < 0$, we can replace $(\boldsymbol{c}^T, c_0)$ and $(\boldsymbol{d}^T, d_0)$ by $(-\boldsymbol{c}^T, -c_0)$ and $(-\boldsymbol{d}^T, -d_0)$, respectively. The value of the objective function will not change and we get $-\boldsymbol{d}^T\boldsymbol{x}^* - d_0 > 0$. Thus the results of Theorem 2.1.2 remain valid in this case too. As a result, we can state the following result:

Theorem 2.1.3. *(Variation of Theorem 3.4.2 in Stancu-Minasian [58])*

If D is a bounded convex set, then for solving (2.5) *it is sufficient to solve the following two LP's:*

$$\max\{\boldsymbol{c}^T\boldsymbol{y} + c_0t | \boldsymbol{A}\boldsymbol{y} - \boldsymbol{b}t \leq 0, \boldsymbol{d}^T y + d_0 t = 1, t \geq 0\} \tag{2.13}$$

$$\max\{-\boldsymbol{c}^T\boldsymbol{y} - c_0t | \boldsymbol{A}\boldsymbol{y} - \boldsymbol{b}t \leq 0, -\boldsymbol{d}^T y - d_0 t = 1, t \geq 0\} \tag{2.14}$$

If the sign of the denominator at the optimum solution is not known, then both problems, (2.13) and (2.14), must be solved. The solution which gives the largest value is selected. But if the sign of the denominator at the optimum solution is known, then according to Theorem 2.1.2, it is sufficient to solve only one of the problems (2.13) or (2.14), depending on the sign of the denominator. Discussions for the case where denominator is zero at optimum solution is omitted here because a risk measure of zero rarely occur in real life applications. Readers are encouraged to refer to Stancu-Minasian [58] for more details.

2.2 Karush-Kuhn-Tucker Optimality Conditions

In theory of nonlinear programming, the Karush-Kuhn-Tucker (KKT) optimality conditions (also known as the Kuhn-Tucker or KKT conditions[2]) are necessary conditions for a solution to be optimal, provided that some regularity conditions are satisfied[3]. Unless

[2]The KKT conditions were originally named after Harold W. Kuhn, and Albert W. Tucker, who first published the conditions in Kuhn and Tucker [38]. It was discovered later that the necessary conditions for this problem had been stated by William Karush in Karush [34].

[3]Examples of regularity conditions include linear independence constraint qualification, constant rank constraint qualification, etc. The required set of regularity conditions varies depending on the structure of programming problem of interest. Discussions on regularity conditions can be found in Bazaraa et al. [7] and Peterson [47].

specified otherwise, we assume throughout this thesis that all regularity conditions are satisfied.

Consider a general nonlinear programming(NLP) minimization problem in *standard inequality form* given by

$$\begin{array}{lrcll} \text{minimize} & f(\boldsymbol{x}) & & & \\ \text{subject to} & h_i(x) & \leq & 0, & i = 1, \cdots, m \end{array} \tag{2.15}$$

Problem (2.15) may have several local minima and the necessary conditions for optimality can be obtained by using the *Lagrange multiplier method.* The Lagrangian function for (2.15) is given by:

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{u}) = f(x) + \sum_{i=1}^{m} u_i h_i(x), \tag{2.16}$$

where $\boldsymbol{u}^T = (u_1, \cdots, u_m)$ is called the Lagrange dual multiplier vector.

Theorem 2.2.1. *(Karush-Kuhn-Tucker Necessary Optimality Conditions)*

If x^* is a regular[4]*, local minimizer to* (2.15)*, then there exists a Lagrange multiplier vector u^* such that*[5]

$$-\nabla f(\boldsymbol{x}^*) = \sum_{i=1}^{m} u_i^* \nabla h_i(\boldsymbol{x}^*) \tag{2.17}$$

$$h_i(\boldsymbol{x}^*) \leq 0, \quad i = 1, \cdots, m \tag{2.18}$$

$$u_i^* \geq 0, \quad i = 1, \cdots, m \tag{2.19}$$

$$u_i^* h_i(\boldsymbol{x}^*) = 0, \quad i = 1, \cdots, m \tag{2.20}$$

Condition (2.17) is referred to as the *Lagrangian stationarity condition* because it is derived from setting the gradient of the Lagrangian function (2.16) equal to zero. The Lagrangian Stationarity Condition gives a set n equations and is a necessary condition for $\boldsymbol{x}^*$ to be a stationary point (could be a minimum, maximum, or saddle point). Condition (2.18) is called the *primal feasibility condition* which restricts $\boldsymbol{x}^*$ to be a feasible solution for problem (2.15). This set of condition is merely a replication of the constraints of (2.15).

[4] x^* satisfies all regularity conditions.

[5] ∇ denotes the gradient operator

Condition (2.19) is called the *dual feasibility condition*. For every nonlinear programming problem, there exists a corresponding Lagrangian dual problem. Any feasible solution of a Lagrangian dual problem produces an objective value that is greater than or equal to that of its primal problem. The dual feasibility condition guarantees the feasibility of u^* for the Lagrangian dual of (2.15). Last but not least, condition (2.20) is commonly referred to as the *complementary slackness condition*. This condition provides linkage between (2.15) to its Lagrangian dual problem. The complementary slackness condition is of critical importance in establishing optimality of x^* and in developing solution methods for various programming problem.

The necessary conditions are sufficient for optimality if the objective function f is continuously differentiable concave function and the inequality constraints h_i are continuously differentiable convex functions. In the special case of LP and LFP, KKT conditions are both necessary and sufficient conditions for optimality. Moreover, KKT conditions can be reduced to a simpler set of conditions for LP and are referred to as duality. Nevertheless, we adhere to KKT conditions as optimality in order to maintain consistencies when comparing optimality conditions for LP and LFP. The necessity and sufficiency of KKT conditions for LP and LFP are both important in establishing equivalences among different formulations of portfolio selection problems.

2.3 CVaR Minimization Short-cut

The linear representation of CVaR developed in Rockafellar and Uryasev [50] is of fundamental importance for the CDRM optimization framework studied in this thesis. In line with their work, we review the definition of CVaR and present several theoretical results that are important to discussions in this thesis. Krokhmal et al. [36] extended the applicability of such linear minimization of CVaR by considering portfolio selection problems with CVaR constraints. Rockafellar and Uryasev [51] considered portfolio selection problems with multiple CVaR constraints, termed the *portfolio risk-shaping* with CVaR. In this section we review their results and will adhere as much as possible to the notations therein.

Let $\boldsymbol{l}(\boldsymbol{x}, \boldsymbol{y})$ be the loss associated with the decision vector $\boldsymbol{x}$, to be chosen from a set $\boldsymbol{D} \subseteq \mathbb{R}^n$, and the random vector $\boldsymbol{y} \in \mathbb{R}^m$. The vector $\boldsymbol{x}$ represents what we may generally call a portfolio, with $\boldsymbol{D}$ expressing the set of all feasible portfolios subject to certain portfolio constraints. $\boldsymbol{y}$ is a probability mass vector, i.e., y_i denotes the probability for realizing the i^{th} scenario. For every $\boldsymbol{x}$, the loss $\boldsymbol{l}(\boldsymbol{x}, \boldsymbol{y})$ is a random variable having a distribution in $\mathbb{R}$ induced by the distribution of $\boldsymbol{y}$. Unless indicated otherwise, the underlying probability distribution of $\boldsymbol{y}$ in $\mathbb{R}^m$ is assumed to be discrete uniform distribution, i.e. the probability of realizing the loss $l_i = l(\boldsymbol{x}, y_i)$ is $p_i = \frac{1}{m}, i = 1, \cdots, m$. This is not a very limiting assumption if we restrict ourselves to discrete portfolio loss distributions, which is always the case if we are obtaining distributional information via scenario generation or from historical data samples. In addition, given any arbitrary discrete distribution representable with rational numbers, we may always convert it to discrete uniform distribution for some large enough m.

For every portfolio $\boldsymbol{x}$, we denote by $\Psi(x, \cdot)$ the cumulative distribution function for the portfolio loss $\boldsymbol{l} = \boldsymbol{l}(\boldsymbol{x}, \boldsymbol{y})$, i.e.,

$$\Psi(\boldsymbol{x}, \zeta) = \sum_{i=1}^{m} p_i \mathbf{1}_{\{l_i \leq \zeta\}} \tag{2.21}$$

Definition 2.3.1. *Definitions of α-VaR and α-CVaR for scenario models(Proposition 8 in Rockafellar et al. [51])*

Suppose for each $\boldsymbol{x} \in \boldsymbol{D}$ the distribution of the portfolio loss $\boldsymbol{l} = \boldsymbol{l}(\boldsymbol{x}, \boldsymbol{y})$ is concentrated in $m < \infty$ points, and $\Psi(\boldsymbol{x}, \cdot)$ is a step function with jumps at those points. Fixing $\boldsymbol{x}$, let those corresponding portfolio loss points be ordered as $l_{(1)} \leq l_{(2)} \leq \cdots \leq l_{(m)}$. Denote the probability of realizing the loss $l_{(i)}$ by $p_{(i)} > 0$ for $i = 1, \cdots, m$.

Let i_α be the unique index such that

$$\sum_{i=1}^{i_\alpha} p_{(i)} \geq \alpha > \sum_{i=1}^{i_\alpha - 1} p_{(i)} \tag{2.22}$$

The α-VaR of the portfolio loss is given by

$$\zeta_\alpha(\boldsymbol{x}) = l_{(i_\alpha)}, \tag{2.23}$$

and the α-CVaR of the portfolio loss is given by

$$\phi_\alpha(\boldsymbol{x}) = \frac{1}{1-\alpha}[(\sum_{i=1}^{i_\alpha} p_{(i)} - \alpha) l_{i_\alpha} + \sum_{i=i_\alpha+1}^{m} p_{(i)} l_{(i)}] \tag{2.24}$$

The key to the CVaR linear minimization formulation is a characterization of $\phi_\alpha(\boldsymbol{x})$ and $\zeta_\alpha(\boldsymbol{x})$ in terms of a special function $F_\alpha(\boldsymbol{x}, \zeta)$ given by

$$F_\alpha(\boldsymbol{x}, \zeta) = \zeta + \frac{1}{1-\alpha} E[(\boldsymbol{l}(\boldsymbol{x}, \boldsymbol{y}) - \zeta)^+] = \zeta + \frac{1}{1-\alpha} \sum_{i=1}^{m} p_i (l_i - \zeta)^+ \tag{2.25}$$

Theorem 2.3.1. *Convex formulation for CVaR minimization problems (Theorem 14 in Rockafellar et al. [51])*

Minimizing $\phi_\alpha(\boldsymbol{x})$ *with respect to* $\boldsymbol{x} \in \boldsymbol{D}$ *is equivalent to minimizing* $F_\alpha(\boldsymbol{x}, \zeta)$ *over all* $(\boldsymbol{x}^T, \zeta)^T \in \boldsymbol{D} \times \mathbb{R}$*, in the sense that*

$$\min_{\boldsymbol{x}\in\boldsymbol{D}} \phi_\alpha(\boldsymbol{x}) = \min_{(\boldsymbol{x}^T,\zeta)^T\in\boldsymbol{D}\times\mathbb{R}} F_\alpha(\boldsymbol{x}, \zeta) \tag{2.26}$$

where moreover

$$(\boldsymbol{x}^{*T}, \zeta^*)^T \in \arg\min_{(\boldsymbol{x}^T,\zeta)^T\in\boldsymbol{D}\times\mathbb{R}} F_\alpha(\boldsymbol{x}, \zeta) \Longleftrightarrow \boldsymbol{x}^* \in \arg\min_{\boldsymbol{x}\in\boldsymbol{D}} \phi_\alpha(\boldsymbol{x}), \zeta^* \in \arg\min_{\zeta\in\mathbb{R}} F_\alpha(\boldsymbol{x}^*, \zeta) \tag{2.27}$$

Theorem 2.3.1 provides a convex representation of CVaR as the optimal objective value of a convex programming problem. Furthermore, such convex programming problem calculates VaR and CVaR simultaneously. Such convex representation of CVaR can be further simplified as linear representation via LP modeling techniques we have discussed in Section 2.1. With such linear representation we can cast any portfolio selection problem with CVaR objective and linear constraints as an LP.

Moreover, the following result consider applying the linear representation of CVaR as a constraint in portfolio selection problems. Such conversion can be shown in various ways. For example, one can apply the third LP modeling technique in Section 2.1 to convert a CVaR objective into a CVaR constraint. One can also derive the same result by using KKT conditions.

Theorem 2.3.2. *Portfolio optimization with CVaR constraint (Theorem 4 in Krokhmal et al. [36])*

For any portfolio selection problem with a continuously differentiable convex objective function $f(\cdot) : \mathbb{R}^n \mapsto \mathbb{R}$*, the two minimization problems below*

$$\min\{f(\boldsymbol{x}) | \boldsymbol{x} \in \boldsymbol{D}, \phi_\alpha(\boldsymbol{x}) \leq \eta\} \tag{2.28}$$

and

$$\min\{f(\boldsymbol{x}) | (\boldsymbol{x}, \zeta) \in \boldsymbol{D} \times \mathbb{R}, F_\alpha(\boldsymbol{x}, \zeta) \leq \eta\} \tag{2.29}$$

are equivalent in the sense that their objectives achieve the same minimum values. Moreover $(\boldsymbol{x}^{*T}, \zeta^*)^T$ *solves problem* (2.29) *if and only if* $\boldsymbol{x}^*$ *solves problem* (2.28)*.*

Rockafellar and Uryasev [51] considered a further extension to Theorem 2.3.2 in which multiple CVaR constraints are incorporated into one problem. Handling several CVaR at different probability thresholds helps portfolio manager in shaping the risk profile of their portfolio losses, hence the term *portfolio risk-shaping* is given to such formulation.

Theorem 2.3.3. *Risk shaping with CVaR (Theorem 16 in Rockafellar et al. [51])*

For any selection of probability thresholds α_i *and loss tolerances* η_i, $i = 1, \cdots, m$, *the problem*

$$\begin{array}{llcll} \textit{minimize} & f(\boldsymbol{x}) & & & \\ \textit{subject to} & \phi_{\alpha_i}(\boldsymbol{x}) & \leq & \eta_i & i = 1, \cdots, m \\ & \boldsymbol{x} & \in & \boldsymbol{D} & \end{array} \tag{2.30}$$

is equivalent to the problem

$$\begin{array}{llcll} \textit{minimize} & f(\boldsymbol{x}) & & & \\ \textit{subject to} & F_{\alpha_i}(\boldsymbol{x}, \zeta_i) & \leq & \eta_i & i = 1, \cdots, m \\ & (\boldsymbol{x}^T, \zeta_1, \cdots, \zeta_m)^T & \in & \boldsymbol{D} \times \mathbb{R} \times \cdots \times \mathbb{R} & \end{array} \tag{2.31}$$

in the sense that their objectives achieve the same minimum values.

Moreover $(\boldsymbol{x}^{*T}, \zeta_1^*, \cdots, \zeta_m^*)^T$ *solves problem* (2.31) *if and only if* $\boldsymbol{x}^*$ *solves problem* (2.30) *and the inequality* $F_{\alpha_i}(\boldsymbol{x}^*, \zeta_i^*) \leq \eta_i$ *holds for* $i = 1, \cdots, m$.

Moreover one then has $\phi_{\alpha_i}(\boldsymbol{x}^*) \leq \eta_i$ *for* $i = 1, \cdots, m$. *And* $\phi_{\alpha_i}(\boldsymbol{x}^*) = \eta_i$ *for every* i *such that* $F_{\alpha_i}(\boldsymbol{x}^*, \zeta_i^*) = \eta_i$.

If the loss $\boldsymbol{l}(\boldsymbol{x}, \boldsymbol{y})$ is a linear function of $\boldsymbol{x}$, then the linearization scheme in Krokhmal et al. [36] can be employed to convert any CVaR objective/constraint into linear functions of $\boldsymbol{x}$. A brief summary of their results is given below

- Replace a CVaR objective function by a linear objective and $2m$ additional linear constraints:

$$\min\{\phi_\alpha(\boldsymbol{x}) | \boldsymbol{x} \in \boldsymbol{D}\}$$
$$\Updownarrow$$
$$\min\{F_\alpha(\boldsymbol{x}, \zeta) | \boldsymbol{x} \in \boldsymbol{D}\}$$
$$\Updownarrow$$

$$
\begin{array}{llll}
\text{minimize} & \zeta + \frac{1}{1-\alpha}\sum\limits_{j=1}^{m} p_j z_j & & \\
\text{subject to} & z_j \ \geq \ l_j(\boldsymbol{x}, \boldsymbol{y}) - \zeta & j = 1, \cdots, m & \\
& z_i \ \geq \ 0 & j = 1, \cdots, m & \\
& \boldsymbol{x} \ \in \ \boldsymbol{D} & &
\end{array} \tag{2.32}
$$

- Replace a CVaR constraint by $2m+1$ linear constraints:

$$
\min\{f(\boldsymbol{x}) | \phi_\alpha(\boldsymbol{x}) \leq \eta, \boldsymbol{x} \in \boldsymbol{D}\}
$$
$$
\Updownarrow
$$
$$
\min\{f(x) | F_\alpha(\boldsymbol{x}, \zeta) \leq \eta, \boldsymbol{x} \in \boldsymbol{D}\}
$$
$$
\Updownarrow
$$

$$
\begin{array}{llll}
\text{minimize} & f(\boldsymbol{x}) & & \\
\text{subject to} & \zeta + \frac{1}{1-\alpha}\sum\limits_{j=1}^{m} p_j z_j \ \leq \ \eta & & \\
& z_j \ \geq \ l_j(\boldsymbol{x}, \boldsymbol{y}) - \zeta & j = 1, \cdots, m & \\
& z_i \ \geq \ 0 & j = 1, \cdots, m & \\
& \boldsymbol{x} \ \in \ \boldsymbol{D} & &
\end{array} \tag{2.33}
$$

- Replace k CVaR constraints by $(2m+1)k$ linear constraints:

$$
\min\{f(\boldsymbol{x}) | \phi_{\alpha_i}(\boldsymbol{x}) \leq \eta_i, i = 1, \cdots, k, \boldsymbol{x} \in \boldsymbol{D}\}
$$
$$
\Updownarrow
$$
$$
\min\{f(\boldsymbol{x}) | F_{\alpha_i}(\boldsymbol{x}, \zeta_i) \leq \eta_i, i = 1, \cdots, k, \boldsymbol{x} \in \boldsymbol{D}\}
$$
$$
\Updownarrow
$$

$$
\begin{array}{llll}
\text{minimize} & f(\boldsymbol{x}) & & \\
\text{subject to} & \zeta_i + \frac{1}{1-\alpha}\sum\limits_{j=1}^{m} z_{ij} \ \leq \ \eta_i & i = 1, \cdots, k & \\
& z_{ij} \ \geq \ l_j(\boldsymbol{x}, \boldsymbol{y}) - \zeta_i & i = 1, \cdots, k; j = 1, \cdots, m & \\
& z_{ij} \ \geq \ 0 & i = 1, \cdots, k; j = 1, \cdots, m & \\
& \boldsymbol{x} \ \in \ \boldsymbol{D} & &
\end{array} \tag{2.34}
$$

Last but not least, in cases where $f(\boldsymbol{x})$ is a linear function or a fraction of two linear functions of $\boldsymbol{x}$, Problem (2.31) reduces to an LP or an LFP respectively. Discussions and results in Chapter 4 rely heavily on the linear CVaR formulation presented in this section.

Chapter 3

Coherent Distortion Risk Measures

Quantifying the risk or the uncertainty in the future value of a portfolio is one of the most important tasks in risk managements. This quantification is usually achieved by modeling the uncertain payoff as a random variable, to which then a certain functional is applied. Such functionals are usually called risk measures. A risk measure attempts to assign a single numerical value to a random potential financial loss. Applications of such risk measures include margin requirements in financial trading, insurance risk premiums, and government regulatory deposit requirements for banking regulation. Conventionally, portfolios' return distributions are considered in finance and investment applications while portfolios' loss distributions are concerned in insurance applications. To avoid ambiguity, this thesis considers a portfolio's loss distribution, or equivalently the negative of return distributions in quantifying its underlying riskiness.

3.1 Basic Risk Measures

The uncertainty for future value of an investment position is usually described by a function $X : \Omega \mapsto \mathbb{R}$, where Ω is a fixed set of scenarios with a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{X}$ be a linear space of random variables on Ω, i.e., a set of functions $X : \Omega \mapsto \mathbb{R}$. It is assumed that X is bounded. In particular, $\mathcal{X} \subseteq L^{\infty}(\Omega, \mathcal{F}, \mathbb{P})$[1]. For introduction, X can be thought of as a loss from an uncertain position. For $X, Y \in \mathcal{X}$, we denote the state-wise dominance by $X \geq Y$, i.e., $X \geq Y \Leftrightarrow X(\omega) \geq Y(\omega)$ for all $\omega \in \Omega$.

[1]When $|\Omega|$ is finite and supported by finite elements, this is automatically satisfied

Definition 3.1.1. *A mapping* $\rho : \mathcal{X} \mapsto \mathbb{R}$ *is called a* risk measure *if it satisfies, for all* $X, Y \in \mathcal{X}$*:*

A1 *Monotonicity: If* $X \geq Y$*, then* $\rho(X) \geq \rho(Y)$*.*

A2 *Translation invariance: If* $c \in \mathbb{R}$*, then* $\rho(X + c) = \rho(X) + c$*.*

The financial meaning of monotonicity is clear: The risk of a portfolio is at least as much as another one if the former incurs at least as much losses as the latter in very state of economy. Translation invariance is motivated by the interpretation of $\rho(X)$ as a reserve requirement, i.e., $\rho(X)$ is the amount which should be raised in order to make X acceptable from the point of view of a supervising agency. Thus, if there is a constant loss added to all future state of economy, then the reserve requirement is increased by the same amount.

In accord to this definition, expectations, variance and higher moments, partial moments, quantile measures such as VaR, and conditional expectations such as CVaR are all risk measures. Our task is to look for particular classes of risk measures that have more desirable properties.

Definition 3.1.2. *A risk measure* ρ *is called a* convex risk measure *if it satisfies:*

A3 *Convexity:* $\rho(\lambda X + (1 - \lambda)Y) \leq \lambda\rho(X) + (1 - \lambda)\rho(Y), \qquad \lambda \in [0, 1]$

Consider two investment strategies that lead to two loss random variables, X and Y. If one takes only a fraction λ of loss X and $1 - \lambda$ of loss Y, one obtains a loss random variable $\lambda X + (1 - \lambda)Y$. The axiom of convexity gives a precise meaning to the idea that diversification should not increase the risk. This is a desirable property for both economic reasons (convex preferences) and computational reasons (ensuring that optimization such over risk measures induces convex optimization problems). The idea of risk diversification becomes even clearer when we note that convexity implies a weaker requirement of quasiconvexity:

$$\rho(\lambda X + (1 - \lambda)Y) \leq \max\{\rho(X), \rho(Y)\}, \qquad \lambda \in [0, 1]$$

Note that quantile risk measures such as VaR is not a convex risk measure. Consider the example shown in Table 3.1. We see that $0.5VaR_{0.95}(X) + 0.5VaR_{0.95}(Y) = 0 < 50 = VaR_{0.95}(0.5X + 0.5Y)$ which violates the convexity axiom. Therefore VaR is not a member of convex risk measure in general. The lack of convexity for VaR has been a strong critique because it is counterintuitive and it is computationally difficult to deal with.

Scenarios	Probability	Loss Profile X	Loss Profile Y	Loss Profile X+Y
1	0.0384	100	0	100
2	0.0384	0	100	100
3	0.0016	100	100	200
4	0.9216	0	0	0
$VaR_{0.95}$		0	0	100

Table 3.1: Example for showing VaR is not convex

3.2 Coherent Risk Measure

Artzner et al. [3] defines the class of *Coherent Risk Measures(CRM)* in terms of four axioms: monotonicity, subadditivity, positive homogeneity, and translation invariance. With the definition of convex risk measure, we can reach an alternative and equivalent[2] definition of CRM:

Definition 3.2.1. *A convex risk measure ρ is called a coherent risk measure if it satisfies:*

A4 *Positive homogeneity: If $\lambda \geq 0$, then $\rho(\lambda X) = \lambda\rho(X)$*

The positive homogeneity axiom states that risk scales linearly with the size of a position.

The following is a representation theorem for CRM. In essence, it states that we can describe any coherent risk measure equivalently in terms of expectations over a family of distributions or a family of "generalized" scenarios.

Theorem 3.2.1. *(Representation Theorem of Coherent Risk Measures) A risk measure ρ is coherent if and only if there exists a family of probability measures $\mathcal{Q}$ on $(\Omega, \mathcal{F})$ with $\mathbb{Q} \ll \mathbb{P}\ \forall \mathbb{Q} \in \mathcal{Q}$ [3] such that*

$$\rho(X) = \sup_{\mathbb{Q}\in\mathcal{Q}} E^{\mathbb{Q}}[X], \qquad \forall X \in \mathcal{X} \tag{3.1}$$

where $E^{\mathbb{Q}}[X]$ denotes the expectation of the random variable X under the measure $\mathbb{Q}$ (as opposed to the measure of X itself).

[2] Under the axiom of positive homogeneity, convexity is equivalent to subadditivity.

[3] $\ll$ denotes absolute continuity, i.e. $\mathbb{Q} \ll \mathbb{P} \Leftrightarrow \mathbb{Q}(A) = 0$ s.t. $\mathbb{P}(A) = 0$ for $A \in \mathcal{F}$

Note that this representation theorem actually predates the introduction of CRM by Artner et al. [3]. See Chapter 10 of Huber [28] for one version of the proof.

The representation theorem says that all coherent risk measures may be represented as the worst-case expectation over a family of "generalized scenarios". For example, the generating family for $CVaR_\alpha(X)$ is $\mathcal{Q}_\alpha = \{\mathbb{Q} \ll \mathbb{P} | \frac{d\mathbb{Q}}{d\mathbb{P}} \leq \frac{1}{1-\alpha}\}$[4].

Given a vector of data observations on portfolio losses $\boldsymbol{l} = (l_1, \cdots, l_m)^T$, coherent risk measure of this set of observations can be obtained based on the representation theorem of CRM. A risk measure is a coherent risk measure if and only if there exists a set of weights $\mathcal{Q} = \{\boldsymbol{q} = (q_1, \cdots, q_m)^T | q_i \geq 0, i = 1, \cdots, m; \sum_{i=1}^m q_i = 1\}$ such that

$$\rho(\boldsymbol{l}) = \sup_{\boldsymbol{q} \in \mathcal{Q}} \{\sum_{i=1}^m q_i l_i\}, \qquad \forall \boldsymbol{l} \in \mathbb{R}^m \tag{3.2}$$

Note that if historical data is used to calculate the CRM for the future, we have implicitly made the assumption that the future losses will have the same distributions as the past's.

Although linear optimization framework for CVaR discussed in Chapter 2 provides connections for CRM and portfolio selection through the particular case of CVaR. Yet we cannot make further claims without imposing more structural properties for CRM.

3.3 Distortion Risk Measure

Another well known class of risk measures is the class of *Distortion Risk Measures(DRM)*, proposed by Wang et al. [69] in the context of calculating insurance risk premiums. This set of axioms also underlies the definition of DRM.

Definition 3.3.1. *A mapping $\rho : \mathcal{X} \mapsto \mathbb{R}$ is called a* distortion risk measure *if it satisfies, for all $X, Y \in \mathcal{X}$*

B1 Conditional state independence: $\rho(X) = \rho(Y)$ if X and Y have the same distribution. This means that the risk of a position is determined only by the loss distribution.

B2 Monotonicity: $\rho(X) \leq \rho(Y)$ if $X \leq Y$.

[4] $\frac{d\mathbb{Q}}{d\mathbb{P}}$ denotes the Radon-Nikondym derivative of $\mathbb{Q}$ with respect to $\mathbb{P}$

B3 Comonotonic additivity: $\rho(X+Y) = \rho(X) + \rho(Y)$ *if* X *and* Y *are comonotonic, where random variables* X *and* Y *are comonotonic if and only*

$$(X(\omega_1) - X(\omega_2))(Y(\omega_1) - Y(\omega_2)) \geq 0 \qquad a.s. \text{ for } \omega_1, \omega_2 \in \Omega$$

B4 Continuity:

$$\begin{aligned}\lim_{d\to 0} \rho((X-d)^+) &= \rho(X^+), \\ \lim_{d\to\infty} \rho(\min\{X, d\}) &= \rho(X), \\ \lim_{d\to-\infty} \rho(\max\{X, d\}) &= \rho(X).\end{aligned}$$

where $(X-d)^+ = \max(X-d, 0)$

If two random variable X and Y are comonotonic, then $X(\omega)$ and $Y(\omega)$ always move in the same direction as the state ω changes. The notion of comonotonicity is central in risk measures. See discussions on comonotonicity in Dhaene et al. [21] and Dhaene et al. [20]. Wang et al. [69] imposed axiom B3 based on the argument that the comonotonic random variables do not hedge against each other, leading to additivity of risks. They also proved(Theorem 3 in Wang et al. [69]) that if $\mathcal{X}$ contains all the *Bernoulli*(p) random variables, $0 \leq p \leq 1$, then risk measure ρ satisfies axioms B1-B4 and $\rho(1) = 1$ if and only if ρ has a Choquet integral representation with respect to a distorted probability:

$$\rho_g(X) = \int X d(g \circ \mathbb{P}) = \int_{-\infty}^{0} [g(\mathbb{P}(X > x)) - 1] dx + \int_{0}^{\infty} g(\mathbb{P}(X > x)) dx \qquad (3.3)$$

where $g(\cdot)$ is called the distortion function which is nondecreasing with $g(0) = 0$ and $g(1) = 1$, and $g \circ \mathbb{P}(A) := g(\mathbb{P}(A))$ is called the distorted probability.

Definition 3.3.2. *A risk measure* $\rho : \mathcal{X} \mapsto \mathbb{R}$ *that satisfies* $\rho(X) = \rho(Y)$ *for all* $X, Y \in \mathcal{X}$ *such that* X *and* Y *have the same distribution under* $\mathbb{P}$ *is called a law invariant risk measure.*

All the standard results about Choquet integrals apply to distortion risk measures:

1. $\rho_g(X) \geq 0$ if $X \geq 0$.

2. $\rho_g(X+c) = \rho_g(X) + c$, $\forall c \in \mathbb{R}$.

3. $\rho_g(\lambda X) = \lambda \rho_g(X)$, $\forall \lambda \geq 0$.

4. $\rho_g(-X) = -\rho_{g^*}(X)$, where $g^*(x) = 1 - g(1-x)$ is the dual distortion of g.

5. $\rho_g(X+Y) = \rho(X) + \rho(Y)$ if X and Y are comonotone.

6. If a random variable X_n has a finite number of values, X_n converges to X, i.e., $X_n \overset{W}{\to} X$, and $\rho_g(X)$ exists, then $\rho_g(X_n) \overset{W}{\to} \rho(X)$. This property implies that it is enough to prove the statement for the discrete random variables, and then carry over the result to the general continuous case.

Furthermore, by construction a distortion risk measure depends only on the distribution of the random variable X hence it is law invariant.

The Choquet integral representation of DRM is used to explore its mathematical properties, calculations of DRMs can be easily done by taking the expected value of X under probability measure $\mathbb{P}^* := g \circ \mathbb{P}$[5].

Some well-known distortion functions g in the literature include:

1. $CVaR_\alpha$ distortion: $g_{CVaR}(x, \alpha) = \min\{\frac{x}{1-\alpha}, 1\}$. This was first observed by Wirch and Hardy [71] .

2. Wang Transform(WT) distortion: $g_{WT}(x, \beta) = \Phi[\Phi^{-1}(x) - \Phi^{-1}(\beta)]$. This was introduced by Wang [67]

3. Proportional hazard(PH) distortion: $g_{PH}(x, \gamma) = x^\gamma$ with $\gamma \in (0, 1]$. This was proposed by Wang et al. [69]

4. Lookback(LB) distortion: $g_{LB}(x, \delta) = x^\delta(1 - \delta \ln x)$ with $\delta \in (0, 1]$. This was proposed by Hürlimann [29]

For discretely distributed portfolio losses random variable $\boldsymbol{l} = (l_1, \cdots, l_m)^T$ and its probability masses $Pr[\boldsymbol{l} = l_i] = p_i$ for $i = 1, \cdots, m$, we can obtain the cumulative distribution function $F_{\boldsymbol{l}}(l) = \sum_{i=1}^{m} p_i \mathbf{1}_{\{l_i \leq l\}}$. Then discrete survival function is given by

[5]See Theorem 1 in Wang [66] and Definition 4.2 in Wang [67]

$\mathbb{P}(X > x) = S_{\boldsymbol{l}}(l) = 1 - F_{\boldsymbol{l}}(l)$ and is applied in the distorted probability representation of (3.3). We have

$$\rho_g(\boldsymbol{l}) = \int_{-\infty}^{0} [g(S_{\boldsymbol{l}}(l)) - 1]dl + \int_{0}^{\infty} g(S_{\boldsymbol{l}}(l))dl = E^*[\boldsymbol{l}] = \sum_{i=1}^{m} p^*_{(i)} l_{(i)} \tag{3.4}$$

where $p^*_{(1)} = 1 - g(S_{\boldsymbol{l}}(l_{(1)}))$ and $p^*_{(i)} = g(S_{\boldsymbol{l}}(l_{(i-1)})) - g(S_{\boldsymbol{l}}(l_{(i)}))$, $i = 2, \cdots, m$.

Since g is non-decreasing, $g(0) = 0$ and $g(1) = 1$, it follows that $p^*_i \geq 0$ for $i = 1, \cdots, m$, and $\sum_{i=1}^{m} p^*_i = 1 - g(S_{\boldsymbol{l}}(l_{(m)})) = 1$. In the special case where $\mathbb{P}(\boldsymbol{l} = l_i) = \frac{1}{m}$ for $i = 1, \cdots, m$, then the construction of q is given by $q_i = g(\frac{m-i+1}{m}) - g(\frac{m-i}{m})$, $i = 1, \cdots, m$. This special case is the emphasis of this thesis.

Similar with CRM, we need to impose more conditions on DRM hence its corresponding distortion functions in order to incorporate it into portfolio selection problems.

3.4 Coherent Distortion Risk Measure

From the previous discussions about CRM and DRM, it is of interest to explore the *Coherent Distortion Risk Measure(CDRM)* which is the intersection of these two classes of risk measures in order to develope CDRM portfolio selection problems.

One crucial property that a DRM lacks to be a CRM is subadditivity. Wirch and Hardy [72] proved that a DRM is subadditive, hence coherent if and only if its distortion function g is concave[6]. Kusuoka [40] showed that DRM is a particular case of law-invariant coherent risk measures.

There are two ways to derive CDRM: Bellini et al. [8] defined CDRM as a subclass of DRM, namely DRM with concave distortion function g; Bertsimas et al. [11] defined CDRM as a subclass of CRM, namely CRM that is also comonotonic and law invariant[7]. These two definitions are indeed equivalent since it is shown in Bellini et al. [8] that the class of coherent distortion risk measures coincides with the class of comonotonic law invariant coherent risk measures.

Definition 3.4.1. *We say ρ is a* coherent distortion risk measure *if:*

[6]Wang and Dhaene [68] proved sufficiency and Wirch and Hardy [72] finished the proof by showing necessity

[7]Definition 4.5 in Bertsimas et al. [11] should be defining CDRM as oppose to defining DRM.

- ρ_g *is a distortion risk measure with a concave distortion function* g*, or equivalently,*
- ρ *is a coherent risk measure that is also comonotonic and law-invariant.*

The following representation theorem for CDRM is the key result that enables us to develope a convex optimization framework for any CDRM portfolio selection problem.

Theorem 3.4.1. *Representation Theorem of Coherent Distortion Risk Measure (Variation of Lemma 4.1 in Bertsimas et al. [11])*

For any random variable X *and a given concave distortion function* g*, risk measure* ρ_g *is a coherent distortion risk measure if and only if there exists a function* $w : [0,1] \mapsto [0,1]$*, satisfying* $\int_{\alpha=0}^{1} w(\alpha)d\alpha = 1$*, such that:*

$$\rho_g(X) = \int_{\alpha=0}^{1} w(\alpha)\phi_\alpha(X)d\alpha \tag{3.5}$$

where $\phi_\alpha(X)$ is the α-CVaR of X.

This representation theorem says any CDRM can be represented as a convex combination of $CVaR_\alpha(X)$, $\alpha \in [0,1]$ and we can construct any CDRM based on a convex combination of $CVaR_\alpha(X)$, $\alpha \in [0,1]$. This result was proved by Kusuoka [40] for continuous portfolio loss distributions. Bertsimas et al. [11] proved this result for discrete portfolio loss distributions.

For portfolio losses random variable $\boldsymbol{l} = (l_1, \cdots, l_m)^T$ and the special case of discrete uniform probability masses $Pr[\boldsymbol{l} = l_i] = \frac{1}{m}$ for $i = 1, \cdots, m$, Bertsimas et al. [11] strengthened the representation theorem that any CDRM can be represented as a convex combination of *finite number* of $CVaR_\alpha(X)$s. The main results of their work are presented below.

Definition 3.4.2. *Define the restricted simplex in* m*-dimension as:*

$$\hat{\Delta}^m \equiv \{\boldsymbol{q} \in \mathbb{R}^m | \sum_{i=1}^{m} q_i = 1, q_1 \leq \cdots \leq q_m\}$$

Definition 3.4.3. *For a given loss observation* $\boldsymbol{l} = (l_1, \cdots, l_m)$ *and the corresponding ordered losses* $l_{(1)}, \cdots, l_{(m)}$*, define a CVaR-Matrix* $\mathbf{Q} \in \mathbb{R}^m \times \mathbb{R}^m$ *with columns* $\boldsymbol{Q}_{.i} \in \mathbb{R}^m$*,* $i = 1, \cdots, m$*:*

$$\mathbf{Q} = [\boldsymbol{Q}_{\cdot 1}, \boldsymbol{Q}_{\cdot 2}, \cdots, \boldsymbol{Q}_{\cdot m}] = \begin{bmatrix} \frac{1}{m} & 0 & 0 & \cdots & 0 \\ \frac{1}{m} & \frac{1}{m-1} & 0 & \cdots & 0 \\ \frac{1}{m} & \frac{1}{m-1} & \frac{1}{m-2} & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{m} & \frac{1}{m-1} & \frac{1}{m-2} & \cdots & 1 \end{bmatrix}$$

Observe that

$$\phi_{\frac{j-1}{m}}(\boldsymbol{l}) = \sum_{i=1}^{m} \frac{1}{m-j+1} l_{(i)} = \sum_{i=1}^{m} \boldsymbol{Q}_{ij} l_{(i)}, \qquad j = 1, \cdots, m \tag{3.6}$$

Therefore $\boldsymbol{Q}_{\cdot j}$ is the restricted simplex to for calculating $CVaR_{\frac{j-1}{m}}(\boldsymbol{l})$.

Theorem 3.4.2. *Finite Generation Theorem for Coherent Distortion Risk Measures (Theorem 4.2 in Bertsimas et al. [11])*

For a give portfolio loss sample $\boldsymbol{l} = (l_1, \cdots, l_m)$*, the corresponding ordered losses* $l_{(1)}, \cdots, l_{(m)}$ *and a given concave distortion function* g*, a risk measure* ρ_g *is a CDRM if and only if there exists a* $\boldsymbol{q} \in \hat{\Delta}^m$ *such that*

$$\rho_g(\boldsymbol{l}) = \sum_{i=1}^{m} q_i l_{(i)} \tag{3.7}$$

Moreover, every such $\boldsymbol{q} \in \hat{\Delta}^m$ *can be written in the form*

$$\boldsymbol{q} = \sum_{j=1}^{m} w_j \boldsymbol{Q}_{\cdot j} \tag{3.8}$$

where $w_i \geq 0$*,* $i = 1, \cdots, m$*,* $\sum_{i=1}^{m} w_i = 1$*, and* $\boldsymbol{Q}_{\cdot j}$ *is the restricted simplex for calculating* $CVaR_{\frac{j-1}{m}}(\boldsymbol{l})$*.*

Moreover, the convex weights are given by

$$\begin{cases} w_1 &= mq_1 \\ w_i &= (m-i+1)(q_i - q_{i-1}) \quad i = 2, \cdots, m \end{cases} \tag{3.9}$$

By (3.7) and (3.8) and by the construction of $\mathbf{Q}$, any CDRM for a give loss sample $\boldsymbol{l}$ can be expressed as a convex combination of $|\boldsymbol{l}|$ CVaRs. In particular, we have

$$\rho_g(\boldsymbol{l}) = \sum_{i=1}^{m} q_i l_{(i)} = \sum_{i=1}^{m}(\sum_{j=1}^{m} w_j \boldsymbol{Q}_{ij}) l_{(i)} = \sum_{j=1}^{m} w_j (\sum_{i=1}^{m} \boldsymbol{Q}_{ij} l_{(i)}) = \sum_{j=1}^{m} w_j \phi_{\frac{j-1}{m}}(\boldsymbol{l}) \tag{3.10}$$

where $w_i \geq 0$, $i = 1, \cdots, m$, $\sum_{i=1}^{m} w_i = 1$.

A complete proof of the finite generation theorem can be found in Bertsimas et al. [11]. We provide here a heuristic proof and verifications of their results.

It is well known that $CVaR$ is a CDRM and CDRM is closed under convex combinations, hence $\rho_g(\boldsymbol{l})$ is a CDRM because it can be written as convex combination of CVaRs. For the other direction, since CDRM is a subclass of DRM, we see from (3.4) that $q_i = g(\frac{m-i+1}{m}) - g(\frac{m-i}{m})$ for $i = 1, \cdots, m$. Let $w_1 = mq_1$, $w_i = (m-i+1)(q_i - q_{i-1})$, $i = 2, \cdots, m$, we can verify the following conditions on $\boldsymbol{q}$ and $\boldsymbol{w}$:

1. Since g is nondecreasing, $q_i \geq 0$, $i = 1, \cdots, m$.
2. Since $g(0) = 0$ and $g(1) = 1$, $\sum_{i=1}^{m} q_i = \sum_{i=1}^{m} g(\frac{m-i+1}{m}) - g(\frac{m-i}{m}) = g(1) - g(0) = 1$.
3. Since g is concave, $g(0.5\frac{m-i+1}{m} + 0.5\frac{m-i-1}{m}) \geq 0.5g(\frac{m-i+1}{m}) + 0.5g(\frac{m-i-1}{m})$. We have

$$\begin{array}{rcll} g(\frac{m-i}{m}) - g(\frac{m-i-1}{m}) & \geq & g(\frac{m-i+1}{m}) - g(\frac{m-i}{m}), & i = 2, \cdots, m \\ q_i & \geq & q_{i-1}, & i = 2, \cdots, m \end{array}$$

   Since $q_1 \geq 0$ and $q_i \geq q_{i-1}$, $i = 2, \cdots, m$, $w_i \geq 0$, $i = 1, \cdots, m$ we have $\sum_{i=1}^{m} w_i = \sum_{i=1}^{m} q_i = 1$.

Definition 3.4.4. *Definition of WT-measure (Definition 4.3 in Wang [67]) For a loss random variable X with cumulative distribution function $F_X(x)$, we define the WT risk measure with parameter β as:*

1. *For a preselected confidence level β, let $\varphi = \Phi^{-1}(\beta)$.*
2. *Apply the Wang Transform: $F^* = \Phi[\Phi^{-1}(F(x)) - \varphi]$.*
3. *The WT-measure is the expected value under F^*.*

$$WT_\beta(X) = E^*[X]$$

WT-measure is a well-known member of coherent distortion risk measure. We will illustrate by numerical examples how WT-measure is expressed as a convex combination of CVaRs.

Consider a hypothetical portfolio with the following discrete uniform loss distribution.

Loss x	PF $f(x)$	CDF $F(x)$	Distorted CDF $F^*(x)$	Weight $\boldsymbol{w}$
1	0.2	0.2	0.006450784	0.03225392
2	0.2	0.4	0.028834822	0.06373302
3	0.2	0.6	0.082035941	0.09245124
4	0.2	0.8	0.210920213	0.15136630
5	0.2	1	1.000000000	0.66019552

Table 3.2: Calculation of $\boldsymbol{w}$ for uniformly distributed losses (WT-measure with $\beta = 0.95$)

On one hand, the expected value under the distorted distribution is given by

$$WT_{0.95} = E^*[X] = 4.671758$$

On the other hand, the convex weight vector is calculated as Equation (3.9) and is given in the last column of Table 3.2. observe that $CVaR$ at confidence level 0, 0.2, 0.4, 0.6 and 0.8 are 3, 3.5, 4, 4.5 and 5 respectively. Then the convex combination of CVaRs weighted by $\boldsymbol{w}$ is given by

$$\sum_{i=1}^{5} w_i \phi_{\frac{i-1}{5}}(X) = 4.671758 = WT_{0.95}$$

Consider another hypothetical portfolio whose losses are not uniformly distributed.

We can convert such a loss distribution into a discrete uniform distribution by enlarging the sample space, i.e., artificially split the mass point(s) that is(are) “too heavy”.

The second and third columns of Table 3.4 are not theoretically rigorous yet they only serve as intermediate steps to calculate our weighting factors $\boldsymbol{w}$.

On one hand, the expected value under distorted distribution is given by

$$WT_{0.95} = E^*[X] = 3.700593$$

Loss x	PF $f(x)$	CDF $F(x)$
1	0.2	0.2
2	0.4	0.6
3	0.2	0.8
4	0.2	1

Table 3.3: Non-uniformly distributed portfolio losses.

Loss x	PF $f(x)$	CDF $F(x)$	Distorted CDF $F^*(x)$	Weight w
1	0.2	0.2	0.006450784	0.03225392
2	0.2	0.4	0.028834822	0.06373302
2^*	0.2	0.6	0.082035941	0.09245124
3	0.2	0.8	0.210920213	0.15136630
4	0.2	1	1.000000000	0.66019552

Table 3.4: Conversion into uniform loss distribution and calculation of $\boldsymbol{w}$ (WT-measure with $\beta = 0.95$).

On the other hand, the convex combination of CVaRs, weighted by $\boldsymbol{w}$, is given by

$$\sum_{i=1}^{5} w_i \phi_{\frac{i-1}{5}}(X) = 3.700593 = WT_{0.95}$$

The above two examples show how to calculate the weighting factors for CVaRs in order to express WT-measure as a convex combination of CVaRs at different confidence levels. Such a calculation is valid for any CDRM in general. These two examples also show how to convert a non-uniform discrete loss distribution into discrete uniform distributions provided that the original probability masses are representable by rational numbers. Moreover, we see that the weighting factors are independent of actual portfolio loss amounts, only the number of scenarios and the distortion function matter. This independence is useful in formulating CDRM portfolio selection problems into programming problems. Once these two pieces of information is known, any optimization problem over CDRM reduces to a optimization problem over convex combination of CVaRs. Taking advantage of the linear optimization framework for CVaR, any portfolio selection problem over CDRM can be similarly formulated as an LP and can be solved efficiently.

Chapter 4

Formulations and Equivalences

Portfolio selection problems can be formulated in different ways. One can consider minimizing CVaR while requiring a minimum level of expected return. Alternatively, one can maximize returns while limiting the maximum level risks. Last but not least, one can maximize a utility function expressed as a linear combination of return and risk. Krokhmal et al.[36] considered these three equivalent formulations of optimization problems under some conditions on risk and return measures. Their equivalent results can be applied to our portfolio selection problems over CDRM. In addition, we show that a CDRM-based Sharpe ratio maximization problem is also equivalent to the aforementioned three formulations for CDRM portfolio selection problems.

4.1 CDRM in Convex Programming

Inspired by the convex representation $F_\alpha(\boldsymbol{x}, \zeta)$ for $\phi_\alpha(\boldsymbol{x})$ and the Representation Theorem of Coherent Distortion Risk Measure(Theorem 3.4.1), we consider the following special function:

$$M_g(\boldsymbol{x}, \boldsymbol{\zeta}) = \int_{\alpha=0}^{1} w(\alpha) F_\alpha(\boldsymbol{x}, \zeta_\alpha) d\alpha \tag{4.1}$$

where $w(\alpha) \geq 0$ and $\int_{\alpha=0}^{1} w(\alpha) d\alpha = 1$.

The convex weights $w(\alpha)$, $\alpha \in [0, 1]$ depend on both the original portfolio loss distribution and the choice of distortion function. The Representation Theorem of Coherent

Distortion Risk Measure(Theorem 3.4.1) ensures the existence of such weights and defines $CDRM$ for a given set of weights.

For each $\alpha \in [0,1]$ there is a corresponding auxiliary variable ζ_α. Taking partial derivatives with respect to all ζ_α for $\alpha \in [0,1]$ and setting them equal to zeros give the extremal properties of $M_g(\boldsymbol{x}, \boldsymbol{\zeta})$ and can provide more insights about the connection between a particular CDRM, $\rho_g(\boldsymbol{x})$, and its convex representation $M_g(\boldsymbol{x}, \boldsymbol{\zeta})$. Yet $\boldsymbol{\zeta}$ may have infinite many entries ζ_α. Taking partial derivative with respect to all ζ_α for $\alpha \in [0,1]$ requires calculus of variations, which is outside the scope of this thesis. We solve this problem by applying properties of Choquet integrals because CDRM is a subclass of DRM.

Theorem 4.1.1. *Convex formulation for CDRM minimization problems*

Minimizing $\rho_g(\boldsymbol{x})$ with respect to $\boldsymbol{x} \in \boldsymbol{D}$ is equivalent to minimizing $M_g(\boldsymbol{x}, \boldsymbol{\zeta})$ over all $(\boldsymbol{x}^T, \boldsymbol{\zeta}^T) \in D \times \mathbb{R}^{|\zeta|}$, in the sense that

$$\min_{\boldsymbol{x} \in \boldsymbol{D}} \rho_g(\boldsymbol{x}) = \min_{(\boldsymbol{x}^T, \boldsymbol{\zeta}^T)^T \in \boldsymbol{D} \times \mathbb{R}^{|\zeta|}} M_g(\boldsymbol{x}, \boldsymbol{\zeta}) \tag{4.2}$$

where moreover

$$(\boldsymbol{x}^{*T}, \boldsymbol{\zeta}^{*T})^T \in \arg\min_{(\boldsymbol{x}^T, \boldsymbol{\zeta}^T)^T \in \boldsymbol{D} \times \mathbb{R}} M_g(\boldsymbol{x}, \boldsymbol{\zeta}) \Longleftrightarrow \boldsymbol{x}^* \in \arg\min_{\boldsymbol{x} \in \boldsymbol{D}} \rho_g(\boldsymbol{x}), \boldsymbol{\zeta}^* \in \arg\min_{\boldsymbol{\zeta} \in \mathbb{R}} M_g(\boldsymbol{x}^*, \boldsymbol{\zeta}) \tag{4.3}$$

Proof. Since $CDRM$ is a subclass of DRM, all results of DRM and of Choquet integrals can be applied. The sixth result of Choquet integral presented in Section 3.3 implies that it is enough to prove the statement for the discrete portfolio loss random variables, and then carry over the result to the continuous case.

Consider a discrete portfolio loss random variable $\boldsymbol{l} = l(\boldsymbol{x}, \boldsymbol{y}) = (l_1, \cdots, l_m)^T$ induced by the choice of portfolio $\boldsymbol{x} \in \mathbb{R}^n$ and the random vector $\boldsymbol{y} \in \mathbb{R}^m$. Denote the probability of realizing $l_i = l(\boldsymbol{x}, y_i)$ by $p_i > 0$, $i = 1, \cdots, m$ and denote the corresponding cumulative distribution function and survival function by $F_{\boldsymbol{l}}(l)$ and $S_{\boldsymbol{l}}(l)$ respectively. Let the portfolio loss points be ordered as $l_{(1)} \leq l_{(2)} \leq \cdots \leq l_{(m)}$. Denote the probability of realizing the loss $l_{(i)}$ by $p_{(i)} > 0$ for $i = 1, \cdots, m$. Given a nondecreasing concave distortion function $g : [0,1] \mapsto [0,1]$ such that $g(0) = 0$ and $g(1) = 1$, the corresponding $CDRM$, $\rho_g(\boldsymbol{x})$, is given by

$$\rho(\boldsymbol{x}) = \sum_{i=1}^{m} q_i l_{(i)}$$

where $q_1 = g(1) - g(1 - p_(1))$ and $q_i = g(S_{\boldsymbol{l}}(l_{(i-1)})) - g(S_{\boldsymbol{l}}(l_{(i)}))$, $i = 2, \cdots, m$.

Since portfolio losses are discretely distributed at m points, there are m jumps in $F_{\boldsymbol{l}}(l)$. Denote these probability levels at these m mass points by:

$$\alpha_i = \begin{cases} 0 & \text{for } i = 1 \\ \sum_{j=1}^{i-1} p_{(j)} & \text{for } i = 2, \cdots, m \end{cases} \tag{4.4}$$

The m CVaRs at these probability levels are given by

$$\phi_{\alpha_i}(\boldsymbol{x}) = \frac{1}{1 - \alpha_i} \sum_{j=i}^{m} p_{(j)} l_{(j)}, \qquad i = 1, \cdots, m \tag{4.5}$$

Consider the weights w_i, $i = 1, \cdots, m$ as follows:

$$w_i = \begin{cases} \frac{q_1}{p_{(1)}} & \text{if} \quad i = 1 \\ (q_i - \frac{p_{(i)}}{p_{(i-1)}} q_{i-1}) \frac{S_{\boldsymbol{l}}(l_{(i-1)})}{p_{(i)}} & \text{if} \quad i = 2, \cdots, m \end{cases} \tag{4.6}$$

One can verify that[1] $w_i \geq 0$ for $i = 1, \cdots, m$ and $\sum_{i=1}^{m} w_i = 1$. Furthermore, we have

$$\rho(\boldsymbol{x}) = \sum_{i=1}^{m} w_i \phi_{\alpha_i}(\boldsymbol{x})$$

where $\phi_{\alpha_i}(\boldsymbol{x})$, $i = 1, \cdots, m$ are defined in Equation (2.24).

Consider the following special function

$$M_g(\boldsymbol{x}, \boldsymbol{\zeta}) = \sum_{i=1}^{m} w_i F_{\alpha_i}(\boldsymbol{x}, \zeta_{\alpha_i})$$

where $F_{\alpha_i}(\boldsymbol{x}, \zeta_{\alpha_i})$ and α_i, $i = 1, \cdots, m$ are defined in Equation (2.25) and Equation (4.4) respectively.

Since $F_{\alpha_i}(\boldsymbol{x}, \zeta_{\alpha_i})$, $i = 1, \cdots, m$ are all joint convex functions of $\boldsymbol{x}$ and ζ_{α_i} and $M_g(\boldsymbol{x}, \boldsymbol{\zeta})$ is a convex combination of $F_{\alpha_i}(\boldsymbol{x}, \zeta_{\alpha_i})$ for $i = 1, \cdots, m$, then $M_g(\boldsymbol{x}, \boldsymbol{\zeta})$ is a joint convex function of $\boldsymbol{x}$ and $\boldsymbol{\zeta}$.

[1] See proof in Appendix A.1

For a given portfolio $\boldsymbol{x}$, we want to find $\boldsymbol{\zeta}^*$ that minimizes $M(\boldsymbol{x},\boldsymbol{\zeta})$. Since $M_g(x,\zeta)$ is a convex function of $\boldsymbol{\zeta}$, we can simply set the gradient of $M_g(\boldsymbol{x},\boldsymbol{\zeta})$ with respect to ζ equal to zero. We have

$$\begin{array}{rcll}
0 & = & \frac{\partial M_g(\boldsymbol{x},\boldsymbol{\zeta})}{\partial \zeta} & \\
0 & = & \frac{\partial}{\partial \zeta_{\alpha_i}} w_j[\zeta_{\alpha_i} + \frac{1}{1-\alpha_i}\sum_{i=1}^{m} p_i(l_i-\zeta_{\alpha_i})^+], & i = 1,\cdots,m \\
0 & = & w_j[1-\frac{1}{1-\alpha_i}\sum_{i=1}^{m} p_i \mathbf{1}_{(l_i-\zeta_{\alpha_i})}], & i = 1,\cdots,m \\
 & \Leftrightarrow & \left\{ \begin{array}{lll} \zeta^*_{\alpha_i} \in [l_i, l_{i+1}) & \text{if} & w_i \neq 0 \\ \zeta^*_{\alpha_i} \text{ unconstrainted} & \text{if} & w_i = 0 \end{array} \right. &
\end{array}$$

Substituting these extremal conditions into $M(\boldsymbol{x},\boldsymbol{\zeta})$, we have

$$\begin{aligned}
\min_{\boldsymbol{\zeta}\in\mathbb{R}^m} M(\boldsymbol{x},\boldsymbol{\zeta}) &= \sum_{i=1}^{m} w_i[\zeta^*_{\alpha_i} + \frac{1}{1-\alpha_i}\sum_{j=1}^{m} p_j(l_j-\zeta^*_{\alpha_i})^+] \\
&= \sum_{i=1}^{m} w_i[\zeta^*_{\alpha_i} + \frac{1}{1-\alpha_i}\sum_{j=i}^{m} p_j(l_j-\zeta^*_{\alpha_i})^+] \\
&= \sum_{i=1}^{m} w_i[\zeta^*_{\alpha_i} + \frac{1}{1-\alpha_i}\sum_{j=i}^{m} p_{(j)}l_{(j)} - \frac{\sum_{j=i}^{m} p_{(j)}}{1-\alpha_i}\zeta^*_{\alpha_i}] \\
&= \sum_{i=1}^{m} w_i[\frac{1}{1-\alpha_i}\sum_{j=i}^{m} p_{(j)}l_{(j)}] \\
&= \sum_{i=1}^{m} w_i\phi_{\alpha_i}(\boldsymbol{x}) \\
&= \rho_g(\boldsymbol{x})
\end{aligned}$$

The minimum value of $M(\boldsymbol{x},\boldsymbol{\zeta})$ is precisely $\rho_g(\boldsymbol{x})$ and such result holds for any portfolio $\boldsymbol{x}$. Therefore the equivalences in Theorem 4.1.1 hold. □

Theorem 4.1.1 is an extension of the convex formulation of CVaR in Rockafellar et al. [50]. It is also an extension to the finite generation of CDRM in Bertsimas ea al.[11] by considering general discrete loss distributions as opposed to discrete uniform loss distribution only. Discrete uniform distribution for portfolio losses is a special case in the above convex formulation hence we can apply our result straightforwardly. Furthermore, it is often easier to prove statements using discrete uniform distribution and carry the results over to continuous distribution by taking limits of the sample size.

4.2 Return-CDRM Formulations and Equivalences

Based on the linear optimization framework for CVaR, we find two ways to develope a linear optimization framework for CDRM portfolio selection problems. Firstly, CVaR has a convex representation w.r.t portfolio weights and any CDRM can be expressed as a convex combination of CVaRs. Since convex combination of convex functions is also a convex function, one can formulate CDRM problems as convex programming problems. Secondly, we can consider the so-called risk-shaping problem with CVaRs and show that it is equivalent to a portfolio selection problem with CDRM and vice versa. The former is straightforward and it was shown partially in Section 4.1. Chapter 3 provides discussions on the convex representation of CDRM and the current chapter derives different formulations for CDRM portfolio selection problems based on risk-shaping problem over CVaR.

Theorem 4.2.1. *Consider the following programming problems with continuously differentiable convex objective function f and constraints h_i*

$$\begin{array}{llcll} \textit{minimize} & f(\boldsymbol{x}) & & & \\ \textit{s.t.} & F_{\alpha_i}(\boldsymbol{x},\zeta_i) & \leq & \eta_i, & i=1,\cdots,m \\ & h_i(\boldsymbol{x}) & \leq & 0 & i=m+1,\cdots,m+k \end{array} \tag{4.7}$$

and

$$\begin{array}{llcll} \textit{minimize} & f(\boldsymbol{x})+M_g(\boldsymbol{x},\boldsymbol{\zeta}) & & & \\ \textit{s.t.} & h_i(\boldsymbol{x}) & \leq & 0 & i=m+1,\cdots,m+k \end{array} \tag{4.8}$$

where $F_{\alpha_i}(\boldsymbol{x},\zeta_i)$ for $i=1,\cdots,m$ are defined in Equation (2.25). $M_g(\boldsymbol{x},\boldsymbol{\zeta})=\sum_{i=1}^{m} F_{\alpha_i}(\boldsymbol{x},\zeta_i)$ for some weight vector $\boldsymbol{w}=(w_1,\cdots,w_m)^T \geq \mathbf{0}$.

Let u_i^1, $i=1,\cdots,m+k$ be the KKT multipliers for constraints in Problem (4.7) and let u_i^2, $i=m+1,\cdots,m+k$ be the KKT multipliers for constraints in Problem (4.8), then these two problems are equivalent in the following sense:

- *If* $(\boldsymbol{x}^{*T}, \boldsymbol{\zeta}^{*T})^T$ *is an optimal solution to Problem* (4.7) *with KKT multipliers* u^{1*}*, then it is also an optimal solution for Problem* (4.8) *provided that* $w_i = \frac{u_i^{1*}}{\sum_{i=1}^{m} u_i^{1*}}$ *for* $i = 1, \cdots, m$*. If* $\sum_{i=1}^{m} u^{1*_i} = 0$*, then set* $w_i = 0$ *for* $i = 1, \cdots, m$*.*
- *If* $(\boldsymbol{x}^{*T}, \boldsymbol{\zeta}^{*T})^T$ *is an optimal solution to Problem* (4.8) *with KKT multipliers* u^{2*}*, then it is also an optimal solution for Problem* (4.7) *provided that* $\eta_i = F_{\alpha_i}(x^*, \zeta_i^*)$ *for* $i = 1, \cdots, m$*.*

Proof. Let u_i^1, $i = 1, \cdots, m+k$ be the KKT multipliers for constraints in Problem (4.7) and let u_i^2, $i = m+1, \cdots, m+k$ be the KKT multipliers for constraints in Problem (4.8). Moreover, let $\boldsymbol{u}_i^{1*}$ and $\boldsymbol{u}_i^{2*}$ be the vectors of KKT multipliers that correspond to the optimal solutions in their respective problems. Then the KKT optimality conditions for Problem (4.7) are given by

$$-\nabla f(\boldsymbol{x}^*) = \sum_{i=1}^{m} u_i^{1*} \nabla F_{\alpha_i}(\boldsymbol{x}^*, \zeta_i^*) + \sum_{i=m+1}^{m+k} u_i^{1*} \nabla h_i(\boldsymbol{x}^*) \tag{4.9a}$$

$$F_{\alpha_i}(\boldsymbol{x}^*, \zeta_i^*) \leq \eta_i, \qquad i = 1, \cdots, m \tag{4.9b}$$

$$h_i(\boldsymbol{x}^*) \leq 0, \qquad i = m+1, \cdots, m+k \tag{4.9c}$$

$$u_i^{1*} \geq 0, \qquad i = 1, \cdots, m+k \tag{4.9d}$$

$$u_i^{1*}[F_{\alpha_i}(\boldsymbol{x}^*, \zeta_i^*) - \eta_i] = 0, \qquad i = 1, \cdots, m \tag{4.9e}$$

$$u_i^{1*} h_i(x^*) = 0, \qquad i = m+1, \cdots, m+k \tag{4.9f}$$

and the KKT optimality conditions for *CDRM* Minimization Problem (4.8) are given by

$$-\nabla f(\boldsymbol{x}^*) - \sum_{i=1}^{m} w_i F_{\alpha_i}(\boldsymbol{x}^*, \zeta_i^*) = \sum_{i=m+1}^{m+k} u_i^{2*} \nabla h_i(\boldsymbol{x}^*) \tag{4.10a}$$

$$h_i(\boldsymbol{x}^*) \leq 0, \qquad i = m+1, \cdots, m+k \tag{4.10b}$$

$$u_i^{2*} \geq 0, \qquad i = m+1, \cdots, m+k \tag{4.10c}$$

$$u_i^{2*} h_i(\boldsymbol{x}^*) = 0, \qquad i = m+1, \cdots, m+k \tag{4.10d}$$

If $(\boldsymbol{x}^{*T}, \boldsymbol{\zeta}^{*T})^T$ is an optimal solution for Problem (4.7), then it satisfies conditions (4.9) by the necessity of KKT optimality conditions. Setting $w_i = \frac{u_i^{1*}}{\sum_{i=1}^{m} u_i^{1*}}$ or $w_i = 0$ if

$\sum_{i=1}^{m} u_i^{1*}$ for $i = 1, \cdots, m$ it is clear that conditions (4.10) are satisfied for $u_i^{2*} = u_i^{1*}$ for $i = m+1, \cdots, m+k$. By the sufficiency of KKT optimality conditions, $(\boldsymbol{x}^{*T}, \boldsymbol{\zeta}^{*T})^T$ is also an optimal solution for Problem (4.8).

If $(\boldsymbol{x}^{*T}, \boldsymbol{\zeta}^{*T})^T$ is an optimal solution for Problem (4.8), then it satisfies conditions (4.10) by the necessity of KKT optimality conditions. Setting $\eta_i = F_{\alpha_i}(\boldsymbol{x}^*, \zeta_i^*)$ then conditions (4.9) are satisfied with $u_i^{1*} = w_i$ for $i = 1, \cdots, m$ and $u_i^{1*} = u_i^{2*}$ for $i = m+1, \cdots, m+k$. By the sufficiency of KKT optimality conditions, $(\boldsymbol{x}^{*T}, \boldsymbol{\zeta}^{*T})^T$ is also an optimal solution for Problem (4.7). □

Theorem 4.2.1 is stated on a mathematical programming basis but we can make connections between CVaR risk shaping problems and *CDRM* optimization problems by using results in Rockafellar et al. [51].

Theorem 4.2.2. *Equivalence between CVaR Risk-Shaping Problems and CDRM Portfolio Selection Problems*

For any selection of probability thresholds α_i, $i = 1, \cdots, m$, consider a CVaR risk-shaping problem

$$\min_{\boldsymbol{x} \in \boldsymbol{D}} \{f(\boldsymbol{x}) | \phi_{\alpha_i}(\boldsymbol{x}) \leq \eta_i, i = 1, \cdots, m\} \tag{4.11}$$

Optimization with respect to a CDRM $\rho_g(\boldsymbol{x}) = \sum_{i=1}^{m} w_i \phi_{\alpha_i}(\boldsymbol{x})$

$$\min_{\boldsymbol{x} \in \boldsymbol{D}} \{f(\boldsymbol{x}) + \rho_g(\boldsymbol{x})\} \tag{4.12}$$

Suppose the feasible sets of both problems satisfy all constraint qualifications of KKT optimality conditions. Varying the parameters η_i and w_i, $i = 1, \cdots, m$ traces the multidimensional efficient frontiers for their respective problem. Moreover, varying these parameters for $\eta_i \in \mathbb{R}$ and $w_i \geq 0$, $i = 1, \cdots, m$ traces out the same efficient frontier.

Proof. By Theorem 2.3.3, Problem (4.11) can be formulated into a convex programming Problem (4.7). Similarly, by Theorem 4.1.1 Problem (4.12) can be formulated into a convex programming Problem (4.8).

Let $(\boldsymbol{x}^{*T}, \boldsymbol{\zeta}^{*T})^T$ be an optimal solution for Problem (4.7), by Theorem 2.3.3 one has $\phi_{\alpha_i}(\boldsymbol{x}^*) \leq \eta_i$ for every i and actually $\phi_{\alpha_i}(\boldsymbol{x}^*) = \eta_i$ for all i such that $F_{\alpha_i}(\boldsymbol{x}^*, \zeta_i^*) = \eta_i$. By

KKT conditions (4.9), we see that $u_i^{1*} = 0$ for all i such that $F_{\alpha_i}(\boldsymbol{x}^*, \zeta_i^*) < \eta_i$. If $u_i^{1*} > 0$ for some $i = 1, \cdots, m$, then based on Theorem 4.2.1 we know that $(\boldsymbol{x}^{*T}, \boldsymbol{\zeta}^{*T})^T$ is also an optimal solution for Problem (4.8) provided that $w_i = \frac{u_i^{1*}}{\sum_{i=1}^{m} u_i^{1*}}$ for $i = 1, \cdots, m$. If $u_i^{1*} = 0$ for all $i = 1, \cdots, m$, then based on Theorem 4.2.1 we know that $(\boldsymbol{x}^{*T}, \boldsymbol{\zeta}^{*T})^T$ is also an optimal solution for Problem (4.8) provided that $w_i = 0$ for $i = 1, \cdots, m$.

Conversely, let $(\boldsymbol{x}^{*T}, \boldsymbol{\zeta}^{*T})^T$ be an optimal solution for Problem (4.8), Theorem 4.2.2 implies that it is also an optimal solution to Problem (4.7) with optimal KKT multipliers $u^{1*} = w_i$ for $i = 1, \cdots, m$ and $u^{1*} = u^{2*}$ for $i = m+1, \cdots, m+k$.

Furthermore, we always have

$$
\begin{aligned}
\sum_{i=1}^{m} w_i F_{\alpha_i}(\boldsymbol{x}^*, \zeta_i^*) &= \sum_{\substack{i=1 \\ w_i > 0}}^{m} w_i F_{\alpha_i}(\boldsymbol{x}^*, \zeta_i^*) \\
&= \sum_{\substack{i=1 \\ u_i^{1*} > 0}}^{m} w_i F_{\alpha_i}(\boldsymbol{x}^*, \zeta_i^*) \\
&= \sum_{\substack{i=1 \\ F_{\alpha_i}(\boldsymbol{x}^*, \zeta_i^*) = \eta_i}}^{m} w_i F_{\alpha_i}(\boldsymbol{x}^*, \zeta_i^*) \\
&= \sum_{\substack{i=1 \\ F_{\alpha_i}(\boldsymbol{x}^*, \zeta_i^*) = \phi_{\alpha_i}(\boldsymbol{x}^*)}}^{m} w_i F_{\alpha_i}(\boldsymbol{x}^*, \zeta_i^*) \\
&= \sum_{\substack{i=1 \\ w_i > 0}}^{m} w_i \phi_{\alpha_i}(\boldsymbol{x}^*) \\
&= \sum_{i=1}^{m} w_i \phi_{\alpha_i}(\boldsymbol{x}^*) \\
&= \rho_g(\boldsymbol{x}^*)
\end{aligned}
$$

Hence any point on the efficient frontier of the $CVaR$ risk-shaping problem (4.11) corresponds to a point on the efficient frontier of the $CDRM$ optimization problem (4.12) and vice versa. Therefore, both problems generate the same efficient frontiers by varying their respective parameters. □

In the special case that $u_i^{1*} = 0$ for all $i = 1, \cdots, m$ for Problem (4.11) or equivalently $w_i = 0$ for all $i = 1, \cdots, m$ in Problem (4.12), both problems in fact reduce to

$$\min_{\boldsymbol{x} \in \boldsymbol{D}} \{f(\boldsymbol{x})\} \tag{4.13}$$

Theorem 4.2.2 provides new perspectives for CDRM and CVaR optimization problems. On one hand, although we make use of the special function $M_g(\boldsymbol{x}, \boldsymbol{\zeta})$ as a convex representation of $CDRM$ hence formulate Problem (4.8), we can view it as a CVaR risk-shaping problem. As a result, we are able to shape the portfolio loss distribution with only one single CDRM. On the other hand, portfolio selection problems with a single CVaR objective/constraint is a special case of CVaR risk-shaping problems and hence a special case of CDRM optimization problem.

Moreover, Theorem 4.2.2 offers an extra way to tackle CVaR risk-shaping problems. Portfolio managers now have the freedom to choose whether to trace the efficient frontier of Problem (4.7) or Problem (4.8) depending on which one can be solved more efficiently with their software.

If $f(\boldsymbol{x})$ is a loss measure (negate of a return measure), then Problem (4.12) is in fact a CDRM utility maximization problem and Problem (4.8) is its convex formulation. There are clearly more formulations for portfolio selection problems. In a general setting, Krokhmal et al. [36] established equivalences among three formulations of a portfolio selection with given risk and reward measures.

Theorem 4.2.3. *(Variation of Theorem 3 in Krokhmal[36])*

Consider a risk functional $\rho(\boldsymbol{x})$ and a reward functional $R(\boldsymbol{x})$ dependent on the portfolio weights $\boldsymbol{x}$ and the following three problems:

$$\max\{R(\boldsymbol{x}) - \tau\rho(\boldsymbol{x}) | \tau_1 \geq 0, \boldsymbol{x} \in \boldsymbol{D}\} \tag{4.14}$$

$$\min\{\rho(\boldsymbol{x}) | R(\boldsymbol{x}) \geq \mu, \boldsymbol{x} \in \boldsymbol{D}\} \tag{4.15}$$

$$\max\{R(\boldsymbol{x}) | \rho(\boldsymbol{x}) \leq \rho, \boldsymbol{x} \in \boldsymbol{D}\} \tag{4.16}$$

Suppose that constraints $R(\boldsymbol{x}) \geq \mu$, $\rho(\boldsymbol{x}) \leq \rho$ are regular constraints[2]*. Varying the parameters τ_1, μ, and ρ traces the efficient frontiers for the problems* (4.14)-(4.16)*, accordingly. If $\rho(x)$ is convex, $R(x)$ is concave and the set D is convex, then the three problems,* (4.14)-(4.16)*, generate the same efficient frontier.*

[2]These constraints satisfy constraint qualifications for KKT conditions.

The equivalence among problems (4.14)-(4.16) is well known for mean-variance efficient (see Steinbach [59]) and mean-regret efficient frontiers (see Dembo and Rosen [17]). For discrete loss distributions, we will further show equivalences among the above three formulations and the CDRM-based Sharpe ratio maximization.

Theorem 4.2.4. *Given a CDRM $\rho_g(\boldsymbol{x})$ that depends on the portfolio weights $\boldsymbol{x}$ and a reward function $R(\boldsymbol{x})$ that is linear function or can be converted into linear functions of $\boldsymbol{x}$, consider the following four problems:*

- *Return maximization subject to CDRM constraint*

$$\max\{R(\boldsymbol{x})|\boldsymbol{x} \in \boldsymbol{D}, \rho_g(\boldsymbol{x}) \leq \eta\} \tag{4.17}$$

- *CDRM minimization subject to return constraint*

$$\min\{\rho_g(\boldsymbol{x})|\boldsymbol{x} \in \boldsymbol{D}, R(\boldsymbol{x}) \geq \mu\} \tag{4.18}$$

- *Return-CDRM utility maximization for given risk aversion parameter $\tau \geq 0$*

$$\max\{R(\boldsymbol{x}) - \tau\rho_g(\boldsymbol{x})|\boldsymbol{x} \in \boldsymbol{D}\} \tag{4.19}$$

- *CDRM-based Sharpe Ratio maximization for given benchmark return level ν*

$$\max\{\frac{R(\boldsymbol{x}) - \nu}{\rho_g(\boldsymbol{x})}|\boldsymbol{x} \in \boldsymbol{D}\} \tag{4.20}$$

Assume $\rho_g(\boldsymbol{x}) > 0$ for all $\boldsymbol{x} \in D$ and $\nu < R(\boldsymbol{x})$ for some $\boldsymbol{x} \in \boldsymbol{D}$. Suppose that constraints $\rho_g(\boldsymbol{x}) \leq \eta$, $R(\boldsymbol{x}) \geq \mu$ are regular constraints. Varying the parameters η, μ, τ, and ν traces the efficient frontiers for the problems (4.17), (4.18), (4.19), *and* (4.20) *respectively. Moreover, these four problems generate the same efficient frontier.*

In addition, suppose $\boldsymbol{x}^*$ is an optimal solution for one of the problems (4.17), (4.18), (4.19), *or* (4.20) *with a preselected parameter. Let u^{3*}_{m+1} and u^{4*}_{m+1} be the KKT multipliers for the return and risk constraints in problems* (4.17) *and* (4.18) *at the optimal solutions, then $\boldsymbol{x}^*$ is also an optimal solution for the other three problems provided that the parameters in those problems equal to the implied values given in Table 4.1.*

Problem	(4.17)	(4.18)	(4.19)	(4.20)
Preselected Parameter	η	μ	τ	ν
Implied Parameters				
For (4.17), $\eta =$	N/A	$\rho_g(x^*)$	$\rho_g(x^*)$	$\rho_g(x^*)$
For (4.18), $\mu =$	$R(x^*)$	N/A	$R(x^*)$	$R(x^*)$
For (4.19), $\tau =$	u^{3*}_{k+1}	$\tau = \frac{1}{u^{4*}_{k+1}}$	N/A	$\frac{R(x^*)-\nu}{\rho_g(x^*)}$
For (4.20), $\nu =$	$R(x^*) - u^{3*}_{k+1}\rho_g(x^*)$	$R(x^*) - \frac{1}{u^{4*}_{k+1}}\rho_g(x^*)$	$R(x^*) - \tau\rho_g(x^*)$	N/A

Table 4.1: Relationships between Preselected Parameter of One Formulation and Implied Parameters of the Other Formulations

Proof. Since $\phi_{\alpha_i}(\boldsymbol{x})$ for any given probability level α_i, $i = 1, \cdots, m$ are convex with respect to $\boldsymbol{x}$ and $\rho_g(\boldsymbol{x})$ is a convex combination of $\phi_{\alpha_i}(\boldsymbol{x})$, $i = 1, \cdots, m$, then $\rho_g(\boldsymbol{x})$ is convex with respect to $\boldsymbol{x}$. Moreover, we have assumed that the return function is linear or representable by linear functions of $\boldsymbol{x}$ hence is concave. Therefore problems (4.17), (4.18)and (4.19) satisfy the assumptions in Theorem 4.2.3 hence their equivalences follow straightforwardly. It remains to show the equivalences among CDRM-based Sharpe Ratio maximization problem (4.20) and the other three problems.

Let $\boldsymbol{D} = \{\boldsymbol{x} | h_i(\boldsymbol{x}) \leq 0, i = 1, \cdots, m\}$ be the set of all feasible portfolios except risk and return constraints. Let $\boldsymbol{u}^{3*}$, $\boldsymbol{u}^{4*}$, $\boldsymbol{u}^{5*}$, and $\boldsymbol{u}^{6*}$ be the KKT multipliers for problems (4.17)-(4.20) at their respective optimal solutions. Note that $\boldsymbol{u}^{3*}$ and $\boldsymbol{u}^{4*}$ have $m+1$ entries while $\boldsymbol{u}^{5*}$, and $\boldsymbol{u}^{6*}$ have m entries. Since $R(\boldsymbol{x})$ is representable by linear functions of $\boldsymbol{x}$ and $M_g(\boldsymbol{x}, \boldsymbol{\zeta})$ can be used for $\rho_g(\boldsymbol{x})$ and can be linearized, problems (4.17)-(4.20) are all convertible into LPs and LFPs for which KKT conditions are both necessary and sufficient conditions for optimality.

- KKT optimality conditions for Problem (4.17)

$$\nabla R(\boldsymbol{x}^*) = \sum_{i=1}^{m} u_i^{3*} \nabla h_i(\boldsymbol{x}^*) + u_{m+1}^{3*} \nabla \rho_g(\boldsymbol{x}^*) \tag{4.21a}$$

$$h_i(\boldsymbol{x}^*) \leq 0, \qquad i = 1, \cdots, m \tag{4.21b}$$

$$\rho_g(\boldsymbol{x}^*) \leq \eta \tag{4.21c}$$

$$u_i^{3*} \geq 0, \qquad i = 1, \cdots, m+1 \tag{4.21d}$$

$$u_i^{3*} h_i(\boldsymbol{x}^*) = 0, \qquad i = 1, \cdots, m \tag{4.21e}$$

$$u_{m+1}^{3*}[\rho_g(\boldsymbol{x}^*) - \eta] = 0 \tag{4.21f}$$

- KKT conditions for Problem (4.18)

$$-\nabla \rho_g(\boldsymbol{x}^*) = \sum_{i=1}^{m} u_i^{4*} \nabla h_i(\boldsymbol{x}^*) - u_{m+1}^{4*} \nabla R(\boldsymbol{x}^*) \tag{4.22a}$$

$$h_i(\boldsymbol{x}^*) \leq 0, \qquad i = 1, \cdots, m \tag{4.22b}$$

$$R(\boldsymbol{x}^*) \geq \mu \tag{4.22c}$$

$$u_i^{4*} \geq 0, \qquad i = 1, \cdots, m+1 \tag{4.22d}$$

$$u_i^{4*} h_i(\boldsymbol{x}^*) = 0, \qquad i = 1, \cdots, m \tag{4.22e}$$

$$u_{m+1}^{4*}[R(\boldsymbol{x}^*) - \mu] = 0 \tag{4.22f}$$

- KKT conditions for Problem (4.19)

$$\nabla R(\boldsymbol{x}^*) - \tau \nabla \rho_g(\boldsymbol{x}^*) = \sum_{i=1}^{m} u_i^{5*} \nabla h_i(\boldsymbol{x}^*) \tag{4.23a}$$

$$h_i(\boldsymbol{x}^*) \leq 0, \qquad i = 1, \cdots, m \tag{4.23b}$$

$$u_i^{5*} \geq 0, \qquad i = 1, \cdots, m \tag{4.23c}$$

$$u_i^{5*} h_i(\boldsymbol{x}^*) = 0, \qquad i = 1, \cdots, m \tag{4.23d}$$

(4.23e)

- KKT conditions for Problem (4.20)

$$\nabla R(\boldsymbol{x}^*)\frac{1}{\rho_g(\boldsymbol{x}^*)} - \nabla\rho_g(\boldsymbol{x}^*)\frac{R(\boldsymbol{x}^*)-\nu}{[\rho_g(\boldsymbol{x}^*)]^2} = \sum_{i=1}^{m} u_i^{6*}\nabla h_i(\boldsymbol{x}^*) \tag{4.24a}$$

$$h_i(\boldsymbol{x}^*) \le 0, \qquad i = 1,\cdots,m \tag{4.24b}$$

$$u_i^{6*} \ge 0, \qquad i = 1,\cdots,m \tag{4.24c}$$

$$u_i^{6*}h_i(\boldsymbol{x}^*) = 0, \qquad i = 1,\cdots,m \tag{4.24d}$$

$$\tag{4.24e}$$

For a chosen benchmark return level ν, let $(\boldsymbol{x}^{*T}, \boldsymbol{\zeta}^{*T})^T$ be an optimal solution for Problem (4.20) with KKT multipliers u^{6*}, then

- $(\boldsymbol{x}^{*T}, \boldsymbol{\zeta}^{*T})^T$ is an optimal solution for problem (4.17) provided that $\eta = \rho_g(\boldsymbol{x}^*)$. Moreover, the optimal KKT multipliers are $u_i^{3*} = u_i^{6*}\rho_g(\boldsymbol{x}^*)$ for $i = 1,\cdots,k$ and $u_{k+1}^{3*} = \frac{R(\boldsymbol{x}^*)-\nu}{\rho_g(\boldsymbol{x}^*)}$.

- $(\boldsymbol{x}^{*T}, \boldsymbol{\zeta}^{*T})^T$ is an optimal solution for problem (4.18) provided that $\mu = R(\boldsymbol{x}^*)$. Moreover, the optimal KKT multipliers are $u_i^{4*} = u_i^{6*}\frac{[\rho_g(\boldsymbol{x}^*)]^2}{R(\boldsymbol{x}^*)-\nu}$ for $i = 1,\cdots,k$ and $u_{k+1}^{4*} = \frac{\rho_g(\boldsymbol{x}^*)}{R(\boldsymbol{x}^*)-\nu}$.

- $(\boldsymbol{x}^{*T}, \boldsymbol{\zeta}^{*T})^T$ is an optimal solution for problem (4.19) provided that $\tau = \frac{R(\boldsymbol{x}^*)-\nu}{\rho_g(\boldsymbol{x}^*)}$. Moreover, the optimal KKT multipliers are $u_i^{5*} = u_i^{6*}\rho_g(\boldsymbol{x}^*)$ for $i = 1,\cdots,k$.

Conversely,

- Let $(\boldsymbol{x}^{*T}, \boldsymbol{\zeta}^{*T})^T$ be an optimal solution for problem (4.17) for given η and u^{3*} be the corresponding KKT multipliers. $(\boldsymbol{x}^{*T}, \boldsymbol{\zeta}^{*T})^T$ is also an optimal solution for problem (4.20) provided that $\nu = R(\boldsymbol{x}^*) - u_{k+1}^{3*}\rho_g(\boldsymbol{x}^*)$. Moreover, the optimal KKT multipliers are $u_i^{6*} = u_i^{3*}\frac{1}{\rho_g(\boldsymbol{x}^*)}$ for $i = 1,\cdots,k$.

- Let $(\boldsymbol{x}^{*T}, \boldsymbol{\zeta}^{*T})^T$ be an optimal solution for problem (4.18) for given μ and u^{4*} be the corresponding KKT multipliers. $(\boldsymbol{x}^{*T}, \boldsymbol{\zeta}^{*T})^T$ is also an optimal solution for problem (4.20) provided that $\nu = R(\boldsymbol{x}^*) - \frac{1}{u_{k+1}^{4*}}\rho_g(\boldsymbol{x}^*)$. Moreover, the optimal KKT multipliers are $u_i^{6*} = u_i^{4*}\frac{1}{\rho_g(\boldsymbol{x}^*)}$ for $i = 1,\cdots,k$.

- Let $(\boldsymbol{x}^{*T}, \boldsymbol{\zeta}^{*T})^T$ be an optimal solution for problem (4.19) for given τ and u^{5*} be the corresponding KKT multipliers. $(\boldsymbol{x}^{*T}, \boldsymbol{\zeta}^{*T})^T$ is also an optimal solution for problem (4.20) provided that $\nu = R(\boldsymbol{x}^*) - \tau\rho_g(\boldsymbol{x}^*)$. Moreover, the optimal KKT multipliers are $u_i^{6*} = \frac{1}{\rho_g(\boldsymbol{x}^*)}$ for $i = 1, \cdots, k$.

Therefore, any point on the efficient frontier of each of the problems (4.17)-(4.20) is correspond to a point on the efficient frontiers of the other three problems. These four problems generate the same efficient frontier. □

Observe that the return maximization problem (4.17) and the CDRM minimization problem (4.18) can be viewed as a pair in the sense that the implied parameter for one problem is equal to the optimal objective value of the other. Similar paring relationship is true for the CDRM utility maximization problem (4.19) and the CDRM-based Sharpe Ratio maximization problem (4.20). We will explore such paring relationships in Chapter 5 by considering the geometric interpretations of the parameters in these four problems.

Theorem 4.1.1 implies that the function $M_g(\boldsymbol{x}, \boldsymbol{\zeta})$ can be used instead of $\rho_g(\boldsymbol{x})$ in optimization problems and by construction $M_g(\boldsymbol{x}, \boldsymbol{\zeta})$ is a convex combination of $F_{\alpha_i}(\boldsymbol{x}, \zeta_i)$. We can employ the linearization scheme for $F_{\alpha_i}(\boldsymbol{x}, \zeta_i)$, which is readily available, repeatedly to linearize $M_g(\boldsymbol{x}, \boldsymbol{\zeta})$.

Consider scenario generations with m scenarios for n financial instruments. One can generate future losses for different insurance/reinsurance contracts or future prices/returns of corporate stocks. Let $\boldsymbol{L} = [L_{ij}]_{i=1}^{m}{}_{j=1}^{n}$ be the loss matrix or the negate of return matrix generated, where L_{ij} denotes the loss in scenario i from instrument i. Let the probability of realizing scenario i be p_i for $i = 1, \cdots, m$. In line with the previous notations, the loss function of a portfolio $\boldsymbol{x}$ is then

$$\boldsymbol{l}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{L}\boldsymbol{x} \Leftrightarrow l(\boldsymbol{x}, y_i) = \boldsymbol{L}_{i\cdot}^T \boldsymbol{x} \tag{4.25}$$

Let $\boldsymbol{c} = (c_1, \cdots, c_n)^T$ be the premium or price vector of the instruments. The vector $\boldsymbol{c}$ can be generated separately from the L, can be obtained from market data, or can be the expected return based on L, i.e., $c_j = \bar{\boldsymbol{L}}_{\cdot i} = \sum_{i=1}^{m} p_i L_{ij}$. Then the LP formulations of portfolio selection problems over CDRM are given by

- Return maximization subject to CDRM constraint, Problem (4.17)

$$\begin{array}{rrcll}
\text{maximize} & \boldsymbol{c}^T\boldsymbol{x} & & & \\
\text{subject to} & \boldsymbol{A}\boldsymbol{x} & \leq & \boldsymbol{b} & \\
& \sum_{j=1}^{m} w_j[\zeta_j + \frac{1}{1-\alpha_j}\sum_{i=1}^{m} p_i z_{ij}] & \leq & \eta & \\
& z_{ij} & \geq & \boldsymbol{L}_{i\cdot}^T x - \zeta_j & i = 1,\cdots,m; j = 1,\cdots,m \\
& z_{ij} & \geq & 0 & i = 1,\cdots,m; j = 1,\cdots,m
\end{array} \tag{4.26}$$

- CDRM minimization subject to return constraint, Problem (4.18)

$$\begin{array}{rrcll}
\text{minimize} & \sum_{j=1}^{m} w_j[\zeta_j + \frac{1}{1-\alpha_j}\sum_{i=1}^{m} p_i z_{ij}] & & & \\
\text{subject to} & \boldsymbol{A}\boldsymbol{x} & \leq & \boldsymbol{b} & \\
& \boldsymbol{c}^T\boldsymbol{x} & \geq & \mu & \\
& z_{ij} & \geq & \boldsymbol{L}_{i\cdot}^T\boldsymbol{x} - \zeta_j & \forall i, j = 1,\cdots,m \\
& z_{ij} & \geq & 0 & \forall i, j = 1,\cdots,m
\end{array} \tag{4.27}$$

- Return-CDRM utility maximization for a given risk aversion parameter $\tau \geq 0$, Problem (4.19)

$$\begin{array}{rrcll}
\text{maximize} & \boldsymbol{c}^T\boldsymbol{x} - \tau\sum_{j=1}^{m} w_j[\zeta_j + \frac{1}{1-\alpha_j}\sum_{i=1}^{m} p_i z_{ij}] & & & \\
\text{subject to} & \boldsymbol{A}\boldsymbol{x} & \leq & \boldsymbol{b} & \\
& z_{ij} & \geq & \boldsymbol{L}_{i\cdot}^T\boldsymbol{x} - \zeta_j & \forall i, j = 1,\cdots,m \\
& z_{ij} & \geq & 0 & \forall i, j = 1,\cdots,m
\end{array} \tag{4.28}$$

The formulation for CDRM-based Sharpe Ratio maximization for a given benchmark return level ν, Problem (4.20), is an LFP

$$\begin{array}{rrcll}
\text{maximize} & \frac{\boldsymbol{c}^T\boldsymbol{x} - \nu}{\sum_{j=1}^{m} w_j[\zeta_j + \frac{1}{1-\alpha_j}\sum_{i=1}^{m} p_i z_{ij}]} & & & \\
\text{subject to} & \boldsymbol{A}\boldsymbol{x} & \leq & \boldsymbol{b} & \\
& z_{ij} & \geq & \boldsymbol{L}_{i\cdot}^T\boldsymbol{x} - \zeta_j & \forall i, j = 1,\cdots,m \\
& z_{ij} & \geq & 0 & \forall i, j = 1,\cdots,m
\end{array} \tag{4.29}$$

We can utilize the variable transformation technique by letting

$$\hat{\boldsymbol{x}} = t\boldsymbol{x}, \hat{\boldsymbol{\zeta}} = t\boldsymbol{\zeta}, \hat{\boldsymbol{z}} = t\boldsymbol{z}$$

where $t = \frac{1}{\sum_{j=1}^{m} w_j[\zeta_j + \frac{1}{1-\alpha_j}\sum_{i=1}^{m} p_i z_{ij}]}$..

Then Problem (4.29) can be solved by solving at most two LPs given by

$$\begin{array}{llcll}
\text{maximize} & \boldsymbol{c}^T\hat{\boldsymbol{x}} - \nu t & & & \\
\text{subject to} & \boldsymbol{A}\hat{\boldsymbol{x}} - \boldsymbol{b}t & \leq & 0 & \\
 & \hat{z}_{ij} & \geq & \boldsymbol{L}_{i\cdot}^T\hat{\boldsymbol{x}} - \zeta_j & \forall i, j = 1, \cdots, m \\
 & \hat{z}_{ij} & \geq & 0 & \forall i, j = 1, \cdots, m \\
 & \sum_{j=1}^{m} w_j[\hat{\zeta}_j + \frac{1}{1-\alpha_j}\sum_{i=1}^{m} p_i\hat{z}_{ij}] & = & 1 & \\
 & t & \geq & 0 &
\end{array} \tag{4.30}$$

and

$$\begin{array}{llcll}
\text{maximize} & -\boldsymbol{c}^T\hat{\boldsymbol{x}} + \nu t & & & \\
\text{subject to} & \boldsymbol{A}\hat{\boldsymbol{x}} - \boldsymbol{b}t & \leq & 0 & \\
 & \hat{z}_{ij} & \geq & \boldsymbol{L}_{i\cdot}^T\hat{\boldsymbol{x}} - \zeta_j & \forall i, j = 1, \cdots, m \\
 & \hat{z}_{ij} & \geq & 0 & \forall i, j = 1, \cdots, m \\
 & -\sum_{j=1}^{m} w_j[\hat{\zeta}_j + \frac{1}{1-\alpha_j}\sum_{i=1}^{m} p_i\hat{z}_{ij}] & = & 1 & \\
 & t & \geq & 0 &
\end{array} \tag{4.31}$$

It is not necessary to solve both Problem (4.30) and Problem (4.31) in many cases. For example, CDRM is a risk measure so it is rarely negative in practice. If we know the optimal portfolio has a positive CDRM, then it is sufficient to solve Problem (4.30) only. In fact, it is sufficient to solve only one of Problem (4.30) or Problem (4.31) if we have prior knowledge about the sign of either risk or return at optimal portfolio.

Chapter 6 Numerical works

Chapter 5

Case Studies and Numerical Results

The properties of CRM and DRM ease our exploitation of CDRM's theoretical properties. Discussions in Chapter 3 reveal that CDRM is both conceptually intuitive and computationally amenable. Various linear formulation for CDRM problems their equivalences as discussed in Chapter 4 largely enhanced the applicability of CDRM in practice. Two case studies are conducted in this chapter to illustrate the CDRM portfolio optimization methodologies presented in previous chapters. A reinsurance portfolio selection problem with simulated insurance losses is studied and numerical verifications of the equivalences among different formulations are provided in Section 5.1. An investment portfolio selection problem with real historical data is studied, empirical analyses and comparisons among different members of CDRM are provided in Section 5.2. The actual choice of CDRM objective/constraints in any portfolio selection problem varies and should be derived on a case by case basis. The emphasis of these two examples is the illustration of the linear optimization methodologies for CDRM portfolio selection problems, rather than a practical recommendation for industrial applications. All programming problems in this chapter are solved with AMPL using the Gurobi 4.5.1 solver.

5.1 Case Study: Reinsurance Portfolio Construction

Suppose a reinsurance company wants to construct a portfolio from n risk contracts. The risk premiums for these contracts are publicly available in the market. The company believes that there are m possible future states of economy and has sufficient information to simulate the future losses for all contracts. Then the company has the loss matrix $\boldsymbol{L}$ and the premium vector $\boldsymbol{c}$ and hence is able to formulate CDRM portfolio selection problems.

An anonymous reinsurance company provides us with the premiums of 10 risk contracts and the simulated losses for these contracts in $10,000$ future states of economy. For simplicity, we assume equal probability of realization for each of these $10,000$ future states. In addition, we assume that the company can purchase (assume risks and collect premiums) or sell (transfer risks and pay premiums) any number of units or fraction of these risk contracts. The losses incurred/reimbursed and premiums collected/paid will be proportional to the number of units purchased/sold. The return measure of interest is the difference between portfolio premium and expected portfolio loss, i.e., the expected portfolio profit. The preselected risk measure is the 95%-CVaR of portfolio losses. Table 5.1 shows some statistics[1] of the 10 risk contracts.

Contract	Premium	Losses					
		Mean	STD	Skew	Kurt	95%VaR	95%CVaR
1	554271	311388	1377843	5	31	2613161	5885442
2	364272	222117	1172497	6	41	588329	4338214
3	91763	55953	739026	16	274	0	1119065
4	867176	437968	1806626	4	21	3845685	7937610
5	798005	438464	2913258	7	55	0	8769284
6	107585	43381	263019	6	42	0	867624
7	878525	375438	1375166	4	18	3160679	5974087
8	3081188	1283828	2199151	2	5	5661191	8442634
9	65162	29352	324061	12	134	0	587044
10	885897	385173	1047454	6	48	1506500	3693435

Table 5.1: Premiums and Summary Statistics of Simulated Losses for the 10 Risk Contracts

We observe from Table 5.1 that

1. The market premium for each contract is greater than its expected loss. We can view such difference as the risk loading for a contract which reflects the risk appetite of the market towards each contract.

2. The skewness coefficients of the simulated losses are positive. This implies that all simulated loss distributions are skewed to the right, which corresponding to large losses.

[1] Rounded to the nearest integer.

3. The kurtosis coefficients of the simulated losses are high. This implies the existence of extremely large losses with small probabilities. Such characteristic is also reflected from the 95%-VaR and 95%-CVaR columns. For example, contract 3, 5, 6 and 9 have zero 95%-VaR yet their 95%-CVaR are large, which mean that with lower than 5% chance there would be huge losses.

Additionally, the empirical correlation matrix for the simulated losses of these 10 risk contracts is provided in Table 5.2. We can see that the maximum correlation between two contracts is 0.55028 while the minimum is -0.0069. Note that there are contracts whose losses are negatively correlated, which enables effective risk hedging against each other.

	1	2	3	4	5	6	7	8	9	10
1	1									
2	0.0004	1								
3	0.0105	-0.0037	1							
4	0.4173	0.0247	-0.0013	1						
5	0.0218	0.1486	0.0026	0.1791	1					
6	0.1746	0.2689	-0.0035	0.1773	0.4436	1				
7	0.0205	0.5503	-0.0069	0.0521	0.2199	0.3304	1			
8	0.4620	0.0216	0.0011	0.2226	0.0469	0.0975	0.0499	1		
9	-0.0031	0.1602	0.0058	0.0068	0.1751	0.2352	0.3691	0.0159	1	
10	0.3313	0.0030	-0.0029	0.2043	0.0255	0.0711	0.0155	0.4258	0.0023	1

Table 5.2: Correlation Matrix of Risk Contracts

In order to get a benchmark portfolio for later comparisons, we first consider a balanced portfolio, i.e., a portfolio consists of $\frac{1}{10}$ unit of each of the 10 contracts. Portfolio premium, summary statistics of portfolio losses, and expected portfolio profit for balanced portfolio[2] are given in Table 5.3.

Premium	Losses						Expected Profit
	Mean	STD	Skew	Kurt	95%-VaR	95%-CVaR	
769384	358306	667647	3	12	1716458	2656764	40578

Table 5.3: Premium, Summary of Losses, and Expected Profit of Balanced Portfolio

[2]Rounded to the nearest integer.

Note that the portfolio premium, the expected loss, and hence the expected profit of balanced portfolio are the average of original 10 risk contracts' premiums, expected losses and expected profits. However, all other statistics of the portfolio losses such as standard deviation, skewness kurtosis, $95\%VaR$, and $95\%CVaR$ are all lower than the average of the 10 risk contracts corresponding measures. This reveals that the balanced portfolio enjoys the benefit of risk diversification hence achieves the average expected profit while bearing lower than average risks.

Although the balanced portfolio illustrates the benefit of diversification, we will show that such a simple portfolio construction is far from being optimal in the sense of risk-reward trade off. We solve the utility maximization problem (4.19) hence LP (4.28) repeatedly by varying the risk-aversion parameter τ from 0 to 3 with increments of 0.005 and hence trace efficient frontier in expected profit v.s. 95%-CVaR plane for these 10 risk contracts. Note that the decision variables x_i, $i = 1, \cdots, 10$ in this example denote the number of unit(s) of risk contract i purchased (or sold if $x_i < 0$) by the reinsurance company. A budget constraint is imposed, all feasible portfolios collect the same amount of premium as the balanced portfolio does. The resulting efficient frontier is shown in Figure 5.1. The balanced portfolio in expected profit v.s. 95%-CVaR plane is shown in the enlarged graph as diamond at the bottom right.

We see that the balanced portfolio is far away from the efficient frontier therefore is highly inefficient in terms of the trade off between expected profit and 95%-CVaR. In particular, at the same 95%-CVaR level as that of the balanced portfolio, one can achieve a much higher expected profit than that of the balanced portfolio. This inefficiency suggests the importance to active risk management in portfolio construction. Despite the benefit of diversification, simple portfolio construction scheme such as the balanced strategy could result in bearing unnecessary risks and/or losing possible profit.

Figure 5.1 also helps us to verify the equivalences established in Theorem 4.2.4 numerically and provides us with geometric interpretations of the parameters in problems (4.17)-(4.20).

To verify the equivalences established in Theorem 4.2.4, we can first solve any one of the problems (4.17)-(4.17) with a preselected parameter then use the implied parameters to solve the other problems. According to Theorem 4.2.4, the parameters and the optimal objective values of these four problems are interrelated as specified in Table 4.1.

We set $\tau = 0.2$ and solve the utility maximization problem (4.19). The optimal solution is referred to as the target portfolio hereafter and is highlighted as a triangle in the enlarged efficient frontier in Figure 5.1. By optimality of the target portfolio it must be a point on the efficient frontier and Figure 5.1 confirms this property. Premium, summary statistics

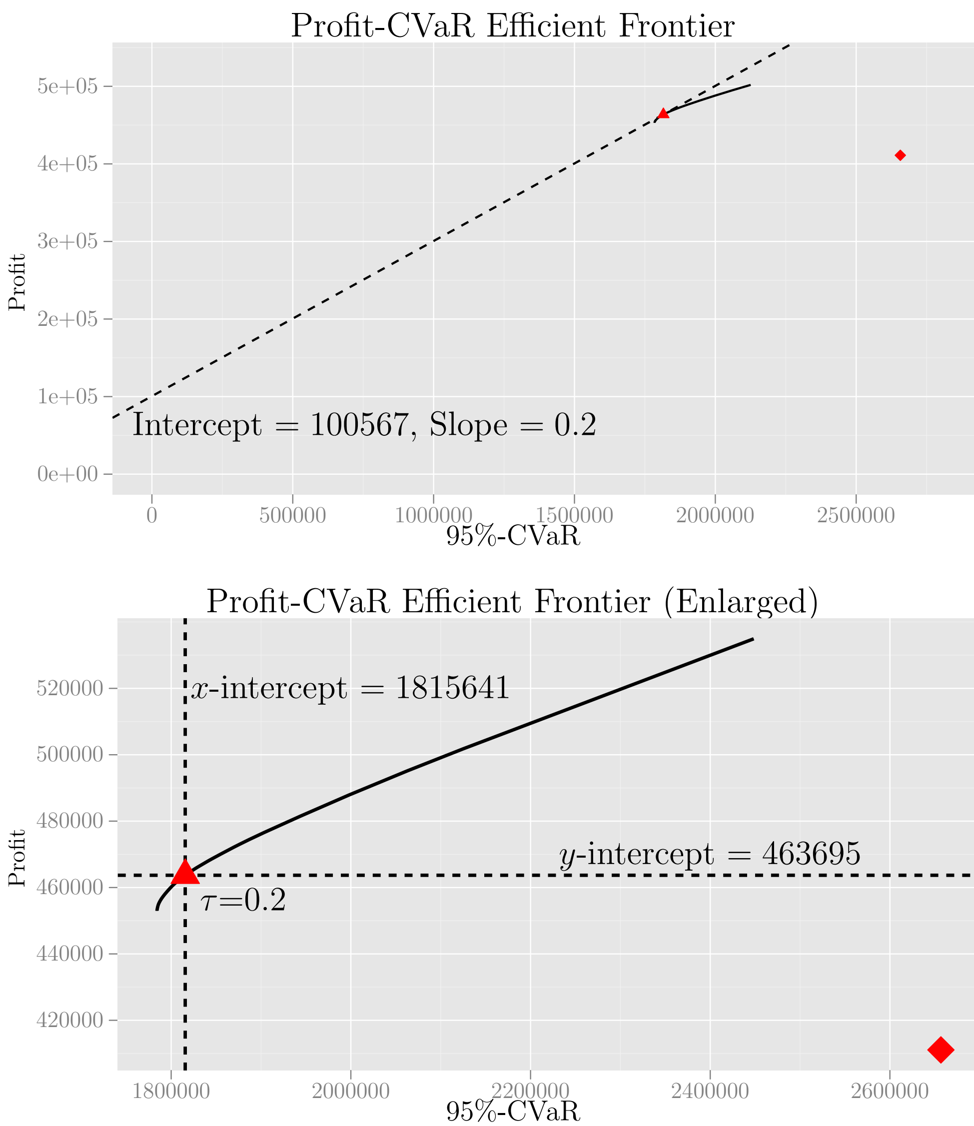

Figure 5.1: Efficient Frontier of Given 10 Risk Contracts

of losses and the expected profit[3] of the target portfolio are shown in Table 5.4.

Optimal Utility	Premium	Losses						Expected Profit
		Mean	STD	Skew	Kurt	95%VaR	95%CVaR	
100567	769384	305689	492425	2	4	1313074	1815641	463695

Table 5.4: Premium and Summary Statistics of Simulated Losses for Target Portfolio

Due to our budget constraint, the total premium collected from target portfolio is the same as that from balanced portfolio. However, comparing Table 5.4 and Table 5.3 we see that the target portfolio's loss distribution is better than the balanced portfolio's in expected loss, standard deviation, skewness, kurtosis, $95\%VaR$, and $95\%CVaR$. Such superiority illustrates once more the insufficiency of balanced portfolio.

We then solve problems (4.17), (4.18) and (4.20) by setting $\eta = 1815641$, $\mu = 769384 - 305689 = 463695$, and $\nu = 463695 - 0.2 \times 1815641 = 100567$ and imposing the same budget constraint. The optimal solutions, the optimal objective values and other related program outputs are summarized in Table 5.5

	Utility Maximization	Profit Maximization	CDRM Minimization	Sharpe Ratio Maximization
Preselected Parameter	$\tau = 0.2$	$\eta = 1815641$	$\mu = 463695$	$\nu = 100567$
Program Outputs				
Optimal Utility[4]	100567	100567	100567	100567
Optimal Profit[5]	463695	463695	463695	463695
$CDRM$[6]	1815641	1815641	1815641	1815641
Optimal Sharpe Ratio[7]	0.2	0.2	0.2	0.2

Table 5.5: Optimal Portfolio Information for Different Formulations of Profit-$CDRM$ Problems

We can see from Table 5.5 that profits, 95%-CVaRs, utilities, and $CDRM$-based Sharpe Ratios are all the same for the optimal portfolios for these three problems with parameters selected according to Table 4.1.

[3] Rounded to the nearest integer.
[4] $R(x^*) - 0.2 \times \rho_g(x^*)$ at the optimal solution for respective problem.
[5] $R(x^*)$ at the optimal solution for respective problem.
[6] $\rho_g(x^*)$ at the optimal solution for respective problem.
[7] $\frac{R(x^*)-\nu}{\rho_g(x^*)}$ at the optimal solution for respective problem.

As mentioned in Chapter 4, the parameter ν in $CDRM$-based Sharpe Ratio maximization problem is equal to the optimal utility and the parameter τ in utility maximization problem is equal to the optimal Sharpe Ratio. Moreover, the parameter η in the profit maximization problem is equal to the optimal $CDRM$ and the parameter μ in the $CDRM$ minimization problem is equal to the optimal profit. These pairing relationships are proven in Theorem 4.2.4 and verified in this numerical example.

We can explore geometric interpretations of these four parameters via the efficient frontier. The dashed line in the upper graph in Figure 5.1 passes through $(0, 100567)$ and $(1815641, 463695)$ and has a slope of 0.2. Note that the intercept is the optimal utility and the slope equals to the optimal Sharpe Ratio. The vertical dashed line in the enlarged efficient frontier graph in Figure 5.1 has an x-intercept of 1815641, the optimal CDRM value. The horizontal dashed line in the same graph has a y-intercept of 463695, the amount of optimal profit.

A CDRM constraint restricts the feasible set to be the area under the efficient frontier and to the left of the maximum acceptable CDRM level η, i.e., the area below the curve and to the left of the vertical dashed line in Figure 5.1. If we look for the highest point in this area, which is equivalent to maximizing profit subject to a CDRM constraint, we will reach at the target portfolio with optimal objective value μ. Similarly, a profit constraint restricts the feasible set to be the area under the efficient frontier and above the minimum acceptable profit level μ, i.e., the area below the curve and above the horizontal dashed line in Figure 5.1. If we search for the leftmost point in this area, which is equivalent to a minimizing CDRM subject to a profit constraint, we will again reach at the target portfolio with optimal objective η, as shown in Figure 5.1. Such geometric interpretations of η and μ explain the relationship between the optimal objective values and parameters in profit maximization problem (4.17) and $CDRM$ minimization problem (4.18).

For any feasible portfolio $\boldsymbol{x}$ with profit $R(\boldsymbol{x})$ and $CDRM$ $\rho_g(\boldsymbol{x})$ and a given risk aversion parameter τ, the utility $R(\boldsymbol{x}) - \tau\rho_g(\boldsymbol{x})$ is the y-intercept for a straight line going through $(\rho_g(\boldsymbol{x}), R(\boldsymbol{x}))$ with slope τ in the profit-CDRM plane. Therefore a utility maximization problem is equivalent to shifting a fixed-slope straight line upward with at least one point within the efficient frontier, highest possible line is the dashed line in the upper graph of Figure 5.1. For any feasible portfolio $\boldsymbol{x}$ with profit $R(\boldsymbol{x})$ and CDRM $\rho_g(\boldsymbol{x})$ and a given benchmark profit level ν, the CDRM-based Sharpe Ratio $\frac{R(\boldsymbol{x})-\nu}{\rho_g(\boldsymbol{x})}$ is the slope of a straight line that goes through $(0, \nu)$ and $(\rho_g(\boldsymbol{x}), R(\boldsymbol{x}))$ in the profit-CDRM plane. Therefore a CDRM-based Sharpe Ratio maximization problem is equivalent to rotating a straight line that is pivoted at $(0, \nu)$ counterclockwise with at least one point of the line within the efficient frontier, the resulting straight line is again the dashed line in the upper graph of

Figure 5.1.

Theoretical aspects such as monotonicity and concavity of the return-CDRM efficient frontier requires a more rigorous exploration and is outside the scope of this thesis. Inspired by the usual mean-variance efficient frontier in canonical Markowitz settings and in view of Figure 5.1, we conjecture that the return-CDRM efficient frontier is an increasing and concave curve in the return-CDRM plane.

5.2 Case Study: Investment Portfolio Construction

Consider an investment portfolio selection problem using historical data. We construct our portfolios from 20 stocks[9], 2 from each of the 10 sectors defined in Global Industry Classification Standard(GICS). The stocks chosen and their corresponding GICS sectors are shown in Appendix B.1. Weekly prices[10] from 02/01/2001 to 31/05/2011 (a total of 543 weeks, hence 542 weekly returns) for these stocks were obtained from finance.yahoo.com[11]. Figure 5.2 shows the time series plot for the sum of these 20 stocks' prices. Since we have confined our portfolios to be constructed from these stocks Figure 5.2 represents our a time series plot for the "market portfolio".

We see from Figure 5.2 that there had been market declines from 2001 to 2003 as the aftershock from $9-11$ terrorist attack in 2001. We also see market declines from 2007 to 2009 resulted from the so-called "sub-prime mortgage financial crisis". Moreover, the "market portfolio" increases gradually from mid-2003 to 2005, from mid-2005 to 2007, and after 2009. We replace scenario generation by historical data of stock returns and assume equal probability for each scenario. Therefore the loss matrix $\boldsymbol{L}_{100\times 20}$ represents the negative returns of 20 stocks in 100 scenarios, and the return vector $\boldsymbol{c} = (\bar{\boldsymbol{L}}_{\cdot 1}, \cdots, \bar{\boldsymbol{L}}_{\cdot 20})$. Note that both quantities changes over time. The portfolio's expected return and 95%-CVaR of the negative returns are the return and risk measures of interest.

We first examine the performance of a simple yet common investment portfolio, the equally weighted portfolio, also known as the $\frac{1}{n}$-portfolio (also known as the naive strategy). In an $\frac{1}{n}$-portfolio, initial wealth is invested equally, in monetary amount, in all available stocks. Benartzi and Thaler [9] observed that many participants in defined contribution plans used this simple strategy. Windcliff and Boyle [70] explored this simple investment

[9]Stocks with ticker symbols AEP, AIG, C, CI, ETR, F, FMC, GD, GE, HUM, IP, KO, MCD, MRO, MSFT, NSM, T, VZ, WMT, and XOM.

[10]Closing prices adjusted for dividends and splits.

[11]Last access on 25/07/2011.

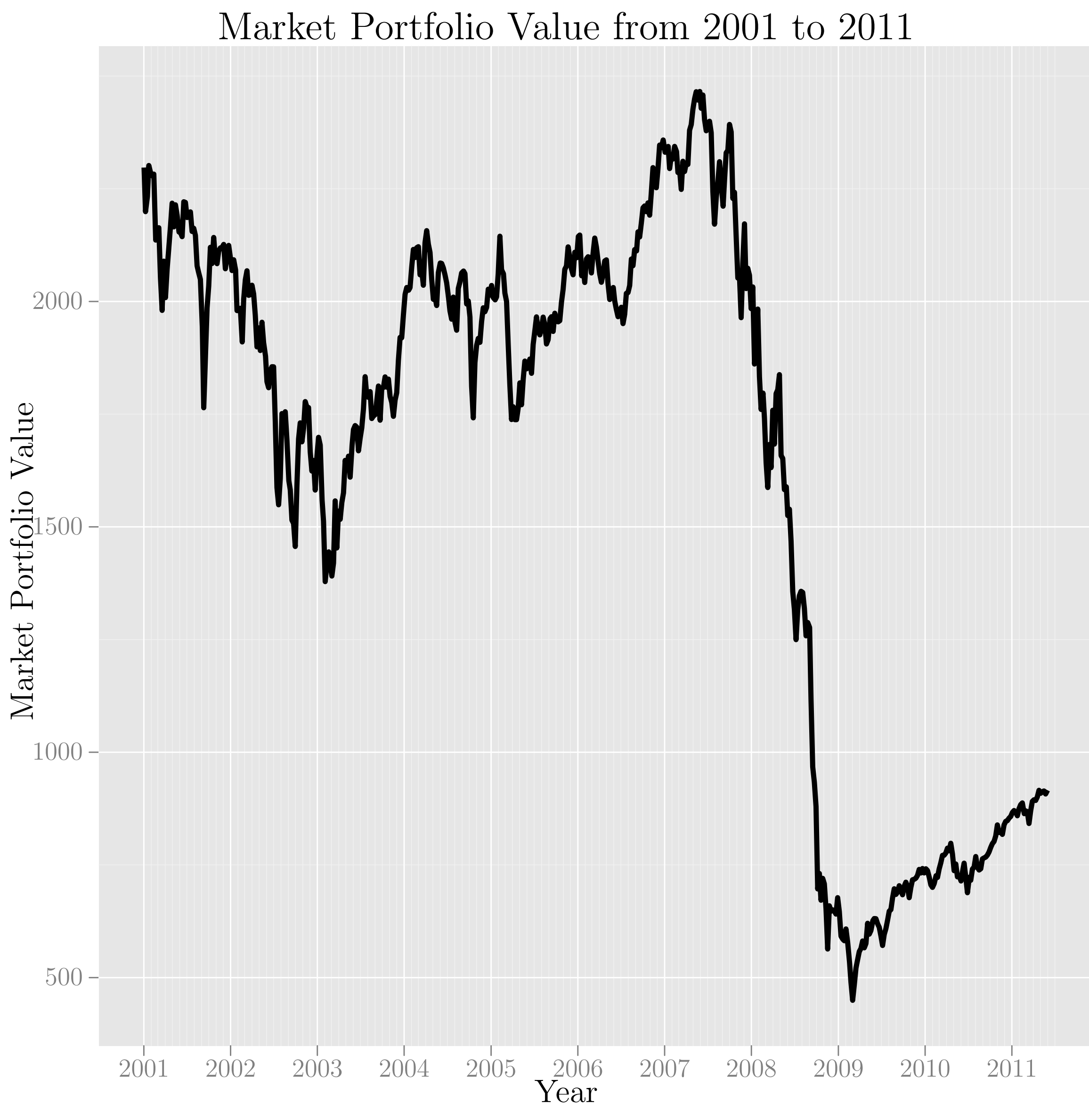

Figure 5.2: Time Series Plot of the “Market Portfolio”

strategy in the classical Markowitz framework and gave merits to this diversification rule when parameter estimation risks and parameter estimation errors are considered. DeMiguel et al. [18] preformed extensive empirical study across 14 portfolio selection models and found that none of them consistently outperforms the $\frac{1}{n}$-portfolio in terms of the Sharpe Ratio. We traced the efficient frontiers in return v.s. 95%-CVaR plane at the beginning of year 2003, 2005, 2007, as well as 2009 with the same methodologies in previous section. Figures 5.3-5.6 these four efficient frontiers receptively. Moreover, the risk-reward trade off for $\frac{1}{n}$-portfolios at the same times are shown as solid diamonds in corresponding figures.

We can see that the $\frac{1}{n}$-portfolio lies far away from all the efficient frontiers. In three out of four figures (Figures 5.3,5.4 and 5.6) the $\frac{1}{n}$-portfolio lies below the minimum-risk portfolio. In Figure 5.6 even bears a risk that exceeds the risk of the maximum return portfolio (and hence maximum risk portfolio). All these figures suggest the inefficiency of the $\frac{1}{n}-portfolio$ and the importance of active risk management. Moreover, the $\frac{1}{n}$-portfolio lies significantly farther from the efficient frontier in periods of market declines (see Figure 5.3 and Figure 5.6) than in periods of market increases (see Figure 5.4 and Figure 5.5). One possible explanation is that, assets are more correlated when the market performs poorly hence the benefit of risk diversification for $\frac{1}{n}$-portfolio becomes the disadvantage of risk aggregation.

5.2.1 Portfolio Optimization with Different Members of CDRM

We examine four members of CDRMs: the Conditional Value-at-Risk measures (CVaR), the Wang Transform measures (WT), the Proportional Hazards transform measures (PH), and the lookback distortion measures (LB). We examine these CDRMs within subclasses (risk measures with the same distortion function but different parameters) and across subclasses (risk measures with different distortion functions).

Recall that for any portfolio loss sample $\boldsymbol{l} = (l_1, \cdots, l_m)$ whose ordered losses are denoted by $l_{(1)}, \cdots, l_{(m)}$, a CDRM with respect to a given distortion function $g(\boldsymbol{x})$ is given by $\rho_g(\boldsymbol{l}) = \sum_{i=1}^{m} q_i l_{(i)} = \sum_{i=1}^{m} w_i \phi_{\frac{i-1}{m}}(\boldsymbol{l})$ where $q_i = g(\frac{m-i+1}{m}) - g(\frac{m-i}{m})$, $i = 1, \cdots, m$ and $\boldsymbol{w}$ is defined in Equation (3.9).

We acknowledge that assuming the one-week-ahead return distribution to be discrete uniform distribution with the past 100 weeks' returns as the support is unrealistic. Estimation errors of using such a distributional assumption can be large hence hinder the usefulness of our results. These issues are addressed in many academic studies such as Jorion [31] and the references therein. Our main purpose is to illustrate CDRM portfolio

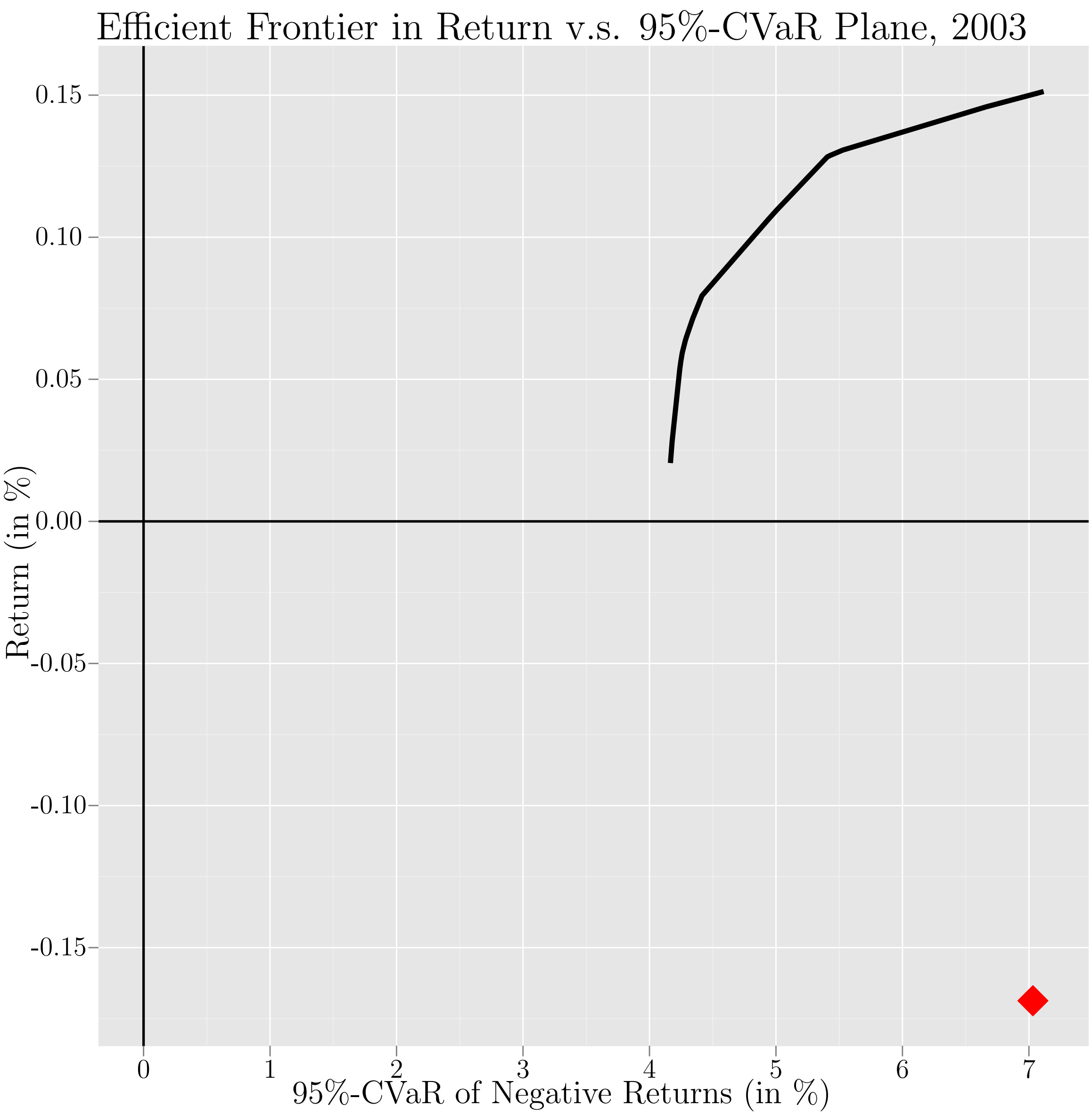

Figure 5.3: Efficient Frontier at the Beginning of 2003

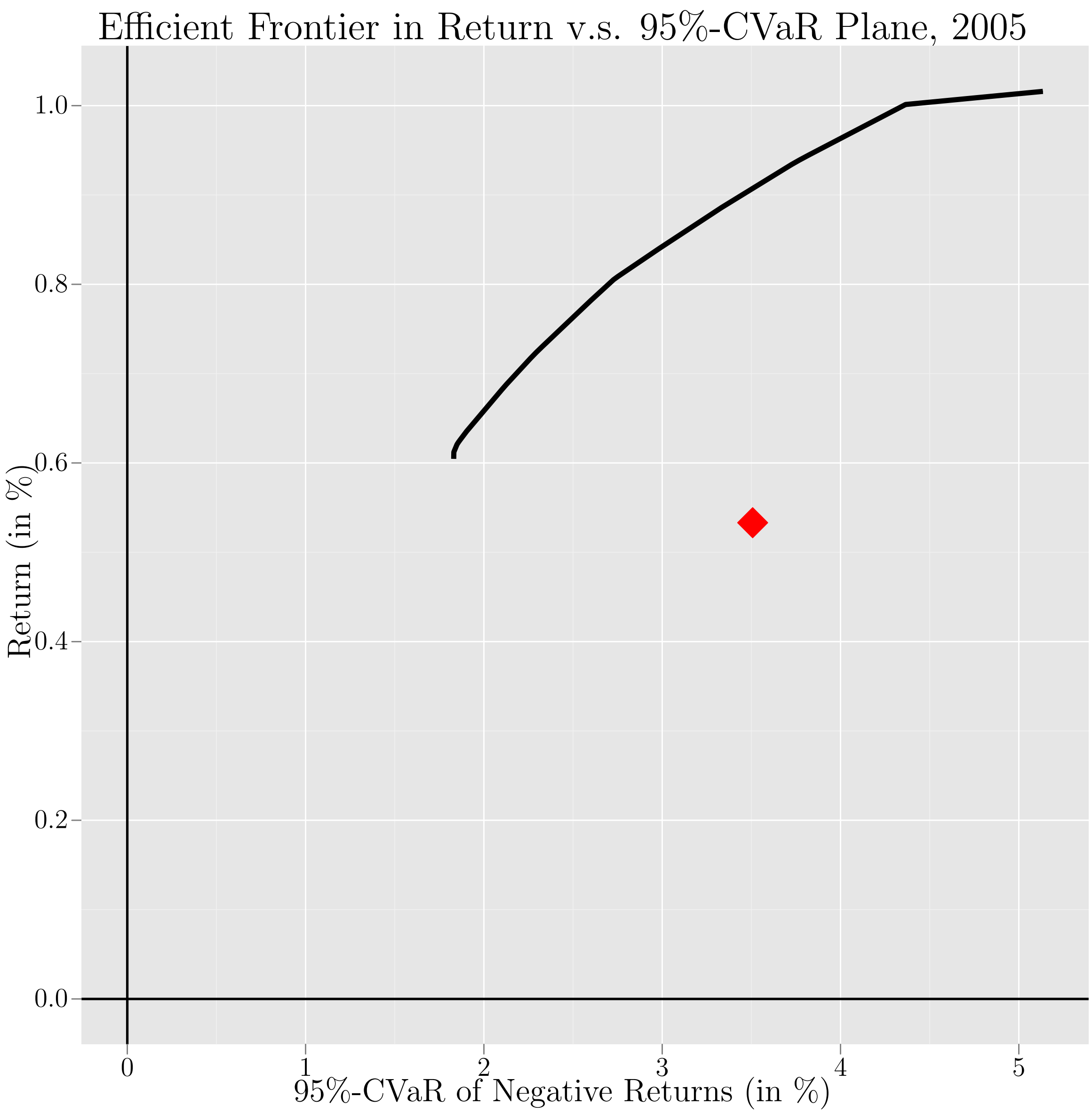

Figure 5.4: Efficient Frontier at the Beginning of 2005

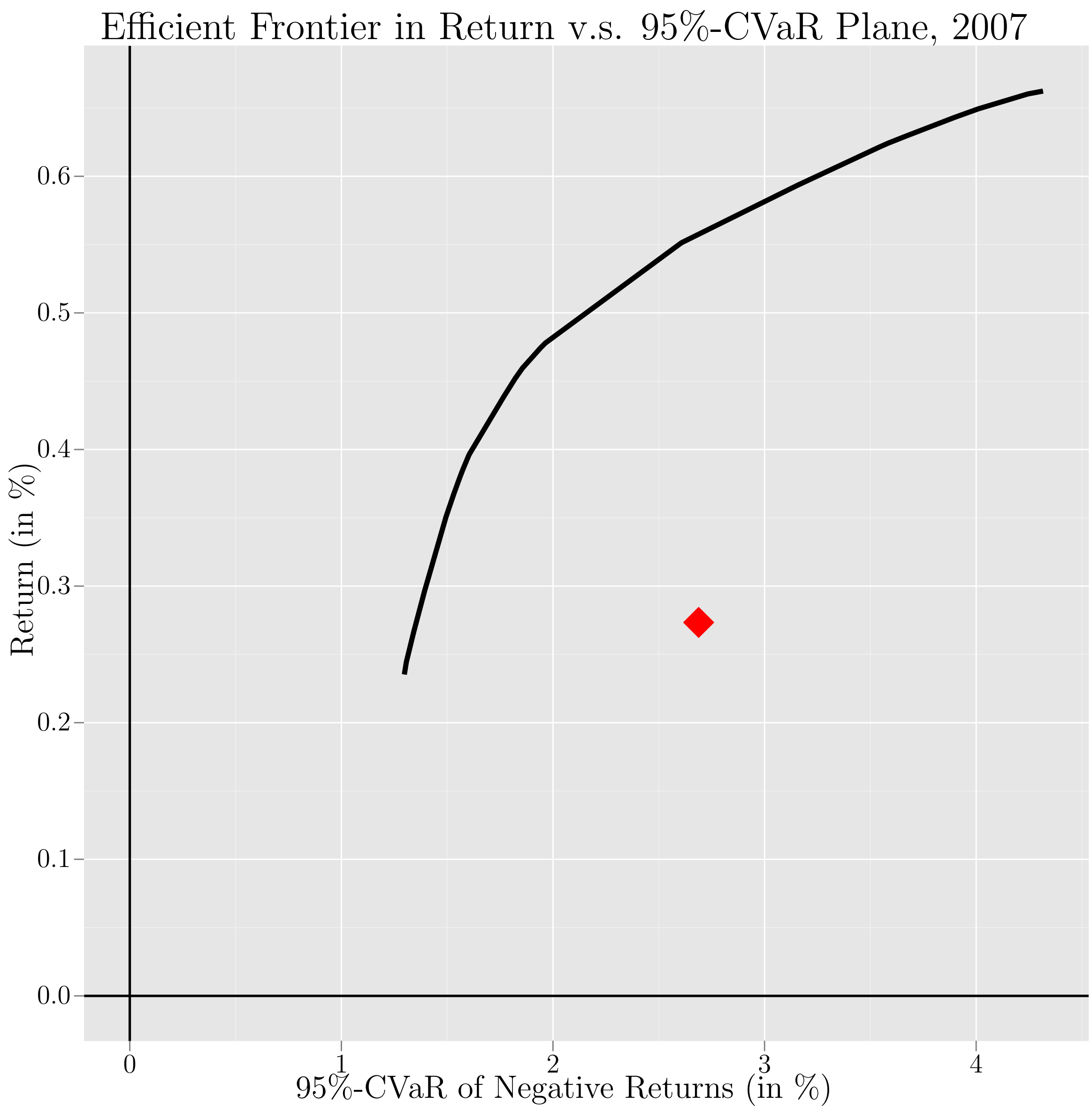

Figure 5.5: Efficient Frontier at the Beginning of 2007

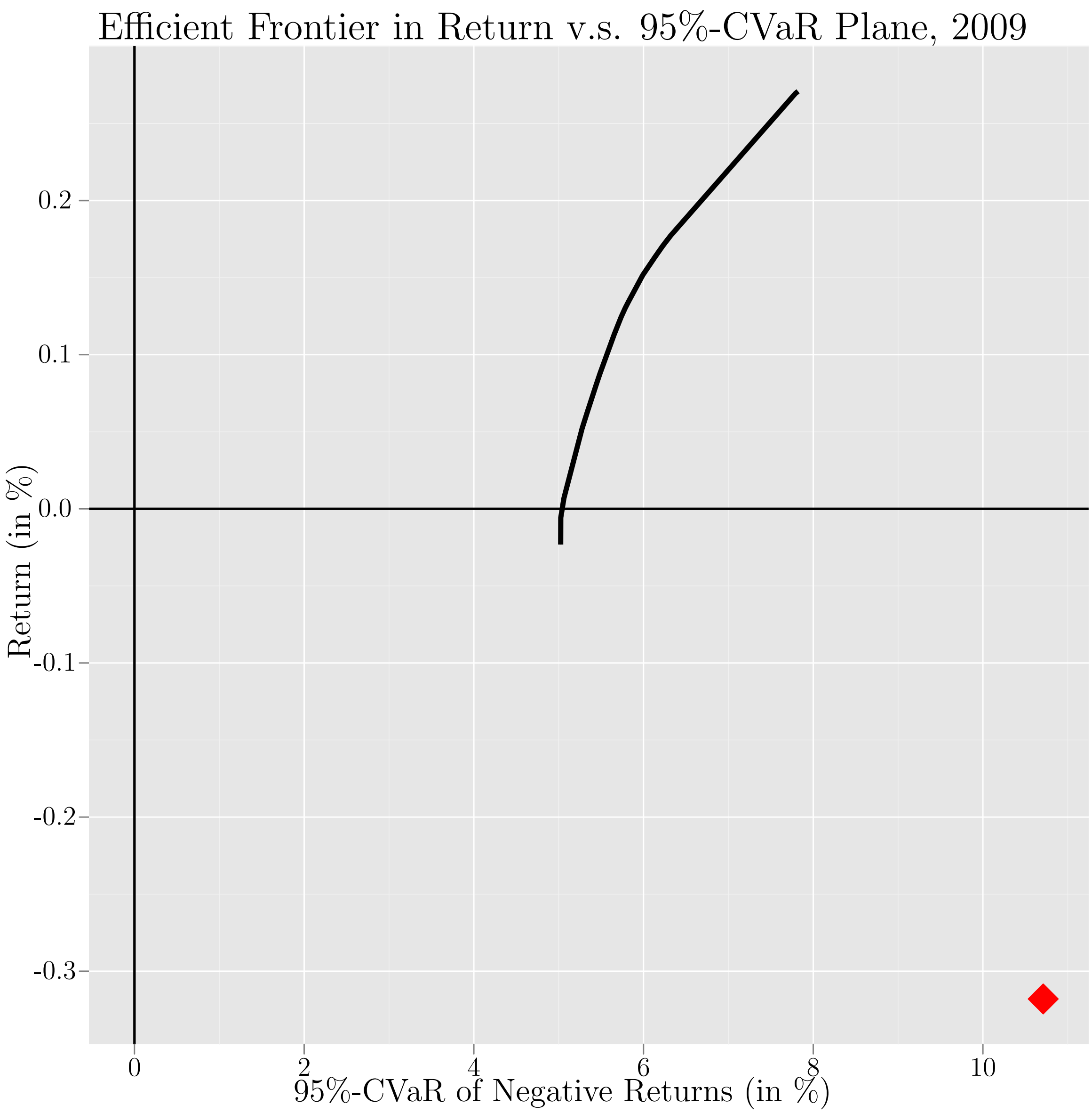

Figure 5.6: Efficient Frontier at the Beginning of 2009

optimization methodologies and comparisons of different of CDRMs. More adequate scenario generation procedures and more stochastic estimation methods may be applied to enhance the practical performance of our optimization framework.

The initial portfolio consists of \$100 cash and the portfolio is rebalanced weekly according to our CDRM optimal portfolios. The optimization was run for 442 overlapping 100-week periods and should determine optimal investment portfolios that minimizes CDRM subject to various constraints. The interpretation of decision variables in this example is different to that in the previous reinsurance example. Decision variables x_i, $i = 1, \cdots, 10$ in this example denote the fraction of wealth, in monetary amount, invested in stock i. We impose a budget constraint $\sum_{i=1}^{20} x_i = 1$, no-short selling constraints $\boldsymbol{x} \geq \boldsymbol{0}$, upper-limit constraints $\boldsymbol{x} \leq \boldsymbol{0.2}$ so that no more than 20% of the total portfolio value should be invested in one single stock, and a return constraint $R(\boldsymbol{x}) \geq \mu$ where μ is the expected return of the $\frac{1}{n}$-portfolio.

Optimization with CVaR

$CVaR_\alpha$ is the fundamental building block in our CDRM optimization framework. The distortion function for CVaR measures is given by

$$g_{CVaR}(x, \alpha) = min\{\frac{x}{1-\alpha}, 1\}, \qquad \alpha \in [0, 1] \tag{5.1}$$

We examine g_{CVaR} for $\alpha = 0.9$, $\alpha = 0.95$, and $\alpha = 0.99$. The weights q_i, $i = 1, \cdots, 100$ for ordered statistics and the weights w_i, $i = 1, \cdots, 100$ for $CVaR_\alpha$ are shown in Figure 5.7.

$CVaR_\alpha = CVaR_\alpha(\boldsymbol{l})$ represents the average of l_is that exceed $VaR_\alpha(\boldsymbol{l})$ and the weights $i_{(i)}$s (as shown in the left column of Figure 5.7) confirm such representation. Graphs in the right column of Figure 5.7 shows the weights for $CVaR_\alpha$s at different αs such that the resulting convex combination of $CVaR_\alpha$ reconciles our desired CDRM. Since $CVaR_\alpha$ is the basic unit, hence the convex weights for $CVaR_\alpha$ measure consist of only one non-zero entry which corresponds to α. The convex representation of $CVaR_\alpha$ is identical to itself therefore our CDRM optimization scheme does not incur extra work should a simple representation be detected.

After optimizations over three different $CVaR_\alpha$s, we compared the performances of their optimal portfolios. Portfolio values with out-of-sample returns (the returns that were realized) and portfolio values and portfolio values with expected returns (the returns based

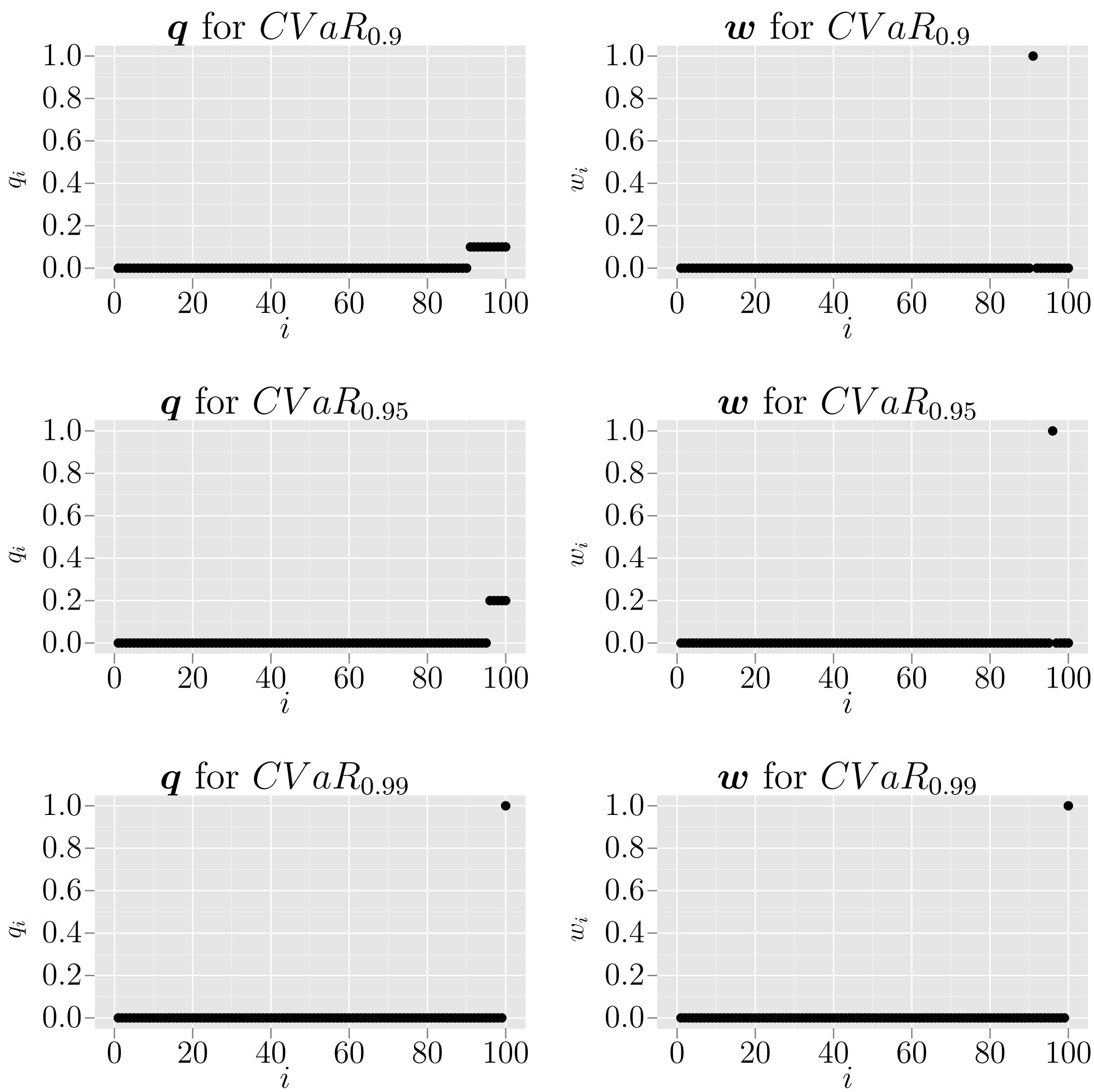

Figure 5.7: $\boldsymbol{q}$ and $\boldsymbol{w}$ for $CVaR_{\alpha}$

on the average over the past 100 weeks) are shown in Figure 5.8. The time series for portfolio values with realized returns is much more spiky compared to that for portfolio values with expected returns. This is not surprising because the expected returns were calculated as an average over consecutive overlapping 100-week periods and hence effectively a smoothing technique had been applied.

Summary statistics for the expected and realized returns for optimal portfolios over $CVaR_{0.9}$, $CVaR_{0.95}$, and $CVaR_{0.99}$ are given in Table 5.6.

Optimal Returns	$CVaR_{0.9}$		$CVaR_{0.95}$		$CVaR_{0.99}$	
	Expected	Realized	Expected	Realized	Expected	Realized
Mean	0.00245	0.00148	0.00228	0.00117	0.00205	0.00139
Standard Deviation	0.00210	0.01891	0.00234	0.02050	0.00262	0.02243
Skewness	0.39751	-0.93697	0.19613	-0.56738	-0.17578	-0.20805
Kurtosis	0.58559	6.08202	0.13816	4.96513	0.21126	4.47107
Sharpe Ratio	1.17098	0.07833	0.97380	0.05718	0.78529	0.06219

Table 5.6: Summary Statistics for $CVaR_\alpha$ Minimization Optimal Portfolios

From a theoretical point of view, optimization of $CVaR_\alpha$ with a lager α represents a more risk averse decision making process that with a smaller α. Therefore optimal portfolios with larger α should have a smaller expected return on average. The bottom graph in Figure 5.8 and the summary statistics of expected returns in Table 5.6 both confirm this theoretical property.

In this example, the mean realized returns are lower the mean expected returns in all three cases. The portfolios with realized returns have terminal wealths than those with expected returns at all confidence levels. Nevertheless, the realized returns and the resulting portfolios display some consistencies with the expected returns and their portfolios. For example, the optimal portfolios w.r.t $CVaR_{0.9}$ (solid lines in Figure 5.8) achieves higher terminal wealth with both realized returns and with expected returns. We also observe that the lower the confidence level is, the higher the canonical Sharpe Ratio is for both expected returns and realized returns.

Optimization with WT-Measures

The WT-measure is a well-known member of DRM introduced in Wang [66]. There have been extensions and variations of the WT since its introduction. In the following example, we follow the definition of WT in Wang [67], whose distortion function is given by

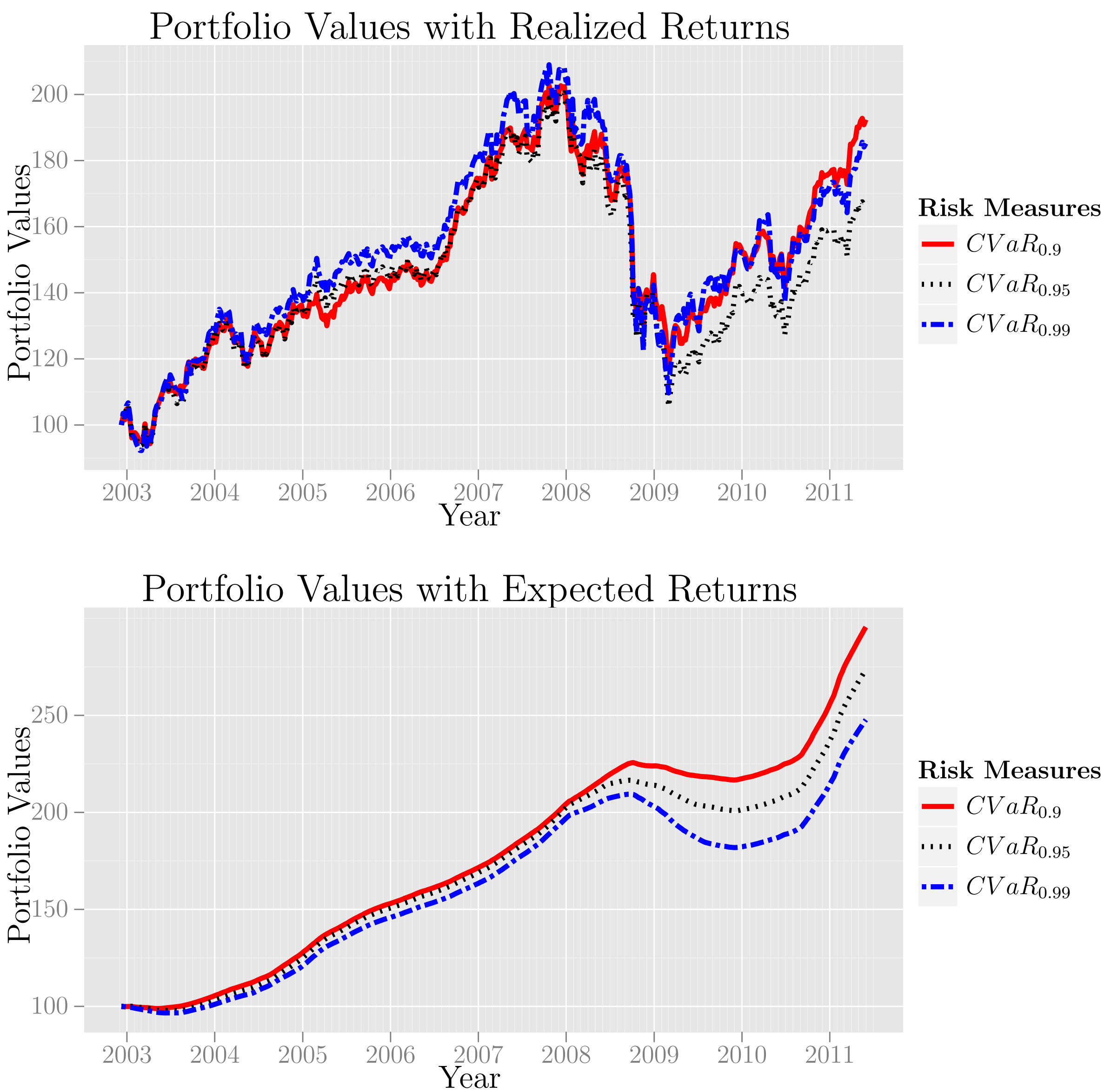

Figure 5.8: Time Series Plots for Optimal Portfolios w.r.t $CVaR_\alpha$s

$$g_{WT}(x,\beta) = \Phi[\Phi^{-1}(x) - \Phi^{-1}(\beta)], \qquad \beta \in [0,1] \tag{5.2}$$

where Φ is the standard normal cumulative distribution function.

Wang [66] showed that g_{WT} is concave for $\beta > 0.5$ and is convex if $\beta < 0.5$. We examine g_{WT} for $\beta = 0.75$, $\beta = 0.85$, and $\beta = 0.95$. The weights q_i, $i = 1, \cdots, 100$ for ordered statistics and the weights w_i, $i = 1, \cdots, 100$ for WT_βs are shown in Figure 5.9.

We see that β in WT_β plays similar role as α in $CVaR\alpha$. More emphasis is placed to minimizing upper tail of the distribution for larger β, as we can see from the graphs in the left column of Figure 5.9. From the graphs in the right column of Figure 5.9, we see that the minimization of $WT_{0.75}$ is emphasizing in minimizing $CVaR_\alpha$s for small αs. Since we calculated $CVaR_\alpha$ using the negative returns and $CVaR_0(\boldsymbol{l}) = E[\boldsymbol{l}]$, minimizing $CVaR_\alpha$s for small αs is in some sense maximizing return. Minimization of $WT_{0.85}$ takes a balanced approach to minimize $CVaR_\alpha$s for all αs non-significant emphasis in both low and high tails of the distribution. Minimization of $WT_{0.95}$ has strong focus on minimizing $CVaR_\alpha$s for large αs. In all three cases, even though there are noticeable preferences towards either return maximization or risk minimization, the preferences is not strong. For example, even though $WT_{0.95}$ has a clear preference towards minimization of $CVaR\alpha$s for large αs, the largest weight among all $CVaR\alpha$s is around 0.15 and is much smaller than 1, in which case becomes the minimization of $CVaR_{0.99}$ itself.

The time series plots for optimal portfolios with realized and expected returns w.r.t the above three WT-measures are shown in Figure 5.10. The corresponding summary statistics for the expected and realized returns are given in Table 5.7.

Optimal Returns	$WT_{0.75}$		$WT_{0.85}$		$WT_{0.95}$	
	Expected	Realized	Expected	Realized	Expected	Realized
Mean	0.00310	0.00164	0.00261	0.00143	0.00232	0.00140
Standard Deviation	0.00217	0.01919	0.00209	0.01915	0.00233	0.02107
Skewness	-0.00193	-1.00243	0.37370	-0.77534	0.10951	-0.30517
Kurtosis	-0.23044	7.06069	0.38650	5.88635	0.31371	5.46812
Sharpe Ratio	1.42672	0.08560	1.24928	0.07477	0.99638	0.06628

Table 5.7: Summary Statistics for WT_β Minimization Optimal Portfolios

Similar to the $CVaR_\alpha$ minimization example, the realized returns has lower mean value than that of the expected returns. Optimal portfolios with expected returns achieve higher terminal wealth than those with realized returns for all β. Nevertheless, for both realized

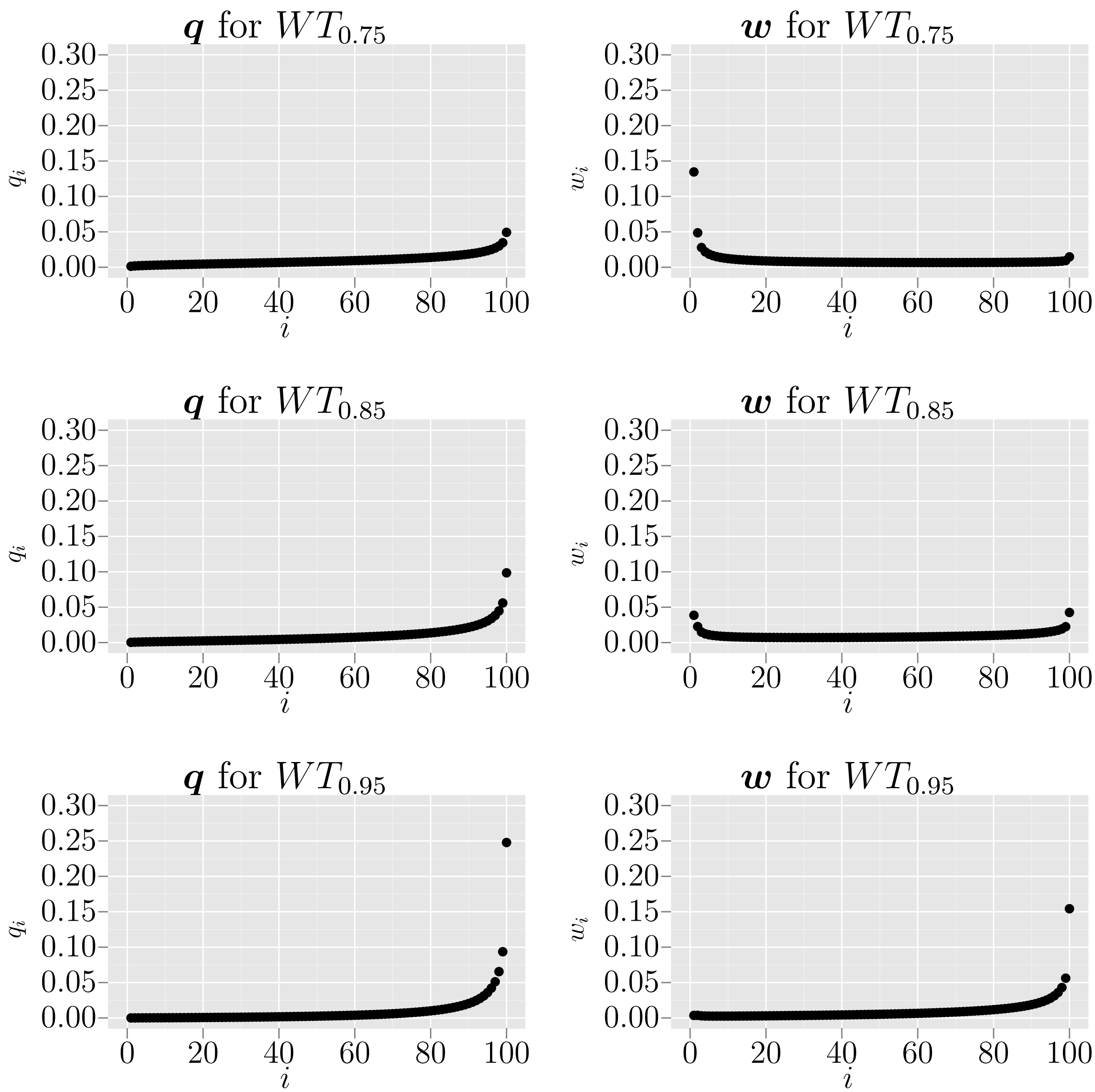

Figure 5.9: $\boldsymbol{q}$ and $\boldsymbol{w}$ for WT_β

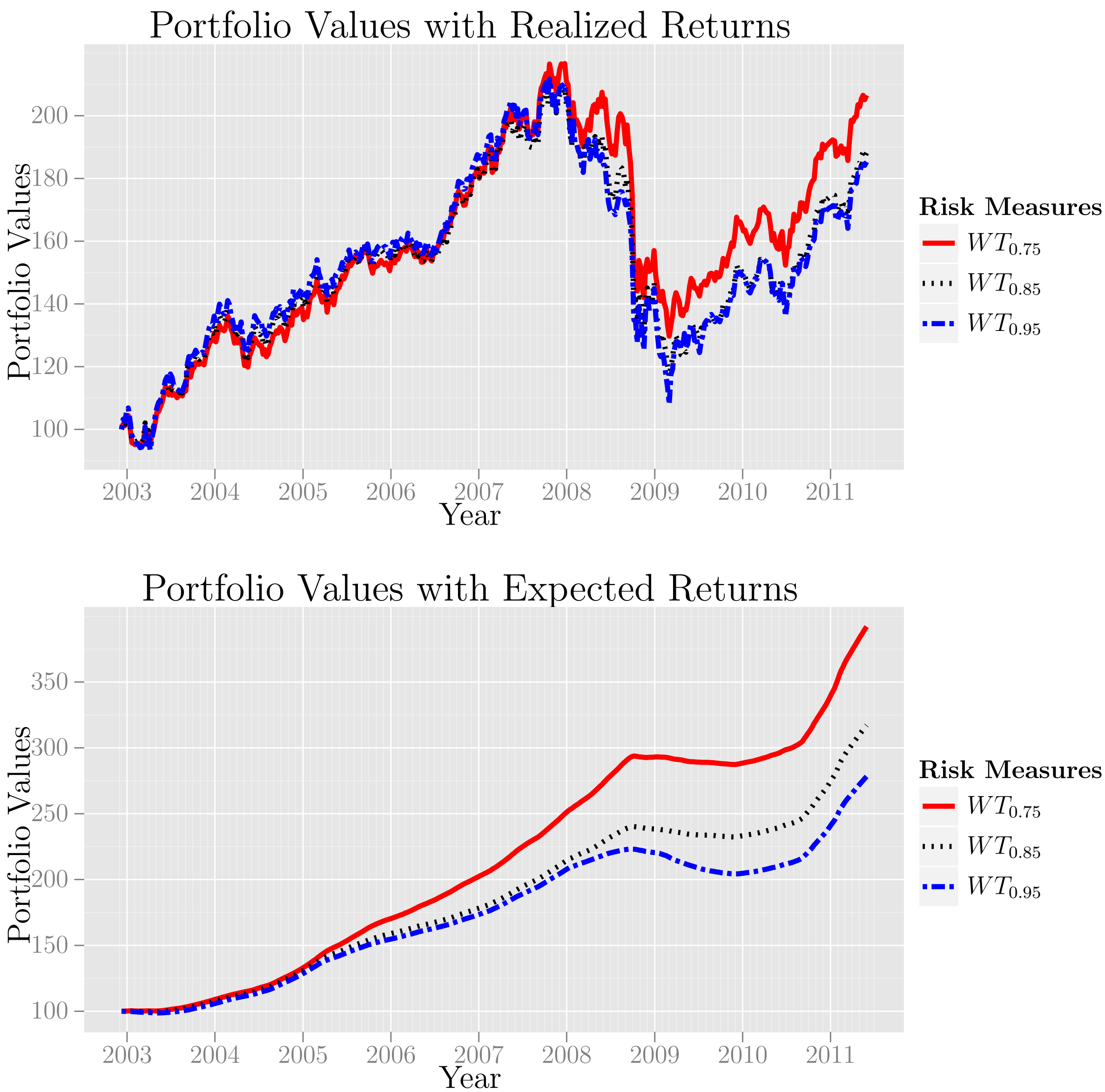

Figure 5.10: Time Series Plots for Optimal Portfolios w.r.t WT_βs

and expected returns, the optimal portfolios w.r.t $WT_{0.75}$ outperform those w.r.t $WT_{0.95}$, which in turn outperform those w.r.t $WT_{0.95}$. The mean of expected returns for $WT_{0.75}$ is higher than that for $WT_{0.85}$, which in turn is higher than that for $WT_0.85$. In our example, the lower β is, the higher the canonical Sharpe Ratio is for both expected and realized returns.

Optimization with PH-Transform Measures

The PH-transform measure was introduced in Wang et al. [69]. The distortion function for PH-transform is given by

$$g_{PH}(x, \gamma) = x^{\gamma}, \qquad \gamma \in (0, 1] \tag{5.3}$$

It is clear that g_{PH} is concave for $\gamma \in (0, 1]$ and hence the PH-transform measure is a member of CDRM. We examine PH_γ for $\gamma = 0.1$, $\gamma = 0.5$, and $\gamma = 0.9$. The weights q_i, $i = 1, \cdots, 100$ for ordered statistics and the weights w_i, $i = 1, \cdots, 100$ for $CVaR_\alpha$s are shown in Figure 5.11.

We see that portfolio optimizations over PH_γ are very close to two-point optimizations: minimization $CVaR_\alpha$ for $\alpha = 0.99$ and hence minimization of extreme losses as well as minimization of $CVaR_\alpha$ for $\alpha = 0$ and hence maximization of expected return. The PH_γ cares little about other percentiles of the loss distribution, as revealed in graphs on the left of Figure 5.11. In our numerical example, optimization w.r.t $PH_{0.9}$ is almost the same as return maximization.

The time series plots for optimal portfolios with realized and expected returns w.r.t the above the PH_γs are given in Figure 5.12. The corresponding summary statistics for the expected and realized returns are given in Table 5.8.

As we have discussed previously, minimization of PH_γ places more emphasis on maximizing the portfolio returns with lager γ. As a result, we see the expected returns for optimal portfolios increases with γ. Moreover, the difference in the mean expected return is surprisingly significant. The mean expected return for optimal portfolios w.r.t $PH_{0.9}$ is more than twice as much as the mean expected return for those w.r.t $PH_{0.5}$. Although the actual realized returns are lower than the expected returns, but we can still observe the same relationship between expected returns and the value of γ: the higher the γ, the higher the mean. We see in our example that the higher the γ, the higher the canonical Sharpe Ratio.

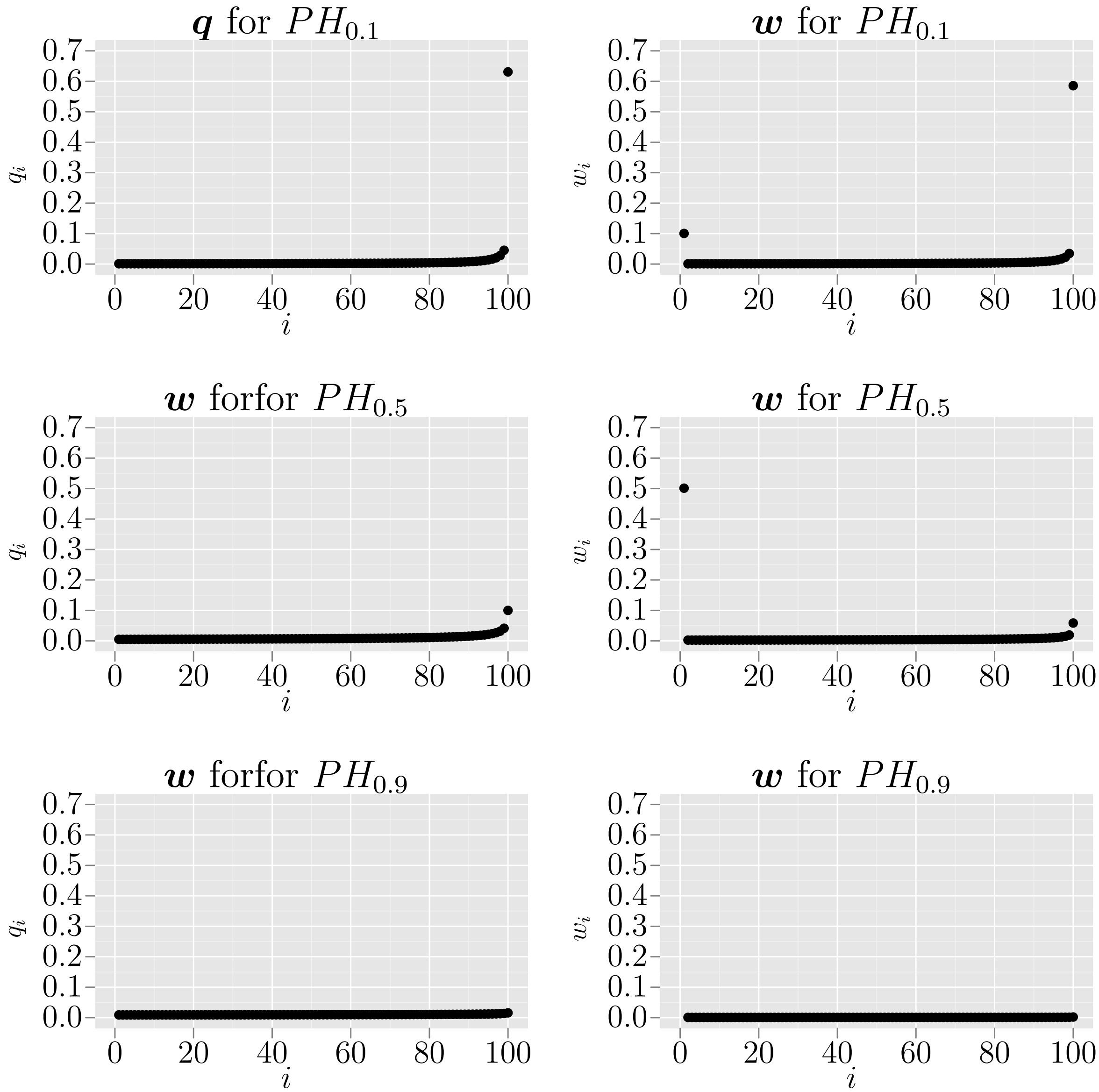

Figure 5.11: $\boldsymbol{q}$ and $\boldsymbol{w}$ for PH_γ

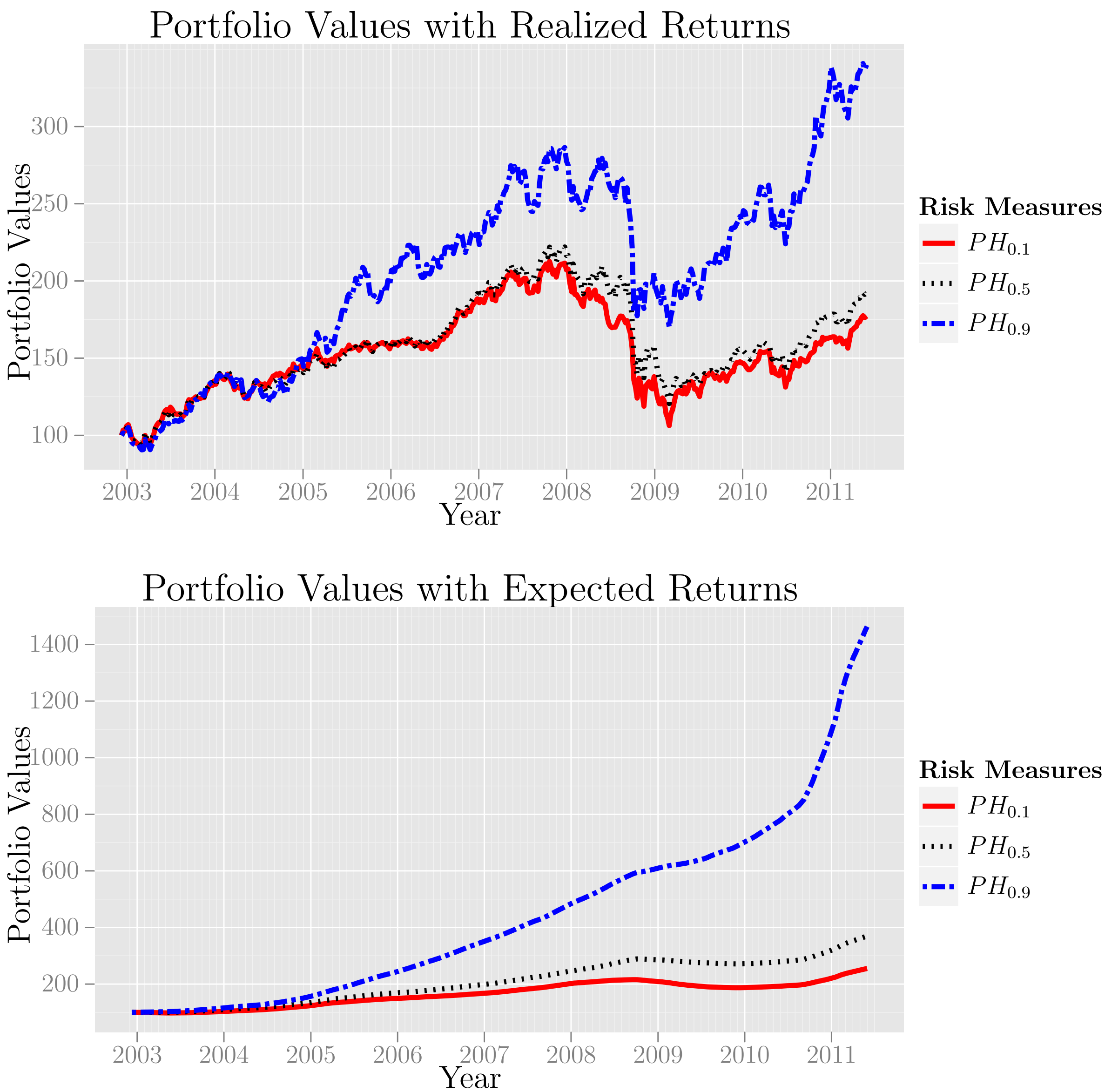

Figure 5.12: Time Series Plots for Optimal Portfolios w.r.t PH_γs

Optimal Returns	$PH_{0.1}$		$PH_{0.5}$		$PH_{0.9}$	
	Expected	Realized	Expected	Realized	Expected	Realized
Mean	0.00212	0.00130	0.00296	0.00148	0.00609	0.00277
Standard Deviation	0.00257	0.02218	0.00231	0.02091	0.00352	0.02622
Skewness	-0.19087	-0.26156	-0.19624	-0.83421	1.51062	-0.95739
Kurtosis	0.36133	5.14293	-0.13181	8.50931	3.96695	6.78000
Sharpe Ratio	0.82426	0.05844	1.28245	0.07091	1.72943	0.10574

Table 5.8: Summary Statistics for PH_γ Minimization Optimal Portfolios

Optimization with LB-Transform Measures

The LB-transform measure was introduced by Hürlimann[29] to investigate option strategies and for exchange risk modeling. The distortion function for LB-transform is given by

$$g_{LB}(x,\delta) = x^\delta, \qquad \delta \in (0,1] \tag{5.4}$$

g_{LB} is a concave function for $\delta \in (0,1]$ hence LB_γ is a member of CDRM. We examine g_{LB} for $\delta = 0.1$, $\delta = 0.5$, and $\delta = 0.9$. The weights q_i, $i = 1, \cdots, 100$ for ordered statistics and the weights w_i, $i = 1, \cdots, 100$ for $CVaR_\alpha$s are shown in Figure 5.13.

We see from Figure 5.14 that LB-transform measure places the main focus in minimizing $CVaR_\alpha$s for large αs. The parameter δ can be viewed as a measure of risk aversion. The smaller the δ, the closer the LB_δ minimization is to a $CVaR_{0.99}$ minimization. For large values of δ, we can see that the weights for $CVaR_\alpha$s are almost the same for all αs, which implies there is no emphasis on minimizing any one $CVaR_\alpha$ over the others.

The time series plots for expected and actual portfolio values with respect to the above three LB-transform measures are given in . The corresponding summary statistics for the expected and realized returns are given in Table 5.9.

Since smaller values of δ correspond to stronger emphasis on minimizing extreme losses, the underlying decision making process is more risk averse. As a result, we shall expect low mean return for small δs. Both the expected and actual portfolios in Figure 5.14 validate such property. We see that the higher the δ, the higher the canonical Sharpe Ratio. However, such observation is merely based on our empirical results and rigorous justifications is absent.

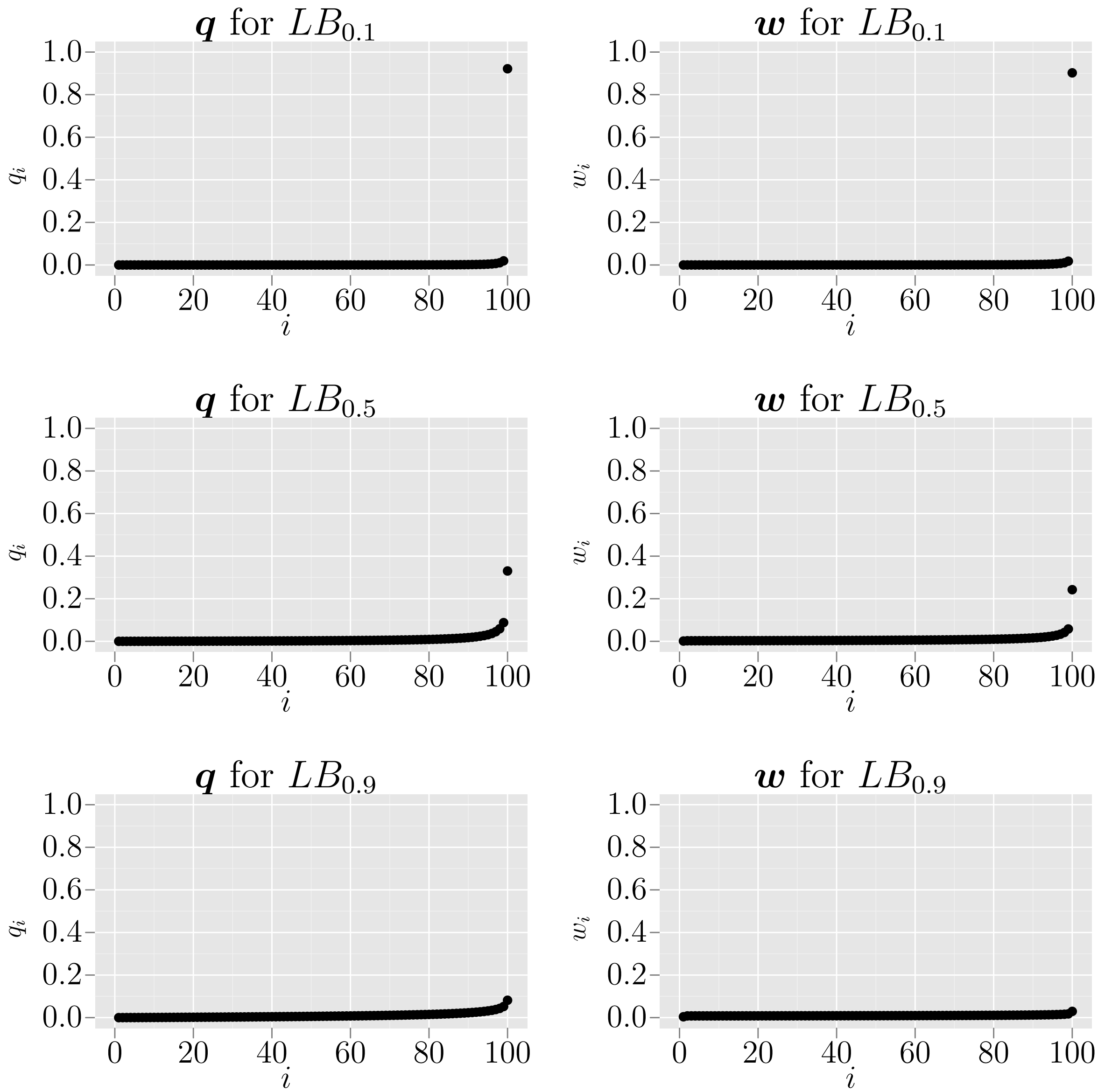

Figure 5.13: $\boldsymbol{q}$ and $\boldsymbol{w}$ for LB_δ

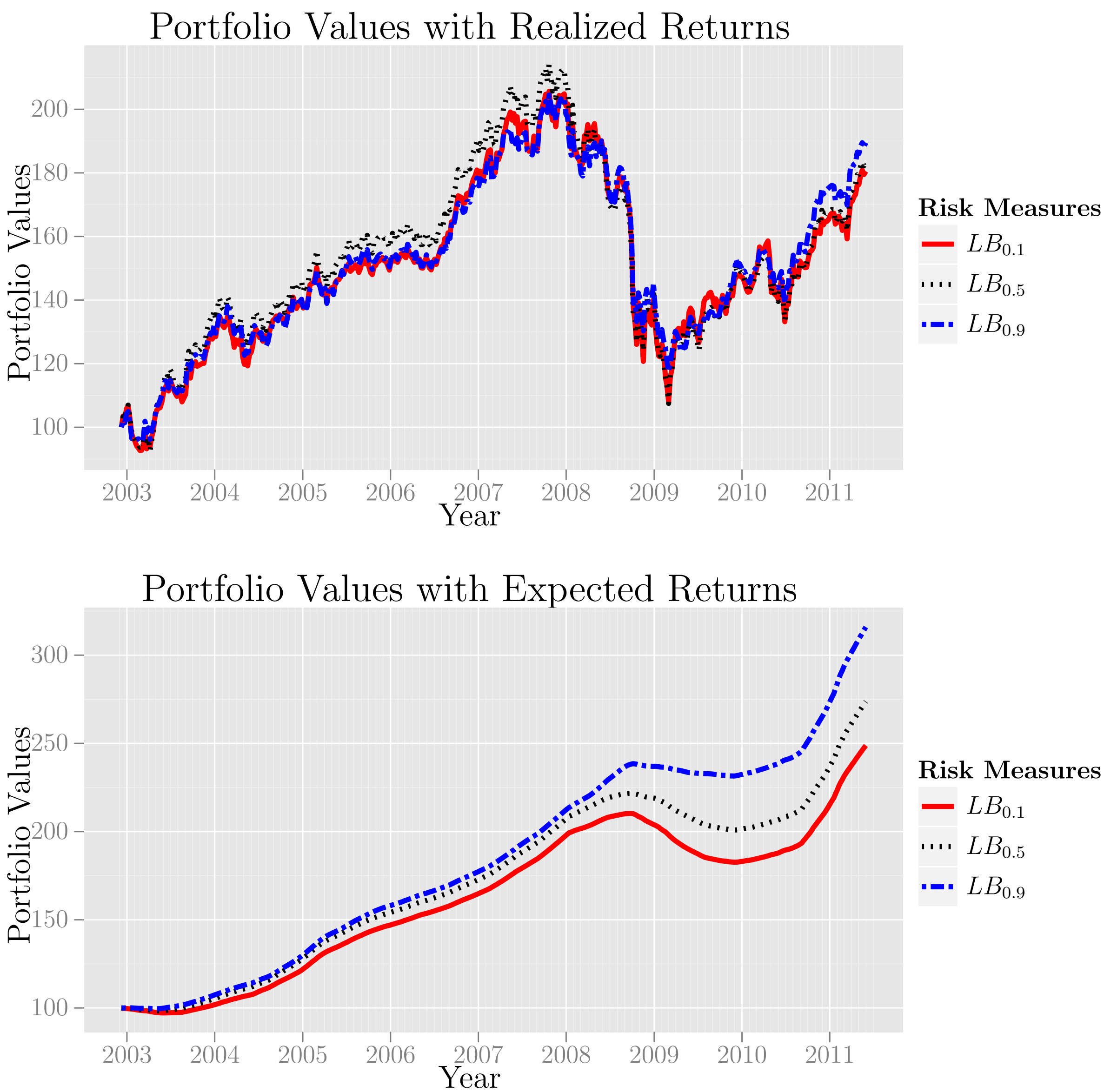

Figure 5.14: Time Series Plots for Optimal Portfolios w.r.t LB_δs

Optimal Returns	$LB_{0.1}$		$LB_{0.5}$		$LB_{0.9}$	
	Expected	Realized	Expected	Realized	Expected	Realized
Mean	0.00206	0.00134	0.00228	0.00137	0.00261	0.00145
Standard Deviation	0.00259	0.02230	0.00239	0.02130	0.00207	0.01893
Skewness	-0.16833	-0.22880	0.03581	-0.34008	0.41528	-0.80400
Kurtosis	0.29438	4.59996	0.28190	5.15387	0.47363	6.04230
Sharpe Ratio	0.79704	0.05995	0.95589	0.06439	1.26086	0.07645

Table 5.9: Summary Statistics for LB_δ Minimization Optimal Portfolios

In summary, we perform a final comparison among the best performing portfolios in the aforementioned four members of CDRMs, the $\frac{1}{n}$-portfolio, and the return maximization portfolio. Note that we can also view the return maximization problem as a CDRM minimization problem by minimizing $CVaR_0$ of the negative returns. The resulting time series are plotted in Figure 5.15 the summary statistics for $\frac{1}{n}$ portfolio and profit maximization portfolio is given in Table 5.10.

Optimal Returns	$\frac{1}{n}$-Portfolio		Max Return Portfolio	
	Expected	Realized	Expected	Realized
Mean	0.00168	0.00208	0.00639	0.00279
Standard Deviation	0.00279	0.03038	0.00352	0.02947
Skewness	-0.33372	0.25175	1.52513	-0.73268
Kurtosis	0.64192	13.73943	4.02045	4.95543
Sharpe Ratio	0.60129	0.06854	1.81415	0.09480

Table 5.10: Summary Statistics for $\frac{1}{n}$-Portfolio and Return Maximization Portfolio

We see that for our selection of stocks, the return maximization produces the best portfolio in terms of its terminal wealth and the mean portfolio returns. However, this should not be a practical recommendation for portfolio manager because it might bear unacceptably high risks. For instance, the standard deviation of the expected returns for return maximization portfolios is the highest among all portfolios we have discussed in this section. Investors should consider their own risk appetites and choose an appropriate risk measure in every investment decision.

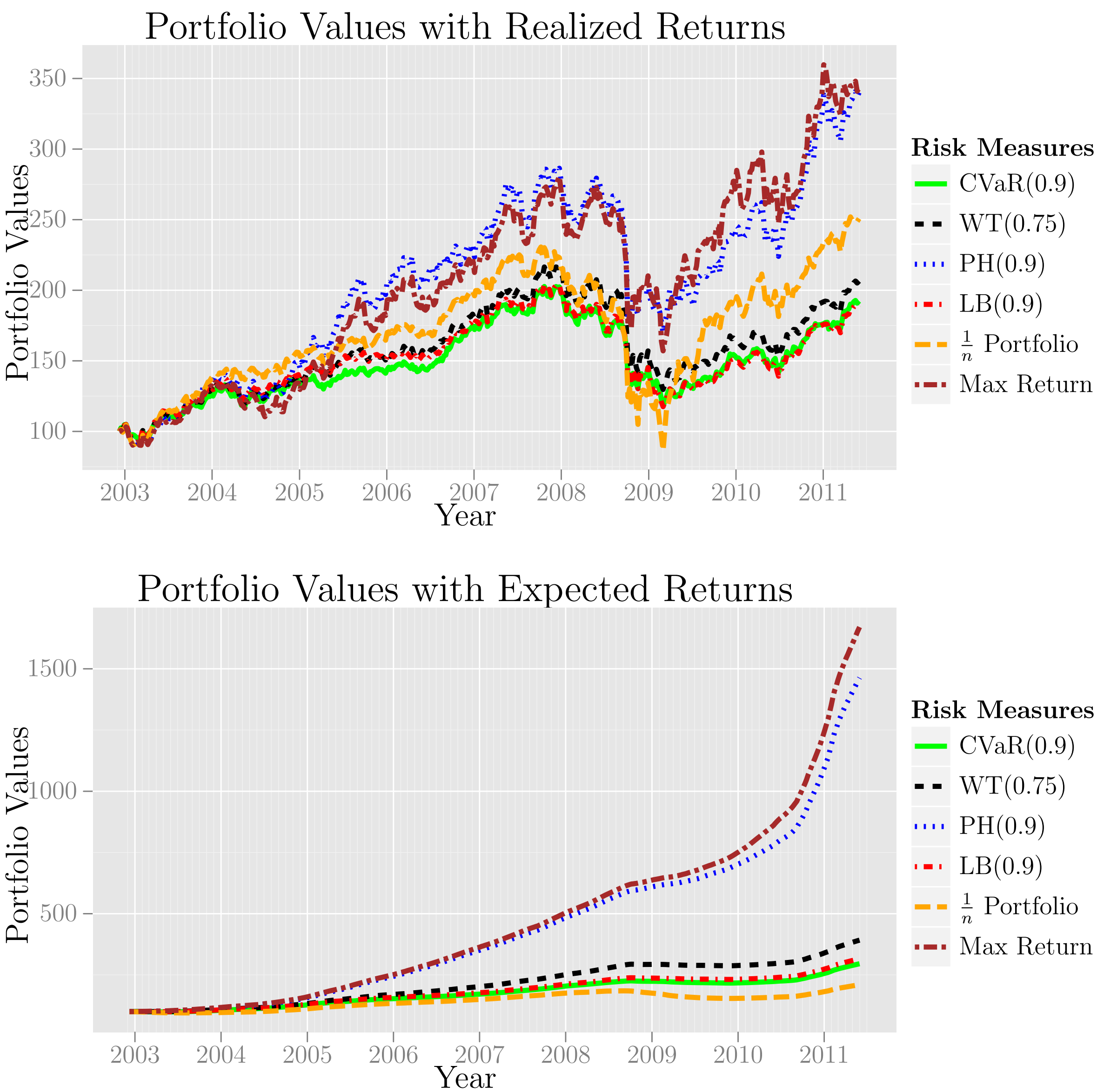

Figure 5.15: Time Series Plots for Optimal Portfolios for Various CDRMs

Chapter 6

Conclusions and Future Directions

In this thesis we extend the convex formulation of CVaR and its linearization scheme developed in Rockafellar et al. [50] [51] and Krokhmal et al. [36] to a general class of risk measure, the CDRM. The class of risk measure includes members such as CVaR, WT-measures, PH-transform measures, and lookback-distortion measures. The risk measures were mainly used for calculating risk levels (or insurance premiums) for a given portfolio. With our CDRM optimization framework, these risk measures can be used for optimizing risk levels (or insurance premiums) within a give set of feasible portfolios.

The Finite Generation Theorem for CDRM in Bertsimas et al. [11] assumes discrete uniform distribution for portfolio losses and states that any CDRM can be expressed as a convex combination of finite many CVaRs at different confidence levels. We extend the Finite Generation Theorem to general discrete distributions portfolio loss and the convex combination representation is still valid.

We show that any CVaR risk-shaping problem is equivalent to a CDRM utility maximization problem in a sense that both problem traces out the same multi-dimensional efficient frontier. Furthermore, portfolio selection problems with CVaR objective can be converted to problems with one CVaR constraint, which are special cases of CVaR risk-shaping problems. As a result, any portfolio problem with CVaR objective/constraint(s) can be viewed as a CDRM utility maximization problem. Conversely, any CDRM utility maximization problem can be viewed as a CVaR risk-shaping problem.

Besides the CDRM utility maximization problem, we propose three other formulations of CDRM portfolio selection problems: return maximization with a CDRM constraint, CDRM minimization with a return constraint, and CDRM-based Sharpe Ratio maximization. We establish equivalences among these four formulations in a sense that they trace out the same efficient frontier. For any of the four problems with a preselected parameter, we provide the expressions of the implied parameters for the other three problems. Last but not least, we explore the paring relationships between CDRM utility maximization problems and CDRM-based Sharpe Ratio maximization problems as well as between return maximization problems and CDRM minimization problems, both analytically and geometrically.

We perform cases studies to provide numerical examples and verifications of the aforementioned relationships and paring properties. In addition, we perform empirical analysis of the efficiency of $\frac{1}{n}$-portfolio in terms of the trade off between portfolio return and 95%-CVaR of portfolio losses. We find that the $\frac{1}{n}$-portfolio is highly inefficient in the sense that it is far away from the efficient frontiers at different times. Our numerical results show that $\frac{1}{n}$-portfolio is farther away from the efficient frontier during periods of market declines than it is during market increases.

Finally, we explore the properties of four members of CDRM, CVaRs, WT-measures, PH-measures, and LB-measures by comparing the performance optimal portfolios and their returns w.r.t these members. Although all these members are risk measures, optimizations over these CDRMs in fact provide risk minimization and return maximization simultaneously. However, each of these four member has unique risk appetite in terms of their relative weights towards minimizing CVaRs at different confidence levels. We also find several relationships between the parameter value for each of these CDRMs and the corresponding Sharpe Ratios in our numerical results.

Results in this thesis are the initial steps and there are still many to take. Some possible extensions of our work for future research are:

1. Explore more efficient solution methods for CDRM optimization problems by exploiting the structural properties of its programming formulation. To CDRM portfolio selection problem with m scenarios via LP we need m CVaR constraints in general, each of which requires m additional auxiliary variables and $2m$ auxiliary constraints. With a reasonable number of scenarios such as $m = 100$ or $m = 1,000$, we may need of solve an LP with at least $10,000$ and $1,000,000$ constraints, which could be time consuming. One can explore the structural properties of CDRM optimization problems (diagonal matrix for auxiliary constraints) and employ decomposition solution methods such as the Bender's Decomposition to solve it more efficiently.

2. Solve CDRM optimization problems via stochastic programming. Künzi et al. [39] considered solving CVaR minimization problems using stochastic programming. Our

CDRM optimization framework is based largely on the CVaR optimization framework, it is of interest to consider solving CDRM optimization problems via similar methodologies.

3. Consider CDRM in multi-period models. We develope our CDRM optimization framework as an one-period model. Yet in practice portfolio managers may want to incorporate information for periods ahead and hence want to make decisions accordingly. A multi-periods model is needed in those situations. For example, Fábián [23] considered CVaR portfolio selection problems in two-stage stochastic models. It is of interest to extend the models therein to CDRM portfolio selection problems.

4. Explore other members of CDRM. The CDRMs considered in this thesis are DRMs with concave distortion functions. It is of interest to consider some known subclasses of CRMs such as higher moment coherent risk measures defined in Krokhmal [35] and to verify whether they are members of CDRM.

APPENDICES

Appendix A

Auxiliary Proof

A.1 Properties of Weights in Equation (4.6)

1. $w_i \geq 0$ for $i = 1, \cdots, m$ By assumption $p_1 > 0$ and $g : [0,1] \mapsto [0,1]$ is nondecreasing, thus $q_1 \geq 0$ and

$$w_1 = \frac{q_1}{p_1} \geq 0$$

Furthermore, since g is concave, for $i = 2, \cdots, m$

$$\begin{aligned}
& (q_i - \frac{p_i}{p_{i-1}} q_{i-1}) \\
= & \{[g(1 - F(l_{i-1})) - g(1 - F(l_i))] - \frac{p_i}{p_{i-1}}[g(1 - F(l_{i-2})) - g(1 - F(l_{i-1}))]\} \\
= & \frac{p_i + p_{i-1}}{p_{i-1}}\{g(1 - F(l_{i-1})) - [\frac{p_{i-1}}{p_i + p_{i-1}} g(1 - F(l_i)) + \frac{p_i}{p_i + p_{i-1}} g(1 - F(l_{i-2}))]\} \\
\geq & \frac{p_i + p_{i-1}}{p_{i-1}}\{g(1 - F(l_{i-1})) - g(1 - F(l_{i-1}))\} \\
\geq & 0
\end{aligned}$$

2. $\sum_{i=1}^{m} w_i = 1$

By construction

$$
\begin{aligned}
& \sum_{i=1}^{m} w_i \\
= \quad & \frac{q_1}{p_1} + \sum_{i=2}^{m} (q_i - \frac{p_i}{p_{i-1}} q_{i-1}) \frac{1 - F(l_{i-1})}{p_i} \\
= \quad & \frac{q_1}{p_1} + \sum_{i=2}^{m} q_i \frac{1 - F(l_{i-1})}{p_i} - \sum_{i=2}^{m} q_{i-1} \frac{1 - F(l_{i-1})}{p_{i-1}} \\
= \quad & \frac{q_1}{p_1} + [\sum_{i=2}^{m-1} q_i \frac{1 - \sum_{j=1}^{i-1} p_j}{p_i} + q_m \frac{1 - \sum_{j=1}^{m-1} p_j}{p_m}] - [\sum_{i=3}^{m} q_{i-1} \frac{1 - \sum_{j=1}^{i-1} p_j}{p_{i-1}} + q_1 \frac{1 - p_1}{p_1}] \\
= \quad & q_1 (\frac{1}{p_1} - \frac{1 - p_1}{p_1}) + [\sum_{i=2}^{m-1} q_i \frac{1 - \sum_{j=1}^{i-1} p_j}{p_i} + \sum_{i=2}^{m-1} q_i \frac{1 - \sum_{j=1}^{i} p_j}{p_i}] + q_m \frac{p_m}{p_m} \\
= \quad & \sum_{i=1}^{m} q_m \\
= \quad & 1
\end{aligned}
$$

Appendix B

Auxiliary Table

B.1 The Companies Selected for Investment Portfolio Construction

GICS Sector	Company	Ticker Symbol
Consumer Discretionary	Ford Motor	F
	McDonald's Corp.	MCD
Consumer Staples	Coca Cola Co.	KO
	Wal-Mart Stores	WMT
Energy	Marathon Oil Corp.	MRO
	Exxon Mobil Corp.	XOM
Financials	American Intl Group Inc	AIG
	Citigroup Inc.	C
Health Care	CIGNA Corp.	CI
	Humana Inc.	HUM
Industrials	General Dynamics	GD
	General Electric	GE
Information Technology	Microsoft Corp.	MSFT
	National Semiconductor	NSM
Materials	FMC Corporation	FMC
	International Paper	IP
Telecommunication Services	AT&T Inc	T
	Verizon Communications	VZ
Utilities	American Electric Power	AEP
	Entergy Corp.	ETR

Table B.1: Companies Selected for Investment Portfolio Construction Case Study

References

[1] C. Acerbi, C. Nordio, and C. Sirtori. Expected shortfall as a tool for financial risk management. *Arxiv preprint cond-mat/0102304*, 2001.

[2] C. Acerbi and P. Simonetti. Portfolio optimization with spectral measures of risk. *Preprint*, 2002.

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

[4] S. Basak and A. Shapiro. Value-at-risk based risk management: Optimal policies and asset prices. *Review of Financial Studies*, 14(2):371, 2001.

[5] W.J. Baumol. An expected gain-confidence limit criterion for portfolio selection. *Management Science*, 10(1):174–182, 1963.

[6] V.S. Bawa and E.B. Lindenberg. Capital market equilibrium in a mean-lower partial moment framework. *Journal of Financial Economics*, 5(2):189–200, 1977.

[7] M.S. Bazaraa, J.J. Goode, and C.M. Shetty. Constraint qualifications revisited. *Management Science*, 18(9):567–573, 1972.

[8] F. Bellini and C. Caperdoni. Coherent distortion risk measures and higher-order stochastic dominances. *North American Actuarial Journal*, 11(2):35, 2007.

[9] S. Benartzi and R.H. Thaler. Naive diversification strategies in defined contribution saving plans. *American Economic Review*, pages 79–98, 2001.

[10] S. Benati and R. Rizzi. A mixed integer linear programming formulation of the optimal mean/value-at-risk portfolio problem. *European Journal of Operational Research*, 176(1):423–434, 2007.

[11] D. Bertsimas and D.B. Brown. Constructing uncertainty sets for robust linear optimization. *Operations Research*, 57(6):1483–1495, 2009.

[12] D. Bertsimas, G.J. Lauprete, and A. Samarov. Shortfall as a risk measure: Properties, optimization and applications. *Journal of Economic Dynamics and Control*, 28(7):1353–1381, 2004.

[13] A. Campana and P. Ferretti. Distortion risk measures and discrete risks. *Game Theory and Information*, 2005.

[14] A. Charnes and W.W. Cooper. Programming with linear fractional functionals. *Naval Research Logistics Quarterly*, 9(3-4):181–186, 1962.

[15] P. Chunhachinda, K. Dandapani, S. Hamid, and A.J. Prakash. Portfolio selection and skewness: Evidence from international stock markets. *Journal of Banking & Finance*, 21(2):143–167, 1997.

[16] G.B. Dantzig. Maximization of a linear function of variables subject to linear inequalities. *New York*, 1951.

[17] R. Dembo and D. Rosen. The practice of portfolio replication. a practical overview of forward and inverse problems. *Annals of Operations Research*, 85:267–284, 1999.

[18] V. DeMiguel, L. Garlappi, and R. Uppal. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *Review of Financial Studies*, 22(5):1915, 2009.

[19] D. Denneberg. *Non-Additive Measure and Integral*. Springer, 1994.

[20] J. Dhaene, S. Vanduffel, Q. Tang, M.J. Goovaerts, R. Kaas, and D. Vyncke. Capital requirements, risk measures and comonotonicity. *Belgian Actuarial Bulletin*, 4:53–61, 2004.

[21] J. Dhaene, S.S. Wang, V.R. Young, and M.J. Goovaerts. Comonotonicity and maximal stop-loss premiums. *Bulletin of the Swiss Association of Actuaries*, 2:99–113, 2000.

[22] W. Dinkelbach. On nonlinear fractional programming. *Management Science*, 13(7):492–498, 1967.

[23] C.I. Fábián. Handling cvar objectives and constraints in two-stage stochastic models. *European Journal of Operational Research*, 191(3):888–911, 2008.

[24] A.A. Gaivoronski and G. Pflug. Value-at-risk in portfolio optimization: Properties and computational approach. *Journal of Risk*, 7(2):1–31, 2005.

[25] M.J. Goovaerts, F. de Vylder, and J. Haezendonck. *Insurance Premiums: Theory and Applications*. North-Holland, 1984.

[26] C. Gourieroux and W. Liu. Sensitivity analysis of distortion risk measures. *Working Papers*, 2006.

[27] C.R. Harvey, J. Liechty, M. Liechty, and P. Müller. Portfolio selection with higher moments. *Quantitative Finance*, 10(5):469–485, 2010.

[28] P.J. Huber, E. Ronchetti, and MyiLibrary. *Robust Statistics*, volume 1. Wiley Online Library, 1981.

[29] W. Hürlimann. Inequalities for lookback option strategies and exchange risk modelling. In *Proceedings of the 1st Euro-Japanese Workshop on Stochastic Modelling for Finance, Insurance, Production and Reliability, Brussels*, 1998.

[30] P. Jorion. Value at risk: A new benchmark for measuring derivatives risk. *Irwin Professional Pub*, 1996.

[31] P. Jorion. Risk management lessons from long-term capital management. *European Financial Management*, 6(3):277–300, 2000.

[32] E. Jurczenko, B. Maillet, and P. Merlin. Hedge funds portfolio selection with higher-order moments: A non-parametric mean-variance-skewness-kurtosis efficient frontier. *Multi Moment Asset Allocation and Pricing Models, pp. 51S66. John Wiley & Sons*, 2006.

[33] L.V. Kantorovich. Mathematical methods of organizing and planning production. *Management Science*, 6(4):366–422, 1960.

[34] W. Karush. Minima of functions of several variables with inequalities as side constraints. *Department of Mathematics. University of Chicago*, 1939.

[35] P. Krokhmal. Higher moment coherent risk measures. *Quantitative Fund Management*, page 271, 2008.

[36] P.A. Krokhmal, J. Palmquist, and S. Uryasev. Portfolio optimization with conditional value-at-risk objective and constraints. *Journal of Risk*, 4:43–68, 2002.

[37] P.A. Krokhmal and P. Soberanis. Risk optimization with p-order conic constraints: A linear programming approach. *European Journal of Operational Research*, 201(3):653–671, 2010.

[38] H.W. Kuhn and A.W. Tucker. Nonlinear programming. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability. July 31-August 12, 1950. Statistical Laboratory of the University of California, Berkeley. Berkeley, Calif.: University of California Press, 1951. 666 pp. Editor: Jerzy Neyman, p. 481-492*, volume 1, pages 481–492, 1951.

[39] A. Künzi-Bay and J. Mayer. Computational aspects of minimizing conditional value-at-risk. *Computational Management Science*, 3(1):3–27, 2006.

[40] S. Kusuoka. On law invariant coherent risk measures. *Advances in Mathematical Economics*, 3:83–95, 2001.

[41] T.Y. Lai. Portfolio selection with skewness: A multiple-objective approach. *Review of Quantitative Finance and Accounting*, 1(3):293–305, 1991.

[42] C. Lin. Comments on "a mixed integer linear programming formulation of the optimal mean/value-at-risk portfolio problem". *European Journal of Operational Research*, 194(1):339–341, 2009.

[43] J. Lintner. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics*, 47(1):13–37, 1965.

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

[45] H.M. Markowitz. *Portfolio Selection: Efficient Diversification of Investments.* Wiley, 1991.

[46] R.C. Merton. An analytic derivation of the efficient portfolio frontier. *Journal of financial and quantitative analysis*, 7(04):1851–1872, 1972.

[47] D.W. Peterson. A review of constraint qualifications in finite-dimensional spaces. *SIAM Review*, 15(3):639–654, 1973.

[48] M. Pritsker. Evaluating value at risk methodologies: Accuracy versus computational time. *Journal of Financial Services Research*, 12(2):201–242, 1997.

[49] A. Puelz. Value-at-risk based portfolio optimization. *Stochastic Optimization: Algorithms and Applications*, page 279, 2001.

[50] R.T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 2000.

[51] R.T. Rockafellar and S. Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, 26(7):1443–1471, 2002.

[52] A.D. Roy. Safety first and the holding of assets. *Econometrica: Journal of the Econometric Society*, pages 431–449, 1952.

[53] S. Schaible and T. Ibaraki. Fractional programming. *European Journal of Operational Research*, 12(4):325–338, 1983.

[54] W.F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3):425–442, 1964.

[55] C. Singh. Optimality conditions in fractional programming. *Journal of Optimization Theory and Applications*, 33(2):287–294, 1981.

[56] P.A. Soberanis. Risk optimization with p-order conic constraints. 2009.

[57] A.L. Soyster. Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research*, 21(5):1154–1157, 1973.

[58] I.M. Stancu-Minasian. *Fractional Programming*. Kluwer academic publishers, 1997.

[59] M.C. Steinbach. Markowitz revisited: Mean-variance models in financial portfolio analysis. *SIAM Review*, 43(1):31–85, 2001.

[60] S.F. Tantawy. Using feasible directions to solve linear fractional programming problems. *Australian Journal of Basic and Applied Sciences*, 1(2):109–114, 2007.

[61] S.F. Tantawy. An iterative method for solving linear fraction programming (lfp) problem with sensitivity analysis. *Mathematical and Computational Applications*, 13(3):147–151, 2008.

[62] Ito Tomoyuki. *Multiperiod Portfolio Selection with Second-order Cone Constraints*. PhD thesis, Kyoto University, 2001.

[63] J. Von Neumann and O. Morgenstern. Theory of games and economical behavior. *Princeton (1944, 1947, 1953)*.

[64] S.S. Wang. Insurance pricing and increased limits ratemaking by proportional hazards transforms. *Insurance: Mathematics and Economics*, 17(1):43–54, 1995.

[65] S.S. Wang. Premium calculation by transforming the premium layer density. *ASTIN Bulletin*, 26(1):71–92, 1996.

[66] S.S. Wang. A class of distortion operators for pricing financial and insurance risks. *The Journal of Risk and Insurance*, 67(1):15–36, 2000.

[67] S.S. Wang. A risk measure that goes beyond coherence. *Research Report*, pages 01–18, 2001.

[68] S.S. Wang and J. Dhaene. Comonotonicity, correlation order and premium principles. *Insurance: Mathematics and Economics*, 22(3):235–242, 1998.

[69] S.S. Wang, V.R. Young, and H.H. Panjer. Axiomatic characterization of insurance prices. *Insurance: Mathematics and Economics*, 21(2):173–183, 1997.

[70] H. Windcliff and P.P. Boyle. The 1/n pension investment puzzle. *North American Actuarial Journal*, 8:32–45, 2004.

[71] J.L. Wirch and M.R. Hardy. A synthesis of risk measures for capital adequacy. *Insurance: Mathematics and Economics*, 25(3):337–347, 1999.

[72] J.L. Wirch and M.R. Hardy. Distortion risk measures. coherence and stochastic dominance. *Unpublished Working Paper (http://www. gloriamundi. org)*, 2006.