Imperial College London
Department of Computing

Distributionally Robust Optimization with Applications to Risk Management

Steve Zymler

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy in Computing of Imperial College London and
the Diploma of Imperial College, June 2010

Abstract

Many decision problems can be formulated as mathematical optimization models. While deterministic optimization problems include only known parameters, real-life decision problems almost invariably involve parameters that are subject to uncertainty. Failure to take this uncertainty under consideration may yield decisions which can lead to unexpected or even catastrophic results if certain scenarios are realized.

While stochastic programming is a sound approach to decision making under uncertainty, it assumes that the decision maker has complete knowledge about the probability distribution that governs the uncertain parameters. This assumption is usually unjustified as, for most realistic problems, the probability distribution must be estimated from historical data and is therefore itself uncertain. Failure to take this distributional modeling risk into account can result in unduly optimistic risk assessment and suboptimal decisions. Furthermore, for most distributions, stochastic programs involving chance constraints cannot be solved using polynomial-time algorithms.

In contrast to stochastic programming, distributionally robust optimization explicitly accounts for distributional uncertainty. In this framework, it is assumed that the decision maker has access to only partial distributional information, such as the first- and second-order moments as well as the support. Subsequently, the problem is solved under the worst-case distribution that complies with this partial information. This worst-case approach effectively immunizes the problem against distributional modeling risk.

The objective of this thesis is to investigate how robust optimization techniques can be used for quantitative risk management. In particular, we study how the risk of large-scale derivative portfolios can be computed as well as minimized, while making minimal assumptions about the probability distribution of the underlying asset returns. Our interest in derivative portfolios stems from the fact that careless investment in derivatives can yield large losses or even bankruptcy. We show that by employing robust optimization techniques we are able to capture the substantial risks involved in derivative investments. Furthermore, we investigate how distributionally robust chance constrained programs can be reformulated or approximated as tractable optimization problems. Throughout the thesis, we aim to derive tractable models that are scalable to industrial-size problems.

Acknowledgements

I would like to begin by thanking my supervisor, Prof. Berç Rustem, for the constant motivation and accessibility he has given during the four years we have worked together. I benefited tremendously from his encouragement, insights, and challenging research interests. Aside from the research related support he has given me, I will also always remember the interesting discussions we had regarding many socio-political issues. Unfortunately, we have yet to discover a tractable approach to solve the Middle-East conflict.

I also cannot overstate my gratitude to my second supervisor, Dr. Daniel Kuhn. His constant energy and research interests have been an example and immense inspiration for me. He has been a perfect mentor who taught me how to approach challenging problems using well founded mathematical techniques. Furthermore, he pushed me to discover relationships between mathematical models, which has greatly benefited this thesis.

I am thankful to Prof. Nicos Christofides and Dr. Michael Bartholomew-Biggs for being on my defense committee. I also want to thank Prof. Aharon Ben-Tal for valuable discussions on topics related to this thesis.

I am grateful to my friends and colleagues, Wolfram Wiesemann, Paul-Amaury Matt, Paul Bilokon, Raquel Fonseca, Panos Parpas, Polyxeni Kleniati, George Tzallas, Angelos Tsoukalas, Nikos Papadakos, Dimitra Bampou, Phebe Vayanos, Michael Hadjiyiannis, Michalis Kapsos, Angelos Georghiou, Fook Kong, Paula Rocha, Chan Kian, Kye Ye, and many others, who made my time as a graduate student an extremely enjoyable period in my life.

I am blessed with a loving family. Without their unconditional love, support, and encouragement this thesis would not have been possible. I fondly thank my parents, for always believing in me, for their love, and for their immense support throughout my studies. I also thank my brother who was there to cheer me up during difficult days. Finally, I would like to thank my girlfriend. Her love, encouragement, and patience were an immense help towards the completion of this thesis.

Dedication

To my parents – their unconditional love and support made this thesis possible.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation and Objectives

At the time of writing, the world has gone through a period of unprecedented financial turbulence. The crisis has resulted in the collapse of large financial institutions, the bailout of banks by national governments and severe downturns in global stock markets. Indeed, many economists consider it to be the worst financial crisis since the Great Depression of the 1930s. The crisis resulted in the stagnation of worldwide economies due to the tightening of credit and decline in international trade. It is now often referred to as the Great Recession. While the global economies are starting to recover from the crisis, its ripple effects are still propagating through the system and investors are exposed to considerable *uncertainty*.

The crisis serves us to illustrate the importance of reliable risk management. Investors face the challenging problem of how to distribute their current wealth over a set of available assets with the goal to earn the highest possible future wealth. However, in order to decide on the portfolio allocations, the investor must take into consideration that the future asset returns are uncertain. The investor's portfolio allocation problem is traditionally solved using *stochastic programming*. Stochastic programming implicitly assumes that the investor has complete knowledge about the probability distribution of the asset returns. The framework offers a large variety of *risk measures*, which are functions that estimate the risk of a given portfolio. Popular examples of risk measures are the variance of the portfolio return, and the Value-at-Risk, which is equal to a given quantile of the portfolio loss distribution. Subsequently, stochastic programming aims to find the portfolio with the lowest associated risk that satisfies additional constraints imposed by the investor on the portfolio allocations.

While stochastic programming is a sound framework that effectively enables the investor to trade off risk and return, the underlying assumption that the investor has full and accurate knowledge about the probability distribution of the asset returns is often unjustified. Indeed, typically the investor must estimate the probability distribution from historical realizations of the asset returns. After observing a limited amount of relevant historical observations, the investor is often unable to accurately determine the probability distribution of the asset returns. This drawback is a serious concern when it comes to estimating the risk associated with a given portfolio. For example, when estimating the Value-at-Risk of a portfolio, we are usually interested in the losses that occur in the "tails" of the portfolio loss distribution, that is, the extreme events that occur with a very low probability. However, it is unlikely that we can accurately estimate these events after observing a limited amount of historical observations. In fact, the recent market crash, discussed above, is precisely one of such low probability events that would have been very difficult, if not impossible, to predict statistically. Thus, using stochastic programming on the basis of inaccurate probabilistic information can yield careless and overly optimistic decisions.

In contrast to stochastic programming, *robust optimization* is an alternative modeling framework for decision making under uncertainty that does not require strong assumptions about the probability distribution of the uncertain parameters in the problem. In the context of the asset allocation problem, the asset returns are assumed to be unknown but confined to an *uncertainty set*, which reflects the decision maker's uncertainty about the asset returns. Although the investor is free to choose the shape and size of the uncertainty set, it often constructed on the basis of some partial distributional information, such as the first- and second-order moments as well as the support of the random asset returns. Robust optimization models aim to find the best decision in view of the *worst-case realization* of the asset returns within this uncertainty set. It is important to note that this worst-case optimization approach offers us guarantees on the portfolio return: whenever the asset returns are realized within the prescribed uncertainty set, the realized portfolio return will be greater than or equal to the calculated worst-case portfolio return.

A closely related modeling paradigm to robust optimization is *distributionally robust optimization*. Distributionally robust optimization is similar to stochastic programming, but explicitly accounts for distributional uncertainty. In this framework, it is assumed that the decision maker has access to only partial distributional information, such as the first- and second-order moments as well as the support of the random asset returns. The investor then considers the set of *all* probability distributions of the asset returns that match the known partial distributional information. Subsequently, the problem is solved under the *worst-case distribution* within this set. Whenever the "true" (but unknown) distribution lies somewhere within this set, the investor is guaranteed that the actual risk will be lower that the calculated worst-case risk. This worst-case approach effectively immunizes the problem against distributional modeling risk.

The main aim of this thesis is to employ (distributionally) robust optimization techniques to elaborate new decision making models for investment problems that: (i) avoid making strong assumptions about the probability distribution of the random parameters in the problem, (ii) provide guarantees about the risk the investor is exposed to, and (iii) are tractably solvable and therefore scalable to realistic problem sizes. More specifically, the objectives of this thesis are to address the following problems:

(1) How can derivatives be incorporated into the robust portfolio optimization framework without compromising the tractability of the problem? Furthermore, robust portfolio optimization only provides weak guarantees when the asset returns are realized *within* the uncertainty sets. We therefore wish to explore how the derivatives can provide insurance against unexpected events when the asset returns are realized *outside* the uncertainty sets. How does the insurance affect the portfolio performance and what can be said about the tradeoff between these weak and strong guarantees?

(2) Value-at-Risk is a popular financial risk measure, but it assumes that the probability distribution of the underlying asset returns is known precisely. Furthermore, it is a non-convex

function of the portfolio weights, which makes it intractable to optimize. These difficulties are further compounded when the portfolio contains derivatives. We wish to investigate how the Value-at-Risk of large-scale derivative portfolios can be optimized in a tractable manner, while making few assumptions about the probability distribution of the underlying assets.

(3) In stochastic programming, we often wish to express that a system of constraints must be satisfied with a given probability. The arising chance constrained programs are usually intractable to solve. We wish to explore how distributionally robust optimization techniques can be used to find conservative but tractable approximations of such chance constrained programs.

1.2 Contributions and Structure of the Thesis

In this thesis, we investigate how robust optimization techniques can be used for quantitative risk management. In particular, we study how the risk of large-scale derivative portfolios can be computed as well as minimized, while making minimal assumptions about the probability distribution of the underlying asset returns. Our interest in derivative portfolios stems from the fact that careless investment in derivatives can yield large losses or even bankruptcy. We show that by employing robust optimization techniques we are able to capture as well as minimize the substantial risks involved in derivative investments. Furthermore, we investigate how distributionally robust chance constrained programs can be reformulated or approximated as tractable optimization problems. Throughout the thesis, we aim to derive tractable models that are scalable to industrial-size problems.

Apart from a review of the background theory in Chapter 2 and conclusions in Chapter 6, the thesis is divided into three chapters, which can be summarized as follows.

In Chapter 3 we investigate how simple derivatives, such as put and call options, can be incorporated into the robust portfolio optimization framework. Robust portfolio optimization aims to maximize the worst-case portfolio return given that the asset returns are allowed to vary within a prescribed uncertainty set. If the uncertainty set is not too large, the resulting portfolio performs well under normal market conditions. However, its performance may substantially degrade in the presence of market crashes, that is, if the asset returns materialize far outside of the uncertainty set. We propose a novel robust optimization model for designing portfolios that include European-style options. This model trades off weak and strong guarantees on the worst-case portfolio return. The weak guarantee applies as long as the asset returns are realized within the prescribed uncertainty set, while the strong guarantee applies for all possible asset returns, including those that are realized outside the uncertainty set. The resulting model constitutes a convex second-order cone program, which is amenable to efficient numerical solution procedures. We evaluate the model using simulated and empirical backtests and analyze the impact of the insurance guarantees on the portfolio performance. The contents of this chapter are published in

1. S. Zymler, B. Rustem, and D. Kuhn. *Robust portfolio optimization with derivative insurance guarantees.* Under revision for the European Journal of Operations Research, 2010.

In Chapter 4 we study how the Value-at-Risk (VaR), a popular financial risk measure, of large-scale derivative portfolios can be minimized while making weak assumptions about the probability distribution of the underlying asset returns. Portfolio optimization problems involving VaR are often computationally intractable and require complete information about the return distribution of the portfolio constituents, which is rarely available in practice. These difficulties are further compounded when the portfolio contains derivatives. We develop two tractable conservative approximations for the VaR of a derivative portfolio by evaluating the worst-case VaR over all return distributions of the derivative underliers with given first- and second-order moments. The derivative returns are modelled as convex piecewise linear or—by using a delta-gamma approximation (a second-order Taylor expansion)—as (possibly non-convex) quadratic functions of the returns of the derivative underliers. These models lead to new Worst-Case Polyhedral VaR (WCPVaR) and Worst-Case Quadratic VaR (WCQVaR) approximations, respectively. WCPVaR is a suitable VaR approximation for portfolios containing long positions in European options expiring at the end of the investment horizon, whereas WCQVaR is suitable for portfolios containing long and/or short positions in European and/or exotic options expiring beyond the investment horizon. We prove that WCPVaR and WCQVaR optimization can be formulated as tractable second-order cone and semidefinite programs, respectively, and reveal interesting connections to robust portfolio optimization. Numerical experiments demonstrate the benefits of incorporating non-linear relationships between the asset returns into a worst-case VaR model. The contents of this chapter are based on

2. S. Zymler, D. Kuhn, and B. Rustem. *Worst-Case Value-at-Risk of Non-linear Portfolios.* Under revision for Operations Research, 2010.

In Chapter 5 we develop tractable semidefinite programming based approximations for distributionally robust individual and joint chance constraints, assuming that only the first- and second-order moments as well as the support of the uncertain parameters are given. It is known that robust chance constraints can be conservatively approximated by Worst-Case Conditional Value-at-Risk (CVaR) constraints. We first prove that this approximation is exact for robust individual chance constraints with concave or (not necessarily concave) quadratic constraint functions. We also show that robust individual chance constraints are equivalent to robust semi-infinite constraints with uncertainty sets that can be interpreted as ellipsoids lifted to the space of positive semidefinite matrices. By using the theory of moment problems we then obtain a conservative approximation for joint chance constraints. This approximation affords intuitive dual interpretations and is provably tighter than two popular benchmark approximations. The tightness depends on a set of scaling parameters, which can be tuned via a sequential convex optimization algorithm. We show that the approximation becomes in fact exact when the scaling parameters are chosen optimally. We further demonstrate that joint chance constraints can be reformulated as robust semi-infinite constraints with uncertainty sets that are reminiscent of the lifted ellipsoidal uncertainty sets characteristic for individual chance constraints. We evaluate our joint chance constraint approximation in the context of a dynamic water reservoir control problem and numerically demonstrate its superiority over the two benchmark approximations. The contents of this chapter are based on

3. S. Zymler, D. Kuhn, and B. Rustem. *Distributionally Robust Joint Chance Constraints with*

Second-Order Moment Information. Under review for Mathematical Programming, 2010.

1.3 Statement of Originality

This thesis is the result of my own work and no other person's work has been used without due acknowledgement in the main text of the thesis. This thesis has not been submitted for the award of any degree or diploma in any other tertiary institution.

Chapter 2

Background Theory

In this chapter we summarize various definitions and results relating to convex optimization and decision making under uncertainty. In particular, we give an overview of stochastic programming, robust optimization, and distributionally robust optimization. We also give a general description of portfolio optimization and risk measures. The selection of presented topics is dictated entirely by their use in subsequent chapters. For a thorough review of convex and robust optimization the reader is referred to [BV04] and [BTEGN09], respectively. We emphasize that each of the subsequent chapters also contain introductions with more specific background references.

2.1 Notation

Throughout this thesis, we will use the following notation. We use lower-case bold face letters to denote vectors and upper-case bold face letters to denote matrices. The space of symmetric matrices of dimension n is denoted by $\mathbb{S}^n$. For any two matrices $\mathbf{X}, \mathbf{Y} \in \mathbb{S}^n$, we let $\langle \mathbf{X}, \mathbf{Y} \rangle = \mathrm{Tr}(\mathbf{XY})$ be the trace scalar product, while the relation $\mathbf{X} \succcurlyeq \mathbf{Y}$ ($\mathbf{X} \succ \mathbf{Y}$) implies that $\mathbf{X} - \mathbf{Y}$ is positive semidefinite (positive definite). Random variables are always represented by symbols with tildes, while their realizations are denoted by the same symbols without tildes. For $x \in \mathbb{R}$, we define $x^+ = \max\{x, 0\}$. Unless stated otherwise, equations involving random variables are assumed to hold almost surely.

2.2 Convex Optimization

A *convex optimization problem* is a minimization problem of the form

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}} \quad & f_0(\boldsymbol{x}) \\ \text{subject to} \quad & f_i(\boldsymbol{x}) \leq 0, \quad \forall i = 1,\dots,m \\ & \mathbf{A}\boldsymbol{x} = \boldsymbol{b}, \end{aligned} \tag{2.1}$$

where $\mathbf{A} \in \mathbb{R}^{p\times n}$ and each of the functions $f_i : \mathbb{R}^n \to \mathbb{R}$ is convex. The function f_0 is referred to as the *objective* or *cost* function. As usual, (2.1) describes the problem of finding an $\boldsymbol{x}$ that minimizes $f_0(\boldsymbol{x})$ among all the $\boldsymbol{x}$ that satisfy the *constraints* $f_i(\boldsymbol{x}) \leq 0,\ i = 1,\dots,m$ and $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$.

In the remainder of this section, we review important classes of convex optimization problems which we will focus on throughout this thesis.

2.2.1 Linear Programming

A *linear program* or LP is a problem of the form

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}} \quad & \boldsymbol{c}^\mathsf{T}\boldsymbol{x} \\ \text{subject to} \quad & \mathbf{G}\boldsymbol{x} \leq \boldsymbol{f} \\ & \mathbf{A}\boldsymbol{x} = \boldsymbol{b}, \end{aligned} \tag{2.2}$$

where $\mathbf{A} \in \mathbb{R}^{p\times n}$, $\mathbf{G} \in \mathbb{R}^{m\times n}$, $\boldsymbol{c} \in \mathbb{R}^n$, $\boldsymbol{f} \in \mathbb{R}^m$, and $\boldsymbol{b} \in \mathbb{R}^p$. Problem (2.2) is convex since it only involves linear constraints.

2.2.2 Second-Order Cone Programming

A *Second-Order Cone Program* or SOCP is a convex optimization problem of the form

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}} \quad & \boldsymbol{c}^\mathsf{T}\boldsymbol{x} \\ \text{subject to} \quad & \|\mathbf{B}_i\boldsymbol{x} + \boldsymbol{d}_i\|_2 \leq \boldsymbol{f}_i^\mathsf{T}\boldsymbol{x} + g_i, \quad \forall i = 1,\ldots,m \\ & \mathbf{A}\boldsymbol{x} = \boldsymbol{b}, \end{aligned} \tag{2.3}$$

where $\mathbf{B}_i \in \mathbb{R}^{m_i\times n}$, $\boldsymbol{d}_i \in \mathbb{R}^{m_i}$, $\boldsymbol{f}_i \in \mathbb{R}^n$, $g_i \in \mathbb{R}$, and $\|\boldsymbol{y}\|_2 = \sqrt{\boldsymbol{y}^\mathsf{T}\boldsymbol{y}}$ denotes the L^2 norm of $\boldsymbol{y}$. Note that when $\mathbf{B}_i$ and $\boldsymbol{d}_i$ are zero for $i = 1,\ldots,m$ then the SOCP (2.3) reduces to a linear program. Thus, the class of SOCPs encapsulates the class of LPs as a special case.

For any $i = 1,\ldots,m$, the constraint

$$\|\mathbf{B}_i\boldsymbol{x} + \boldsymbol{d}_i\|_2 \leq \boldsymbol{f}_i^\mathsf{T}\boldsymbol{x} + g_i \tag{2.4}$$

is referred to as a *second-order cone constraint*, since it is the same as requiring the affine function $(\mathbf{B}_i\boldsymbol{x} + \boldsymbol{d}_i, \boldsymbol{f}_i^\mathsf{T}\boldsymbol{x} + g_i)$ to lie in the second-order cone in $\mathbb{R}^{m_i+1}$, see Figure 2.1.

It is known that SOCPs can be solved in polynomial-time using interior point algorithms, thus, SOCPs are *tractable* problems, see [AG03]. Furthermore, the reader is refered to [LVBL98] for a detailed survey on the applications of second-order cone programming.

2.2.3 Semidefinite Programming

A *Semidefinite Program* or SDP is a convex optimization problem of the form

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}} \quad & \boldsymbol{c}^\mathsf{T}\boldsymbol{x} \\ \text{subject to} \quad & \mathbf{F}_0 + \sum_{i=1}^{n} \mathbf{F}_i x_i \succcurlyeq \mathbf{0} \\ & \mathbf{A}\boldsymbol{x} = \boldsymbol{b}, \end{aligned} \tag{2.5}$$

where each of the matrices $\mathbf{F}_i \in \mathbb{R}^{n\times n}$ is symmetric.

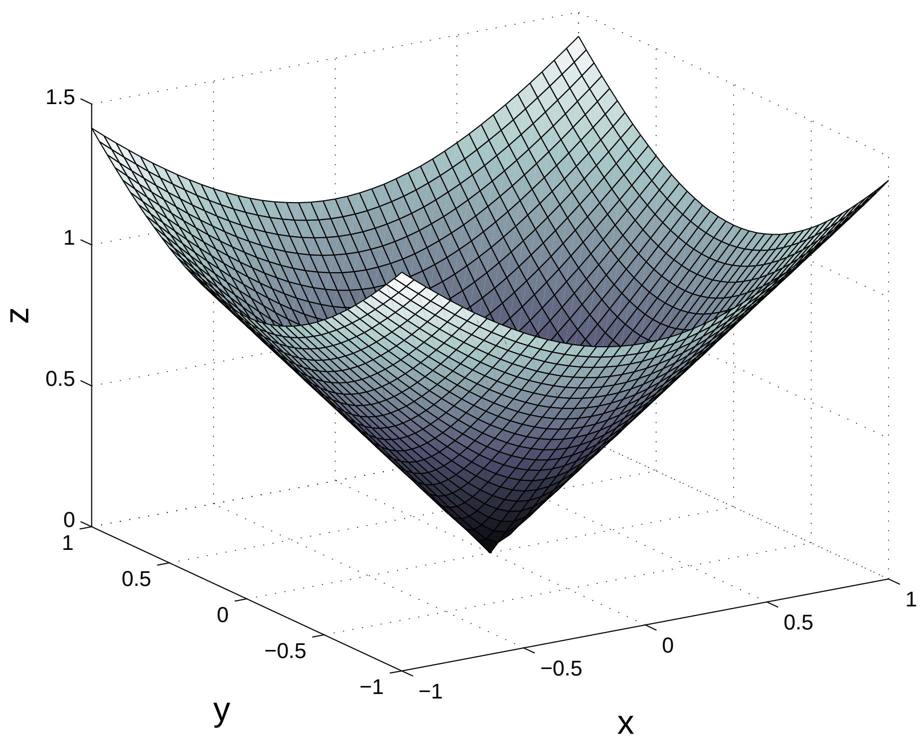

Figure 2.1: Boundary of the convex second-order cone $\{(x, y, z) \ : \ \sqrt{x^2 + y^2} \leq z\}$ in $\mathbb{R}^3$.

The constraint

$$\mathbf{F}(\boldsymbol{x}) = \mathbf{F}_0 + \sum_{i=1}^{n} \mathbf{F}_i x_i \succcurlyeq \mathbf{0} \tag{2.6}$$

requires that the linear combination $\mathbf{F}(\boldsymbol{x})$ of the matrices $\mathbf{F}_i$ is *positive semidefinite* and is refered to as a *linear matrix inequality* or LMI. An LMI constraint of the form (2.6) is a convex constraint on $\boldsymbol{x}$ since $\{\boldsymbol{x} \in \mathbb{R}^n \ : \ \mathbf{F}(\boldsymbol{x}) \succcurlyeq \mathbf{0}\}$ is a closed and convex set. In Figure 2.2 we plot the boundary of the positive semidefinite cone

$$\begin{bmatrix} x & y \\ y & z \end{bmatrix} \succcurlyeq \mathbf{0} \quad \Longleftrightarrow \quad x \geq 0, \quad y \geq 0, \quad xz \geq y^2.$$

The following lemma is often useful to rewrite general matrix inequalities as LMIs or to simplify SDPs.

Lemma 2.2.1 (Schur Complement) *Consider the matrix* $\mathbf{X} \in \mathbb{S}^n$*, which can be partitioned*

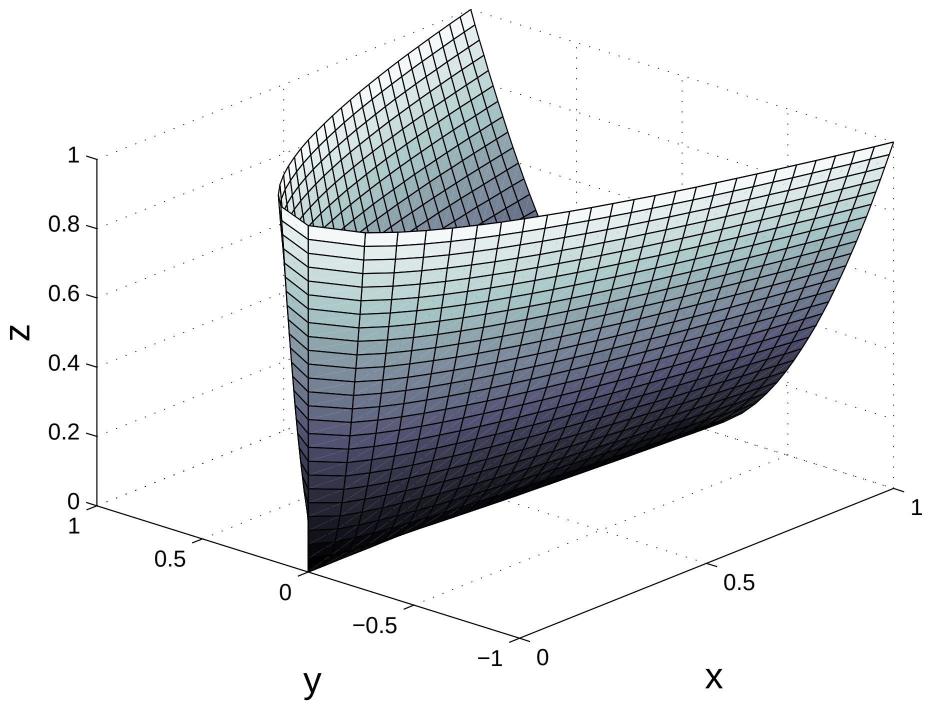

Figure 2.2: Boundary of the convex positive semidefinite cone in $\mathbb{S}^2$.

as

$$\mathbf{X} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^{\mathsf{T}} & \mathbf{C} \end{bmatrix},$$

then the following results hold:

(i) $\mathbf{X} \succ \mathbf{0}$ *if and only if* $\mathbf{A} \succ \mathbf{0}$ *and* $(\mathbf{C} - \mathbf{B}^{\mathsf{T}}\mathbf{A}^{-1}\mathbf{B}) \succ \mathbf{0}$.

(ii) If $\mathbf{A} \succ \mathbf{0}$, *then* $\mathbf{X} \succcurlyeq \mathbf{0}$ *if and only if* $(\mathbf{C} - \mathbf{B}^{\mathsf{T}}\mathbf{A}^{-1}\mathbf{B}) \succcurlyeq \mathbf{0}$.

It is known that SDPs can also be solved in polynomial-time using interior point algorithms, see [VB96]. Furthermore, any LP and SOCP can be formulated as an SDP. However, it is generally recommended to reduce SDPs to LPs or SOCPs if it is possible to do so, since they exhibit better scalability properties than SDPs [AG03], and the solver implementations for these problems are more mature.

2.3 Decision Making under Uncertainty

Many real-world optimization problems involve data parameters which are subject to uncertainty or cannot be estimated accurately. Failure to take this uncertainty into account may lead to suboptimal decisions. Consider for example the following convex optimization problem.

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}} \quad & f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) \\ \text{subject to} \quad & g(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) \leq 0 \\ & \boldsymbol{x} \in \mathcal{X}, \end{aligned} \tag{2.7}$$

where $\tilde{\boldsymbol{\xi}}$ denotes the uncertain or random vector of data parameters and $\mathcal{X} \subseteq \mathbb{R}^n$ is some convex set that is not affected by uncertainty. Note that the cost function f and constraint function g depend on the random vector $\tilde{\boldsymbol{\xi}}$. This model essentially represents a whole family of optimization problems, one for each possible realization of $\tilde{\boldsymbol{\xi}}$. Therefore, (2.7) fails to provide a unique solution. In the remainder of this section we briefly review alternative modeling paradigms to disambiguate (2.7).

2.3.1 Stochastic Programming

Stochastic Programming assumes that the decision maker has full and accurate information about the probability distribution $\mathbb{Q}$ of the random vector $\tilde{\boldsymbol{\xi}}$. Subsequently, stochastic programming enables us to disambiguate problem (2.7) as follows.

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}} \quad & \mathbb{E}_{\mathbb{Q}}\left(f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})\right) \\ \text{subject to} \quad & \mathbb{Q}\left(g(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) \leq 0\right) \geq 1 - \epsilon \\ & \boldsymbol{x} \in \mathcal{X}, \end{aligned} \tag{2.8}$$

where $\mathbb{E}_{\mathbb{Q}}(\cdot)$ denotes the expectation with respect to the random vector $\tilde{\boldsymbol{\xi}}$ given that it follows the probability distribution $\mathbb{Q}$. The stochastic program (2.8) aims to find the optimal solution $\boldsymbol{x} \in \mathcal{X}$ that minimizes the expected value $\mathbb{E}_{\mathbb{Q}}(f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}))$ of the cost function. Furthermore, the problem requires that the uncertain constraint $g(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) \leq 0$ is satisfied with some high probability $1-\epsilon$. This is formulized by the *chance-constraint*

$$\mathbb{Q}\left(g(\boldsymbol{x},\tilde{\boldsymbol{\xi}})\leq 0\right)\geq 1-\epsilon, \tag{2.9}$$

where $\epsilon\in(0,1)$ denotes the *risk factor* that is specified by the decision maker. Note that as the value of ϵ decreases, the chance constraint has to be satisfied with a higher probability. Chance-constrained programs of the type (2.8) were first discussed by Charnes *et al.* [CCS58], Miller and Wagner [MW65] and Prékopa [Pre70].

Computing the optimal solution of a chance-constrained program is notoriously difficult. In fact, even checking the feasibility of a fixed decision $\boldsymbol{x}$ requires the computation of a multi-dimensional integral, which becomes increasingly difficult as the dimension of the random vector $\tilde{\boldsymbol{\xi}}$ increases. Moreover, even though the constraint function g is convex in $\boldsymbol{x}$, the feasible set of chance constraint (2.9) is typically nonconvex and sometimes even disconnected [Pre70, NS06]. Thus, chance-constrained programs are generically *intractable* to solve.

Furthermore, in order to evaluate the chance constraint (2.9), full and accurate information about the probability distribution $\mathbb{Q}$ of the random vector $\tilde{\boldsymbol{\xi}}$ is required. However, in many practical situations $\mathbb{Q}$ must be estimated from historical data and is therefore itself uncertain. Typically, one has only partial information about $\mathbb{Q}$, e.g. about its moments or its support. Replacing the unknown distribution $\mathbb{Q}$ in (2.8) by an estimate $\hat{\mathbb{Q}}$ corrupted by measurement errors may lead to over-optimistic solutions which often fail to satisfy the chance constraint under the true distribution $\mathbb{Q}$.

2.3.2 Robust Optimization

In order to disambiguate the problem (2.7), robust optimization adopts a worst-case perspective, see Ben-Tal *et al.* [BTEGN09] for a thorough exposition on robust optimization. In this modelling framework, the random vector $\tilde{\boldsymbol{\xi}}$ remains unknown, but it is believed to materialize within an *uncertainty set* $\mathcal{U}$. To *immunize* problem (2.7) against the inherent uncertainty in $\tilde{\boldsymbol{\xi}}$, we minimize the worst-case cost, where the worst-case is calculated with respect to all realizations $\boldsymbol{\xi}$ within the uncertainty set $\mathcal{U}$. This can be formalized as a min-max problem

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}} \quad & \max_{\boldsymbol{\xi}\in\mathcal{U}} f(\boldsymbol{x},\boldsymbol{\xi}) \\ \text{subject to} \quad & g(\boldsymbol{x},\boldsymbol{\xi}) \le 0 \quad \forall \boldsymbol{\xi}\in\mathcal{U} \\ & \boldsymbol{x}\in\mathcal{X}. \end{aligned} \tag{2.10}$$

Problem (2.10) is often refered to as the *robust counterpart* of problem (2.7). For any fixed $\boldsymbol{x}$, the function $\max_{\boldsymbol{\xi}\in\mathcal{U}} f(\boldsymbol{x},\boldsymbol{\xi})$ computes the worst-case *realized* cost given that $\boldsymbol{\xi}$ can obtain values within $\mathcal{U}$. Note that this quantity depends in a non-trivial way on the decision variable $\boldsymbol{x}$. Thus, the aim of the above problem is to minimize the worst-case cost. Furthermore, problem (2.10) requires that the constraint $g(\boldsymbol{x},\boldsymbol{\xi}) \le 0$ is satisfied for *all* realizations of $\boldsymbol{\xi}\in\mathcal{U}$. This is formulized by the *semi-infinite constraint*

$$g(\boldsymbol{x},\boldsymbol{\xi}) \le 0 \quad \forall \boldsymbol{\xi}\in\mathcal{U}, \tag{2.11}$$

which, in the context of a robust optimization problem of type (2.10), is sometimes refered to as a *robust constraint.*

The shape of the uncertainty set $\mathcal{U}$ should reflect the modeller's knowledge about the distribution of the random vector $\tilde{\boldsymbol{\xi}}$, e.g., full or partial information about the support and certain moments of the random vector $\tilde{\boldsymbol{\xi}}$. Moreover, the size of $\mathcal{U}$ determines the degree to which the user wants to safeguard feasibility of the corresponding explicit inequality constraint. The robust semi-infinite constraint (2.11) is therefore closely related to the chance constraint (2.9).

For a large class of convex uncertainty sets, the semi-infinite constraint (2.11) can be reformulated in terms of a small number of tractable (i.e., linear, second-order conic, or LMI) constraints [BTN98, BTN99]. Consider, for example, the rectangular uncertainty set defined as

$$\mathcal{U}_{\text{box}} = \{\boldsymbol{\xi}\in\mathbb{R}^n \ : \ \boldsymbol{l} \le \boldsymbol{\xi} \le \boldsymbol{u}\},$$

where $\boldsymbol{l}, \boldsymbol{u} \in \mathbb{R}^n$ and $\boldsymbol{l} < \boldsymbol{u}$. Then, the following equivalences hold.

$$
\begin{aligned}
& \boldsymbol{x}^\mathsf{T}\boldsymbol{\xi} \leq 0 \quad \forall \boldsymbol{\xi} \in \mathcal{U}_{\text{box}} \\
\iff \quad & 0 \geq \max_{\boldsymbol{\xi} \in \mathbb{R}^n} \left\{ \boldsymbol{x}^\mathsf{T}\boldsymbol{\xi} \;:\; \boldsymbol{l} \leq \boldsymbol{\xi} \leq \boldsymbol{u} \right\} \\
\iff \quad & 0 \geq \min_{\boldsymbol{\lambda} \in \mathbb{R}^n} \left\{ \boldsymbol{x}^\mathsf{T}\boldsymbol{u} + \boldsymbol{\lambda}^\mathsf{T}(\boldsymbol{l} - \boldsymbol{u}) \;:\; \boldsymbol{\lambda} \leq \boldsymbol{x},\ \boldsymbol{\lambda} \leq \boldsymbol{0} \right\} \\
\iff \quad & \exists \boldsymbol{\lambda} \in \mathbb{R}^n \;:\; \boldsymbol{x}^\mathsf{T}\boldsymbol{u} + \boldsymbol{\lambda}^\mathsf{T}(\boldsymbol{l} - \boldsymbol{u}) \leq 0,\ \boldsymbol{\lambda} \leq \boldsymbol{x},\ \boldsymbol{\lambda} \leq \boldsymbol{0}
\end{aligned}
$$

The equivalence in the third line in the above expression follows from strong linear programming duality, which holds since the primal maximization problem has a nonempty feasible set, see [BV04, §5] for a thorough review on convex duality. Note that by employing this dualization technique, we effectively reformulated the semi-infinite constraint in terms of a tractable system of linear constraints. Similar dualization techniques will be employed throughout this thesis to find tractable reformulations of robust constraints.

2.3.3 Distributionally Robust Optimization

Distributionally robust optimization is closely related to both stochastic programming and robust optimization. In contrast to stochastic programming, the distributionally robust optimization framework assumes that the decision maker only has partial information about the probability distribution $\mathbb{Q}$ of the random vector $\tilde{\boldsymbol{\xi}}$, such as the first and second moments and its support. Let $\mathcal{P}$ denote the set of all probability distributions that are consistent with the known distributional properties of $\mathbb{Q}$. Similar to the robust optimization framework discussed above, distributionally robust optimization adopts a worst-case approach. Only now the worst-case is computed over all probability distributions within the set $\mathcal{P}$. Thus, distributionally robust optimization disambiguates problem (2.7) as follows.

$$
\begin{aligned}
\underset{\boldsymbol{x} \in \mathbb{R}^n}{\text{minimize}} \quad & \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left(f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) \right) \\
\text{subject to} \quad & \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left(g(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) \leq 0 \right) \geq 1 - \epsilon \\
& \boldsymbol{x} \in \mathcal{X}
\end{aligned}
\tag{2.12}
$$

For any fixed $\boldsymbol{x}$, the function $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}(f(\boldsymbol{x},\tilde{\boldsymbol{\xi}}))$ computes the worst-case *expected* cost, that is, the highest expected cost evaluated over all probability distributions $\mathbb{P}$ within the set $\mathcal{P}$. The aim of problem (2.12) is to minimize the worst-case expected cost. Furthermore, problem (2.12) requires that the uncertain constraint $g(\boldsymbol{x},\tilde{\boldsymbol{\xi}}) \leq 0$ is satisfied with some high probability $1-\epsilon$ under any probability distribution $\mathbb{P} \in \mathcal{P}$. This is formulized by the *distributionally robust chance constraint*

$$\mathbb{P}\left(g(\boldsymbol{x},\tilde{\boldsymbol{\xi}}) \leq 0\right) \geq 1-\epsilon \quad \forall \mathbb{P}\in\mathcal{P} \iff \inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(g(\boldsymbol{x},\tilde{\boldsymbol{\xi}}) \leq 0\right) \geq 1-\epsilon. \tag{2.13}$$

It is easily verified that whenever $\boldsymbol{x}$ satisfies (2.13) and $\mathbb{Q} \in \mathcal{P}$, then $\boldsymbol{x}$ also satisfies the chance constraint (2.9) under the true probability distribution $\mathbb{Q}$. Thus, by adopting a worst-case appoach, distributionally robust optimization effectively immunizes the stochastic program (2.8) against uncertainty about the probability distribution $\mathbb{Q}$.

For certain choices of $\mathcal{P}$, the distributionally robust optimization problem (2.12) can be reformulated as a tractable convex optimization problem. Scarf [Sca58] applies distributionally robust optimization to a single-product newsboy problem and shows that, when only the first- and second-order moments of the demand are known, the problem can be reformulated as a tractable optimization problem. More recently, Bertsimas and Popescu [BP02] use semidefinite programming to derive tight upper and lower bounds on option prices given that only the moments of the underlying asset prices are known. El Ghaoui *et al.* [EGOO03] prove that the worst-case Value-at-Risk of a financial portfolio can be optimized by solving tractable SOCPs and SDPs by assuming that only the first- and second-order moments as well as the support of the asset returns are known. Delage *et al.* [DY10] incorporate confidence intervals for the first- and second-order moments within the distributionally robust optimization framework. We refer the reader to Ben-Tal *et al.* [BTEGN09] for an overview on tractable reformulations of distributionally robust chance constraints.

2.4 Portfolio Optimization and Risk Measures

Investors face the challenging problem of how to distribute their current wealth over a set of available assets, such as stocks, bonds, and derivatives, with the goal to earn the highest possible future wealth. One of the first mathematical models for this problem was formulated by Harry Markowitz [Mar52]. In his Nobel prize-winning work, he observed that a rational investor does not aim solely at maximizing the expected return of an investment, but also at minimizing its risk. In the Markowitz model, which is also refered to as mean-variance optimization, the risk of a portfolio is measured by the variance of the portfolio return.

Although mean-variance optimization is appropriate when the asset returns are symmetrically distributed, it is known to result in counter intuitive asset allocations when the portfolio return is skewed [FKD07]. This shortcoming triggered extensive research on downside risk measures. In this section we give a brief overview on portfolio optimization, describe some popular risk measures that will be used in this thesis, and review the concept of *coherent risk measures*.

2.4.1 Portfolio Optimization

A general portfolio optimization problem can be formulized as

$$\begin{aligned} \underset{\boldsymbol{w}\in\mathbb{R}^n}{\text{minimize}} \quad & \rho(\boldsymbol{w}^\top\tilde{\boldsymbol{r}}) \\ \text{subject to} \quad & \boldsymbol{w}\in\mathcal{W}. \end{aligned} \tag{2.14}$$

In the above problem, the vector $\boldsymbol{w}\in\mathbb{R}^n$ denotes the *portfolio allocation weights*, namely the percentages of wealth allocated in different assets, and $\tilde{\boldsymbol{r}}$ denotes the $\mathbb{R}^n$-valued random vector of asset returns. The set $\mathcal{W}\subseteq\mathbb{R}^n$ denotes the set of admissible portfolios. The inclusion $\boldsymbol{w}\in\mathcal{W}$ usually implies the budget constraint $\boldsymbol{w}^\top\boldsymbol{e}=1$ (where $\boldsymbol{e}$ denotes the vector of 1s). Optionally, the set $\mathcal{W}$ may account for bounds on the allocation vector $\boldsymbol{w}$ and/or a constraint enforcing a minimum expected portfolio return. The random return of the portfolio is computed as $\boldsymbol{w}^\top\tilde{\boldsymbol{r}}$. The risk measure ρ maps the random portfolio return to a real number which represents the risk of the portfolio $\boldsymbol{w}$. Thus, problem (2.14) aims to determine the portfolio with the lowest risk from the set $\mathcal{W}$ of admissible portfolios. A recent survey of portfolio optimization is provided in the monograph [FKD07].

2.4.2 Popular Risk Measures

In finance, risk measures can be subdivided into two main categories: moment-based and quantile-based risk measures, see [NPS09]. Moment-based risk measures are related to classical utility theory, whereas the theory of stochastic dominance has spurred interest in quantile-based risk measures [Lev92]. In this subsection we review three commonly used risk measures: mean-variance, Value-at-Risk (VaR), and Conditional Value-at-Risk (CVaR). The VaR and CVaR measures will be used throughout this thesis.

Mean-Variance

The most popular moment-based risk measure trades off the expected portfolio return and variance of the portfolio return. It is defined as

$$\rho(\boldsymbol{w}^\mathsf{T}\tilde{\boldsymbol{r}}) = -\boldsymbol{w}^\mathsf{T}\boldsymbol{\mu} + \lambda\boldsymbol{w}^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{w},$$

where $\boldsymbol{\mu}$ denotes the vector of mean asset returns, $\boldsymbol{\Sigma}$ represents the covariance matrix of the asset returns, and the parameter λ characterizes the risk-aversion level of the investor. As λ increases, the risk measure puts more weight on the variance of the portfolio return and therefore results in a higher risk estimate. The use of the mean-variance risk measure can be traced back to Markowitz' seminal work [Mar52]. Although mean-variance optimization is appropriate when the asset returns are symmetrically distributed, it is known to result in counter intuitive asset allocations when the portfolio return is skewed. This shortcoming triggered extensive research on quantile-based risk measures, which we discuss next.

Value-at-Risk

The most popular quantile-base risk measure is the Value-at-Risk [Jor01]. The VaR at level ϵ is defined as the $(1-\epsilon)$-percentile of the portfolio loss distribution, where ϵ is typically chosen as 1% or 5%. Put differently, $\mathrm{VaR}_\epsilon(\boldsymbol{w})$ is defined as the smallest real number γ with the property that $-\boldsymbol{w}^\mathsf{T}\tilde{\boldsymbol{r}}$ exceeds γ with a probability not larger than ϵ, that is,

$$\mathrm{VaR}_\epsilon(\boldsymbol{w}) = \min\left\{\gamma \ : \ \mathbb{P}\{\gamma \leq -\boldsymbol{w}^\mathsf{T}\tilde{\boldsymbol{r}}\} \leq \epsilon\right\}, \tag{2.15}$$

where $\mathbb{P}$ denotes the distribution of the asset returns $\tilde{\boldsymbol{r}}$. Note that (2.15) constitutes a chance-constrained stochastic program which is non-convex under general probability distributions $\mathbb{P}$, see Section 2.3.1. Thus, VaR optimization is generically intractable. We shall investigate this issue in much greater detail in Chapter 4.

Conditional Value-at-Risk

The Conditional Value-at-Risk, proposed by Rockafellar and Uryasev [RU02], is an alternative quantile-based risk measure which has been gaining popularity due to its desirable computational properties. The CVaR evaluates the conditional expectation of loss above the $(1-\epsilon)$-quantile of the portfolio loss distribution, and can be formulized as

$$\mathrm{CVaR}_\epsilon(\boldsymbol{w}) = \min_{\beta\in\mathbb{R}}\left\{\beta + \frac{1}{\epsilon}\mathbb{E}_\mathbb{P}\left(-\boldsymbol{w}^\mathsf{T}\tilde{\boldsymbol{r}} - \beta\right)^+\right\}. \tag{2.16}$$

In contrast to VaR, the CVaR is a convex function of the portfolio weights $\boldsymbol{w}$. Moreover, it is known that $\mathrm{CVaR}_\epsilon(\boldsymbol{w}) \geq \mathrm{VaR}_\epsilon(\boldsymbol{w})$ for any portfolio $\boldsymbol{w} \in \mathcal{W}$. Thus, CVaR can be used to conservatively approximate the VaR of a portfolio. We will use this property in Chapter 5 to derive tractable approximations for chance constrained optimization problems. Furthermore, CVaR is known to be a coherent risk measure. The next subsection reviews what coherent risk measures are.

2.4.3 Coherent Risk Measures

Consider the linear space of random variables

$$\mathcal{V} = \left\{ \boldsymbol{w}^\top \tilde{\boldsymbol{r}} : \boldsymbol{w} \in \mathbb{R}^n \right\}. \tag{2.17}$$

The function $\rho : \mathcal{V} \to \mathbb{R}$ is said to be a *coherent risk measure* if it satisfies the following four axioms:

(i) **Subadditivity:** For all $\tilde{v}_1, \tilde{v}_2 \in \mathcal{V}$, $\rho(\tilde{v}_1 + \tilde{v}_2) \leq \rho(\tilde{v}_1) + \rho(\tilde{v}_2)$.

(ii) **Translation Invariance:** For all $\tilde{v} \in \mathcal{V}$ and $a \in \mathbb{R}$, $\rho(\tilde{v} + a) = \rho(\tilde{v}) - a$.

(iii) **Positive Homogeneity:** For all $\tilde{v} \in \mathcal{V}$ and $\alpha \geq 0$, $\rho(\alpha\tilde{v}) = \alpha\rho(\tilde{v})$.

(iv) **Monotonicity:** For all $\tilde{v}_1, \tilde{v}_2 \in \mathcal{V}$ such that $\tilde{v}_1 \geq \tilde{v}_2$, $\rho(\tilde{v}_1) \leq \rho(\tilde{v}_2)$ (where $\tilde{v}_1 \geq \tilde{v}_2$ means that $\tilde{v}_1(\omega) \geq \tilde{v}_2(\omega)$ for all elements ω of the corresponding sample space).

The four axioms that define coherency were introduced and justified by Artzner *et al.* [ADEH99]. The subadditivity axiom ensures that the risk associated with the sum of two assets cannot be larger than the sum of their individual risk quantities. This property entails that financial diversification can only reduce the risk. Translation invariance means that receiving a sure amount of a reduces the risk quantity by a. Positive homogeneity implies that the risk measure scales proportionally with the size of the investment. Finally, monotonicity implies that when one investment almost surely outperforms another investment, its risk must be smaller.

From all the risk measures discussed in the previous section, only the CVaR is a coherent risk measure. VaR fails to satisfy the subadditivity axiom and the mean-standard deviation risk measure does not satisfy the monotonicity axiom.

Chapter 3

Robust Portfolio Optimization with Derivative Insurance Guarantees

Robust portfolio optimization aims to maximize the worst-case portfolio return given that the asset returns are allowed to vary within a prescribed uncertainty set. If the uncertainty set is not too large, the resulting portfolio performs well under normal market conditions. However, its performance may substantially degrade in the presence of market crashes, that is, if the asset returns materialize far outside of the uncertainty set. In this chapter, we propose a novel robust optimization model for designing portfolios that include European-style options. This model trades off weak and strong guarantees on the worst-case portfolio return. The weak guarantee applies as long as the asset returns are realized within the prescribed uncertainty set, while the strong guarantee applies for all possible asset returns. The resulting model constitutes a convex second-order cone program, which is amenable to efficient numerical solution procedures. We evaluate the model using simulated and empirical backtests and analyze the impact of the insurance guarantees on the portfolio performance.

3.1 Introduction

Investors face the challenging problem of how to distribute their current wealth over a set of available assets, such as stocks, bonds, and derivatives, with the goal to earn the highest possible future wealth. One of the first mathematical models for this problem was formulated by Harry Markowitz [Mar52]. In his Nobel prize-winning work, he observed that a rational investor does not aim solely at maximizing the expected return of an investment, but also at minimizing its risk. In the Markowitz model, the risk of a portfolio is measured by the variance of the portfolio return. A practical advantage of the Markowitz model is that it reduces to a convex quadratic program, which can be solved efficiently.

Although the Markowitz model has triggered a tremendous amount of research activities in the field of finance, it has serious disadvantages which have discouraged practitioners from using it. The main problem is that the means and covariances of the asset returns, which are important inputs to the model, have to be estimated from noisy data. Hence, these estimates are not accurate. In fact, it is fundamentally impossible to estimate the mean returns with statistical methods to within workable precision, a phenomenon which is sometimes referred to as *mean blur* [Lue98, Mer80]. Unfortunately, the mean-variance model is very sensitive to the distributional input parameters. As a result, the model amplifies any estimation errors, yielding extreme portfolios which perform badly in out-of-sample tests [CZ93, Bro93, Mic01, DN09].

Many attempts have been undertaken to ease this amplification of estimation errors. Black and Litterman [BL91] suggest Bayesian estimation of the means and covariances using the market portfolio as a prior. Jagannathan and Ma [JM03] as well as Chopra [Cho93] impose portfolio constraints in order to guide the optimization process towards more intuitive and diversified portfolios. Chopra *et al.* [CHT93] use a James-Steiner estimator for the means which tilts the optimal allocations towards the minimum-variance portfolio, while DeMiguel *et al.* [DN09] employ robust estimators.

In recent years, *robust optimization* has received considerable attention. Robust optimization is a powerful modeling paradigm for decision problems subject to non-stochastic data uncertainty

[BTN98]. The uncertain problem parameters are assumed to be unknown but confined to an *uncertainty set*, which reflects the decision maker's uncertainty about the parameters. Robust optimization models aim to find the best decision in view of the worst-case parameter values within these sets, see also Section 2.3.2 for an introduction to robust optimization. Ben-Tal and Nemirovski [BTN99] propose a robust optimization model to immunize a portfolio against the uncertainty in the asset returns. They show that when the asset returns can vary within an ellipsoidal uncertainty set determined through their means and covariances, the resulting optimization problem is reminiscent of the Markowitz model. This robust portfolio selection model still assumes that the distributional input parameters are known precisely. Therefore, it suffers from the same shortcomings as the Markowitz model.

Robust portfolio optimization can also be used to immunize a portfolio against the uncertainty in the distributional input parameters. Goldfarb and Iyengar [GI03] use statistical methods for constructing uncertainty sets for factor models of the asset returns and show that their robust portfolio problem can be reformulated as a second-order cone program. Tütüncü and Koenig [TK04] propose a model with box uncertainty sets for the means and covariances and show that the arising model can be reduced to a smooth saddle-point problem subject to semidefinite constraints. Rustem and Howe [RH02] describe algorithms to solve general continuous and discrete minimax problems and present several applications of worst-case optimization for risk management. Rustem *et al.* [RBM00] propose a model that optimizes the worst-case portfolio return under rival risk and return forecasts in a discrete minimax setting. El Ghaoui *et al.* [EGOO03] show that the worst-case Value-at-Risk under partial information on the moments can be formulated as a semidefinite program. Ben-Tal *et al.* [BTMN00] as well as Bertsimas and Pachamanova [BP08] suggest robust portfolio models in a multi-period setting. Recently, the relationship between uncertainty sets in robust optimization and *coherent risk measures* [ADEH99] has been described in Natarajan *et al.* [NPS08] and Bertsimas and Brown [BB08], see also Section 2.4.3 for an introduction to coherent risk measures. A recent survey of applications of robust portfolio optimization is provided in the monograph [FKD07]. Robust portfolios of this kind are relatively insensitive to the distributional input parameters and typically outperform classical Markowitz portfolios [CS06].

Robust portfolios exhibit a *non-inferiority* property [RBM00]: whenever the asset returns are realized within the prescribed uncertainty set, the realized portfolio return will be greater than or equal to the calculated worst-case portfolio return. Note that this property may fail to hold when the asset returns happen to fall outside of the uncertainty set. In this sense, the non-inferiority property only offers a *weak guarantee.* When a rare event (such as a market crash) occurs, the asset returns can materialize far beyond the uncertainty set, and hence the robust portfolio will remain unprotected. A straightforward way to overcome this problem is to enlarge the uncertainty set to cover also the most extreme events. However, this can lead to robust portfolios that are too conservative and perform poorly under normal market conditions.

In this chapter we will use portfolio insurance to hedge against rare events which are not captured by a reasonably sized uncertainty set. Classical portfolio insurance is a well studied topic in finance. The idea is to enrich a portfolio with specific derivative products in order to obtain a deterministic lower bound on the portfolio return. The insurance holds for all possible realizations of the asset returns and can therefore be qualified as a *strong guarantee.* Numerous studies have investigated the integration of options in portfolio optimization models. Ahn *et al.* [ABRW99] minimize the Value-at-Risk of a portfolio consisting of a single stock and a put option by controlling the portfolio weights and the option strike price. Dert and Oldenkamp [DO00] propose a model that maximizes the expected return of a portfolio consisting of a single index stock and several European options while guaranteeing a maximum loss. Howe *et al.* [HRS94] introduce a risk management strategy for the writer of a European call option based on minimax using box uncertainty. Lutgens *et al.* [LSK06] propose a robust optimization model for option hedging using ellipsoidal uncertainty sets. They formulate their model as a second-order cone program which may have, in the worst-case, an exponential number of conic constraints.

By combining robust portfolio optimization and classical portfolio insurance, we aim to provide two layers of guarantees. The weak non-inferiority guarantee applies as long as the returns are realized within the uncertainty set, while the strong portfolio insurance guarantee also covers cases in which the returns are realized outside of the uncertainty set. The ideas set out in this chapter are related to the concept of Comprehensive Robustness proposed by Ben-Tal *et* *al.* [BTBN06]. Comprehensive Robustness aims to control the deterioration in performance when the uncertainties materialize outside of the uncertainty set. Our work establishes the relationship between offering guarantees beyond the uncertainty set and portfolio insurance. Indeed, we will show that in order to control the deterioration in portfolio return, our model will allocate wealth in put and call options. The premia of these options will determine the cost to satisfy the guarantee levels. The contributions in this chapter can be summarized as follows:

(1) We extend the existing robust portfolio optimization models to include options as well as stocks. Because option returns are convex piece-wise linear functions of the underlying stock returns, options cannot be treated as additional stocks, and the use of an ellipsoidal uncertainty set is no longer adequate. Under a no short-sales restriction on the options, we demonstrate how our model can be reformulated as a convex second-order cone program that scales gracefully with the number of stocks and options. We also show that our model implicitly minimizes a *coherent risk measure* [ADEH99]. Coherency is a desirable property from a risk management viewpoint.

(2) We describe how the options in the portfolio can be used to obtain additional strong guarantees on the worst-case portfolio return even when the stock returns are realized outside of the uncertainty set. We show that the arising *Insured Robust Portfolio Optimization* model trades off the guarantees provided through the non-inferiority property and the derivative insurance strategy. Using conic duality, we reformulate this model as a tractable second-order cone program.

(3) We perform a variety of numerical experiments using simulated as well as real market data. In our simulated tests we illustrate the tradeoff between the non-inferiority guarantee and the strong insurance guarantee. We also evaluate the performance of the Insured Robust Portfolio Optimization model under “normal” market conditions, in which the asset prices are governed by geometric Brownian motions, as well as in a market environment in which the prices experience significant downward jumps. The impact of the insurance guarantees on the portfolio performance is also analyzed using real market prices.

The rest of the chapter is organized as follows. In Section 3.2 we review robust portfolio optimization and elaborate on the non-inferiority guarantee. In Section 3.3 we show how a portfolio that contains options can be modelled in a robust optimization framework and how strong insurance guarantees can be imposed on the worst-case portfolio return. We also demonstrate how the resulting model can be formulated as a tractable second-order cone program. In Section 3.4 we report on numerical tests in which we compare the insured robust model with the standard robust model as well as the classical mean-variance model. We run simulated as well as empirical backtests. Conclusions are drawn in Section 3.5, and a notational reference table is provided in Appendix 3.6.1.

3.2 Robust Portfolio Optimization

Consider a market consisting of n stocks. Moreover, denote the current time as $t = 0$ and the end of investment horizon as $t = T$. A portfolio is completely characterized by a vector of weights $\boldsymbol{w} \in \mathbb{R}^n$, whose elements add up to 1. The component w_i denotes the percentage of total wealth which is invested in the ith stock at time $t = 0$. Furthermore, let $\tilde{\boldsymbol{r}}$ denote the random vector of *total* stock returns over the investment horizon, which takes values in $\mathbb{R}^n_+$.[1] By definition, the investor will receive $\tilde{r}_i$ dollars at time T for every dollar invested in stock i at time 0. The return vector $\tilde{\boldsymbol{r}}$ is representable as

$$\tilde{\boldsymbol{r}} = \boldsymbol{\mu} + \tilde{\boldsymbol{\epsilon}}, \tag{3.1}$$

where $\boldsymbol{\mu} = \mathbb{E}[\tilde{\boldsymbol{r}}] \in \mathbb{R}^n_+$ denotes the vector of mean returns and $\tilde{\boldsymbol{\epsilon}} = \tilde{\boldsymbol{r}} - \mathbb{E}[\tilde{\boldsymbol{r}}]$ stands for the vector of residual returns. We assume that $\mathbf{Cov}[\tilde{\boldsymbol{r}}] = \mathbb{E}[\tilde{\boldsymbol{\epsilon}}\tilde{\boldsymbol{\epsilon}}^\mathsf{T}] = \boldsymbol{\Sigma} \in \mathbb{S}^n$ is strictly positive definite. The return $\tilde{r}_p$ on some portfolio $\boldsymbol{w}$ is given by

$$\tilde{r}_p = \boldsymbol{w}^\mathsf{T}\tilde{\boldsymbol{r}} = \boldsymbol{w}^\mathsf{T}\boldsymbol{\mu} + \boldsymbol{w}^\mathsf{T}\tilde{\boldsymbol{\epsilon}}.$$

[1] In this chapter, we will only use total returns because doing so considerably simplifies the notation and mathematical derivations. In Chapter 4, however, we will use relative returns, which are more commonly used in the literature.

Markowitz suggested to determine an optimal tradeoff between the expected return $\mathbb{E}[\tilde{r}_p]$ and the risk $\mathbb{V}\mathrm{ar}[\tilde{r}_p]$ of the portfolio [Mar52]. The optimal portfolio can thus be found by solving the following convex quadratic program

$$\max_{\boldsymbol{w} \in \mathbb{R}^n} \left\{ \boldsymbol{w}^\top \boldsymbol{\mu} - \lambda \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w} \mid \boldsymbol{w}^\top \boldsymbol{e} = 1,\ \boldsymbol{l} \leq \boldsymbol{w} \leq \boldsymbol{u} \right\}, \tag{3.2}$$

where the parameter λ characterizes the investor's risk-aversion, the constant vectors $\boldsymbol{l}, \boldsymbol{u} \in \mathbb{R}^n$ are used to model portfolio constraints, and $\boldsymbol{e} \in \mathbb{R}^n$ denotes a vector of 1s.

3.2.1 Basic Model

Robust optimization offers a different interpretation of the classical Markowitz problem. Ben-Tal and Nemirovski [BTN99] argue that the investor wishes to maximize the portfolio return and thus attempts to solve the uncertain linear program

$$\max_{\boldsymbol{w} \in \mathbb{R}^n} \left\{ \boldsymbol{w}^\top \tilde{\boldsymbol{r}} \mid \boldsymbol{w}^\top \boldsymbol{e} = 1,\ \boldsymbol{l} \leq \boldsymbol{w} \leq \boldsymbol{u} \right\}.$$

However, this problem is not well-defined. It constitutes a whole family of linear programs. In fact, for each return realization we obtain a different optimal solution. In order to disambiguate the investment decisions, robust optimization adopts a worst-case perspective. In this modeling framework, the return vector $\tilde{\boldsymbol{r}}$ remains unknown, but it is believed to materialize within an *uncertainty set* $\mathcal{U}_{\boldsymbol{r}}$. To immunize the portfolio against the inherent uncertainty in $\tilde{\boldsymbol{r}}$, we maximize the worst-case portfolio return, where the worst-case is calculated with respect to all asset returns in $\mathcal{U}_{\boldsymbol{r}}$. This can be formalized as a max-min problem

$$\max_{\boldsymbol{w} \in \mathbb{R}^n} \left\{ \min_{\boldsymbol{r} \in \mathcal{U}_{\boldsymbol{r}}} \boldsymbol{w}^\top \boldsymbol{r} \mid \boldsymbol{w}^\top \boldsymbol{e} = 1,\ \boldsymbol{l} \leq \boldsymbol{w} \leq \boldsymbol{u} \right\}. \tag{3.3}$$

The objective function in (3.3) represents the worst-case portfolio return should $\tilde{\boldsymbol{r}}$ be realized within $\mathcal{U}_{\boldsymbol{r}}$. Note that this quantity depends in a non-trivial way on the portfolio vector $\boldsymbol{w}$.

There are multiple ways to specify $\mathcal{U}_{\boldsymbol{r}}$. A natural choice is to use an ellipsoidal uncertainty set

$$\mathcal{U}_{\boldsymbol{r}} = \left\{ \boldsymbol{r} : (\boldsymbol{r} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{r} - \boldsymbol{\mu}) \leq \delta^2 \right\}. \tag{3.4}$$

As shown in an influential paper by El Ghaoui *et al.* [EGOO03], when $\tilde{\boldsymbol{r}}$ has finite second-order moments, then, the choice

$$\delta = \sqrt{\frac{p}{1-p}} \text{ for } p \in [0, 1) \quad \text{and} \quad \delta = +\infty \text{ for } p = 1 \tag{3.5}$$

implies the following probabilistic guarantee for any portfolio $\boldsymbol{w}$.[2]

$$\mathbb{P} \left\{ \boldsymbol{w}^\top \tilde{\boldsymbol{r}} \geq \min_{\boldsymbol{r} \in \mathcal{U}_{\boldsymbol{r}}} \boldsymbol{w}^\top \boldsymbol{r} \right\} \geq p \tag{3.6}$$

The investor controls the size of the uncertainty set by choosing the parameter p. For p close to 0, the ellipsoid shrinks to $\{\boldsymbol{\mu}\}$, and therefore the investor is only concerned about the average performance of the portfolio. When p is close to 1, the ellipsoid becomes very large, which implies that the investor wants to safeguard against a large set of possible return outcomes. Thus, the choice of uncertainty set size depends on the risk attitude of the investor.

It is shown in [BTN99] that for ellipsoidal uncertainty sets of the type (3.4), problem (3.3) reduces to a convex second-order cone program [LVBL98].

$$\max_{\boldsymbol{w} \in \mathbb{R}^n} \left\{ \boldsymbol{w}^\top \boldsymbol{\mu} - \delta \left\| \boldsymbol{\Sigma}^{1/2} \boldsymbol{w} \right\|_2 \,\middle|\, \boldsymbol{w}^\top \boldsymbol{e} = 1,\ \boldsymbol{l} \leq \boldsymbol{w} \leq \boldsymbol{u} \right\} \tag{3.7}$$

Note that (3.7) is very similar to the classical Markowitz model (3.2). The main difference is that the standard deviation $\left\| \boldsymbol{\Sigma}^{1/2} \boldsymbol{w} \right\|_2 = \sqrt{\boldsymbol{w}^\top \Sigma \boldsymbol{w}}$ replaces the variance. The parameter δ is the analogue of λ, which determines the risk-return tradeoff. It can be shown that (3.2) and (3.7) are equivalent problems in the sense that for every λ there is some δ for which the two problems have the same optimal solution.

[2] In Chapter 4, we will go into much greater detail about the probabilistic guarantees associated with the size of the uncertainty set. For now, we only use (3.5) as a rule to select the uncertainty set size, without emphasizing the probabilistic interpretation.

3.2.2 Parameter Uncertainty

In the Introduction we outlined the shortcomings of the Markowitz model, which carry over to the equivalent mean-standard deviation model (3.7): both models are highly sensitive to the distributional input parameters $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. These parameters, in turn, are difficult to estimate from noisy historical data. The optimization problems (3.2) and (3.7) amplify these estimation errors, yielding extreme portfolios that perform poorly in out-of-sample tests. It turns out that robust optimization can also be used to immunize the portfolio against uncertainties in $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. The starting point of such a robust approach is to assume that the true parameter values are unknown but contained in some uncertainty sets which reflect the investor's confidence in the parameter estimates.

Assume that the true (but unobservable) mean vector $\boldsymbol{\mu} \in \mathbb{R}^n_+$ is known to belong to a set $\mathcal{U}_{\boldsymbol{\mu}}$, and the true covariance matrix $\boldsymbol{\Sigma} \in \mathbb{S}^n$ is known to belong to a set $\mathcal{U}_{\boldsymbol{\Sigma}}$. Robust portfolio optimization aims to find portfolios that perform well under worst-case values of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ within the corresponding uncertainty sets. The parameter robust generalization of problem (3.7) can thus be formulated as

$$\max_{\boldsymbol{w} \in \mathbb{R}^n} \left\{ \min_{\boldsymbol{\mu} \in \mathcal{U}_{\boldsymbol{\mu}}} \boldsymbol{w}^\mathsf{T} \boldsymbol{\mu} - \delta \max_{\boldsymbol{\Sigma} \in \mathcal{U}_{\boldsymbol{\Sigma}}} \left\| \boldsymbol{\Sigma}^{1/2} \boldsymbol{w} \right\|_2 \,\middle|\, \boldsymbol{w}^\mathsf{T} \boldsymbol{e} = 1,\ \boldsymbol{l} \leq \boldsymbol{w} \leq \boldsymbol{u} \right\}. \tag{3.8}$$

There are multiple ways to specify the new uncertainty sets $\mathcal{U}_{\boldsymbol{\mu}}$ and $\mathcal{U}_{\boldsymbol{\Sigma}}$. Let $\hat{\boldsymbol{\mu}}$ be the sample average estimate of $\boldsymbol{\mu}$, and $\hat{\boldsymbol{\Sigma}}$ the sample covariance estimate of $\boldsymbol{\Sigma}$. In the remainder, we will assume that the estimate $\hat{\boldsymbol{\Sigma}}$ is reasonably accurate such that there is no uncertainty about it. This assumption is justified since the estimation error in $\hat{\boldsymbol{\mu}}$ by far outweighs the estimation error in $\hat{\boldsymbol{\Sigma}}$, see e.g. [CZ93]. Thus, we may view the uncertainty set for the covariance matrix as a singleton, $\mathcal{U}_{\boldsymbol{\Sigma}} = \{\hat{\boldsymbol{\Sigma}}\}$. We note that all the following results can be generalized to cases in which $\mathcal{U}_{\boldsymbol{\Sigma}}$ is not a singleton. This, however, leads to more convoluted model formulations. If the stock returns are serially independent and identically distributed, we can invoke the Central Limit Theorem to conclude that the sample mean $\hat{\boldsymbol{\mu}}$ is approximately normally distributed.

Henceforth we will thus assume that

$$\hat{\boldsymbol{\mu}} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Lambda}), \quad \boldsymbol{\Lambda} = (1/E)\boldsymbol{\Sigma}, \tag{3.9}$$

where E is the number of historical samples used to calculate $\hat{\boldsymbol{\mu}}$. It is therefore natural to assume an ellipsoidal uncertainty set for the means,

$$\mathcal{U}_{\boldsymbol{\mu}} = \left\{ \boldsymbol{\mu} : (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})^\mathsf{T} \boldsymbol{\Lambda}^{-1} (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) \leq \kappa^2 \right\}, \tag{3.10}$$

where $\kappa = \sqrt{q/(1-q)}$ for some $q \in [0, 1)$. The confidence level q has an analog interpretation as the parameter p in (3.6). Using the above specifications of the uncertainty sets, problem (3.8) reduces to

$$\max_{\boldsymbol{w} \in \mathbb{R}^n} \left\{ \boldsymbol{w}^\mathsf{T} \hat{\boldsymbol{\mu}} - \kappa \left\| \boldsymbol{\Lambda}^{1/2} \boldsymbol{w} \right\|_2 - \delta \left\| \hat{\boldsymbol{\Sigma}}^{1/2} \boldsymbol{w} \right\|_2 \,\middle|\, \boldsymbol{w}^\mathsf{T} \boldsymbol{e} = 1,\ \boldsymbol{l} \leq \boldsymbol{w} \leq \boldsymbol{u} \right\}, \tag{3.11}$$

see [CS06]. By using the relations (3.9), one easily verifies that (3.11) is equivalent to

$$\max_{\boldsymbol{w} \in \mathbb{R}^n} \left\{ \boldsymbol{w}^\mathsf{T} \hat{\boldsymbol{\mu}} - \left(\frac{\kappa}{\sqrt{E}} + \delta \right) \left\| \hat{\boldsymbol{\Sigma}}^{1/2} \boldsymbol{w} \right\|_2 \,\middle|\, \boldsymbol{w}^\mathsf{T} \boldsymbol{e} = 1,\ \boldsymbol{l} \leq \boldsymbol{w} \leq \boldsymbol{u} \right\}.$$

This problem is equivalent to (3.7) with the risk parameter δ shifted by $\kappa/\sqrt{E}$. Therefore, it is also equivalent to the standard Markowitz model. Hence, seemingly nothing has been gained by incorporating parameter uncertainty into the model (3.7).

Ceria and Stubbs [CS06] demonstrate that robust optimization can nevertheless be used to systematically improve on the common Markowitz portfolios (which are optimal in (3.2), (3.7), and (3.11)). The key idea is to replace the elliptical uncertainty set (3.10) by a less conservative one. Since the estimated expected returns $\hat{\boldsymbol{\mu}}$ are symmetrically distributed around $\boldsymbol{\mu}$, we expect that the estimation errors cancel out when summed over all stocks. It may be more natural and less pessimistic to explicitly incorporate this expectation into the uncertainty model. To this end, Ceria and Stubbs set

$$\mathcal{U}_{\boldsymbol{\mu}} = \left\{ \boldsymbol{\mu} : (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})^{\mathsf{T}} \boldsymbol{\Lambda}^{-1} (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) \leq \kappa^2, \; \boldsymbol{e}^{\mathsf{T}} (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) = 0 \right\}. \tag{3.12}$$

With this new uncertainty set problem (3.8) reduces to

$$\max_{\boldsymbol{w} \in \mathbb{R}^n} \left\{ \boldsymbol{w}^{\mathsf{T}} \hat{\boldsymbol{\mu}} - \kappa \left\| \boldsymbol{\Omega}^{1/2} \boldsymbol{w} \right\|_2 - \delta \left\| \hat{\boldsymbol{\Sigma}}^{1/2} \boldsymbol{w} \right\|_2 \;\middle|\; \boldsymbol{w}^{\mathsf{T}} \boldsymbol{e} = 1, \; \boldsymbol{l} \leq \boldsymbol{w} \leq \boldsymbol{u} \right\}, \tag{3.13}$$

where

$$\boldsymbol{\Omega} = \boldsymbol{\Lambda} - \frac{1}{\boldsymbol{e}^{\mathsf{T}} \boldsymbol{\Lambda} \boldsymbol{e}} \boldsymbol{\Lambda} \boldsymbol{e} \boldsymbol{e}^{\mathsf{T}} \boldsymbol{\Lambda}, \tag{3.14}$$

see [CS06]. A formal derivation of the optimization problem (3.13) is provided in Theorem 3.6.1 in Appendix 3.6.

Example 3.2.1 *We demonstrate the significance of parameter uncertainty on the optimal portfolios with a simple example. Consider a market consisting of two stocks. We assume the their returns are jointly normally distributed with mean vector* $\boldsymbol{\mu}$ *and covariance matrix* $\boldsymbol{\Sigma}$ *set to*

$$\boldsymbol{\mu} = \begin{bmatrix} 1.10 \\ 1.10 \end{bmatrix}, \quad \textit{and} \quad \boldsymbol{\Sigma} = \begin{bmatrix} 0.04 & 0.024 \\ 0.024 & 0.04 \end{bmatrix}. \tag{3.15}$$

Thus, both stocks have a mean return of 1.10 *and volatility of* 0.20*, and are positively correlated with coefficient* 0.6*.*

Of course, in reality, these parameters are not known precisely and must be estimated from historical data. To this end, we draw $E = 250$ *samples from the normal distribution with the above parameters and compute the sample means and covariance matrix by which we obtain*

$$\hat{\boldsymbol{\mu}} = \begin{bmatrix} 1.0891 \\ 1.0970 \end{bmatrix}, \quad \textit{and} \quad \hat{\boldsymbol{\Sigma}} = \begin{bmatrix} 0.0427 & 0.0228 \\ 0.0228 & 0.0369 \end{bmatrix}. \tag{3.16}$$

Note that the estimated parameters are close but not equal to the true parameters due to estimation errors.

Next, we assess the impact of the parameter estimation errors on the optimal portfolios. To this end, we first solve problem (3.7) *with* $\lambda = 1$ *using the* true *parameter values in* (3.15) *and by constraining the weights to be nonnegative. Thus, we solve the following problem.*

$$\begin{aligned} \underset{w_1, w_2}{\text{maximize}} \quad & 1.10w_1 + 1.10w_2 - \sqrt{0.04w_1^2 + 0.048w_1w_2 + 0.04w_2^2} \\ \text{subject to} \quad & w_1 + w_2 = 1 \\ & w_1 \geq 0, \quad w_2 \geq 0 \end{aligned}$$

The above problem is solved using the SDPT3 optimization toolkit [TTT03] and we determine the optimal portfolio weights to be $\boldsymbol{w}^*_{true} = [0.5 \quad 0.5]^\mathsf{T}$. *The equally weighted portfolio solution makes sense since both stocks returns have the same mean and standard deviation.*

We now solve the same problem using the estimated *parameters values in* (3.16).

$$\begin{aligned} \underset{w_1, w_2}{\text{maximize}} \quad & 1.0891w_1 + 1.0970w_2 - \sqrt{0.0427w_1^2 + 0.0456w_1w_2 + 0.0369w_2^2} \\ \text{subject to} \quad & w_1 + w_2 = 1 \\ & w_1 \geq 0, \quad w_2 \geq 0 \end{aligned}$$

The optimal portfolio solution of the above problem is determined to be $\boldsymbol{w}^*_{est} = [0.374 \quad 0.626]^\mathsf{T}$. *Note that* $\boldsymbol{w}^*_{est}$ *is significantly different from* $\boldsymbol{w}^*_{true}$ *due to the estimation errors. In fact, the absolute error is* $|\boldsymbol{w}^*_{est} - \boldsymbol{w}^*_{true}| = 25\%$.

We now focus on problem (3.13), *which explicitly accounts for parameter uncertainty in the means. Firstly, we compute the* $\boldsymbol{\Omega}$ *matrix using equation* (3.14) *and we obtain*

$$\boldsymbol{\Omega} = \begin{bmatrix} 3.382 & -3.382 \\ -3.382 & 3.382 \end{bmatrix} \times 10^{-5}.$$

Next, we solve the following instance of problem (3.13) *with* $\kappa = 2$, *which indicates that we are* *uncertain about the mean estimates.*

$$\begin{aligned} \underset{w_1,w_2}{\text{maximize}} \quad & 1.0891w_1 + 1.0970w_2 - 2\times 10^{-5}\sqrt{11.4379w_1^2 - 6.7640w_1w_2 + 11.4379w_2^2} \\ & \qquad - \sqrt{0.0427w_1^2 + 0.0456w_1w_2 + 0.0369w_2^2} \\ \text{subject to} \quad & w_1 + w_2 = 1 \\ & w_1 \geq 0, \quad w_2 \geq 0 \end{aligned}$$

The optimal portfolio solution of the above problem is $\boldsymbol{w}^*_{rob} = [0.495 \quad 0.505]^\mathsf{T}$. *Note that* $\boldsymbol{w}^*_{rob}$ *lies significantly closer to* $\boldsymbol{w}^*_{true}$ *than* $\boldsymbol{w}^*_{est}$. *In fact, the absolute error is* $|\boldsymbol{w}^*_{rob} - \boldsymbol{w}^*_{true}| = 1\%$. *This simple example demonstrates that the robust portfolio optimization model* (3.13) *produces portfolios which are less sensitive to estimation errors.*

3.2.3 Uncertainty Sets with Support Information

For ease of exposition, consider again the basic model of Section 3.2.1. When the uncertainty set $\mathcal{U}_{\boldsymbol{r}}$ becomes excessively large, as is the case when $\delta \to +\infty$ or, equivalently, when $p \to 1$ (see (3.5)), $\mathcal{U}_{\boldsymbol{r}}$ may extend beyond the support of $\tilde{\boldsymbol{r}}$, which coincides with the positive orthant of $\mathbb{R}^n$. The resulting portfolios can then become unnecessarily conservative. To overcome this deficiency, we modify $\mathcal{U}_{\boldsymbol{r}}$ defined in (3.4) by including a non-negativity constraint

$$\mathcal{U}_{\boldsymbol{r}}^+ = \left\{ \boldsymbol{r} \geq \boldsymbol{0} : (\boldsymbol{r} - \boldsymbol{\mu})^\mathsf{T} \boldsymbol{\Sigma}^{-1} (\boldsymbol{r} - \boldsymbol{\mu}) \leq \delta^2 \right\}. \tag{3.17}$$

It can be shown that problem (3.3) with $\mathcal{U}_{\boldsymbol{r}}$ replaced by $\mathcal{U}_{\boldsymbol{r}}^+$ is equivalent to

$$\max_{\boldsymbol{w},\boldsymbol{s}\in\mathbb{R}^n} \left\{ \boldsymbol{\mu}^\mathsf{T}(\boldsymbol{w} - \boldsymbol{s}) - \delta \left\| \boldsymbol{\Sigma}^{1/2}(\boldsymbol{w} - \boldsymbol{s}) \right\|_2 \,\middle|\, \boldsymbol{w}^\mathsf{T}\boldsymbol{e} = 1, \ \boldsymbol{s} \geq \boldsymbol{0}, \ \boldsymbol{l} \leq \boldsymbol{w} \leq \boldsymbol{u} \right\}. \tag{3.18}$$

Remark 3.2.1 (Relation to coherent risk measures) *Problem* (3.18) *can be shown to implicitly minimize a* coherent downside risk measure *[ADEH99] associated with the underlying uncertainty set, see Section 2.4.3 for an overview of coherent risk measures. Natarajan* et al. *[NPS08] show that there exists a one-to-one correspondence between uncertainty sets and risk measures (see also [BB08]). In what follows, we will briefly explain this correspondence in* *the context of problem* (3.18)*. Introduce a linear space of random variables*

$$\mathcal{V} = \left\{ \boldsymbol{w}^\mathsf{T} \tilde{\boldsymbol{r}} : \boldsymbol{w} \in \mathbb{R}^n \right\}, \tag{3.19}$$

and define the risk measure $\rho : \mathcal{V} \to \mathbb{R}$ *through*

$$\begin{aligned} \rho(\boldsymbol{w}^\mathsf{T} \tilde{\boldsymbol{r}}) &= \max_{\boldsymbol{r}} \left\{ -\boldsymbol{w}^\mathsf{T} \boldsymbol{r} \mid \boldsymbol{r} \in \mathcal{U}_{\boldsymbol{r}}^+ \right\} \\ &= \min_{\boldsymbol{s} \geq \boldsymbol{0}} -\boldsymbol{\mu}^\mathsf{T}(\boldsymbol{w} - \boldsymbol{s}) + \delta \left\| \boldsymbol{\Sigma}^{1/2}(\boldsymbol{w} - \boldsymbol{s}) \right\|_2 . \end{aligned} \tag{3.20}$$

It can be seen that problem (3.18) *is equivalent to the risk minimization problem*

$$\min_{\boldsymbol{w}} \left\{ \rho\left(\boldsymbol{w}^\mathsf{T} \tilde{\boldsymbol{r}}\right) \mid \boldsymbol{e}^\mathsf{T} \boldsymbol{w} = 1,\ \boldsymbol{l} \leq \boldsymbol{w} \leq \boldsymbol{u} \right\}. \tag{3.21}$$

Since the feasible set in (3.20) *is a subset of the support of* $\tilde{\boldsymbol{r}}$*, the risk measure* ρ *is coherent, see [NPS08, Theorem 4]. Moreover,* ρ *can be viewed as a downside risk measure since it evaluates to worst-case return over an uncertainty set centered around the expected asset return vector.*

As in Section 3.2.2, model (3.18) may be improved by immunizing it against the uncertainty in the distributional input parameters. Using similar arguments as in Theorem 3.6.1, it can be shown that the parameter robust variant of problem (3.18),

$$\max_{\boldsymbol{w}, \boldsymbol{s}} \left\{ \min_{\boldsymbol{\mu} \in \mathcal{U}_{\boldsymbol{\mu}}} \boldsymbol{\mu}^\mathsf{T}(\boldsymbol{w} - \boldsymbol{s}) - \delta \max_{\boldsymbol{\Sigma} \in \mathcal{U}_{\boldsymbol{\Sigma}}} \left\| \boldsymbol{\Sigma}^{1/2}(\boldsymbol{w} - \boldsymbol{s}) \right\|_2 \,\middle|\, \boldsymbol{w}^\mathsf{T} \boldsymbol{e} = 1,\ \boldsymbol{s} \geq \boldsymbol{0},\ \boldsymbol{l} \leq \boldsymbol{w} \leq \boldsymbol{u} \right\},$$

is equivalent to

$$\max_{\boldsymbol{w}, \boldsymbol{s}, \boldsymbol{v}} \left\{ \hat{\boldsymbol{\mu}}^\mathsf{T} \boldsymbol{v} - \kappa \left\| \boldsymbol{\Omega}^{1/2} \boldsymbol{v} \right\|_2 - \delta \left\| \hat{\boldsymbol{\Sigma}}^{1/2} \boldsymbol{v} \right\|_2 \,\middle|\, \boldsymbol{w}^\mathsf{T} \boldsymbol{e} = 1,\ \boldsymbol{w} - \boldsymbol{s} = \boldsymbol{v},\ \boldsymbol{s} \geq \boldsymbol{0},\ \boldsymbol{l} \leq \boldsymbol{w} \leq \boldsymbol{u} \right\}. \tag{3.22}$$

We note that we could have directly obtained (3.22) from the basic model (3.3) by defining the uncertainty set for the returns as

$$\mathcal{U}_{r,\boldsymbol{\mu}}^+ = \left\{ \boldsymbol{r} \geq \boldsymbol{0} : \exists \boldsymbol{\mu} \in \mathcal{U}_{\boldsymbol{\mu}},\ (\boldsymbol{r} - \boldsymbol{\mu})^\mathsf{T} \boldsymbol{\Sigma}^{-1} (\boldsymbol{r} - \boldsymbol{\mu}) \leq \delta^2 \right\} \tag{3.23}$$

where $\mathcal{U}_{\boldsymbol{\mu}}$ is defined as in (3.12). The uncertainty set $\mathcal{U}^{+}_{\boldsymbol{r},\boldsymbol{\mu}}$ accounts for the uncertainty in the returns whilst taking into consideration that the centroid $\boldsymbol{\mu}$ of $\mathcal{U}^{+}_{\boldsymbol{r}}$, as defined in (3.17), has to be estimated and is therefore also subject to uncertainty.

Problem (3.22) implicitly minimizes a coherent risk measure associated with the uncertainty set $\mathcal{U}^{+}_{\boldsymbol{r},\boldsymbol{\mu}}$. Coherency holds since $\mathcal{U}^{+}_{\boldsymbol{r},\boldsymbol{\mu}}$ is a subset of the support of $\tilde{\boldsymbol{r}}$, see Remark 3.2.1. Some risk-tolerant investors may not want to minimize a risk measure without imposing a constraint on the portfolio return. Taking into account the uncertainty in the expected asset returns motivates us to constrain the worst-case expected portfolio return,

$$\min_{\boldsymbol{\mu} \in \mathcal{U}_{\boldsymbol{\mu}}} \boldsymbol{w}^{\mathsf{T}} \boldsymbol{\mu} \geq \mu_{\text{target}},$$

where μ_{target} represents the return target the investor wishes to attain in average. This semi-infinite constraint can be reformulated as a second-order cone constraint of the form

$$\boldsymbol{w}^{\mathsf{T}} \hat{\boldsymbol{\mu}} - \kappa \left\| \boldsymbol{\Omega}^{1/2} \boldsymbol{w} \right\|_2 \geq \mu_{\text{target}}. \tag{3.24}$$

The optimal portfolios obtained from problem (3.22), with or without the return target constraint (3.24), provide certain performance guarantees. They exhibit a *non-inferiority property* in the sense that, as long as the asset returns materialize within the prescribed uncertainty set, the realized portfolio return never falls below the optimal value of problem (3.22). However, no guarantees are given when the asset returns are realized outside of the uncertainty set.

In Section 3.3 we suggest the use of derivatives to enforce strong performance guarantees, which will complement the weak guarantees provided by the non-inferiority property.

3.3 Insured Robust Portfolio Optimization

Since their introduction in the second half of the last century, options have been praised for their ability to give stock holders protection against adverse market fluctuations [Mac92]. A standard option contract is determined by the following parameters: the premium or price of the option, the underlying security, the expiration date, and the strike price. A put (call) option gives the option holder the right, but not the obligation, to sell to (buy from) the option writer the underlying security by the expiration date and at the prescribed strike price. American options can be exercised at any time up to the expiration date, whereas European options can be exercised only on the expiration date itself. We will only work with European options, which expire at the end of investment horizon, that is, at time T. We restrict attention to these instruments because of their simplicity and since they fit naturally in the single period portfolio optimization framework of the previous section.

We now briefly illustrate how options can be used to insure a stock portfolio. An option's payoff function represents its value at maturity as a function of the underlying stock price S_T. For put and call options with strike price K, the payoff functions are thus given by

$$V_{\mathrm{put}}(S_T) = \max\{0, K - S_T\} \quad \text{and} \quad V_{\mathrm{call}}(S_T) = \max\{0, S_T - K\}, \tag{3.25}$$

respectively. Assume now that we hold a portfolio of a single long stock and a put option on this stock with strike price K. Then, the payoff of the portfolio amounts to

$$V_{\mathrm{pf}}(S_T) = S_T + V_{\mathrm{put}}(S_T) = \max\{S_T, K\}.$$

This shows that the put option with strike price K prevents the portfolio value at maturity from dropping below K. Of course, this insurance comes at the cost of the option premium, which has to be paid at the time when the option contract is negotiated.

Similarly, assume that we hold a portfolio of a single shorted stock and a call option on this stock with strike price K. Then, the payoff function of this portfolio is

$$V_{\mathrm{pf}}(S_T) = -S_T + V_{\mathrm{call}}(S_T) = \max\{-S_T, -K\},$$

which insures the portfolio value at maturity against falling below $-K$.

Although we focus on European options expiring at time T, all models to be developed in this chapter remain valid for American options exercisable at time T. We emphasize that the timing flexibility of American options cannot be exploited in the single-period setting under consideration, and therefore American options are usually too expensive for our purposes. Nevertheless, if there are only very few European options expiring at the end of the investment horizon, it may be beneficial to include American options into our portfolio to increase the spectrum of available strike prices.

3.3.1 Robust Portfolio Optimization with Options

Assume that there are m European options in our market, each of which has one of the n stocks as an underlying security. We denote the initial investment in the options by the vector $\boldsymbol{w^d} \in \mathbb{R}^m$. The component w_i^d denotes the percentage of total wealth which is invested in the ith option at time $t = 0$. A portfolio is now completely characterized by a joint vector $(\boldsymbol{w}, \boldsymbol{w^d}) \in \mathbb{R}^{n+m}$, whose elements add up to 1. In what follows, we will forbid short-sales of options and therefore require that $\boldsymbol{w^d} \geq \boldsymbol{0}$. Short-selling of options can be very risky, and therefore the imposed restriction should be in line with the preferences of a risk-averse investor.

The return $\tilde{r}_p$ of some portfolio $(\boldsymbol{w}, \boldsymbol{w^d})$ is given by

$$\tilde{r}_p = \boldsymbol{w}^\mathsf{T} \tilde{\boldsymbol{r}} + (\boldsymbol{w^d})^\mathsf{T} \tilde{\boldsymbol{r}}^{\boldsymbol{d}}, \tag{3.26}$$

where $\tilde{\boldsymbol{r}}^{\boldsymbol{d}}$ represents the vector of option returns. It is important to note that $\tilde{\boldsymbol{r}}^{\boldsymbol{d}}$ is uniquely determined by $\tilde{\boldsymbol{r}}$, that is, there exists a function $f : \mathbb{R}^n \to \mathbb{R}^m$ such that $\tilde{\boldsymbol{r}}^{\boldsymbol{d}} \equiv f(\tilde{\boldsymbol{r}})$.

Let option j be a call with strike price K_j on the underlying stock i, and denote the return and the initial price of the option by $\tilde{r}_j^d$ and C_j, respectively. If S_0^i denotes the initial price of stock i, then its end-of-period price can be expressed as $S_0^i \tilde{r}_i$. Using the above notation, we can now explicitly express the return $\tilde{r}^d_j$ as a convex piece-wise linear function of $\tilde{r}_i$,

$$\begin{aligned} f_j(\tilde{\boldsymbol{r}}) &= \frac{1}{C_j} \max\left\{0, S^i_0 \tilde{r}_i - K_j\right\} \\ &= \max\left\{0, a_j + b_j \tilde{r}_i\right\}, \text{ with } a_j = -\frac{K_j}{C_j} < 0 \text{ and } b_j = \frac{S^i_0}{C_j} > 0. \end{aligned} \tag{3.27a}$$

Similarly, if $\tilde{r}^d_j$ is the return of a put option with price P_j and strike price K_j on the underlying stock i, then $\tilde{r}^d_j$ is representable as a slightly different convex piece-wise linear function of $\tilde{r}_i$,

$$f_j(\tilde{\boldsymbol{r}}) = \max\left\{0, a_j + b_j \tilde{r}_i\right\}, \text{ with } a_j = \frac{K_j}{P_j} > 0 \text{ and } b_j = -\frac{S^i_0}{P_j} < 0. \tag{3.27b}$$

Using the above notation, we can write the vector of option returns $\tilde{\boldsymbol{r}}^{\boldsymbol{d}}$ compactly as

$$\tilde{\boldsymbol{r}}^{\boldsymbol{d}} = f(\tilde{\boldsymbol{r}}) = \max\left\{\boldsymbol{0}, \boldsymbol{a} + \mathbf{B}\tilde{\boldsymbol{r}}\right\}, \tag{3.28}$$

where $\boldsymbol{a} \in \mathbb{R}^m$, $\mathbf{B} \in \mathbb{R}^{m \times n}$ are known constants determined through (3.27a) and (3.27b), and 'max' denotes the component-wise maximization operator.

As in Section 3.2.3, we adopt the view that the investor wishes to maximize the worst-case portfolio return whilst assuming that the stock returns $\tilde{\boldsymbol{r}}$ will materialize within the uncertainty set $\mathcal{U}^+_{\boldsymbol{r}}$ as defined in (3.17). This problem can be formalized as

$$\max_{\boldsymbol{w}, \boldsymbol{w}^{\boldsymbol{d}}} \left\{ \min_{\substack{\boldsymbol{r} \in \mathcal{U}^+_{\boldsymbol{r}} \\ \boldsymbol{r}^{\boldsymbol{d}} = f(\boldsymbol{r})}} \boldsymbol{w}^\top \boldsymbol{r} + (\boldsymbol{w}^{\boldsymbol{d}})^\top \boldsymbol{r}^{\boldsymbol{d}} \;\middle|\; \boldsymbol{e}^\top \boldsymbol{w} + \boldsymbol{e}^\top \boldsymbol{w}^{\boldsymbol{d}} = 1,\ \boldsymbol{l} \le \boldsymbol{w} \le \boldsymbol{u},\ \boldsymbol{w}^{\boldsymbol{d}} \ge \boldsymbol{0} \right\}, \tag{3.29}$$

which is equivalent to

$$\begin{aligned} &\text{maximize} && \phi && (3.30a) \\ &\text{subject to} && \boldsymbol{w} \in \mathbb{R}^n, \quad \boldsymbol{w}^{\boldsymbol{d}} \in \mathbb{R}^m, \quad \phi \in \mathbb{R} && \\ &&& \boldsymbol{w}^\top \boldsymbol{r} + (\boldsymbol{w}^{\boldsymbol{d}})^\top \boldsymbol{r}^{\boldsymbol{d}} \ge \phi \qquad \forall \boldsymbol{r} \in \mathcal{U}^+_{\boldsymbol{r}},\ \boldsymbol{r}^{\boldsymbol{d}} = f(\boldsymbol{r}) && (3.30b) \\ &&& \boldsymbol{e}^\top \boldsymbol{w} + \boldsymbol{e}^\top \boldsymbol{w}^{\boldsymbol{d}} = 1 && (3.30c) \\ &&& \boldsymbol{l} \le \boldsymbol{w} \le \boldsymbol{u},\ \boldsymbol{w}^{\boldsymbol{d}} \ge \boldsymbol{0}. && (3.30d) \end{aligned}$$

Note that the worst-case objective is reexpressed in terms of the semi-infinite constraint (3.30b), and at optimality, ϕ represents the worst-case portfolio return. In the remainder we will work with the epigraph formulation (3.30) instead of the max-min formulation (3.29) because it enables us to incorporate portfolio insurance constraints in a convenient way, see Section 3.3.2.

The constraint (3.30b) looks intractable, but it can be reformulated in terms of finitely many conic constraints.

Theorem 3.3.1 *Problem* (3.30) *is equivalent to*

$$\begin{aligned}
&\text{maximize} && \phi && \text{(3.31a)}\\
&\text{subject to} && \boldsymbol{w} \in \mathbb{R}^n, \quad \boldsymbol{w^d} \in \mathbb{R}^m, \quad \boldsymbol{y} \in \mathbb{R}^m, \quad \boldsymbol{s} \in \mathbb{R}^n, \quad \phi \in \mathbb{R} && \\
& && \boldsymbol{\mu}^\mathsf{T}(\boldsymbol{w} + \mathbf{B}^T\boldsymbol{y} - \boldsymbol{s}) - \delta \left\| \boldsymbol{\Sigma}^{1/2}(\boldsymbol{w} + \mathbf{B}^T\boldsymbol{y} - \boldsymbol{s}) \right\|_2 + \boldsymbol{a}^\mathsf{T}\boldsymbol{y} \geq \phi && \text{(3.31b)}\\
& && \boldsymbol{e}^\mathsf{T}\boldsymbol{w} + \boldsymbol{e}^\mathsf{T}\boldsymbol{w^d} = 1 && \text{(3.31c)}\\
& && \boldsymbol{0} \leq \boldsymbol{y} \leq \boldsymbol{w^d}, \ \boldsymbol{s} \geq \boldsymbol{0} && \text{(3.31d)}\\
& && \boldsymbol{l} \leq \boldsymbol{w} \leq \boldsymbol{u}, \ \boldsymbol{w^d} \geq \boldsymbol{0}, && \text{(3.31e)}
\end{aligned}$$

which is a tractable second-order cone program.

Proof Assume first that $\delta > 0$. We observe that the semi-infinite constraint (3.30b) can be reexpressed in terms of the solution of a subordinate minimization problem,

$$\min_{\substack{r \in \mathcal{U}_r \\ r^d = f(r)}} \ \boldsymbol{w}^\mathsf{T}\boldsymbol{r} + (\boldsymbol{w^d})^\mathsf{T}\boldsymbol{r^d} \geq \phi. \tag{3.32}$$

By using the definitions of the function f and the set $\mathcal{U}_{\boldsymbol{r}}^+$, we obtain a more explicit representation for this subordinate problem.

$$
\begin{aligned}
\text{minimize} \quad & \boldsymbol{w}^\top \boldsymbol{r} + (\boldsymbol{w^d})^\top \boldsymbol{r^d} \\
\text{subject to} \quad & \boldsymbol{r} \in \mathbb{R}^n, \quad \boldsymbol{r^d} \in \mathbb{R}^m \\
& \left\| \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{r} - \boldsymbol{\mu}) \right\|_2 \leq \delta \\
& \boldsymbol{r} \ \geq \mathbf{0} \\
& \boldsymbol{r^d} \geq \mathbf{0} \\
& \boldsymbol{r^d} \geq \boldsymbol{a} + \mathbf{B}\boldsymbol{r}
\end{aligned}
\tag{3.33}
$$

For any fixed portfolio vector $(\boldsymbol{w}, \boldsymbol{w^d})$ feasible in (3.30), problem (3.33) represents a convex second-order cone program. Note that since $\boldsymbol{w^d} \geq \mathbf{0}$ for any admissible portfolio, (3.33) has an optimal solution $(\boldsymbol{r}, \boldsymbol{r^d})$ which satisfies the relation (3.28). The dual problem associated with (3.33) reads:

$$
\begin{aligned}
\text{maximize} \quad & \boldsymbol{\mu}^\top(\boldsymbol{w} + \mathbf{B}^\top \boldsymbol{y} - \boldsymbol{s}) - \delta \left\| \boldsymbol{\Sigma}^{1/2}(\boldsymbol{w} + \mathbf{B}^T \boldsymbol{y} - \boldsymbol{s}) \right\|_2 + \boldsymbol{a}^\top \boldsymbol{y} \\
\text{subject to} \quad & \boldsymbol{y} \in \mathbb{R}^m, \quad \boldsymbol{s} \in \mathbb{R}^n \\
& \mathbf{0} \leq \boldsymbol{y} \leq \boldsymbol{w^d}, \ \boldsymbol{s} \geq \mathbf{0}
\end{aligned}
\tag{3.34}
$$

Note that strong conic duality holds since the primal problem (3.33) is strictly feasible for $\delta > 0$, see [AG03, LVBL98]. Thus, both the primal and dual problems (3.33) and (3.34) are feasible and share the same objective values at optimality. This allows us to replace the inner minimization problem in (3.32) by the maximization problem (3.34). The requirement that the optimal value of (3.34) be larger than or equal to ϕ is equivalent to the assertion that there exist $\boldsymbol{y} \in \mathbb{R}^m, \boldsymbol{s} \in \mathbb{R}^n$ feasible in (3.34) whose objective value is larger than or equal to ϕ. This justifies the constraints (3.31b) and (3.31d). All other constraints and the objective function in (3.31) are the same as in (3.30), and thus the two problems are equivalent.

We now assume that $\delta = 0$. Then, by definition, the uncertainty set $\mathcal{U}_{\boldsymbol{r}}^+ = \{\boldsymbol{\mu}\}$ and $\boldsymbol{r^d} = f(\boldsymbol{\mu})$.

Therefore, constraint (3.30b) reduces to

$$
\begin{aligned}
& \boldsymbol{\mu}^{\mathsf{T}}\boldsymbol{w} + f(\boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{w^d} \geq \phi \\
\iff\ & \boldsymbol{\mu}^{\mathsf{T}}\boldsymbol{w} + \left(\max\left\{\mathbf{0},\ \boldsymbol{a} + \mathrm{B}\boldsymbol{\mu}\right\}\right)^{\mathsf{T}}\boldsymbol{w^d} \geq \phi \\
\iff\ & \boldsymbol{\mu}^{\mathsf{T}}\boldsymbol{w} + \max_{\mathbf{0}\leq\boldsymbol{y}\leq\boldsymbol{w^d}} \left\{\boldsymbol{a}^{\mathsf{T}}\boldsymbol{y} + \boldsymbol{\mu}^{\mathsf{T}}\mathrm{B}^{\mathsf{T}}\boldsymbol{y}\right\} \geq \phi \\
\iff\ & \max_{\substack{\mathbf{0}\leq\boldsymbol{y}\leq\boldsymbol{w^d}\\ \boldsymbol{s}\geq\mathbf{0}}} \left\{\boldsymbol{\mu}^{\mathsf{T}}(\boldsymbol{w} + \mathrm{B}^{T}\boldsymbol{y} - \boldsymbol{s}) + \boldsymbol{a}^{\mathsf{T}}\boldsymbol{y}\right\} \geq \phi,
\end{aligned}
$$

where the last equivalence holds because $\boldsymbol{\mu} \geq \mathbf{0}$. Constraint (3.30b) is thus equivalent to (3.31b) and (3.31d). ∎

Observe that in the absence of options we must set $\boldsymbol{w^d} = \mathbf{0}$, which implies via constraint (3.31d) that $\boldsymbol{y} = \mathbf{0}$. Thus, (3.31) reduces to (3.18), that is, the robust portfolio optimization problem of a stock only portfolio.

We note that Lutgens *et al.* [LSK06] propose a robust portfolio optimization model that incorporates options and also allows short-sales of options. However, their problem reformulation contains, in the worst case, an exponential amount of second-order constraints whereas our reformulation (3.31) only contains a single conic constraint at the cost of excluding short-sales of options.

Example 3.3.1 *Consider a market consisting of a stock and a European put option written on this stock. Assume that the stock has an expected monthly return of* 1.01 *and monthly volatility of* 9%. *The initial price of the stock is* $S_0 = \$100$. *The option matures in 21 days and has a strike price of* $K = \$100$. *Furthermore, we assume that the price of the put option is* $P = \$3.58$.

In this example we wish to compute the optimal portfolio containing these two assets using model (3.31). *We assume that the modeler assigns* $p = 70\%$ *uncertainty to the stock return. Thus, using equation* (3.5), *we obtain* $\delta = 1.53$. *Now we compute the option specific multipliers* *a and b, see* (3.27b)*. More specifically, we have*

$$a = \frac{K}{P} = \frac{100}{3.58} = 27.93 \quad \textit{and} \quad b = -\frac{S_0}{P} = -\frac{100}{3.58} = -27.93.$$

We now insert the above parameter values into model (3.31) *and we obtain*

$$\begin{array}{ll} \text{maximize} & \phi \\ \text{subject to} & w \in \mathbb{R}, \quad w^d \in \mathbb{R}, \quad y \in \mathbb{R}, \quad s \in \mathbb{R}, \quad \phi \in \mathbb{R} \\ & 1.01(w - 27.93y - s) - 1.53 \left\|0.09(w - 27.93y - s) + 27.93y\right\|_2 \geq \phi \\ & w + w^d = 1 \\ & 0 \leq y \leq w^d, \ s \geq 0 \\ & w \geq 0, \ w^d \geq 0. \end{array}$$

We solve the above problem using SDPT3 [TTT03] and we obtain the optimal solution values $w^* = 0.9654$ *and* $w^{d*} = 0.0346$*. Thus, the majority of the wealth is invested in the stock whereas the remainder is invested in the put option to hedge away the downside risk. In fact, the optimal amount of units of the stock in the portfolio is* $w^*/S_0 = 0.9654/100 = 0.0097$ *and the optimal amount of units of the put option is* $w^{d*}/P = 0.0346/3.58 = 0.0097$*. Thus, the optimal solution is to match the investment of stock with the option precisely.*

As in Section 3.2.3, one can immunize model (3.30) against estimation errors in $\hat{\boldsymbol{\mu}}$. If we replace the uncertainty set $\mathcal{U}_{\boldsymbol{r}}^+$ by $\mathcal{U}_{\boldsymbol{r},\boldsymbol{\mu}}^+$ defined in (3.23), then problem (3.30) reduces to the following second-order cone program similar to (3.31).

$$\begin{array}{ll} \text{maximize} & \phi \\ \text{subject to} & \hat{\boldsymbol{\mu}}^\mathsf{T}\boldsymbol{v} - \kappa \left\|\boldsymbol{\Omega}^{1/2}\boldsymbol{v}\right\|_2 - \delta \left\|\hat{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{v}\right\|_2 + \boldsymbol{a}^\mathsf{T}\boldsymbol{y} \geq \phi \\ & \boldsymbol{w} + \mathbf{B}^T\boldsymbol{y} - \boldsymbol{s} = \boldsymbol{v}, \ \text{and (3.31c), (3.31d), (3.31e)} \end{array} \tag{3.35}$$

This model guarantees the optimal portfolio return to exceed ϕ conditional on the stock returns $\tilde{\boldsymbol{r}}$ being realized within the uncertainty set $\mathcal{U}_{\boldsymbol{r},\boldsymbol{\mu}}^+$. In what follows, we will thus refer to ϕ as the *conditional worst-case return.*

3.3.2 Robust Portfolio Optimization with Insurance Guarantees

We now augment model (3.35) by requiring the realized portfolio return to exceed some fraction $\theta \in [0, 1]$ of ϕ under every possible realization of the return vector $\tilde{\boldsymbol{r}}$. This requirement is enforced through a semi-infinite constraint of the form

$$\boldsymbol{w}^\mathsf{T}\boldsymbol{r} + (\boldsymbol{w}^{\boldsymbol{d}})^\mathsf{T}\boldsymbol{r}^{\boldsymbol{d}} \geq \theta\phi \qquad \forall \boldsymbol{r} \geq \boldsymbol{0},\ \boldsymbol{r}^{\boldsymbol{d}} = f(\boldsymbol{r}). \tag{3.36}$$

Model (3.35) with the extra constraint (3.36) provides two layers of guarantees: the weak non-inferiority guarantee applies as long as the returns are realized within the uncertainty set, while the strong portfolio insurance guarantee (3.36) also covers cases in which the stock returns are realized outside of $\mathcal{U}^+_{\boldsymbol{r},\boldsymbol{\mu}}$.[3] The level of the portfolio insurance guarantee is expressed as a percentage θ of the conditional worst-case portfolio return ϕ, which can be interpreted as the level of the non-inferiority guarantee. This reflects the idea that the derivative insurance strategy only has to hedge against certain extreme scenarios, which are not already covered by the non-inferiority guarantee. It also prevents the portfolio insurance from being overly expensive. The *Insured Robust Portfolio Optimization* model can be formulated as

$$\begin{aligned}
&\text{maximize} && \phi && \text{(3.37a)}\\
&\text{subject to} && \boldsymbol{w} \in \mathbb{R}^n, \quad \boldsymbol{w}^{\boldsymbol{d}} \in \mathbb{R}^m, \quad \phi \in \mathbb{R} &&\\
& && \boldsymbol{w}^\mathsf{T}\boldsymbol{r} + (\boldsymbol{w}^{\boldsymbol{d}})^\mathsf{T}\boldsymbol{r}^{\boldsymbol{d}} \geq \phi \qquad \forall \boldsymbol{r} \in \mathcal{U}^+_{\boldsymbol{r},\boldsymbol{\mu}},\ \boldsymbol{r}^{\boldsymbol{d}} = f(\boldsymbol{r}) && \text{(3.37b)}\\
& && \boldsymbol{w}^\mathsf{T}\boldsymbol{r} + (\boldsymbol{w}^{\boldsymbol{d}})^\mathsf{T}\boldsymbol{r}^{\boldsymbol{d}} \geq \theta\phi \qquad \forall \boldsymbol{r} \geq \boldsymbol{0},\ \boldsymbol{r}^{\boldsymbol{d}} = f(\boldsymbol{r}) && \text{(3.37c)}\\
& && \boldsymbol{e}^\mathsf{T}\boldsymbol{w} + \boldsymbol{e}^\mathsf{T}\boldsymbol{w}^{\boldsymbol{d}} = 1 && \text{(3.37d)}\\
& && \boldsymbol{l} \leq \boldsymbol{w} \leq \boldsymbol{u},\ \boldsymbol{w}^{\boldsymbol{d}} \geq \boldsymbol{0}. && \text{(3.37e)}
\end{aligned}$$

[3] In reality one has to also consider counterparty risk of the options, but this is beyond the scope of this thesis.

Note that the conditional worst-case return ϕ drops when the uncertainty set $\mathcal{U}^+_{\boldsymbol{r},\boldsymbol{\mu}}$ increases. At the same time, the required insurance level decreases, and hence the insurance premium drops as well. This manifests the tradeoff between the non-inferiority and insurance guarantees. In Proposition 3.3.1 below we show that when the highest possible uncertainty is assigned to the returns (by setting $p = 1$, see (3.5)), or the highest insurance guarantee is demanded (by setting $\theta = 1$), the same optimal conditional worst-case return is obtained. Intuitively, this can be explained as follows. When the uncertainty set covers the whole support, then the insurance guarantee adds nothing to the non-inferiority guarantee. Conversely, the highest possible insurance is independent of the size of the uncertainty set.

Proposition 3.3.1 *If $\boldsymbol{u} \geq \boldsymbol{0}$, then the optimal objective value of problem* (3.37) *for $p = 1$ coincides with the optimal value obtained for $\theta = 1$.*

Proof Since $\boldsymbol{u} \geq \boldsymbol{0}$, there are feasible portfolios with $\boldsymbol{w} \geq \boldsymbol{0}$. Thus, $\phi \geq \theta\phi \geq 0$ at optimality. For $p = 1$, the uncertainty sets in (3.37b) and (3.37c) coincide, which implies that (3.37c) becomes redundant. For $\theta = 1$, on the other hand, (3.37b) becomes redundant. In both cases we end up with the same constraint set. Thus, the claim follows. ∎

Although we exclusively use uncertainty sets of the type (3.23), the models in this chapter do not rely on any assumptions about the size or shape of $\mathcal{U}^+_{\boldsymbol{r},\boldsymbol{\mu}}$ and can be extended to almost any other geometry. We note that for the models to be tractable, it must be possible to describe $\mathcal{U}^+_{\boldsymbol{r},\boldsymbol{\mu}}$ through finitely many linear or conic constraints.

Problem (3.37) involves two semi-infinite constraints: (3.37b) and (3.37c). In Theorem 3.3.2 we show that (3.37) still has a reformulation as a tractable conic optimization problem.

Theorem 3.3.2 *Problem* (3.37) *is equivalent to the following second-order cone program.*

$$
\begin{array}{ll}
\text{maximize} & \phi \\
\text{subject to} & \boldsymbol{w} \in \mathbb{R}^n, \quad \boldsymbol{s} \in \mathbb{R}^n, \quad \boldsymbol{w^d} \in \mathbb{R}^m, \quad \boldsymbol{y} \in \mathbb{R}^m, \quad \boldsymbol{z} \in \mathbb{R}^m, \quad \phi \in \mathbb{R} \\
& \hat{\boldsymbol{\mu}}^\top \boldsymbol{v} - \kappa \left\| \boldsymbol{\Omega}^{1/2} \boldsymbol{v} \right\|_2 - \delta \left\| \hat{\boldsymbol{\Sigma}}^{1/2} \boldsymbol{v} \right\|_2 + \boldsymbol{a}^\top \boldsymbol{y} \geq \phi \\
& \boldsymbol{a}^\top \boldsymbol{z} \geq \theta \phi \\
& \boldsymbol{w} + \mathbf{B}^T \boldsymbol{y} - \boldsymbol{s} = \boldsymbol{v} \\
& \boldsymbol{w} + \mathbf{B}^\top \boldsymbol{z} \geq \mathbf{0} \\
& \boldsymbol{e}^\top \boldsymbol{w} + \boldsymbol{e}^\top \boldsymbol{w^d} = 1 \\
& \mathbf{0} \leq \boldsymbol{y} \leq \boldsymbol{w^d}, \; \mathbf{0} \leq \boldsymbol{z} \leq \boldsymbol{w^d}, \\
& \boldsymbol{s} \geq \mathbf{0}, \; \boldsymbol{w^d} \geq \mathbf{0}, \; \boldsymbol{l} \leq \boldsymbol{w} \leq \boldsymbol{u}.
\end{array}
$$

Proof We already know how to reexpress (3.37b) in terms of finitely many conic constraints. Therefore, we now focus on the reformulation of (3.37c).

As usual, we first reformulate (3.37c) in terms of a subordinate minimization problem,

$$
\min_{\substack{\boldsymbol{r} \geq \mathbf{0} \\ \boldsymbol{r^d} = f(\boldsymbol{r})}} \quad \boldsymbol{w}^\top \boldsymbol{r} + (\boldsymbol{w^d})^\top \boldsymbol{r^d} \geq \theta \phi. \tag{3.38}
$$

By using the definition of the function f and the fact that $\boldsymbol{w^d} \geq \mathbf{0}$, the left-hand side of (3.38) can be reexpressed as the linear program

$$
\begin{array}{ll}
\text{minimize} & \boldsymbol{w}^\top \boldsymbol{r} + (\boldsymbol{w^d})^\top \boldsymbol{r^d} \\
\text{subject to} & \boldsymbol{r} \in \mathbb{R}^n, \quad \boldsymbol{r^d} \in \mathbb{R}^m \\
& \boldsymbol{r} \;\; \geq \mathbf{0} \\
& \boldsymbol{r^d} \geq \mathbf{0} \\
& \boldsymbol{r^d} \geq \boldsymbol{a} + \mathbf{B} \boldsymbol{r}.
\end{array} \tag{3.39}
$$

The dual of problem (3.39) reads

$$\begin{aligned} \text{maximize} \quad & \boldsymbol{a}^\top \boldsymbol{z} \\ \text{subject to} \quad & \boldsymbol{z} \in \mathbb{R}^m \\ & \boldsymbol{w} + \mathbf{B}^\top \boldsymbol{z} \geq \mathbf{0} \\ & \mathbf{0} \leq \boldsymbol{z} \leq \boldsymbol{w}^{\boldsymbol{d}}. \end{aligned} \tag{3.40}$$

Strong linear duality holds because the primal problem (3.39) is manifestly feasible. Therefore, the optimal objective value of problem (3.40) coincides with that of problem (3.39), and we can substitute (3.40) into the constraint (3.38). This leads to the postulated reformulation in (3.38). ∎

Note that problem (3.38) implicitly minimizes a coherent risk measure determined through the uncertainty set

$$\{(\boldsymbol{r}, \boldsymbol{r}^{\boldsymbol{d}}) : \boldsymbol{r} \in \mathcal{U}^{+}_{\boldsymbol{r},\boldsymbol{\mu}},\ \boldsymbol{r}^{\boldsymbol{d}} = f(\boldsymbol{r})\}. \tag{3.41}$$

Coherency holds since this uncertainty set is a subset of the support of the random vector $(\tilde{\boldsymbol{r}}, \tilde{\boldsymbol{r}}^{\boldsymbol{d}})$, see Remark 3.2.1. A risk-tolerant investor may want to move away from the minimum risk portfolio. This is achieved by appending an expected return constraint to the problem:

$$\mathbb{E}[\tilde{r}_p] = \boldsymbol{w}^\top \boldsymbol{\mu} + (\boldsymbol{w}^{\boldsymbol{d}})^\top \mathbb{E}[\max\{\mathbf{0}, \boldsymbol{a} + \mathbf{B}\tilde{\boldsymbol{r}}\}] \geq \mu_{\text{target}}. \tag{3.42}$$

For any distribution of $\tilde{\boldsymbol{r}}$, we can evaluate the expected return of the options via sampling. Since sampling is impractical when the expected returns are ambiguous, one may alternatively use a conservative approximation of the return target constraint (3.42),

$$\boldsymbol{w}^\top \boldsymbol{\mu} + (\boldsymbol{w}^{\boldsymbol{d}})^\top (\max\{\mathbf{0}, \boldsymbol{a} + \mathbf{B}\boldsymbol{\mu}\}) \geq \mu_{\text{target}}. \tag{3.43}$$

Indeed, (3.42) is less restrictive than (3.43) by Jensen's inequality. To account for the uncertainty in the estimated means, we can further robustify (3.43) as follows,

$$\left.\begin{array}{ll}\underset{\boldsymbol{q}\in\mathbb{R}^m}{\text{maximize}} & \boldsymbol{\mu}^\mathsf{T}(\boldsymbol{w}+\mathbf{B}^\mathsf{T}\boldsymbol{q})+\boldsymbol{a}^\mathsf{T}\boldsymbol{q}\\ \text{subject to} & \mathbf{0}\leq\boldsymbol{q}\leq\boldsymbol{w}^{\boldsymbol{d}}\end{array}\right\}\geq\mu_{\text{target}}\quad\forall\boldsymbol{\mu}\in\mathcal{U}_{\boldsymbol{\mu}},$$

which is equivalent to

$$\left.\begin{array}{ll}\underset{\boldsymbol{q}\in\mathbb{R}^m}{\text{maximize}} & \hat{\boldsymbol{\mu}}^\mathsf{T}(\boldsymbol{w}+\mathbf{B}^\mathsf{T}\boldsymbol{q})-\kappa\left\|\boldsymbol{\Omega}^{1/2}(\boldsymbol{w}+\mathbf{B}^\mathsf{T}\boldsymbol{q})\right\|_2+\boldsymbol{a}^\mathsf{T}\boldsymbol{q}\\ \text{subject to} & \mathbf{0}\leq\boldsymbol{q}\leq\boldsymbol{w}^{\boldsymbol{d}}\end{array}\right\}\geq\mu_{\text{target}}.$$

As a third alternative, the investor may wish to disregard the expected returns of the options altogether in the return target constraint. Taking into account the uncertainty in the estimated means, we thus obtain the second-order cone constraint

$$\boldsymbol{w}^\mathsf{T}\hat{\boldsymbol{\mu}}-\kappa\left\|\boldsymbol{\Omega}^{1/2}\boldsymbol{w}\right\|_2\geq\mu_{\text{target}}, \tag{3.44}$$

which is identical to (3.24). The advantages of this third approach are twofold.

Firstly, by omitting the options in the expected return constraint, we force the model to use the options for risk reduction and insurance only, but not for speculative reasons. Only the stocks are used to attain the prescribed expected return target. In light of the substantial risks involved in speculation with options, this might be attractive for risk-averse investors.

Secondly, the inclusion of an expected return constraint converts (3.38) to a mean-risk model [Har91], which minimizes a coherent downside risk measure, see Remark 3.2.1. However, Dert and Oldenkamp [DO00] and Lucas and Siegmann [LS08] have identified several pitfalls that may arise when using mean-downside risk models in the presence of highly asymmetric asset classes such as options and hedge funds. The particular problems that occur in the presence of options have been characterized as the *Casino Effect*: Mean-downside risk models typically choose portfolios which use the least amount of money that is necessary to satisfy the insurance constraint, whilst allocating the remaining money in the assets with the highest expected return. In our context, a combination of inexpensive stocks and put options will be used to satisfy the insurance constraint. Since call options are leveraged assets and have expected returns that increase with the strike price [CS02], the remaining wealth will therefore generally be invested in the call options with the highest strike prices available. The resulting portfolios have a high probability of small losses and a very low probability of high returns. Since the robust framework is typically used by risk-averse investors, the resulting portfolios are most likely in conflict with their risk preferences. It should be emphasized that the Casino Effect is characteristic for mean-downside risk models and not a side-effect of the robust portfolio optimization methodology. In order to alleviate its impact, Dert and Oldenkamp propose the use of several Value-at-Risk constraints to shape the distribution of terminal wealth. Lucas and Siegmann propose a modified risk measure that incorporates a quadratic penalty function to the expected losses. In all our numerical tests, we choose to exclude the expected option returns from the return target constraint. This will avoid betting on the options and thus mitigate the Casino Effect. As we will show in the next section, our numerical results indicate that the suggested portfolio model successfully reduces the downside risk and sustains high out-of-sample expected returns.

3.4 Computational Results

In Section 3.4.1 we investigate the optimal portfolio composition for different levels of risk-aversion and illustrate the tradeoff between the weak non-inferiority guarantee and the strong insurance guarantee. In Section 3.4.2 we conduct several tests based on simulated data, while the tests in Section 3.4.3 are performed on the basis of real market data. In both sections, we compare the out-of-sample performance of the insured robust portfolios with that of the non-insured robust and classical mean-variance portfolios. The comparisons are based on the following performance measures: average yearly return, worst-case and best-case monthly returns, yearly variance, skewness, and Sharpe ratio [Sha66]. All computations are performed using the C++ interface of the MOSEK 5.0.0.105 conic optimization toolkit on a 2.0 GHz Core 2 Duo machine running Linux Ubuntu 8.04. The details of the experiments are described in the next sections.

3.4.1 Portfolio Composition and Tradeoff of Guarantees

All experiments in this section are based on the $n = 30$ stocks in the Dow 30 index. We assume that for each stock there are 40 put and 40 call options that mature in one year. The 40 strike prices of the put and call options for one particular stock are located at equidistant points between 70% and 130% of the stock's current price. In total, the market thus comprises 2400 options in addition to the 30 stocks.

In our first simulated backtests, we assume that the stock prices are governed by a multivariate geometric Brownian motion,

$$\begin{aligned} \frac{\mathrm{d}\tilde{S}_t^i}{\tilde{S}_t^i} &= \mu_i^c \,\mathrm{d}t + \sigma_i^c \,\mathrm{d}\widetilde{W}_t^i, \quad i = 1 \dots n, \\ \mathbb{E}\left[\mathrm{d}\widetilde{W}_t^i \,\mathrm{d}\widetilde{W}_t^j\right] &= \rho_{ij}^c \,\mathrm{d}t, \quad i, j = 1 \dots n, \end{aligned} \tag{3.45}$$

where $\tilde{S}^i$ denotes the price process of stock i and $\widetilde{W}^i$ denotes a standard Wiener process. The continuous-time parameters μ_i^c, σ_i^c, and ρ_{ij}^c represent the drift rates, volatilities and correlation rates of the instantaneous stock returns, respectively. We calibrate this stochastic model to match the annualized means and covariances of the total returns of the Dow 30 stocks reported in Idzorek [Idz02]. The transformation which maps the annualized parameters to the continuous-time parameters in (3.45) is described in [Meu05, p. 345]. Furthermore, we assume that the risk-free rate amounts to $r_f = 5\%$ per annum and that the options are priced according to the Black-Scholes formula [BS73].

In the experiments of this section we do not allow short-selling of stocks. Furthermore, we assume that there is no parameter uncertainty. Therefore, we set $q = 0$. In the first set of tests we solve problem (3.38) without an expected return constraint and without a portfolio insurance constraint. We determine the optimal portfolio allocations for increasing sizes of uncertainty sets parameterized by $p \in [0, 1)$. The optimal portfolio weights are visualized in the top left panel of Figure 3.1, and the optimal conditional worst-case returns are displayed in the bottom left panel. For simplicity, we only report the total percentage of wealth allocated in stocks, calls, and put options, and provide no information about the individual asset allocations. All instances of problem (3.38) considered in this test were solved within less than 2 seconds, which manifests the tractability of the proposed model.

Figure 3.1 exhibits three different allocation regimes. For small values of p, the optimal portfolios are entirely invested in call options or a mixture of calls and stocks. This is a natural consequence of the leverage effect of the call options, which have a much higher return potential than the stocks when they mature in-the-money. As a result, the optimal conditional worst-case return is very high. Large investments in call options tend to be highly risky; this is reflected by a sudden decrease in call option allocation at threshold value $p \approx 7\%$.

We also observe a regime which is entirely invested in stocks. Here, the risk is minimized through variance reduction by diversification, and no option hedging is involved.

At higher uncertainty levels, there is a sudden shift to portfolios composed of stocks and put options. This transition takes place when the uncertainty set is large enough such that stock-only portfolios necessarily incur a loss in the worst case. The effect of the put options can be observed in the bottom left panel of Figure 3.1, which shows a constant worst-case return $\phi > 1$ for higher uncertainty levels. Here, risk is not reduced through diversification. Instead, an aggressive portfolio insurance strategy is adopted using deep in-the-money put options. The put options are used to cut away the losses, and thus $\phi > 1$. For high uncertainty levels, maximizing the conditional worst-case return amounts to maximizing the absolute insurance guarantee because the uncertainty set converges to the support of the returns, see Proposition 3.3.1.

The Black-Scholes market under consideration is arbitrage-free. An elementary arbitrage argument implies that the maximum guaranteed lower bound on the return of any portfolio is not larger than the risk-free return $\exp(r_f T)$. The conditional worst-case return in problem (3.38) is therefore bounded above by $\exp(r_f T)$ already for moderately sized uncertainty sets. This risk-free return can indeed be attained, at least approximately, by combining a stock and a put option on that stock with a very large strike price. Note that the put option matures in-the-money with high probability. Thus, the resulting portfolio pays off the strike price in most cases and is almost risk-free. Its conditional worst-case return is only slightly smaller than $\exp(r_f T)$ (for large uncertainty sets with $p \lesssim 1$). However, investing in an almost risk-free portfolio keeps the expected portfolio return fairly low, that is, close to the risk-free return.

In order to bypass this shortcoming, we impose an expected return constraint on the stock part of the portfolio with a target return of 8% per annum, see (3.44). The results of model (3.38) *with* an expected return constraint and *without* a portfolio insurance constraint are visualized on the right hand side of Figure 3.1. Most of the earlier conclusions remain valid, but there are a few differences. Because the stocks are needed to satisfy the return target, we now observe that all portfolios put a minimum weight of nearly 90% in stocks. For higher levels of uncertainty, the allocation in put options increases gradually when higher uncertainty is assigned to the returns.

The optimal conditional worst-case return smoothly degrades for increasing uncertainty levels and now drops below 1. Here, we anticipate a loss in the worst-case. Recall that the (negative) conditional worst-case return can be interpreted as a risk measure, see Remark 3.2.1. In order to satisfy the expected return constraint, the optimal portfolios have to take higher risks than in the absence of an expected return constraint. As a result, the optimal conditional worst-case return is now lower (due to the higher risk) than before. This is a natural consequence of the risk-return tradeoff. For $p \gtrsim 90\%$, the conditional worst-case return saturates at the worst-case return that can be guaranteed with certainty.

Next, we analyze the effects of the insurance constraint on the conditional worst-case return. To this end, we solve problem (3.38) for various insurance levels $\theta \in [0, 1]$ and uncertainty levels $p \in [0, 1)$, whilst still requiring the expected return to exceed 8%. Figure 3.2 shows the conditional worst-case return as a function of p and θ.

For any fixed p, the conditional worst-case return monotonically decreases with θ. Observe that this decrease is steeper for lower values of p. When the uncertainty set is small, the conditional worst-case return is relatively high. Therefore, the inclusion of the insurance guarantee has a significant impact due to the high insurance costs that are introduced. When the uncertainty set size is increased, the conditional worst-case return drops, and portfolio insurance needs to be provided for a lower worst-case portfolio return at an associated lower portfolio insurance cost.

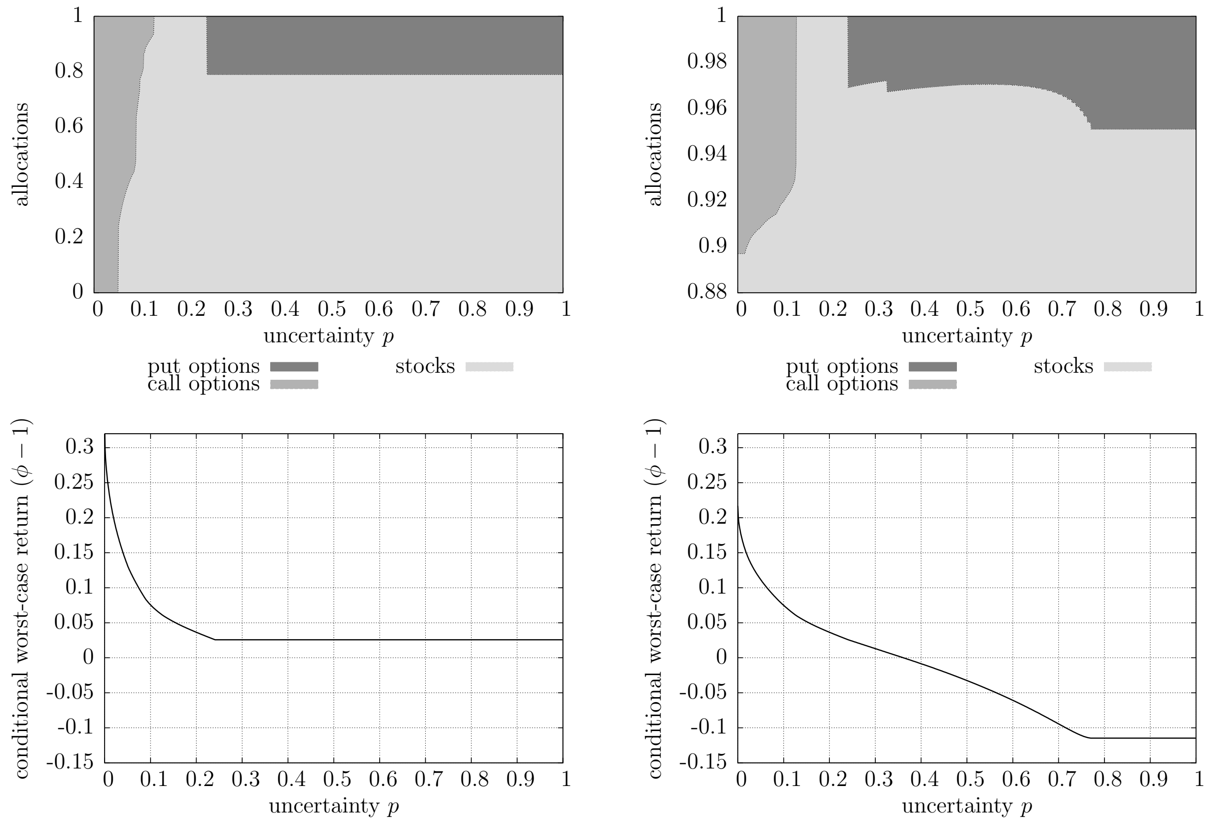

Figure 3.1: Visualization of the optimal portfolio allocations (top) and corresponding conditional worst-case returns (bottom), with (right) and without (left) an expected return constraint.

When $\theta = 1$, the portfolio is insured against dropping below the conditional worst-case return. That is, the optimal portfolio provides the highest possible insurance guarantee that is still compatible with the expected return target. This optimal portfolio is independent of the size of the uncertainty set, and therefore the worst-case return is constant in p. For $p \gtrsim 80\%$, the uncertainty set converges to the support of the returns, and the resulting optimal portfolio is independent of θ, see Proposition 3.3.1. Note that if the expected return target is increased, then the guaranteed worst-case return for $\theta = 1$ decreases. In fact, in order to satisfy the higher expected return constraint the cost of insurance has to be decreased. The cost of insurance can only be lowered by decreasing the allocation in put options, which implies a lower guaranteed worst-case return.

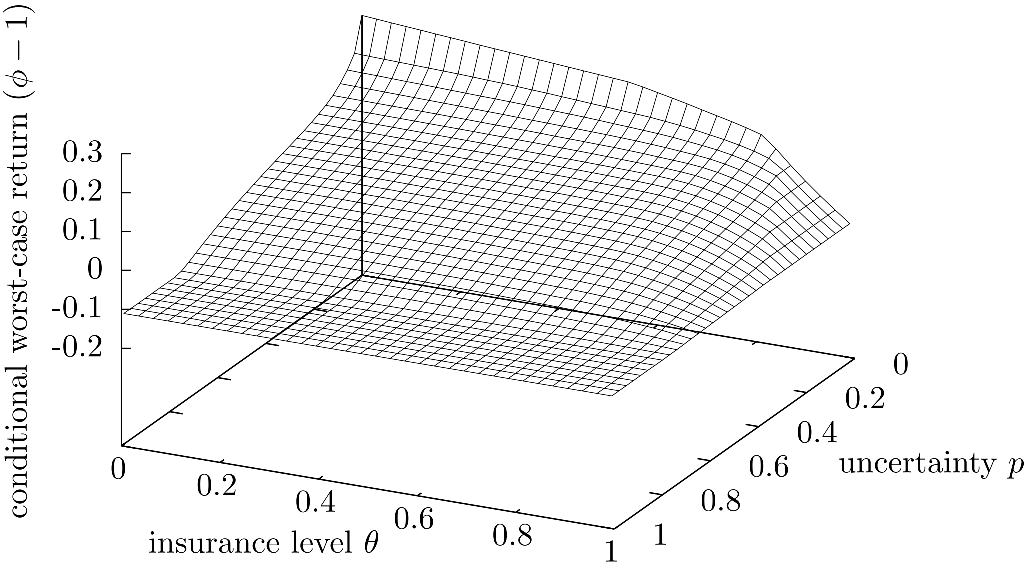

Figure 3.2: Tradeoff of weak and strong guarantees.

3.4.2 Out-of-Sample Evaluation Using Simulated Prices

A series of controlled experiments with simulated data help us to assess the performance of the proposed Insured Robust Portfolio Optimization (IRPO) model under different market conditions. We first generate price paths under a multivariate geometric Brownian motion model to reflect "normal" market conditions. Next, we use a multivariate jump-diffusion process to simulate a volatile environment in which market crashes can occur. In both settings, we compare the performance of the IRPO model to that of the Robust Portfolio Optimization (RPO) model (3.22), and the classical Mean-Variance Optimization (MVO) model. The optimal MVO portfolio is found by minimizing the variance of the portfolio return subject to an expected portfolio return constraint. In this case the estimated means and covariance matrix of the asset returns are used without taking parameter uncertainty into account.

Backtest Procedure and Evaluation

The following experiments are again based on the stocks in the Dow 30 index. The first test series is aimed at assessing the performance of the models under "normal" market conditions. To this end, we assume that the stock prices are governed by the multivariate geometric Brownian motion described in (3.45).

We denote by $\tilde{\boldsymbol{r}}_l$ the vector of the asset returns over the interval $[(l-1)\Delta t, l\Delta t]$, where Δt is set to one month (i.e., $\Delta t = 1/12$) and $l \in \mathbb{N}$. By solving the stochastic differential equations (3.45), we find

$$\tilde{r}_l^i = \exp\left[\left(\mu_i^c - \frac{(\sigma_i^c)^2}{2}\right)\Delta t + \tilde{\epsilon}_l^i\sqrt{\Delta t}\right], \quad i = 1 \dots n, \tag{3.46}$$

where $\{\tilde{\boldsymbol{\epsilon}}_l\}_{l\in\mathbb{N}}$ are independent and identically normally distributed with zero mean and covariance matrix $\boldsymbol{\Sigma^c} \in \mathbb{R}^{n\times n}$ with entries $\Sigma_{ij}^c = \rho_{ij}^c \sigma_i^c \sigma_j^c$ for $i, j = 1 \dots n$.

To evaluate the performance of the different portfolio models, we use the following rolling-horizon procedure:

1. Generate a time-series of L monthly stock returns $\{\boldsymbol{r}_l\}_{l=1}^L$ using (3.46) and initialize the iteration counter at $l = E$. The number $E < L$ determines the size of a moving estimation window.

2. Calculate the sample mean $\hat{\boldsymbol{\mu}}_l$ and sample covariance matrix $\hat{\boldsymbol{\Sigma}}_l$ of the stock returns $\{\boldsymbol{r}_l\}_{l'=l-E+1}^l$ in the current estimation window. We assume that there are 20 put and 20 call options available for each stock that expire after one month. The 20 strike prices of the options are assumed to scale with the underlying stock price: the proportionality factor ranges from 80% to 120% in steps of 2%.[4]

   Next, convert the estimated monthly volatilities to continuous-time volatilities via the transformation in [Meu05, p. 345] and calculate the option prices via the Black-Scholes formula.[5] For the IRPO model we then calculate the necessary option related data $\boldsymbol{a}_l$ and $\mathbf{B}_l$ defined in (3.28).

[4]This set of options is a reasonable proxy for the set available in reality. Depending on liquidity, there might be more or fewer options available, but the use of 20 strike prices oriented around the spot prices seems a good compromise.

[5]In reality, one would use option prices observed in the market instead of calculated ones. An empirical backtest based on real option price data is provided in Section 3.4.3.

Model k	Type	p	q	θ	Model k	Type	p	q	θ
1	MVO	–	–	–					
2	RPO	0.50	0.80	–	17	IRPO	0.70	0.80	0.00
3	RPO	0.60	0.80	–	18	IRPO	0.70	0.80	0.70
4	RPO	0.70	0.80	–	19	IRPO	0.70	0.80	0.80
5	RPO	0.80	0.80	–	20	IRPO	0.70	0.80	0.90
6	RPO	0.90	0.80	–	21	IRPO	0.70	0.80	0.99
7	IRPO	0.50	0.80	0.00	22	IRPO	0.80	0.80	0.00
8	IRPO	0.50	0.80	0.70	23	IRPO	0.80	0.80	0.70
9	IRPO	0.50	0.80	0.80	24	IRPO	0.80	0.80	0.80
10	IRPO	0.50	0.80	0.90	25	IRPO	0.80	0.80	0.90
11	IRPO	0.50	0.80	0.99	26	IRPO	0.80	0.80	0.99
12	IRPO	0.60	0.80	0.00	27	IRPO	0.90	0.80	0.00
13	IRPO	0.60	0.80	0.70	28	IRPO	0.90	0.80	0.70
14	IRPO	0.60	0.80	0.80	29	IRPO	0.90	0.80	0.80
15	IRPO	0.60	0.80	0.90	30	IRPO	0.90	0.80	0.90
16	IRPO	0.60	0.80	0.99	31	IRPO	0.90	0.80	0.99

Table 3.1: Parameter settings of the portfolio models used in the backtests.

3. Determine the optimal portfolios $(\boldsymbol{w}_l^k, \boldsymbol{w}_l^{\boldsymbol{d},k})$ corresponding to the models $k = 1, \ldots, 31$ specified in Table 3.1.

4. For strategy k, the portfolio return r_{l+1}^k over the interval $[l\Delta t, (l+1)\Delta t]$ is given by:

$$r_{l+1}^k = (\boldsymbol{w}_l^k)^\mathsf{T} \boldsymbol{r}_{l+1} + (\max\{\mathbf{0},\ \boldsymbol{a}_l + \mathbf{B}_l\ \boldsymbol{r}_{l+1}\})^\mathsf{T}\ \boldsymbol{w}_l^{\boldsymbol{d},k}.$$

   Since $\boldsymbol{r}_{l+1}$ is outside of the estimation window, this constitutes an out-of-sample evaluation.

5. If $l < L - 1$, then increment l and go to step 2. Otherwise, terminate.

In all backtests we set $L = 240$ and use an estimation window of size $E = 120$. We set the risk-free rate to $r_f = 5\%$ per annum and the expected return target to 8% per annum. We allow short-selling of individual stocks up to -20% and do not impose upper bounds on the portfolio weights.

The rolling-horizon procedure generates $L - E$ returns $\{r_l^k\}_{l=E+1}^L$ for our 31 portfolio strategies indexed by k. For each of these strategies we calculate the following performance measures: the out-of-sample mean, variance, skewness, Sharpe ratio, worst-case and best-case monthly return.

$$
\begin{aligned}
\hat{\mu}^k &= \frac{1}{L-E}\sum_{l=E+1}^{L} r_l^k, && \text{(mean)}\\
(\hat{\sigma}^2)^k &= \frac{1}{L-E-1}\sum_{l=E+1}^{L} (r_l^k - \hat{\mu}^k)^2, && \text{(variance)}\\
\hat{\gamma}^k &= \frac{1}{L-E}\sum_{l=E+1}^{L} ((r_l^k - \hat{\mu}^k)/\hat{\sigma}^k)^3, && \text{(skewness)}\\
\widehat{SR}^k &= \frac{\hat{\mu}^k - r_f}{\hat{\sigma}^k}, && \text{(Sharpe ratio)}\\
\underline{\hat{r}}^k &= \min\ \{r_l^k : E+1 \le l \le L\}, && \text{(worst-case return)}\\
\bar{\hat{r}}^k &= \max\ \{r_l^k : E+1 \le l \le L\}. && \text{(best-case return)}
\end{aligned}
$$

By assuming an initial wealth of 1, we also calculate the final wealth $\hat{\omega}^k$ of strategy k as follows

$$\hat{\omega}^k = \prod_{t=E+1}^{L} r_l^k.$$

We repeat the rolling-horizon procedure described above $R = 300$ times with different random generator seeds and calculate averages of the performance measures. We also estimate the probability of the different portfolio strategies (with $k > 1$) yielding a higher final wealth than the Markowitz strategy (with $k = 1$) by counting the simulation runs in which this outperformance is observed. Finally, we compute the excess return of any strategy k relative to the Markowitz strategy, $\hat{\omega}^k/\hat{\omega}^1 - 1$, averaged over all simulation runs.

A property of the geometric Brownian motion price process is that there are almost surely no discontinuities in the price paths. In reality, rare events such as market crashes can occur, and therefore the Jump-Diffusion model introduced by Merton [Mer76] may be more suitable to describe real price movements. Under Merton's Jump-Diffusion model, the stock prices are governed by the stochastic differential equations

$$\begin{aligned} \frac{\mathrm{d}\tilde{S}_t^i}{\tilde{S}_t^i} &= (\mu_i^c - \lambda^c \eta)\,\mathrm{d}t + \sigma_i^c\,\mathrm{d}\widetilde{W}_t^i + \mathrm{d}\sum_{j=1}^{\tilde{N}_t} (\tilde{Y}_j - 1), \quad && i = 1 \dots n, \\ \mathbb{E}\left[\mathrm{d}\widetilde{W}_t^i\,\mathrm{d}\widetilde{W}_t^j\right] &= \rho_{ij}^c\,\mathrm{d}t, && i, j = 1 \dots n, \end{aligned} \tag{3.47}$$

where $\tilde{N}$ is a Poisson process with arrival intensity λ^c, and $\{\tilde{Y}_j\}_{j\in\mathbb{N}}$ is a sequence of independent identically distributed nonnegative random variables. $\tilde{N}_t$ denotes the number of jumps, or market crashes, between 0 and time t, while the $\tilde{Y}_j$ represent the relative price changes when such crashes occur. $\widetilde{W}_t$ and $\tilde{N}_t$ are assumed to be independent.

For simplicity, we assume that all stock prices jump at the same time. Moreover, instead of making the jump sizes stochastic, as in the general formulation above, we assume that all prices experience a deterministic relative change of $\eta = -15\%$ when a crash occurs. We set $\lambda^c = 2$, indicating that on average there are two crashes per year. Solving the stochastic differential equations (3.47) we obtain the following expression for the stock returns

$$\tilde{r}_l^i = \exp\left[\left(\mu_i^c - \frac{(\sigma_i^c)^2}{2} - \lambda^c \eta\right)\Delta t + \tilde{\epsilon}_l^i \sqrt{\Delta t}\right] \prod_{j=\tilde{N}_{(l-1)\Delta t}+1}^{\tilde{N}_{l\Delta t}} Y_j, \quad i = 1 \dots n, \tag{3.48}$$

where $\tilde{N}_t$ follows a Poisson distribution with parameter $\lambda^c \Delta t$ and $Y_j = e^\eta$ for all j. We now repeat the previously described rolling-horizon backtest by using (3.48) instead of (3.46).

Discussion of Results

The results of our simulated backtests based on the geometric Brownian motion model are summarized in Table 3.3.

In comparison with the nominal MVO portfolio, we observe that the RPO portfolios exhibit a significantly higher average return at the cost of a relatively small increase in variance. This is also reflected by the Sharpe ratio values, which are higher than that of the MVO portfolio for all levels of p. When p increases, we notice a slight decrease in variance and expected return because the portfolios become more conservative. We see that the non-insured RPO portfolios outperform the MVO portfolio with probability 75%. This indicates that taking the uncertainty of the mean estimates into account results in a considerable improvement of out-of-sample performance.

Next, we assess the performance of the IRPO portfolios. For a fixed insurance level θ, we observe that the worst-case monthly return (Min) increases with p. In most cases, it also increases with θ for fixed p. However, this is not always the case. At $p = 80\%$, for instance, the worst-case return for $\theta = 90\%$ is higher than for $\theta = 99\%$. The reason for this is that a large portion of wealth is allocated to the options in order to satisfy the high insurance demands. Because there are no price jumps, these options have a low probability to mature in-the-money. The options have a noticeable effect on the skewness of the portfolio returns, which increases with p and θ. This is because the put options are effectively cutting away the losses and therefore cause the portfolio return distribution to be positively skewed.

Finally, for all tested values of p and θ, the IRPO portfolios accumulate a higher final wealth than the nominal MVO portfolio in about 65% of the cases. In terms of Sharpe ratio, the IRPO portfolios perform comparably to the RPO portfolios. However, the non-insured RPO portfolios have an increased expected return and a higher probability of outperforming the nominal MVO portfolio in terms of realized wealth. Note that, although the IRPO portfolios have a lower probability of outperforming the MVO portfolio, they achieve higher excess returns than the RPO portfolios because the options help preserve wealth over time. We conclude that under normal market conditions the non-insured RPO model seems to generate the most attractive out-of-sample results.

The results of our simulated backtests based on Merton's jump diffusion model are summarized in Table 3.4. The following discussion highlights the differences to the results obtained using the geometric Brownian motion model.

The RPO portfolios still have a significant probability of outperforming the MVO portfolio in terms of realized wealth. Due to the crashes, however, this probability now decreases to 65% (as opposed to 75% in the absence of crashes). Notice that the worst-case monthly returns of the RPO portfolios are of the same order of magnitude as those of the MVO portfolio. We also observe that the realized returns for the RPO and MVO portfolios are highly negatively skewed because of the downward jumps of the prices.

The IRPO portfolios have an increased expected return and lower variance with respect to the MVO portfolio for all tested values of p and θ. This is also reflected by an improvement in Sharpe ratio, which for $p = 60\%$ and $\theta = 99\%$ is 60% higher than that of the MVO portfolio. The IRPO portfolios exhibit increased skewness relative to the MVO and RPO portfolios. The skewness of the IRPO portfolios becomes positive for values of $p \geq 80\%$ and $\theta = 99\%$. The worst-case return gradually improves with increasing values of p and θ, and for $p = 90\%$ the worst-case is 50% higher than that of the nominal MVO portfolio. Finally, the IRPO portfolios achieve a higher realized wealth than the MVO portfolio in about 77% of the simulation runs. Notice also that the excess returns monotonically increase with θ. The increase in realized wealth is due to the option insurance which helps preserve wealth during market crashes. In contrast, the crashes cause large losses of wealth to the MVO and RPO portfolios.

In conclusion, the simulated tests indicate that the IRPO model has advantages over the MVO and RPO models when the market exhibits jumps. It typically results in a higher realized wealth and Sharpe ratio.

3.4.3 Out-of-Sample Evaluation Using Real Market Prices

Simulated stock and option prices may give an unrealistic view of how our portfolio strategies perform in reality due to the following reasons. Firstly, it is known that real stock returns are not serially independent and identically distributed. Secondly, real option prices deviate from those obtained via the Black-Scholes formula by using historical volatilities. Finally, we are restricted to invest in the options traded in the market, and our assumption about the range of available strike prices may not hold.

Therefore, we now evaluate the portfolio strategies under the same rolling-horizon procedure described in the previous section but with real stock and option prices. Historical stock and

Ticker	Name
XMI	AMEX Major Market Index
SPX	S&P 500 Index
MID	S&P Midcap 400 Index
SML	S&P Smallcap 600 Index
RUT	Russell 2000 Index
NDX	NASDAQ 100 Index

Table 3.2: Equity indices used in the historical backtest.

option prices are obtained from the OptionMetrics IvyDB database, which is one of the most complete sources of historical option data available. We limit ourselves to the equity indices shown in Table 3.2. These indices were chosen because they have the most complete time-series in the database. As before, we rebalance on a monthly basis, and at every rebalancing date we consider all available European put and call options that expire in one month.[6] Because the IRPO strategy is long in options, we use the highest option ask prices to make sure that we could have acquired the options at the specified prices.

The time-series covers the period from 18/01/1996 until 18/09/2008. We use an estimation window of 15 months.[7] Moreover, we allow short-selling in every equity index up to -20% of total portfolio value but impose no upper bounds on the weights. The target expected return is set to 8% per annum. The range of tested p and θ values is the same as in the previous section, see Table 3.1.

Discussion of Results

The results of the backtests based on real market prices are given in Table 3.5. Similar to the out-of-sample results based on simulated prices, the RPO portfolios produce higher expected returns than the nominal MVO model, while their Sharpe ratios are more than twice as large as that of the MVO portfolio for all tested values of p.

The IRPO portfolios also outperform the MVO portfolio in terms of expected return and Sharpe

[6] In order to avoid the use of erroneous option data, we only selected those options for which the implied volatility was supplied and which had a bid and ask price greater than 0. We found that this procedure allowed us to filter out incorrect entries.

[7] Different estimation windows yielded slightly different out-of-sample results. However, the general conclusions are independent of the choice of the estimation window.

ratio for all values of p and θ. However, compared to the RPO portfolios, they have a slightly lower expected return on average. This decrease in expected return is due to the cost of insurance. We also observe that the IRPO portfolios have smaller variance than the RPO portfolios for all tested parameter settings. On average the IRPO portfolios also produce slightly higher Sharpe ratios than the RPO portfolios.

In Figure 3.3 we plot the cumulative wealth over time of the MVO portfolio, an RPO portfolio with $p = 50\%$, an IRPO portfolio with $p = 50\%$ and $\theta = 70\%$, and an IRPO portfolio with $p = 50\%$ and $\theta = 99\%$. The IRPO portfolio with $\theta = 70\%$ performed better than the MVO and RPO portfolios. However, we emphasize that the performance of the IRPO model is highly dependent on the values chosen for p and θ. For example, it can be observed that with $p = 50\%$ and $\theta = 99\%$ the IRPO portfolio is outperformed by the RPO portfolio due to the high cost of insurance.

For all tested parameters values, the IRPO model yields a higher worst-case monthly return than the RPO model and a significant increase in skewness for levels of $p \geq 60\%$. The worst-case return monotonically increases with p. However, it is not always increasing in θ. High insurance levels of $\theta \gtrsim 90\%$ lead to large investments in put options which expire worthless with high probability. This is also reflected by a significant drop in expected return and an associated decrease in Sharpe ratio.

The reasons for this are twofold. Firstly, the strong insurance guarantees are more expensive in reality than in the simulations. This is because the Black-Scholes formula underestimates the prices of far out-of-the-money put options when historical volatilities are used. Secondly, we are limited to invest in the options that are traded in the market and are therefore unable to invest in options with strike prices that would have resulted in better portfolios.

To conclude, we note that for this particular data set the RPO and IRPO portfolios systematically outperform the nominal MVO portfolio in terms of expected return and Sharpe ratio. On average the RPO portfolios achieve higher expected returns than the IRPO portfolios, whereas the IRPO portfolios obtain slightly higher average Sharpe ratios. We also conclude that the performance of the IRPO model is highly dependent on the chosen values of p and θ. The

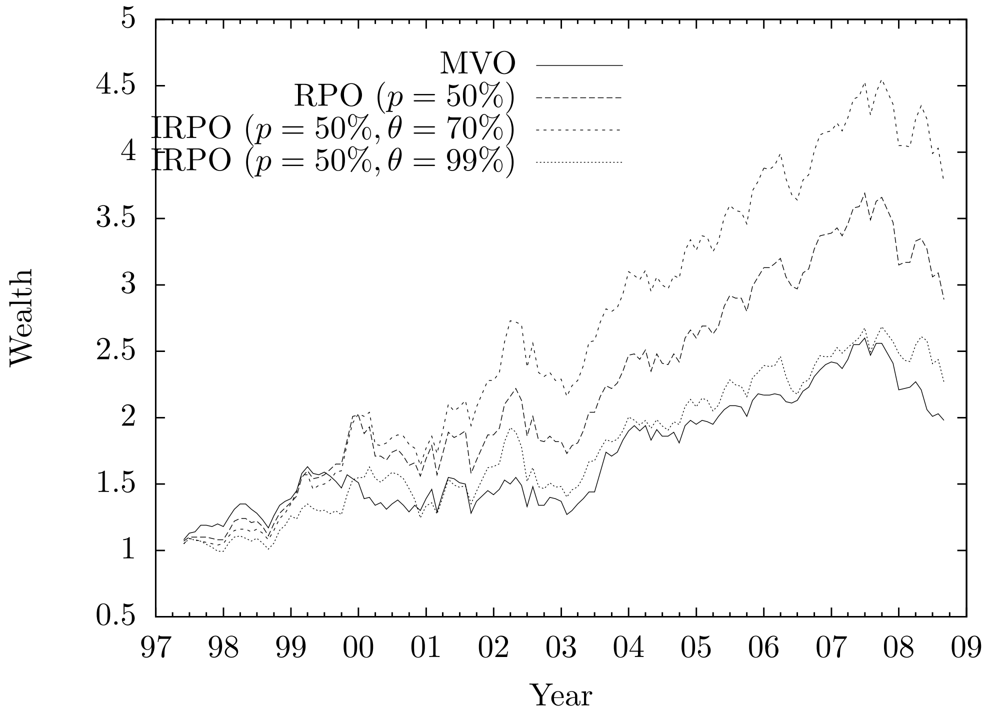

Figure 3.3: Cumulated return of the MVO, RPO, and IRPO portfolios using monthly rebalancing between 19/06/1997 and 18/09/2008.

insurance levels should therefore be tuned to market behavior. Higher insurance levels can help preserve the accumulated portfolio wealth when the market is volatile and experiences jumps. Lower insurance levels are preferable in less volatile periods since unnecessary insurance costs are avoided.

3.5 Conclusions

In this chapter, we extended robust portfolio optimization to accommodate options. Moreover, we showed how the options can be used to provide strong insurance guarantees, which also hold when the stock returns are realized outside of the prescribed uncertainty set. Using conic and linear duality, we reformulated the problem as a convex second-order cone program, which is scalable in the amount of stocks and options and can be solved efficiently with standard optimization packages. The proposed methodology can be applied to a wide range of uncertainty sets and can therefore be seen as a generic extension to the robust portfolio optimization framework.

We first performed backtests on simulated data, in which the asset prices reflect normal market conditions as well as market crashes. In both cases the option premia are calculated using the standard Black-Scholes model. The simulated results indicate that the insured robust portfolios have lower expected returns than the non-insured robust portfolios under normal market conditions but have clear advantages with respect to Sharpe ratio, expected return, as well as cumulative wealth, when the prices experience jumps.

Since the Black-Scholes prices might not reflect realistic option premia, we also performed backtests on historical data. We observed that on average the RPO portfolios achieve higher expected returns than the IRPO portfolios, whereas the IRPO portfolios obtain higher Sharpe ratios. The results also indicate that the performance of the IRPO model is highly dependent on the values chosen for p and θ. When the insurance level is set too high, the cost of insurance causes the performance to deteriorate. Therefore, the level of insurance should be tuned to the market; to preserve wealth, higher insurance levels can benefit the portfolio when the market is volatile and experiences jumps. Lower insurance levels are preferable in less volatile periods since unnecessary insurance costs are avoided.

3.6 Appendix

3.6.1 Notational Reference Table

n	Number of stocks
m	Number of options
$\tilde{r}_p$	Total portfolio return
r_f	Risk-free rate
$\boldsymbol{w}$, $\boldsymbol{w^d}$	Weights of the stocks and options, respectively
$\boldsymbol{e}$	Vector of ones
$\boldsymbol{l}$, $\boldsymbol{u}$	Lower and upper bounds on the weights of the stocks
$\tilde{\boldsymbol{r}}$, $\tilde{\boldsymbol{r}}^{\boldsymbol{d}}$	Total stock and option returns, respectively
$\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$	Mean vector and covariance matrix of $\tilde{\boldsymbol{r}}$, respectively
$\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\Sigma}}$	Sample mean and sample covariance matrix of $\tilde{\boldsymbol{r}}$, respectively
$\boldsymbol{\Lambda}$	Covariance matrix of $\hat{\boldsymbol{\mu}}$
$\boldsymbol{\Omega}$	Modified covariance matrix of $\hat{\boldsymbol{\mu}}$
λ	Risk-aversion parameter
μ_{target}	Portfolio return target
p, q	Probabilities of $\tilde{\boldsymbol{r}}$ and $\hat{\boldsymbol{\mu}}$ to be realized within their respective uncertainty sets, respectively
$\mathcal{U}_{\boldsymbol{r}}$	Uncertainty set for $\tilde{\boldsymbol{r}}$
$\mathcal{U}_{\boldsymbol{r}}^{+}$	Uncertainty set for $\tilde{\boldsymbol{r}}$ including support information
$\mathcal{U}_{\boldsymbol{\mu}}$	Uncertainty set for $\boldsymbol{\mu}$
$\mathcal{U}_{\boldsymbol{r},\boldsymbol{\mu}}^{+}$	Uncertainty set for $\tilde{\boldsymbol{r}}$ and $\boldsymbol{\mu}$ including support information
δ, κ	Size parameters for the uncertainty sets $\mathcal{U}_{\boldsymbol{r},\boldsymbol{\mu}}^{+}$ and $\mathcal{U}_{\boldsymbol{\mu}}$, respectively
ϕ	Conditional worst-case portfolio return
θ	Insurance level
T	End of investment horizon
$\tilde{S}_t^i,\ i = 1, \ldots, n$	Price of stock i at time t
$\widetilde{W}^i,\ i = 1, \ldots, n$	Standard Wiener processes
$\tilde{N}$	Poisson process
λ^c	Arrival intensity
η	Relative price change during crash
μ^c, σ^c, ρ^c	Instantaneous drifts, volatilities and correlation rates, respectively
L	Size of the time-series
E	Size of the estimation window
$K_i,\ i = 1, \ldots, m$	Strike price of option i
$P_i, C_i,\ i = 1, \ldots, m$	Price of option i if it is a call/put option
$\boldsymbol{a}, \mathbf{B}$	Parameters of function f
f	Function relating $\tilde{\boldsymbol{r}}$ and $\tilde{\boldsymbol{r}}^{\boldsymbol{d}}$

3.6.2 Proof of Theorem 3.6.1

Theorem 3.6.1 *For $\mathcal{U}_{\boldsymbol{\mu}}$ defined as in* (3.12)*, and $\mathcal{U}_{\boldsymbol{\Sigma}} = \{\hat{\boldsymbol{\Sigma}}\}$, problem* (3.8) *is equivalent to the following second-order cone program,*

$$\max_{\boldsymbol{w}\in\mathbb{R}^n} \left\{ \boldsymbol{w}^\top \hat{\boldsymbol{\mu}} - \kappa \left\|\boldsymbol{\Omega}^{1/2}\boldsymbol{w}\right\|_2 - \delta \left\|\hat{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{w}\right\|_2 \;\middle|\; \boldsymbol{w}^\top \boldsymbol{e} = 1,\ \boldsymbol{l} \le \boldsymbol{w} \le \boldsymbol{u} \right\},$$

where

$$\boldsymbol{\Omega} = \boldsymbol{\Lambda} - \frac{1}{\boldsymbol{e}^\top \boldsymbol{\Lambda} \boldsymbol{e}} \boldsymbol{\Lambda} \boldsymbol{e} \boldsymbol{e}^\top \boldsymbol{\Lambda}.$$

Proof Because $\mathcal{U}_{\boldsymbol{\Sigma}}$ is a singleton, it is clear that problem (3.8) is equivalent to

$$\max_{\boldsymbol{w}\in\mathbb{R}^n} \left\{ \min_{\boldsymbol{\mu}\in\mathcal{U}_{\boldsymbol{\mu}}} \boldsymbol{w}^\top \boldsymbol{\mu} - \delta \left\|\hat{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{w}\right\|_2 \;\middle|\; \boldsymbol{w}^\top \boldsymbol{e} = 1,\ \boldsymbol{l} \le \boldsymbol{w} \le \boldsymbol{u} \right\}. \tag{3.49}$$

When $\kappa = 0$, the claim is obviously true. In the rest of the proof we thus assume that $\kappa > 0$. Using the definition of the uncertainty set $\mathcal{U}_{\boldsymbol{\mu}}$, the inner minimization problem in (3.49) can be rewritten as

$$\begin{aligned} \min_{\boldsymbol{\mu}\in\mathbb{R}^n} \quad & \boldsymbol{w}^\top \boldsymbol{\mu} \\ \text{s. t.} \quad & \left\|\boldsymbol{\Lambda}^{-1/2}(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})\right\|_2 \le \kappa \\ & \boldsymbol{e}^\top (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) = 0. \end{aligned} \tag{3.50}$$

For any fixed portfolio $\boldsymbol{w}$, problem (3.50) represents a second-order cone program. We proceed by dualizing (3.50). After a few minor simplification steps, we obtain the dual problem

$$\max_{q\in\mathbb{R}} \boldsymbol{w}^\top \hat{\boldsymbol{\mu}} - \kappa \left\|\boldsymbol{\Lambda}^{1/2}(\boldsymbol{w} - q\boldsymbol{e})\right\|. \tag{3.51}$$

Strong conic duality holds since the primal problem (3.49) is strictly feasible for $\kappa > 0$. Thus, both the primal and dual problems (3.49) and (3.50) are feasible and share the same objective values at optimality. Since $\kappa > 0$, the optimal dual solution is given by

$$q^* = \operatorname*{argmin}_{q\in\mathbb{R}} \left\|\boldsymbol{\Lambda}^{1/2}(\boldsymbol{w} - q\boldsymbol{e})\right\| = \frac{\boldsymbol{w}^\top \boldsymbol{\Lambda} \boldsymbol{e}}{\boldsymbol{e}^\top \boldsymbol{\Lambda} \boldsymbol{e}}.$$

By substituting q^* into (3.51) we obtain the optimal value of (3.50), which amounts to

$$\boldsymbol{w}^\mathsf{T}\hat{\boldsymbol{\mu}} - \kappa \left\|\boldsymbol{\Omega}^{1/2}\boldsymbol{w}\right\|_2 . \tag{3.52}$$

We can now substitute (3.52) into (3.49) to obtain the postulated result. ∎

k	Av. Return (%)	Variance (%)	Skewness (%)	Min (%)	Max (%)	Sharpe Ratio	Win (%)	Excess Ret. (%)
1	8.71594	2.00607	2.68932	-10.29675	11.7450	0.264210	0.00	0.00
2	10.3755	2.55062	3.43910	-10.5330	13.5163	0.341027	74.6	10.5
3	10.3113	2.51639	3.39247	-10.4717	13.4097	0.340005	75.3	9.29
4	10.2613	2.49331	3.35748	-10.4334	13.3355	0.338927	74.6	8.29
5	10.2244	2.47738	3.33970	-10.4134	13.2856	0.338070	74.6	7.47
6	10.1915	2.46703	3.32159	-10.4009	13.2520	0.337030	74.6	6.85
7	9.75114	2.37732	6.37693	-9.82254	13.5763	0.312815	65.0	22.7
8	9.76275	2.35325	6.86229	-9.63240	13.5286	0.315774	66.5	21.7
9	9.95955	2.35109	7.14115	-9.51791	13.5508	0.330269	67.9	21.6
10	10.4537	2.58300	9.11501	-9.48632	14.4127	0.350974	70.6	24.7
11	11.1375	3.69973	15.8147	-9.43376	17.9969	0.318184	62.9	34.0
12	9.70266	2.33910	6.74911	-9.70263	13.5173	0.312973	64.9	22.7
13	9.71800	2.32261	7.20889	-9.54075	13.4913	0.315523	65.6	20.9
14	9.89839	2.32169	7.50994	-9.42003	13.5127	0.328530	67.2	21.6
15	10.3990	2.52040	9.03305	-9.44218	14.2191	0.351144	70.9	24.1
16	11.1357	3.64071	15.6969	-9.41962	17.7885	0.321900	63.5	33.5
17	9.70377	2.32215	7.50623	-9.56143	13.5511	0.314441	64.5	22.8
18	9.71463	2.31140	7.94322	-9.41865	13.5345	0.316258	64.9	21.4
19	9.85969	2.31282	8.27716	-9.30624	13.5606	0.326427	65.6	22.0
20	10.3427	2.47753	9.56243	-9.31501	14.1313	0.348963	70.2	23.3
21	11.0978	3.54072	15.6044	-9.36810	17.5359	0.325626	63.2	32.4
22	9.83033	2.37172	9.79252	-9.25168	13.8913	0.317983	63.5	23.3
23	9.83828	2.37367	10.2294	-9.12943	13.9154	0.318421	64.2	24.0
24	9.91845	2.38374	10.5327	-9.04900	13.9499	0.323221	65.9	22.6
25	10.3096	2.53120	11.6889	-8.96272	14.4386	0.339619	69.9	24.3
26	11.0432	3.45595	15.5965	-9.23531	17.3043	0.325910	65.6	30.4
27	10.5568	3.07095	15.6408	-8.67166	16.4439	0.311806	61.9	32.0
28	10.5757	3.09742	15.7754	-8.64120	16.5136	0.311775	62.2	31.2
29	10.5787	3.11860	15.8493	-8.63573	16.5691	0.311006	61.5	30.5
30	10.6744	3.20165	15.9592	-8.67983	16.7601	0.312786	62.2	31.2
31	11.0842	3.61410	15.9569	-9.22338	17.7917	0.317473	63.5	33.0

Table 3.3: Out-of-sample statistics obtained for the various portfolio policies when the asset prices follow a geometric Brownian motion model. All the portfolios were constrained to have an expected return of 8% per annum. We report the out-of-sample yearly average return, variance, skewness, and Sharpe ratio. We also give the worst (Min) and best (Max) monthly return, as well as the probability of the robust policies generating a final wealth that outperforms the standard mean-variance policy (Win). Finally, we report the excess return in final wealth of the robust policies relative to the mean-variance policy. All values represent averages over 300 simulations.

k	Av. Return (%)	Variance (%)	Skewness (%)	Min (%)	Max (%)	Sharpe Ratio	Win (%)	Excess Ret. (%)
1	11.2147	6.34540	-27.5025	-24.3438	15.3838	0.267697	0.00	0.00
2	13.0907	6.83008	-24.2199	-24.5152	16.7323	0.330526	63.9	14.7
3	12.9937	6.76026	-24.7205	-24.4847	16.5833	0.329218	64.2	12.5
4	12.9106	6.70894	-25.0985	-24.4621	16.4574	0.327860	66.5	11.0
5	12.8408	6.67203	-25.3888	-24.4497	16.3637	0.326491	67.2	9.60
6	12.7765	6.64734	-25.5948	-24.4539	16.2904	0.324888	67.2	9.74
7	12.3068	5.82155	-14.2809	-20.2584	17.1322	0.329408	65.9	38.9
8	13.1004	5.45702	-9.56472	-16.5229	17.0942	0.372863	76.3	45.6
9	13.2278	5.42011	-9.05962	-16.0991	17.1440	0.380884	77.6	46.2
10	13.8212	5.34474	-6.03003	-15.0283	17.5517	0.411923	80.3	57.0
11	14.6230	5.71965	6.16843	-13.0636	19.8228	0.427392	78.9	77.0
12	12.3521	5.66294	-11.8597	-19.5035	17.2461	0.336597	66.2	39.5
13	13.1229	5.34435	-7.38688	-16.0240	17.2225	0.378164	76.9	45.7
14	13.2368	5.31367	-6.86320	-15.6296	17.2694	0.385195	77.3	47.0
15	13.6546	5.30885	-5.71704	-15.2192	17.4656	0.406045	80.0	54.1
16	14.6057	5.65646	5.86874	-13.0687	19.6627	0.429978	78.6	76.3
17	12.6586	5.51344	-6.70853	-17.9349	17.6830	0.353456	69.6	43.1
18	13.3544	5.28022	-2.94795	-15.0876	17.6845	0.389341	76.3	49.1
19	13.4568	5.25690	-2.47033	-14.7576	17.6995	0.395346	76.6	50.0
20	13.6629	5.27957	-2.04267	-14.6158	17.8369	0.405240	77.9	53.3
21	14.5529	5.63034	5.57181	-13.1036	19.5813	0.427982	78.9	75.3
22	13.6427	5.56548	2.70522	-14.7963	19.0505	0.389035	74.2	56.6
23	14.0189	5.51099	4.40116	-13.4448	19.1661	0.405404	76.3	60.8
24	14.1036	5.50960	4.69928	-13.2577	19.1834	0.409234	77.6	61.7
25	14.2020	5.53412	4.96087	-13.1608	19.2660	0.413189	77.9	62.8
26	14.5430	5.71816	5.99048	-13.0967	19.7601	0.421727	79.6	74.6
27	14.5493	5.75625	6.90160	-12.8156	20.0099	0.419263	78.6	76.3
28	14.5823	5.75682	6.97966	-12.7264	20.0293	0.420468	79.3	77.1
29	14.5928	5.75793	6.97477	-12.7223	20.0304	0.420899	79.3	77.2
30	14.6103	5.76479	6.93533	-12.7460	20.0186	0.421606	79.3	77.3
31	14.6319	5.79007	6.57951	-12.9931	19.9917	0.422424	79.6	77.4

Table 3.4: Out-of-sample statistics obtained for the various portfolio policies when the asset prices follow Merton's Jump-Diffusion model. The values represent averages over 300 simulations.

k	Av. Return (%)	Variance (%)	Skewness (%)	Min (%)	Max (%)	Sharpe Ratio	Excess Ret. (%)
1	7.31825	2.34178	-9.12004	-16.4139	11.7043	0.151491	0.00
2	11.2927	3.01830	-19.5287	-17.1069	12.3246	0.362207	57.0
3	11.0634	2.87484	-18.3199	-16.9154	11.6224	0.357613	54.4
4	10.9261	2.76526	-17.5850	-16.6864	11.4770	0.356368	53.0
5	10.6595	2.67737	-16.8366	-16.4205	11.4771	0.345881	49.3
6	10.2065	2.59044	-15.8553	-15.9605	11.4771	0.323491	42.7
7	13.1180	2.35028	-6.21748	-11.8770	12.2885	0.529530	80.8
8	12.4548	2.26840	-6.36185	-13.1038	11.6413	0.494968	74.1
9	11.4658	2.29011	-8.42871	-14.8555	11.0644	0.427262	67.6
10	9.69145	2.35543	-9.06444	-14.8548	10.7600	0.305684	41.7
11	8.42547	2.45024	-4.29419	-14.8553	11.9108	0.218834	23.4
12	12.7677	2.25056	-3.73207	-11.4450	11.4077	0.517785	81.4
13	12.1563	2.22157	-4.06279	-12.6428	11.3968	0.480132	68.9
14	11.2310	2.24928	-6.79232	-14.8559	10.9404	0.415466	64.0
15	9.56499	2.30204	-7.52749	-14.8556	10.7594	0.300873	40.0
16	8.44237	2.42517	-4.49501	-14.8549	11.9361	0.221048	23.7
17	12.5775	2.17887	-0.34246	-10.6547	11.4211	0.513348	80.1
18	11.9782	2.16979	-1.84851	-12.3076	11.2077	0.473736	67.4
19	11.0419	2.19461	-5.87528	-14.8556	10.7596	0.407848	62.6
20	9.62988	2.24781	-5.11817	-14.8558	10.9213	0.308809	41.9
21	8.31310	2.43287	-4.38022	-14.8557	11.9830	0.212410	21.9
22	11.5649	2.14677	2.53860	-10.3774	11.6451	0.448060	62.3
23	10.8342	2.16369	0.81466	-12.0324	11.2003	0.396631	53.0
24	10.2051	2.24547	-2.69276	-14.8321	11.6277	0.347358	47.1
25	9.13540	2.25700	-2.25949	-14.8560	12.1547	0.275265	34.0
26	8.19263	2.48381	-2.95234	-14.8556	13.1482	0.202576	20.0
27	10.4035	2.22619	4.54875	-10.2823	13.3358	0.362153	43.9
28	9.87056	2.27988	1.62919	-11.9177	12.6118	0.322569	38.1
29	9.57241	2.32928	-0.41801	-13.9725	12.6016	0.299595	36.8
30	9.07195	2.37072	-1.62395	-14.8547	12.8105	0.264462	33.3
31	8.42897	2.49602	-2.53740	-14.8553	13.3825	0.217040	23.1

Table 3.5: Out-of-sample statistics obtained for the various portfolio policies using the real stock and option prices evaluated using monthly rebalancing between 19/06/1997 and 18/09/2008.

Chapter 4

Worst-Case Value-at-Risk of Non-Linear Portfolios

In Chapter 3, we investigated how to incorporate options within the robust portfolio optimization framework. In this chapter our aim will be to apply *distributionally* robust optimization techniques to minimize the Value-at-Risk (VaR) of derivative portfolios. Portfolio optimization problems involving VaR are often computationally intractable and require complete information about the return distribution of the portfolio constituents, which is rarely available in practice. These difficulties are further compounded when the portfolio contains derivatives. Nevertheless, we will show that by employing duality theory and by solving moment problems, the Worst-Case VaR, which is a distributionally robust version of VaR, can be optimized efficiently even when the portfolio contains derivatives. Interestingly, we will also show that there exists an equivalence between Worst-Case VaR optimization and robust portfolio optimization, which we elaborated in Chapter 3.

4.1 Introduction

Although mean-variance optimization is appropriate when the asset returns are symmetrically distributed, it is known to result in counter intuitive asset allocations when the portfolio return is skewed. This shortcoming triggered extensive research on downside risk measures. Due to its intuitive appeal and since its use is enforced by financial regulators, Value-at-Risk (VaR) remains the most popular downside risk measure [Jor01]. The VaR at level ϵ is defined as the $(1-\epsilon)$-quantile of the portfolio loss distribution.

Despite its popularity, VaR lacks some desirable theoretical properties. Firstly, VaR is known to be a non-convex risk measure. As a result, VaR optimization problems usually are computationally intractable. In fact, they belong to the class of chance-constrained stochastic programs, which are notoriously difficult to solve. Secondly, VaR fails to satisfy the subadditivity property of coherent risk measures [ADEH99], see also Section 2.4.3. Thus, the VaR of a portfolio can exceed the weighted sum of the VaRs of its constituents. In other words, VaR may penalize diversification. Thirdly, the computation of VaR requires precise knowledge of the joint probability distribution of the asset returns, which is rarely available in practice.

A typical investor may know the first- and second-order moments of the asset returns but is unlikely to have complete information about their distribution. Therefore, El Ghaoui *et al.* [EGOO03] propose to maximize the VaR of a given portfolio over all asset return distributions consistent with the known moments. The resulting Worst-Case VaR (WCVaR) represents a conservative (that is, pessimistic) approximation for the true (unknown) portfolio VaR. In contrast to VaR, WCVaR represents a convex function of the portfolio weights and can be optimized efficiently by solving a tractable second-order cone program. El Ghaoui *et al.* [EGOO03] also disclose an interesting connection to robust optimization [BTN98, BTN99, RH02]: WCVaR coincides with the worst-case portfolio loss when the asset returns are confined to an *ellipsoidal uncertainty set* determined through the known means and covariances.

In this chapter we study portfolios containing derivatives, the most prominent examples of which are European call and put options. Sophisticated investors frequently enrich their portfolios with derivative products, be it for hedging and risk management or speculative purposes. In the presence of derivatives, WCVaR still constitutes a tractable conservative approximation for the true portfolio VaR. However, it tends to be over-pessimistic and thus may result in undesirable portfolio allocations. The main reasons for the inadequacy of WCVaR are the following.

- The calculation of WCVaR requires the first- and second-order moments of the derivative returns as an input. These moments are difficult or (in the case of exotic options) almost impossible to estimate due to scarcity of time series data.

- WCVaR disregards perfect dependencies between the derivative returns and the underlying asset returns. These (typically non-linear) dependencies are known in practice as they can be inferred from contractual specifications (payoff functions) or option pricing models. Note that the covariance matrix of the asset returns, which is supplied to the WCVaR model, fails to capture non-linear dependencies among the asset returns, and therefore WCVaR tends to severely *overestimate* the true VaR of a portfolio containing derivatives.

Recall that WCVaR can be calculated as the optimal value of a robust optimization problem with an ellipsoidal uncertainty set, which is highly symmetric. This symmetry hints at the inadequacy of WCVaR from a geometrical viewpoint. An intuitively appealing uncertainty set should be asymmetric to reflect the skewness of the derivative returns. Recently, Natarajan *et al.* [NPS08] included asymmetric distributional information into the WCVaR optimization in order to obtain a tighter approximation of VaR. However, their model requires forward- and backward-deviation measures as an input, which are difficult to estimate for derivatives. In contrast, reliable information about the functional relationships between the returns of the derivatives and their underlying assets is readily available.

In this chapter we develop novel Worst-Case VaR models which explicitly account for perfect non-linear dependencies between the asset returns. We first introduce the *Worst-Case Polyhedral VaR* (WCPVaR), which provides a tight conservative approximation for the VaR of a portfolio containing European-style options expiring at the end of the investment horizon. In this situation, the option returns constitute convex piecewise-linear functions of the underlying asset returns. WCPVaR evaluates the worst-case VaR over all asset return distributions consistent with the given first- and second-order moments of the option underliers and the piecewise linear relation between the asset returns. Under a no short-sales restriction on the options, we are able to formulate WCPVaR optimization as a convex second-order cone program, which can be solved efficiently [AG03]. We also establish the equivalence of the WCPVaR model to a robust optimization model described in Chapter 3.

Next, we introduce the *Worst-Case Quadratic VaR* (WCQVaR) which approximates the VaR of a portfolio containing long and/or short positions in plain vanilla and/or exotic options with arbitrary maturity dates. In contrast to WCPVaR, WCQVaR assumes that the derivative returns are representable as (possibly non-convex) quadratic functions of the underlying asset returns. This can always be enforced by invoking a *delta-gamma approximation*, that is, a second-order Taylor approximation of the portfolio return. The delta-gamma approximation is popular in many branches of finance and is accurate for short investment periods. Moreover, it has been used extensively for VaR estimation, see, e.g., the surveys by Jaschke [Jas02] and Mina and Ulmer [MU99]. However, to the best of our knowledge, the delta-gamma approximation has never been used in a VaR optimization model. We define WCQVaR as the worst-case VaR over all asset return distributions consistent with the known first- and second-order moments of the option underliers and the given quadratic relation between the asset returns. WCQVaR provides a tight conservative approximation for the true portfolio VaR if the delta-gamma approximation is accurate. We show that WCQVaR optimization can be formulated as a convex semidefinite program, which can be solved efficiently [VB96], and we establish a connection to a novel robust optimization problem. The main contributions in this chapter can be summarized as follows:

(1) We generalize the WCVaR model [EGOO03] to explicitly account for the non-linear relationships between the derivative returns and the underlying asset returns. To this end, we develop the WCPVaR and WCQVaR models as described above. We show that in the absence of derivatives both models reduce to the WCVaR model. Moreover, we formulate WCPVaR optimization as a second-order cone program and WCQVaR optimization as a semidefinite program. Both models are polynomial time solvable.

(2) We show that both the WCPVaR and the WCQVaR models have equivalent reformulations as robust optimization problems. We explicitly construct the associated uncertainty sets which are, unlike conventional ellipsoidal uncertainty sets, asymmetrically oriented around the mean values of the asset returns. This asymmetry is caused by the non-linear dependence of the derivative returns on their underlying asset returns. Simple examples

illustrate that the new models may approximate the true portfolio VaR significantly better than WCVaR in the presence of derivatives.

(3) The robust WCQVaR model is of relevance beyond the financial domain because it constitutes a tractable approximation of a chance-constrained stochastic program, see Section 2.3.1, that is affine in the decision variables but (possibly non-convex) *quadratic* in the uncertainties. Although tractable approximations for chance constrained programs with affine perturbations have been researched extensively (see, e.g., [NS06]), the case of quadratic data dependence has remained largely unexplored (with the exception of [BTEGN09, §1.4]).

(4) We evaluate the WCQVaR model in the context of an index tracking application. We show that when investment in options is allowed, the optimal portfolios exhibit vastly improved out-of-sample performance compared to the optimal portfolios based on stocks only.

The remainder of the chapter is organized as follows. In Section 4.2 we review the mathematical definitions of VaR and WCVaR. Moreover, we recall the relationship between WCVaR optimization and robust optimization. In Section 4.3 we highlight the shortcomings of WCVaR in the presence of derivatives. In Section 4.4 we develop the WCPVaR model in which the option returns are modelled as convex piecewise-linear functions of the underlying asset returns. We prove that it can be reformulated as a second-order cone program and construct the uncertainty set which generates the equivalent robust portfolio optimization model. In Section 4.5 we describe the WCQVaR model, which approximates the portfolio return by a quadratic function of the underlying asset returns. We show that it can be reformulated as a semidefinite program and prove its equivalence to an augmented robust optimization problem whose uncertainty set is embedded into the space of positive semidefinite matrices. Section 4.6 evaluates the out-of-sample performance of the WCQVaR model in the context of an index tracking application. Conclusions are drawn in Section 4.7.

4.2 Worst-Case Value-at-Risk Optimization

Consider a market consisting of m assets such as equities, bonds, and currencies. We denote the present as time $t = 0$ and the end of the investment horizon as $t = T$. A portfolio is characterized by a vector of asset weights $\boldsymbol{w} \in \mathbb{R}^m$, whose elements add up to 1. The component w_i denotes the percentage of total wealth which is invested in the ith asset at time $t = 0$. Furthermore, $\tilde{\boldsymbol{r}}$ denotes the $\mathbb{R}^m$-valued random vector of relative assets returns over the investment horizon. By definition, an investor will receive $1 + \tilde{r}_i$ dollars at time T for every dollar invested in asset i at time 0. The return of a given portfolio $\boldsymbol{w}$ over the investment period is thus given by the random variable

$$\tilde{r}_p = \boldsymbol{w}^\mathsf{T} \tilde{\boldsymbol{r}}. \tag{4.1}$$

Loosely speaking, we aim at finding an allocation vector $\boldsymbol{w}$ which entails a high portfolio return, whilst keeping the associated risk at an acceptable level. Depending on how risk is defined, we end up with different portfolio optimization models.

Arguably one of the most popular measures of risk is the *Value-at-Risk* (VaR). The VaR at level ϵ is defined as the $(1-\epsilon)$-percentile of the portfolio loss distribution, where ϵ is typically chosen as 1% or 5%. Put differently, $\mathrm{VaR}_\epsilon(\boldsymbol{w})$ is defined as the smallest real number γ with the property that $-\boldsymbol{w}^\mathsf{T}\tilde{\boldsymbol{r}}$ exceeds γ with a probability not larger than ϵ, that is,

$$\mathrm{VaR}_\epsilon(\boldsymbol{w}) = \min \left\{ \gamma \ : \ \mathbb{P}\{\gamma \le -\boldsymbol{w}^\mathsf{T}\tilde{\boldsymbol{r}}\} \le \epsilon \right\}, \tag{4.2}$$

where $\mathbb{P}$ denotes the distribution of the asset returns $\tilde{\boldsymbol{r}}$.

In this chapter we investigate portfolio optimization problems of the type

$$\begin{aligned} \underset{\boldsymbol{w} \in \mathbb{R}^m}{\text{minimize}} \quad & \mathrm{VaR}_\epsilon(\boldsymbol{w}) \\ \text{subject to} \quad & \boldsymbol{w} \in \mathcal{W}, \end{aligned} \tag{4.3}$$

where $\mathcal{W} \subseteq \mathbb{R}^m$ denotes the set of admissible portfolios. The inclusion $\boldsymbol{w} \in \mathcal{W}$ usually implies the budget constraint $\boldsymbol{w}^\mathsf{T}\boldsymbol{e} = 1$ (where $\boldsymbol{e}$ denotes the vector of 1s). Optionally, the set $\mathcal{W}$

may account for bounds on the allocation vector $\boldsymbol{w}$ and/or a constraint enforcing a minimum expected portfolio return. In this chapter we only require that $\mathcal{W}$ must be a convex polyhedron.

By using (4.2), the VaR optimization model (4.3) can be reformulated as

$$\begin{aligned} \underset{\boldsymbol{w}\in\mathbb{R}^m,\gamma\in\mathbb{R}}{\text{minimize}} \quad & \gamma \\ \text{subject to} \quad & \mathbb{P}\{\gamma + \boldsymbol{w}^\mathsf{T}\tilde{\boldsymbol{r}} \geq 0\} \geq 1-\epsilon \\ & \boldsymbol{w} \in \mathcal{W}, \end{aligned} \tag{4.4}$$

which constitutes a chance-constrained stochastic program, see Section 2.3.1. Optimization problems of this kind are usually difficult to solve since they tend to have non-convex or even disconnected feasible sets. Furthermore, the evaluation of the chance constraint requires precise knowledge of the probability distribution of the asset returns, which is rarely available in practice.

4.2.1 Two Analytical Approximations of Value-at-Risk

In order to overcome the computational difficulties and to account for the lack of knowledge about the distribution of the asset returns, the objective function in (4.3) must usually be approximated. Most existing approximation techniques fall into one of two main categories: *non-parametric approaches* which approximate the asset return distribution by a discrete (sampled or empirical) distribution and *parametric approaches* which approximate the asset return distribution by the best fitting member of a parametric family of continuous distributions. We now give a brief overview of two analytical VaR approximation schemes that are of particular relevance for our purposes.

Both in the financial industry as well as in the academic literature, it is frequently assumed that the asset returns $\tilde{\boldsymbol{r}}$ are governed by a Gaussian distribution with given mean vector $\boldsymbol{\mu}_{\boldsymbol{r}} \in \mathbb{R}^m$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{r}} \in \mathbb{S}^m$. This assumption has the advantage that the VaR can be calculated analytically as

$$\mathrm{VaR}_\epsilon(\boldsymbol{w}) = -\boldsymbol{\mu}_{\boldsymbol{r}}^\mathsf{T}\boldsymbol{w} - \Phi^{-1}(\epsilon)\sqrt{\boldsymbol{w}^\mathsf{T}\boldsymbol{\Sigma}_{\boldsymbol{r}}\boldsymbol{w}}, \tag{4.5}$$

where Φ is the standard normal distribution function. This model is sometimes referred to as *Normal VaR* (see, e.g., [NPS08]). In practice, the distribution of the asset returns often fails to be Gaussian. In these cases, (4.5) can still be used as an approximation. However, it may lead to gross *underestimation* of the actual portfolio VaR when the true portfolio return distribution is leptokurtic or heavily skewed, as is the case for portfolios containing options.

To avoid unduly optimistic risk assessments, El Ghaoui *et al.* [EGOO03] suggest a conservative (that is, pessimistic) approximation for VaR under the assumption that only the mean values and covariance matrix of the asset returns are known. Let $\mathcal{P}_{\boldsymbol{r}}$ be the set of all probability distributions on $\mathbb{R}^m$ with mean value $\boldsymbol{\mu}_{\boldsymbol{r}}$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{r}}$. We emphasize that $\mathcal{P}_{\boldsymbol{r}}$ contains also distributions which exhibit considerable skewness, so long as they match the given mean vector and covariance matrix. The *Worst-Case Value-at-Risk* for portfolio $\boldsymbol{w}$ is now defined as

$$\mathrm{WCVaR}_\epsilon(\boldsymbol{w}) = \min\left\{\gamma \ : \ \sup_{\mathbb{P}\in\mathcal{P}_{\boldsymbol{r}}} \mathbb{P}\{\gamma \leq -\boldsymbol{w}^\mathsf{T}\tilde{\boldsymbol{r}}\} \leq \epsilon\right\}. \tag{4.6}$$

Note that the above problem constitutes a *distributionally robust* chance-constrained program, see Section 2.3.3. El Ghaoui *et al.* demonstrate that WCVaR has the closed form expression

$$\mathrm{WCVaR}_\epsilon(\boldsymbol{w}) = -\boldsymbol{\mu}_{\boldsymbol{r}}^\mathsf{T}\boldsymbol{w} + \kappa(\epsilon)\sqrt{\boldsymbol{w}^\mathsf{T}\boldsymbol{\Sigma}_{\boldsymbol{r}}\boldsymbol{w}}, \tag{4.7}$$

where $\kappa(\epsilon) = \sqrt{(1-\epsilon)/\epsilon}$. WCVaR represents a tight approximation for VaR in the sense that there exists a worst-case distribution $\mathbb{P}^* \in \mathcal{P}_{\boldsymbol{r}}$ such that VaR with respect to $\mathbb{P}^*$ is equal to WCVaR.

When using WCVaR instead of VaR as a risk measure, we end up with the portfolio optimization problem

$$\begin{aligned} \underset{\boldsymbol{w}\in\mathbb{R}^m}{\text{minimize}} \quad & -\boldsymbol{\mu}_{\boldsymbol{r}}^\mathsf{T}\boldsymbol{w} + \kappa(\epsilon)\left\|\boldsymbol{\Sigma}_{\boldsymbol{r}}^{1/2}\boldsymbol{w}\right\|_2 \\ \text{subject to} \quad & \boldsymbol{w} \in \mathcal{W}, \end{aligned} \tag{4.8}$$

which represents a second-order cone program that is amenable to efficient numerical solution procedures.

4.2.2 Robust Optimization Perspective on Worst-Case VaR

Consider the following robust optimization problem (see Section 2.3.2 for an intoduction to robust optimization).

$$\begin{aligned} \underset{\boldsymbol{w}\in\mathbb{R}^m,\gamma\in\mathbb{R}}{\text{minimize}} \quad & \gamma \\ \text{subject to} \quad & \gamma + \boldsymbol{w}^\mathsf{T}\boldsymbol{r} \geq 0 \quad \forall \boldsymbol{r} \in \mathcal{U} \\ & \boldsymbol{w} \in \mathcal{W}. \end{aligned} \tag{4.9}$$

An uncertainty set that enjoys wide popularity in the robust optimization literature is the *ellipsoidal set*,

$$\mathcal{U} = \{\boldsymbol{r} \in \mathbb{R}^m \ : \ (\boldsymbol{r} - \boldsymbol{\mu_r})^\mathsf{T}\boldsymbol{\Sigma}_{\boldsymbol{r}}^{-1}(\boldsymbol{r} - \boldsymbol{\mu_r}) \leq \delta^2\},$$

which is defined in terms of the mean vector $\boldsymbol{\mu_r}$ and covariance matrix $\boldsymbol{\Sigma_r}$ of the asset returns as well as a size parameter δ. By conic duality it can be shown that the following equivalence holds for any fixed $(\boldsymbol{w}, \gamma) \in \mathcal{W} \times \mathbb{R}$.

$$\gamma + \boldsymbol{w}^\mathsf{T}\boldsymbol{r} \geq 0 \quad \forall \boldsymbol{r} \in \mathcal{U} \iff -\boldsymbol{\mu}_{\boldsymbol{r}}^\mathsf{T}\boldsymbol{w} + \delta \left\|\boldsymbol{\Sigma}_{\boldsymbol{r}}^{1/2}\boldsymbol{w}\right\|_2 \leq \gamma \tag{4.10}$$

Problem (4.9) can therefore be reformulated as the following second-order cone program.

$$\begin{aligned} \underset{\boldsymbol{w}\in\mathbb{R}^m}{\text{minimize}} \quad & -\boldsymbol{\mu}_{\boldsymbol{r}}^\mathsf{T}\boldsymbol{w} + \delta \left\|\boldsymbol{\Sigma}_{\boldsymbol{r}}^{1/2}\boldsymbol{w}\right\|_2 \\ \text{subject to} \quad & \boldsymbol{w} \in \mathcal{W} \end{aligned} \tag{4.11}$$

By comparing (4.8) and (4.11), El Ghaoui *et al.* [EGOO03] noticed that optimizing WCVaR at level ϵ is equivalent to solving the robust optimization problem (4.9) under an ellipsoidal uncertainty set with size parameter $\delta = \kappa(\epsilon)$, see also Natarajan *et al.* [NPS08]. This uncertainty set will henceforth be denoted by $\mathcal{U}_\epsilon$.

In this chapter we extend the WCVaR model (4.7) and the equivalent robust optimization model (4.9) to situations in which there are non-linear relationships between the asset returns, as is the case in the presence of derivatives.

4.3 Worst-Case VaR for Derivative Portfolios

From now on assume that our market consists of $n \leq m$ *basic assets* and $m - n$ *derivatives*. We partition the asset return vector as $\tilde{\boldsymbol{r}} = (\tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\eta}})$, where the $\mathbb{R}^n$-valued random vector $\tilde{\boldsymbol{\xi}}$ and $\mathbb{R}^{m-n}$-valued random vector $\tilde{\boldsymbol{\eta}}$ denote the basic asset returns and derivative returns, respectively.

To approximate the VaR of some portfolio $\boldsymbol{w} \in \mathcal{W}$ containing derivatives, one can principally still use the WCVaR model (4.7), which has the advantage of computational tractability and accounts for the absence of distributional information beyond first- and second-order moments. However, WCVaR is not a suitable approximation for VaR in the presence of derivatives due to the following reasons.

The first- and second-order moments of the derivative returns, which must be supplied to the WCVaR model, are difficult to estimate reliably from historical data, see, e.g., [CS02]. Note that the moments of the basic assets returns (i.e., stocks and bonds etc.) can usually be estimated more accurately due to the availability of longer historical time series. However, even if the means and covariances of the derivative returns were precisely known, WCVaR would still provide a poor approximation of the actual portfolio VaR because it disregards known perfect dependencies between the derivative returns and their underlying asset returns. In fact, the returns of the derivatives are uniquely determined by the returns of the underlying assets, that is, there exists a (typically non-linear) measurable function $f : \mathbb{R}^n \to \mathbb{R}^m$ such that $\tilde{\boldsymbol{r}} = f(\tilde{\boldsymbol{\xi}})$.[1] Put differently, the derivatives introduce no new uncertainties in the market; their returns are uncertain only because the underlying asset returns are uncertain. The function f can usually be inferred reliably from contractual specifications (payoff functions) or pricing models of the derivatives.

[1] For ease of exposition, we assume that the returns of the derivative underliers are the only risk factors determining the option returns.

In summary, WCVaR provides a conservative approximation to the actual VaR. However, it relies on first- and second-order moments of the derivative returns, which are difficult to obtain in practice, but disregards the perfect dependencies captured by the function f, which is typically known.

When f is non-linear, WCVaR tends to severely *overestimate* the actual VaR since the covariance matrix $\boldsymbol{\Sigma_r}$ accounts only for *linear* dependencies. The robust optimization perspective on WCVaR manifests this drawback geometrically. Recall that the ellipsoidal uncertainty set $\mathcal{U}_\epsilon$ introduced in Section 4.2.2 is symmetrically oriented around the mean vector $\boldsymbol{\mu_r}$. If the underlying assets of the derivatives have approximately symmetrically distributed returns, then the derivative returns are heavily skewed. An ellipsoidal uncertainty set fails to capture this asymmetry. This geometric argument supports our conjecture that WCVaR provides a poor (over-pessimistic) VaR estimate when the portfolio contains derivatives.

In the remainder of the chapter we assume to know the first- and second-order moments of the basic asset returns as well as the function f, which captures the non-linear dependencies between the basic asset and derivative returns. In contrast, we assume that the moments of the derivative returns are unknown.

In the next sections we derive generic Worst-Case Value-at-Risk models that explicitly account for non-linear (piecewise linear or quadratic) relationships between the asset returns. These new models provide tighter approximations for the actual VaR of portfolios containing derivatives than the WCVaR model, which relies solely on moment information.

Below, we will always denote the mean vector and the covariance matrix of the basic asset returns by $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively. Without loss of generality we assume that $\boldsymbol{\Sigma}$ is strictly positive definite.

4.4 Worst-Case Polyhedral VaR Optimization

In this section we describe a Worst-Case VaR model that explicitly accounts for piecewise linear relationships between option returns and their underlying asset returns. We show that this model can be cast as a tractable second-order cone program and establish its equivalence to a robust optimization model that admits an intuitive interpretation.

4.4.1 Piecewise Linear Portfolio Model

We now assume that the $m-n$ derivatives in our market are *European-style call and/or put options* derived from the basic assets. All these options are assumed to mature at the end of the investment horizon, that is, at time T.

For ease of exposition, we partition the allocation vector as $\boldsymbol{w} = (\boldsymbol{w}^{\boldsymbol{\xi}}, \boldsymbol{w}^{\boldsymbol{\eta}})$, where $\boldsymbol{w}^{\boldsymbol{\xi}} \in \mathbb{R}^n$ and $\boldsymbol{w}^{\boldsymbol{\eta}} \in \mathbb{R}^{m-n}$ denote the percentage allocations in the basic assets and options, respectively. In this section we forbid short-sales of options, that is, we assume that the inclusion $\boldsymbol{w} \in \mathcal{W}$ implies $\boldsymbol{w}^{\boldsymbol{\eta}} \geq \mathbf{0}$. Recall that the set $\mathcal{W}$ of admissible portfolios was assumed to be a convex polyhedron.

We now derive an explicit representation for f by using the known payoff functions of the basic assets as well as the European call and put options. Since the first n components of $\tilde{\boldsymbol{r}}$ represent the basic asset returns $\tilde{\boldsymbol{\xi}}$, we have $f_j(\tilde{\boldsymbol{\xi}}) = \tilde{\xi}_j$ for $j = 1, \ldots, n$. Next, we investigate the option returns $\tilde{r}_j$ for $j = n+1, \ldots, m$.[2] Let asset j be a call option with strike price k_j on the basic asset i, and denote the return and the initial price of the option by $\tilde{r}_j$ and c_j, respectively. If s_i denotes the initial price of asset i, then its end-of-period price amounts to $s_i(1+\tilde{\xi}_i)$. We can now explicitly express the return $\tilde{r}_j$ as a convex piecewise linear function of $\tilde{\xi}_i$,

$$\begin{aligned} f_j(\tilde{\boldsymbol{\xi}}) &= \frac{1}{c_j} \max\left\{0, s_i(1+\tilde{\xi}_i) - k_j\right\} - 1 \\ &= \max\left\{-1, a_j + b_j\tilde{\xi}_i - 1\right\}, \quad \text{where} \quad a_j = \frac{s_i - k_j}{c_j} \quad \text{and} \quad b_j = \frac{s_i}{c_j}. \end{aligned} \tag{4.12a}$$

[2]The following equations are equivalent to those presented in Section 3.3.1 but where we now use relative returns. The equations are repeated for clarity of exposition.

Similarly, if asset j is a put option with price p_j and strike price k_j on the basic asset i, then its return $\tilde{r}_j$ is representable as a different convex piecewise linear function,

$$f_j(\tilde{\boldsymbol{\xi}}) = \max\left\{-1, a_j + b_j\tilde{\xi}_i - 1\right\}, \quad \text{where} \quad a_j = \frac{k_j - s_i}{p_j} \quad \text{and} \quad b_j = -\frac{s_i}{p_j}. \tag{4.12b}$$

Using the above notation, we can write the vector of asset returns $\tilde{\boldsymbol{r}}$ compactly as

$$\tilde{\boldsymbol{r}} = f(\tilde{\boldsymbol{\xi}}) = \begin{pmatrix} \tilde{\boldsymbol{\xi}} \\ \max\left\{-\boldsymbol{e}, \boldsymbol{a} + \mathbf{B}\tilde{\boldsymbol{\xi}} - \boldsymbol{e}\right\} \end{pmatrix}, \tag{4.13}$$

where $\boldsymbol{a} \in \mathbb{R}^{m-n}$, $\mathbf{B} \in \mathbb{R}^{(m-n)\times n}$ are known constants determined through (3.27a) and (3.27b), $\boldsymbol{e} \in \mathbb{R}^{m-n}$ is the vector of 1s, and 'max' denotes the component-wise maximization operator. Thus, the return $\tilde{r}_p$ of some portfolio $\boldsymbol{w} \in \mathcal{W}$ can be expressed as

$$\begin{aligned} \tilde{r}_p &= \boldsymbol{w}^\mathsf{T}\tilde{\boldsymbol{r}} = (\boldsymbol{w}^{\boldsymbol{\xi}})^\mathsf{T}\tilde{\boldsymbol{\xi}} + (\boldsymbol{w}^{\boldsymbol{\eta}})^\mathsf{T}\tilde{\boldsymbol{\eta}} \\ &= \boldsymbol{w}^\mathsf{T} f(\tilde{\boldsymbol{\xi}}) = (\boldsymbol{w}^{\boldsymbol{\xi}})^\mathsf{T}\tilde{\boldsymbol{\xi}} + (\boldsymbol{w}^{\boldsymbol{\eta}})^\mathsf{T} \max\left\{-\boldsymbol{e}, \boldsymbol{a} + \mathbf{B}\tilde{\boldsymbol{\xi}} - \boldsymbol{e}\right\}. \end{aligned} \tag{4.14}$$

4.4.2 Worst-Case Polyhedral VaR Model

For any portfolio $\boldsymbol{w} \in \mathcal{W}$, we define the *Worst-Case Polyhedral VaR* (WCPVaR) as

$$\begin{aligned} \text{WCPVaR}_\epsilon(\boldsymbol{w}) &= \min\left\{\gamma \;:\; \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left\{\gamma \le -\boldsymbol{w}^\mathsf{T} f(\tilde{\boldsymbol{\xi}})\right\} \le \epsilon\right\} \qquad (4.15) \\ &= \min\left\{\gamma \;:\; \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left\{\gamma \le -(\boldsymbol{w}^{\boldsymbol{\xi}})^\mathsf{T}\tilde{\boldsymbol{\xi}} - (\boldsymbol{w}^{\boldsymbol{\eta}})^\mathsf{T} \max\left\{-\boldsymbol{e}, \boldsymbol{a} + \mathbf{B}\tilde{\boldsymbol{\xi}} - \boldsymbol{e}\right\}\right\} \le \epsilon\right\}, \end{aligned}$$

where $\mathcal{P}$ denotes the set of all probability distributions of the *basic* asset returns $\tilde{\boldsymbol{\xi}}$ with a given mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. WCPVaR provides a tight conservative approximation for the VaR of a portfolio whose return constitutes a convex piecewise linear (i.e., polyhedral) function of the basic asset returns.

In the remainder of this section we derive a manifestly tractable representation for WCPVaR.

As a first step to achieve this goal, we simplify the maximization problem

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left\{\gamma \leq -(\boldsymbol{w}^{\boldsymbol{\xi}})^\top\tilde{\boldsymbol{\xi}} - (\boldsymbol{w}^{\boldsymbol{\eta}})^\top \max\left\{-\boldsymbol{e}, \boldsymbol{a} + \mathbf{B}\tilde{\boldsymbol{\xi}} - \boldsymbol{e}\right\}\right\}, \tag{4.16}$$

which can be identified as the subordinate optimization problem in (4.15).

For some fixed portfolio $\boldsymbol{w} \in \mathcal{W}$ and $\gamma \in \mathbb{R}$, we define the set $\mathcal{S}_\gamma \subseteq \mathbb{R}^n$ as

$$\mathcal{S}_\gamma = \{\boldsymbol{\xi} \in \mathbb{R}^n \ : \ \gamma + (\boldsymbol{w}^{\boldsymbol{\xi}})^\top\boldsymbol{\xi} + (\boldsymbol{w}^{\boldsymbol{\eta}})^\top \max\{-\boldsymbol{e}, \boldsymbol{a} + \mathbf{B}\boldsymbol{\xi} - \boldsymbol{e}\} \leq 0\}.$$

For any $\boldsymbol{\xi} \in \mathbb{R}^n$ and nonnegative $\boldsymbol{w}^{\boldsymbol{\eta}} \in \mathbb{R}^{m-n}$ we have

$$\begin{aligned}(\boldsymbol{w}^{\boldsymbol{\eta}})^\top \max\{-\boldsymbol{e}, \boldsymbol{a} + \mathbf{B}\boldsymbol{\xi} - \boldsymbol{e}\} &= \min_{\boldsymbol{g}\in\mathbb{R}^{m-n}} \left\{\boldsymbol{g}^\top\boldsymbol{w}^{\boldsymbol{\eta}} \ : \ \boldsymbol{g} \geq -\boldsymbol{e}, \ \boldsymbol{g} \geq \boldsymbol{a} + \mathbf{B}\boldsymbol{\xi} - \boldsymbol{e}\right\} \\ &= \max_{\boldsymbol{y}\in\mathbb{R}^{m-n}} \left\{\boldsymbol{y}^\top(\boldsymbol{a} + \mathbf{B}\boldsymbol{\xi}) - \boldsymbol{e}^\top\boldsymbol{w}^{\boldsymbol{\eta}} \ : \ \mathbf{0} \leq \boldsymbol{y} \leq \boldsymbol{w}^{\boldsymbol{\eta}}\right\},\end{aligned}$$

where the second equality follows from strong linear programming duality. Thus, the set $\mathcal{S}_\gamma$ can be written as

$$\mathcal{S}_\gamma = \left\{\boldsymbol{\xi} \in \mathbb{R}^n \ : \ \max_{\mathbf{0}\leq\boldsymbol{y}\leq\boldsymbol{w}^{\boldsymbol{\eta}}} \left\{\gamma + (\boldsymbol{w}^{\boldsymbol{\xi}})^\top\boldsymbol{\xi} + \boldsymbol{y}^\top(\boldsymbol{a} + \mathbf{B}\boldsymbol{\xi}) - \boldsymbol{e}^\top\boldsymbol{w}^{\boldsymbol{\eta}}\right\} \leq 0\right\}. \tag{4.17}$$

The optimal value of problem (4.16) can be obtained by solving the worst-case probability problem

$$\pi_{\mathrm{wc}} = \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\{\tilde{\boldsymbol{\xi}} \in \mathcal{S}_\gamma\}. \tag{4.18}$$

The next lemma reviews a general result about worst-case probability problems and will play a key role in many of the following derivations. The proof is due to Calafiore *et al.* [CTEG09] but is repeated in Appendix 4.8.1 to keep this chapter self-contained.

Lemma 4.4.1 *Let $\mathcal{S} \subseteq \mathbb{R}^n$ be any Borel measurable set (which is not necessarily convex), and define the worst-case probability π_{wc} as*

$$\pi_{\mathrm{wc}} = \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\{\tilde{\boldsymbol{\xi}} \in \mathcal{S}\}, \tag{4.19}$$

where $\mathcal{P}$ is the set of all probability distributions of $\tilde{\boldsymbol{\xi}}$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma} \succ \mathbf{0}$. Then,

$$\pi_{\text{wc}} = \inf_{\mathbf{M} \in \mathbb{S}^{n+1}} \left\{ \langle \boldsymbol{\Omega}, \mathbf{M} \rangle \ : \ \mathbf{M} \succcurlyeq \mathbf{0}, \quad \left[\boldsymbol{\xi}^\intercal \ 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\intercal \ 1\right]^\intercal \geq 1 \quad \forall \boldsymbol{\xi} \in \mathcal{S} \right\}, \tag{4.20}$$

where

$$\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\intercal & \boldsymbol{\mu} \\ \boldsymbol{\mu}^\intercal & 1 \end{bmatrix} \tag{4.21}$$

is the second-order moment matrix of $\tilde{\boldsymbol{\xi}}$.

Lemma 4.4.1 enables us to reformulate the worst-case probability problem (4.18) as

$$\begin{aligned} \pi_{\text{wc}} = \inf_{\mathbf{M} \in \mathbb{S}^{n+1}} \quad & \langle \boldsymbol{\Omega}, \mathbf{M} \rangle \\ \text{s.t.} \quad & \left[\boldsymbol{\xi}^\intercal \ 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\intercal \ 1\right]^\intercal \geq 1 \ \forall \boldsymbol{\xi} : \max_{\mathbf{0} \leq \boldsymbol{y} \leq \boldsymbol{w}^{\boldsymbol{\eta}}} \{\gamma + (\boldsymbol{w}^{\boldsymbol{\xi}})^\intercal \boldsymbol{\xi} + \boldsymbol{y}^\intercal (\boldsymbol{a} + \mathbf{B}\boldsymbol{\xi}) - \boldsymbol{e}^\intercal \boldsymbol{w}^{\boldsymbol{\eta}}\} \leq 0 \\ & \mathbf{M} \succcurlyeq \mathbf{0}. \end{aligned} \tag{4.22}$$

We now recall the non-linear Farkas Lemma, which is a fundamental theorem of alternatives in convex analysis and will enable us to simplify the optimization problem (4.22), see, e.g., [PT07, Theorem 2.1] and the references therein.

Lemma 4.4.2 (Farkas Lemma) *Let $f_0, \ldots, f_p : \mathbb{R}^n \to \mathbb{R}$ be convex functions, and assume that there exists a strictly feasible point $\bar{\boldsymbol{\xi}}$ with $f_i(\bar{\boldsymbol{\xi}}) < 0$, $i = 1, \ldots, p$. Then, $f_0(\boldsymbol{\xi}) \geq 0$ for all $\boldsymbol{\xi}$ with $f_i(\bar{\boldsymbol{\xi}}) \leq 0$, $i = 1, \ldots, p$, if and only if there exist constants $\tau_i \geq 0$ such that*

$$f_0(\boldsymbol{\xi}) + \sum_{i=1}^{p} \tau_i f_i(\boldsymbol{\xi}) \geq 0 \quad \forall \boldsymbol{\xi} \in \mathbb{R}^n.$$

We will now argue that problem (4.22) can be reformulated as follows.

$$\begin{array}{ll} \inf & \langle \boldsymbol{\Omega}, \mathbf{M} \rangle \\ \text{s.t.} & \mathbf{M} \in \mathbb{S}^{n+1}, \quad \tau \in \mathbb{R}, \quad \mathbf{M} \succcurlyeq \mathbf{0}, \quad \tau \geq 0 \\ & \left[\boldsymbol{\xi}^\top \; 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\top \; 1\right]^\top - 1 + 2\tau \left(\max\limits_{\mathbf{0} \leq \boldsymbol{y} \leq \boldsymbol{w}^{\boldsymbol{\eta}}} \{\gamma + (\boldsymbol{w}^{\boldsymbol{\xi}})^\top \boldsymbol{\xi} + \boldsymbol{y}^\top (\boldsymbol{a} + \mathbf{B}\boldsymbol{\xi}) - \boldsymbol{e}^\top \boldsymbol{w}^{\boldsymbol{\eta}}\} \right) \geq 0 \;\; \forall \boldsymbol{\xi} \in \mathbb{R}^n \end{array} \tag{4.23}$$

For ease of exposition, we first first define

$$h = \min_{\boldsymbol{\xi} \in \mathbb{R}^n} \max_{\mathbf{0} \leq \boldsymbol{y} \leq \boldsymbol{w}^{\boldsymbol{\eta}}} \{\gamma + (\boldsymbol{w}^{\boldsymbol{\xi}})^\top \boldsymbol{\xi} + \boldsymbol{y}^\top (\boldsymbol{a} + \mathbf{B}\boldsymbol{\xi}) - \boldsymbol{e}^\top \boldsymbol{w}^{\boldsymbol{\eta}}\}.$$

The equivalence of (4.22) and (4.23) is proved case by case. Assume first that $h < 0$. Then, the equivalence follows from the Farkas Lemma. Assume next that $h > 0$. Then, the semi-infinite constraint in (4.22) becomes redundant and, since $\boldsymbol{\Omega} \succ \mathbf{0}$, the optimal solution of (4.22) is given by $\mathbf{M} = \mathbf{0}$ with a corresponding optimal value of 0. The optimal value of problem (4.23) is also equal to 0. Indeed, by choosing $\tau = 1/h$, the semi-infinite constraint in (4.23) is satisfied independently of $\mathbf{M}$. Finally, assume that $h = 0$. In this degenerate case the equivalence follows from a standard continuity argument. Details are omitted for brevity of exposition.

It can be seen that since $\tau \geq 0$, the semi-infinite constraint in (4.23) is equivalent to the assertion that there exists some $\mathbf{0} \leq \boldsymbol{y} \leq \boldsymbol{w}^{\boldsymbol{\eta}}$ with

$$\left[\boldsymbol{\xi}^\top \; 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\top \; 1\right]^\top - 1 + 2\tau \left(\gamma + (\boldsymbol{w}^{\boldsymbol{\xi}})^\top \boldsymbol{\xi} + \boldsymbol{y}^\top (\boldsymbol{a} + \mathbf{B}\boldsymbol{\xi}) - \boldsymbol{e}^\top \boldsymbol{w}^{\boldsymbol{\eta}}\right) \geq 0 \qquad \forall \boldsymbol{\xi} \in \mathbb{R}^n.$$

This semi-infinite constraint can be written as

$$\begin{aligned} & \begin{bmatrix} \boldsymbol{\xi} \\ 1 \end{bmatrix}^\top \left(\mathbf{M} + \begin{bmatrix} \mathbf{0} & \tau(\boldsymbol{w}^{\boldsymbol{\xi}} + \mathbf{B}^\top \boldsymbol{y}) \\ \tau(\boldsymbol{w}^{\boldsymbol{\xi}} + \mathbf{B}^\top \boldsymbol{y})^\top & -1 + 2\tau(\gamma + \boldsymbol{y}^\top \boldsymbol{a} - \boldsymbol{e}^\top \boldsymbol{w}^{\boldsymbol{\eta}}) \end{bmatrix} \right) \begin{bmatrix} \boldsymbol{\xi} \\ 1 \end{bmatrix} \geq \mathbf{0} \quad \forall \boldsymbol{\xi} \in \mathbb{R}^n \\ \iff \quad & \mathbf{M} + \begin{bmatrix} \mathbf{0} & \tau(\boldsymbol{w}^{\boldsymbol{\xi}} + \mathbf{B}^\top \boldsymbol{y}) \\ \tau(\boldsymbol{w}^{\boldsymbol{\xi}} + \mathbf{B}^\top \boldsymbol{y})^\top & -1 + 2\tau(\gamma + \boldsymbol{y}^\top \boldsymbol{a} - \boldsymbol{e}^\top \boldsymbol{w}^{\boldsymbol{\eta}}) \end{bmatrix} \succcurlyeq \mathbf{0}. \end{aligned}$$

Thus, the worst-case probability problem (4.22) can equivalently be formulated as

$$\begin{aligned}
\pi_{\mathrm{wc}} = \inf \quad & \langle \boldsymbol{\Omega}, \mathbf{M} \rangle \\
\text{s.t.} \quad & \mathbf{M} \in \mathbb{S}^{n+1}, \quad \boldsymbol{y} \in \mathbb{R}^{m-n}, \quad \tau \in \mathbb{R} \\
& \mathbf{M} \succcurlyeq \mathbf{0}, \quad \tau \geq 0, \quad \mathbf{0} \leq \boldsymbol{y} \leq \boldsymbol{w}^{\boldsymbol{\eta}} \\
& \mathbf{M} + \begin{bmatrix} \mathbf{0} & \tau(\boldsymbol{w}^{\boldsymbol{\xi}} + \mathbf{B}^{\mathsf{T}} \boldsymbol{y}) \\ \tau(\boldsymbol{w}^{\boldsymbol{\xi}} + \mathbf{B}^{\mathsf{T}} \boldsymbol{y})^{\mathsf{T}} & -1 + 2\tau(\gamma + \boldsymbol{y}^{\mathsf{T}} \boldsymbol{a} - \boldsymbol{e}^{\mathsf{T}} \boldsymbol{w}^{\boldsymbol{\eta}}) \end{bmatrix} \succcurlyeq \mathbf{0}.
\end{aligned} \tag{4.24}$$

By using (4.24) we can express WCPVaR in (4.15) as the optimal value of the following minimization problem.

$$\begin{aligned}
\mathrm{WCPVaR}_{\epsilon}(\boldsymbol{w}) = \inf \quad & \gamma \\
\text{s.t.} \quad & \mathbf{M} \in \mathbb{S}^{n+1}, \quad \boldsymbol{y} \in \mathbb{R}^{m-n}, \quad \tau \in \mathbb{R}, \quad \gamma \in \mathbb{R} \\
& \langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq \epsilon, \quad \mathbf{M} \succcurlyeq \mathbf{0}, \quad \tau \geq 0, \quad \mathbf{0} \leq \boldsymbol{y} \leq \boldsymbol{w}^{\boldsymbol{\eta}} \\
& \mathbf{M} + \begin{bmatrix} \mathbf{0} & \tau(\boldsymbol{w}^{\boldsymbol{\xi}} + \mathbf{B}^{\mathsf{T}} \boldsymbol{y}) \\ \tau(\boldsymbol{w}^{\boldsymbol{\xi}} + \mathbf{B}^{\mathsf{T}} \boldsymbol{y})^{\mathsf{T}} & -1 + 2\tau(\gamma + \boldsymbol{y}^{\mathsf{T}} \boldsymbol{a} - \boldsymbol{e}^{\mathsf{T}} \boldsymbol{w}^{\boldsymbol{\eta}}) \end{bmatrix} \succcurlyeq \mathbf{0}
\end{aligned} \tag{4.25}$$

Problem (4.25) is non-convex due to the bilinear terms in the matrix inequality constraint. It can easily be shown that $\langle \boldsymbol{\Omega}, \mathbf{M} \rangle \geq 1$ for any feasible point with vanishing τ-component. However, since $\epsilon < 1$, this is in conflict with the constraint $\langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq \epsilon$. We thus conclude that no feasible point can have a vanishing τ-component. This allows us to divide the matrix inequality in problem (4.25) by τ. Subsequently we perform variable substitutions in which we replace $1/\tau$ by τ and $\mathbf{M}/\tau$ by $\mathbf{M}$. This yields the following reformulation of problem (4.25).

$$\begin{aligned}
\mathrm{WCPVaR}_{\epsilon}(\boldsymbol{w}) = \inf \quad & \gamma \\
\text{s.t.} \quad & \mathbf{M} \in \mathbb{S}^{n+1}, \quad \boldsymbol{y} \in \mathbb{R}^{m-n}, \quad \tau \in \mathbb{R}, \quad \gamma \in \mathbb{R} \\
& \langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq \tau\epsilon, \quad \mathbf{M} \succcurlyeq \mathbf{0}, \quad \tau \geq 0, \quad \mathbf{0} \leq \boldsymbol{y} \leq \boldsymbol{w}^{\boldsymbol{\eta}} \\
& \mathbf{M} + \begin{bmatrix} \mathbf{0} & \boldsymbol{w}^{\boldsymbol{\xi}} + \mathbf{B}^{\mathsf{T}} \boldsymbol{y} \\ (\boldsymbol{w}^{\boldsymbol{\xi}} + \mathbf{B}^{\mathsf{T}} \boldsymbol{y})^{\mathsf{T}} & -\tau + 2(\gamma + \boldsymbol{y}^{\mathsf{T}} \boldsymbol{a} - \boldsymbol{e}^{\mathsf{T}} \boldsymbol{w}^{\boldsymbol{\eta}}) \end{bmatrix} \succcurlyeq \mathbf{0}
\end{aligned} \tag{4.26}$$

Observe that (4.26) constitutes a semidefinite program (SDP) that can be used to efficiently compute the WCPVaR of a given portfolio $\boldsymbol{w} \in \mathcal{W}$. However, it would be desirable to obtain an equivalent second-order cone program (SOCP) because SOCPs exhibit better scalability properties than SDPs [AG03]. Theorem 4.4.1 shows that such a reformulation exists.

Theorem 4.4.1 *Problem* (4.26) *can be reformulated as*

$$\mathrm{WCPVaR}_\epsilon(\boldsymbol{w}) = \min_{\mathbf{0} \leq \boldsymbol{g} \leq \boldsymbol{w}^{\boldsymbol{\eta}}} \; -\boldsymbol{\mu}^\mathsf{T}(\boldsymbol{w}^{\boldsymbol{\xi}} + \mathbf{B}^\mathsf{T}\boldsymbol{g}) + \kappa(\epsilon) \left\| \boldsymbol{\Sigma}^{1/2}(\boldsymbol{w}^{\boldsymbol{\xi}} + \mathbf{B}^\mathsf{T}\boldsymbol{g}) \right\|_2 - \boldsymbol{a}^\mathsf{T}\boldsymbol{g} + \boldsymbol{e}^\mathsf{T}\boldsymbol{w}^{\boldsymbol{\eta}}, \quad (4.27)$$

which constitutes a tractable SOCP.

Proof The proof follows a similar reasoning as in [EGOO03, Theorem 1] and is therefore relegated to Appendix 4.8.2.

Remark 4.4.1 *In the absence of derivatives, that is, when the market only contains basic assets, then $m = n$ and $\boldsymbol{w} = \boldsymbol{w}^{\boldsymbol{\xi}}$. In this special case we obtain*

$$\mathrm{WCPVaR}_\epsilon(\boldsymbol{w}) = -\boldsymbol{\mu}^\mathsf{T}\boldsymbol{w} + \kappa(\epsilon) \left\| \boldsymbol{\Sigma}^{1/2}\boldsymbol{w} \right\|_2 = \mathrm{WCVaR}_\epsilon(\boldsymbol{w}).$$

Thus, the WCPVaR *model encapsulates the* WCVaR *model* (4.7) *as a special case.*

The problem of minimizing the WCVaR of a portfolio containing European options can now be conservatively approximated by

$$\begin{aligned} \underset{\boldsymbol{w} \in \mathbb{R}^m}{\text{minimize}} \quad & \mathrm{WCPVaR}_\epsilon(\boldsymbol{w}) \\ \text{subject to} \quad & \boldsymbol{w} \in \mathcal{W}, \end{aligned}$$

which is equivalent to the tractable SOCP

$$
\begin{aligned}
\text{minimize} \quad & \gamma \\
\text{subject to} \quad & \boldsymbol{w}^{\boldsymbol{\xi}} \in \mathbb{R}^n, \quad \boldsymbol{w}^{\boldsymbol{\eta}} \in \mathbb{R}^{m-n}, \quad \boldsymbol{g} \in \mathbb{R}^{m-n}, \quad \gamma \in \mathbb{R} \\
& -\boldsymbol{\mu}^\mathsf{T}(\boldsymbol{w}^{\boldsymbol{\xi}} + \mathbf{B}^\mathsf{T}\boldsymbol{g}) + \kappa(\epsilon) \left\| \boldsymbol{\Sigma}^{1/2}(\boldsymbol{w}^{\boldsymbol{\xi}} + \mathbf{B}^\mathsf{T}\boldsymbol{g}) \right\|_2 - \boldsymbol{a}^\mathsf{T}\boldsymbol{g} + \boldsymbol{e}^\mathsf{T}\boldsymbol{w}^{\boldsymbol{\eta}} \le \gamma \\
& \mathbf{0} \le \boldsymbol{g} \le \boldsymbol{w}^{\boldsymbol{\eta}}, \quad \boldsymbol{w} = (\boldsymbol{w}, \boldsymbol{w}^{\boldsymbol{\eta}}), \quad \boldsymbol{w} \in \mathcal{W}.
\end{aligned}
\tag{4.28}
$$

Recall that the set of admissible portfolios $\mathcal{W}$ precludes short positions in options, that is, $\boldsymbol{w} \in \mathcal{W}$ implies $\boldsymbol{w}^{\boldsymbol{\eta}} \ge \mathbf{0}$. Furthermore, note that problem (4.28) bears a strong similarity to the robust portfolio optimization model (3.31), which we derived in Section 3.[3]

4.4.3 Robust Optimization Perspective on WCPVaR

In Section 4.2 we highlighted a known relationship between WCVaR optimization and robust optimization. Moreover, in Section 4.3 we argued that the ellipsoidal uncertainty set related to the WCVaR model is symmetric and as such fails to capture the asymmetric dependencies between options and their underlying assets. In the next theorem we establish that the WCPVaR minimization problem (4.28) can also be cast as a robust optimization problem of the type (4.9). However, the uncertainty set which generates WCPVaR is no longer symmetric.

Theorem 4.4.2 *The* WCPVaR *minimization problem* (4.28) *is equivalent to the robust optimization problem*

$$
\begin{aligned}
\underset{\boldsymbol{w} \in \mathbb{R}^m, \gamma \in \mathbb{R}}{\text{minimize}} \quad & \gamma \\
\text{subject to} \quad & -\boldsymbol{w}^\mathsf{T}\boldsymbol{r} \le \gamma \qquad \forall \boldsymbol{r} \in \mathcal{U}^p_\epsilon \\
& \boldsymbol{w} \in \mathcal{W},
\end{aligned}
\tag{4.29}
$$

where the uncertainty set $\mathcal{U}^p_\epsilon \subseteq \mathbb{R}^m$ *is defined as*

$$
\mathcal{U}^p_\epsilon = \left\{ \boldsymbol{r} \in \mathbb{R}^m \ : \ \exists \boldsymbol{\xi} \in \mathbb{R}^n, \ (\boldsymbol{\xi} - \boldsymbol{\mu})^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\xi} - \boldsymbol{\mu}) \le \kappa(\epsilon)^2, \ \boldsymbol{r} = f(\boldsymbol{\xi}) \right\}. \tag{4.30}
$$

[3] The small differences are due to the facts that we use relative returns and that we do not consider support information in this chapter.

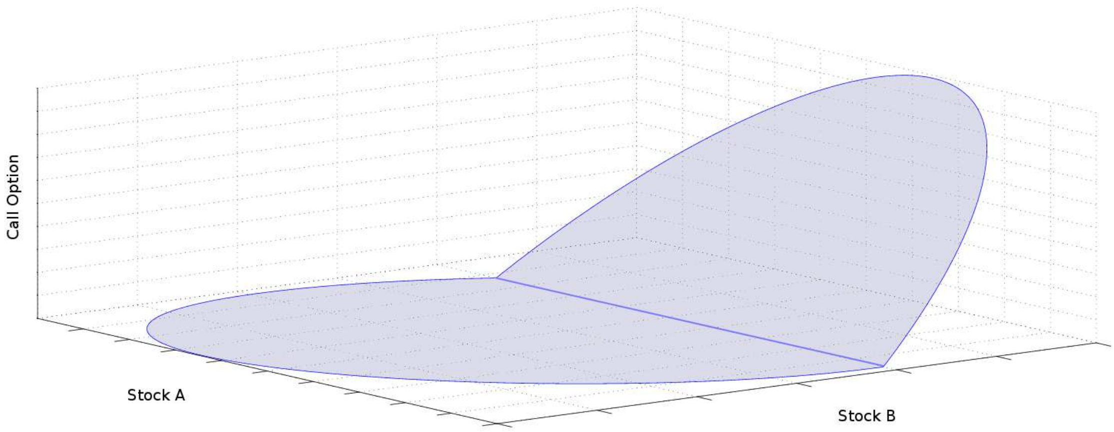

Figure 4.1: Illustration of the $\mathcal{U}_\epsilon^p$ uncertainty set: the classical ellipsoidal uncertainty set has been transformed by the piecewise linear payoff function of the call option written on stock B.

Proof The result is based on conic duality. We refer to Theorem 3.3.1 for an analogous exposition of the proof. ∎

Remark 4.4.2 *Unlike the uncertainty set $\mathcal{U}_\epsilon$ defined in Section 4.2.2, the new uncertainty set $\mathcal{U}_\epsilon^p$ reflects the non-linear relationship between the option returns and their underlying asset returns. Because f is a convex piecewise linear function, the uncertainty set is no longer symmetric around $\boldsymbol{\mu}$. The asymmetry is caused by the option returns, see Figure 4.1.*

Example 4.4.1 *Consider a Black-Scholes economy consisting of stocks A and B, a European call option on stock A, and a European put option on stock B. Furthermore, let $\boldsymbol{w}$ be an equally weighted portfolio of these $m = 4$ assets, that is, set $w_i = 1/m$ for $i = 1, \ldots, m$. Thus we have $\boldsymbol{w} = [0.25\ 0.25\ 0.25\ 0.25]^\mathsf{T}$.*

We assume that the prices of stocks A and B are governed by a bivariate geometric Brownian motion with drift coefficients of 12% and 8%, and volatilities of 30% and 20% per annum, respectively. The correlation between the instantaneous stock returns amounts to 20%. The initial prices of the stocks are \$100. The options mature in 21 days and have strike prices of \$100. We assume that the risk-free rate is 3% per annum and that there are 252 trading days per year. By using the Black-Scholes formulas [BS73], we obtain call and put option prices of \$3.5758 and \$2.1774, respectively.

We want to compute the VaR at confidence level ϵ for portfolio $\boldsymbol{w}$ and a 21-day time horizon.

To this end, we randomly generate L=5,000,000 end-of-period stock prices and corresponding option payoffs. These are used to obtain L asset and portfolio return samples. Figure 4.2 (left) displays the sampled portfolio loss distribution, which exhibits considerable skewness due to the options. The Monte-Carlo VaR *is obtained by computing the* $(1-\epsilon)$*-quantile of the sampled portfolio loss distribution.*

We also compute the 21-day sample means $\boldsymbol{\mu_r}$ *and sample covariance matrix* $\boldsymbol{\Sigma_r}$ *of the asset returns, which are used for the calculation of* WCVaR (4.7) *and* WCPVaR (4.27)*. These values are*

$$\boldsymbol{\mu_r} = \begin{bmatrix} 0.01 \\ 0.0067 \\ 0.1165 \\ -0.0856 \end{bmatrix}, \quad \textit{and} \quad \boldsymbol{\Sigma_r} = \begin{bmatrix} 0.0077 & 0.0010 & 0.1245 & -0.0204 \\ 0.0010 & 0.0034 & 0.0160 & -0.0670 \\ 0.1245 & 0.0160 & 2.5466 & -0.3028 \\ -0.0204 & -0.0670 & -0.3028 & 1.9580 \end{bmatrix},$$

where the first two entries in $\boldsymbol{\mu_r}$ *belong to the stock returns, followed by the call and put option returns. The entries for the covariance matrix obey this ordering.*

Let us now compute WCVaR *at, for example,* $\epsilon = 10\%$*. We have* $\kappa(0.1) = \sqrt{(1-0.1)/0.1} = 3$*. We now insert the above parameter values into equation* (4.7)*, and compute* $\mathrm{WCVaR}_{0.1}(\boldsymbol{w})$ *as*

$$-\begin{bmatrix} 0.01 \\ 0.0067 \\ 0.1165 \\ -0.0856 \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix} + 3\sqrt{\begin{bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} 0.0077 & 0.0010 & 0.1245 & -0.0204 \\ 0.0010 & 0.0034 & 0.0160 & -0.0670 \\ 0.1245 & 0.0160 & 2.5466 & -0.3028 \\ -0.0204 & -0.0670 & -0.3028 & 1.9580 \end{bmatrix} \begin{bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix}},$$

which is equal to 1.4916.

Next, we evaluate WCPVaR *at* $\epsilon = 10\%$*. To this end, we first compute the option specific multipliers* $\boldsymbol{a}$ *and* $\mathbf{B}$*, see* (4.12)*. These are equal to*

$$\boldsymbol{a} = \begin{bmatrix} \frac{100-100}{100} \\ \frac{100-100}{100} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \textit{and} \quad \mathbf{B} = \begin{bmatrix} \frac{100}{3.5758} & 0 \\ 0 & -\frac{100}{2.1774} \end{bmatrix} = \begin{bmatrix} 27.9655 & 0 \\ 0 & -45.9261 \end{bmatrix}.$$

Furthermore, the WCPVaR *model only requires the means* $\boldsymbol{\mu}$ *and covariance matrix* $\boldsymbol{\Sigma}$ *of the*

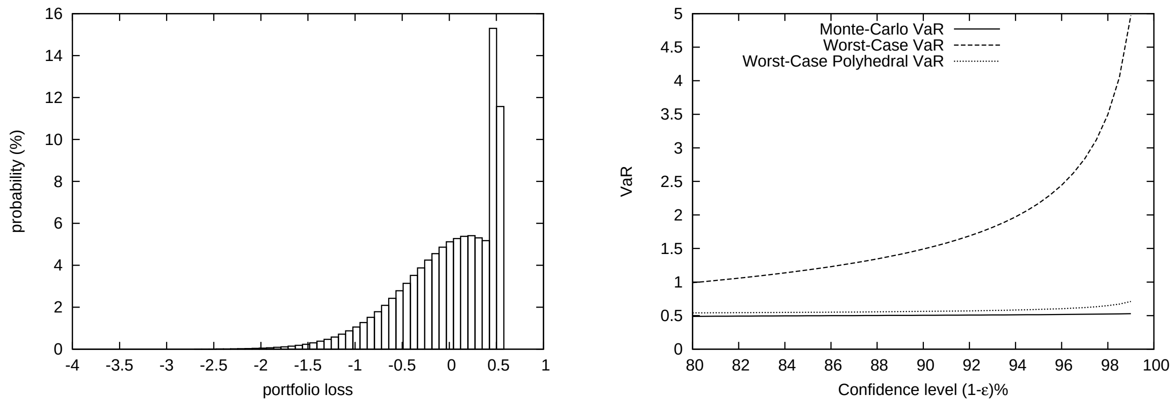

Figure 4.2: Left: The portfolio loss distribution obtained via Monte-Carlo simulation. Note that negative values represent gains. Right: The VaR estimates at different confidence levels obtained via Monte-Carlo sampling, WCVaR, and WCPVaR.

stock returns, which are equal to

$$\boldsymbol{\mu} = \begin{bmatrix} 0.01 \\ 0.0067 \end{bmatrix}, \quad \textit{and} \quad \boldsymbol{\Sigma} = \begin{bmatrix} 0.0077 & 0.0010 \\ 0.0010 & 0.0034 \end{bmatrix}.$$

We now compute $\mathrm{WCPVaR}_{0.1}(\boldsymbol{w})$ *by inserting the above parameter values into problem* (4.27), *by which we obtain*

$$\begin{aligned} \min_{g_1\in\mathbb{R}, g_2\in\mathbb{R}} \quad & -\begin{bmatrix} 0.01 \\ 0.0067 \end{bmatrix}^\mathsf{T} \left(\begin{bmatrix} 0.25 \\ 0.25 \end{bmatrix} + \begin{bmatrix} 27.9655 g_1 \\ -45.9261 g_2 \end{bmatrix} \right) + \begin{bmatrix} 1 \\ 1 \end{bmatrix}^\mathsf{T} \begin{bmatrix} 0.25 \\ 0.25 \end{bmatrix} \\ & +3 \sqrt{ \left(\begin{bmatrix} 0.25 \\ 0.25 \end{bmatrix} + \begin{bmatrix} 27.9655 g_1 \\ -45.9261 g_2 \end{bmatrix} \right)^\mathsf{T} \begin{bmatrix} 0.0077 & 0.0010 \\ 0.0010 & 0.0034 \end{bmatrix} \left(\begin{bmatrix} 0.25 \\ 0.25 \end{bmatrix} + \begin{bmatrix} 27.9655 g_1 \\ -45.9261 g_2 \end{bmatrix} \right) } \\ \text{s.t.} \quad & 0 \le g_1 \le 0.25, \quad 0 \le g_2 \le 0.25. \end{aligned}$$

The above problem is solved using the SDPT3 optimization toolkit [TTT03] and we find an optimal objective value equal to 0.5624.

Figure 4.2 (right) displays the VaR estimates at different levels of $\epsilon \in [0.01, 0.2]$. *We observe that for all values of* ϵ, *the* WCVaR *and* WCPVaR *values exceed the* Monte-Carlo VaR *estimate. This is not surprising since these models are distributionally robust and as such provide a conservative estimate of VaR. Note that the* Monte-Carlo VaR *can only be calculated accu-*

rately if many return samples are available (e.g., if the return distribution is precisely known). However, WCVaR *vastly overestimates* WCPVaR*. This effect is amplified for lower values of ϵ, where the accuracy of the VaR estimate matters most. Indeed, for $\epsilon = 1\%$, the* WCVaR *reports an unrealistically high value of* 497%*, which is 7 times larger than the corresponding* WCPVaR *value.*

4.5 Worst-Case Quadratic VaR Optimization

The WCPVaR model suffers from a number of weaknesses which may make it unattractive for certain investors.

Firstly, in order to obtain a tractable problem reformulation we had to prohibit short-sales of options. Although this is not restrictive for investors who merely want to enrich their portfolios with options in order to obtain insurance benefits (see Chapter 3), it severely constrains the complete set of option strategies that larger institutions might want to include in their portfolios.

Furthermore, we can only calculate and optimize the risk of portfolios comprising options that mature at the end of the investment horizon. As a result, investors cannot use the model, for example, to optimize portfolios including longer term options that mature far beyond the investment horizon.

Finally, the model is only suitable for portfolios containing plain vanilla European options and can not be used when exotic options are included in the portfolio.

In this section we propose an alternative Worst-Case VaR model which mitigates the weaknesses of the WCPVaR model. It is important to note that WCPVaR does not make any assumptions about the pricing model of the options. Only observable market prices and the known payoff functions of the options are used to calculate the option returns. In contrast, the new model proposed in this section requires the availability of a pricing model for the options. Moreover, it approximates the portfolio return using a second-order Taylor expansion which is only accurate for short investment horizons.

4.5.1 Delta-Gamma Portfolio Model

As in Section 4.4, we assume that there are $n \leq m$ *basic assets* and $m - n$ *derivatives* whose values are uniquely determined by the values of the basic assets. However, in contrast to Section 4.4, we do not only focus on European style options but also allow for exotic derivatives. Furthermore, we no longer require that the options mature at the end of the investment horizon.

We let $\tilde{\boldsymbol{s}}(t)$ denote the n-dimensional vector of basic asset prices at time $t \geq 0$ and assume that the prices at time $t = 0$ are known (i.e., deterministic). Moreover, we assume that the value of any (basic or non-basic) asset $i = 1, \ldots, m$ is representable as $v_i(\tilde{\boldsymbol{s}}(t), t)$, where $v_i : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ is twice continuously differentiable.

For a sufficiently short horizon time T, a second-order Taylor expansion accurately approximates the asset values at the end of the investment horizon. For $i = 1, \ldots, m$ we have

$$v_i(\tilde{\boldsymbol{s}}(T), T) - v_i(\boldsymbol{s}(0), 0) \;\approx\; \bar{\theta}_i T + \bar{\boldsymbol{\Delta}}_i^{\mathsf{T}}(\tilde{\boldsymbol{s}}(T) - \boldsymbol{s}(0)) + \frac{1}{2}(\tilde{\boldsymbol{s}}(T) - \boldsymbol{s}(0))^{\mathsf{T}}\bar{\boldsymbol{\Gamma}}_i(\tilde{\boldsymbol{s}}(T) - \boldsymbol{s}(0)),$$

where

$$\bar{\theta}_i = \partial_t v_i(\boldsymbol{s}(0), 0) \in \mathbb{R}, \quad \bar{\boldsymbol{\Delta}}_i = \nabla_{\boldsymbol{s}} v_i(\boldsymbol{s}(0), 0) \in \mathbb{R}^n, \quad \text{and} \quad \bar{\boldsymbol{\Gamma}}_i = \nabla^2_{\boldsymbol{s}} v_i(\boldsymbol{s}(0), 0) \in \mathbb{S}^n. \tag{4.31}$$

The values computed in (4.31) are referred to as the 'greeks' of the assets. We emphasize that the computation of the greeks relies on the availability of a pricing model, that is, the value functions v_i must be known. Note that the values of the functions v_i at $(\boldsymbol{s}(0), 0)$ can be observed in the market. However, the values of v_i in a neighborhood of $(\boldsymbol{s}(0), 0)$ are not observable. The proposed second-order Taylor approximation is very popular in finance and is often referred to as the *delta-gamma approximation*, see [Jas02].

By using the *relative greeks*

$$\theta_i = \frac{T}{v_i(\boldsymbol{s}(0), 0)}\bar{\theta}_i, \quad \boldsymbol{\Delta}_i = \frac{1}{v_i(\boldsymbol{s}(0), 0)} \operatorname{diag}(\boldsymbol{s}(0))\bar{\boldsymbol{\Delta}}_i, \quad \boldsymbol{\Gamma}_i = \frac{1}{v_i(\boldsymbol{s}(0), 0)} \operatorname{diag}(\boldsymbol{s}(0))^{\mathsf{T}}\bar{\boldsymbol{\Gamma}}_i \operatorname{diag}(\boldsymbol{s}(0)),$$

the delta-gamma approximation can be reformulated in terms of relative returns

$$\tilde{r}_i \;\approx\; f_i(\tilde{\boldsymbol{\xi}}) = \theta_i + \boldsymbol{\Delta}_i^{\mathsf{T}}\tilde{\boldsymbol{\xi}} + \frac{1}{2}\tilde{\boldsymbol{\xi}}^{\mathsf{T}}\boldsymbol{\Gamma}_i\tilde{\boldsymbol{\xi}} \quad \forall i = 1, \ldots, m. \tag{4.32}$$

Here we use the (possibly non-convex) quadratic functions f_i to map the basic asset returns $\tilde{\boldsymbol{\xi}}$ to the asset returns $\tilde{\boldsymbol{r}}$.

The return of a portfolio $\boldsymbol{w} \in \mathcal{W}$ can therefore be approximated by

$$\boldsymbol{w}^{\mathsf{T}}\tilde{\boldsymbol{r}} \;\approx\; \theta(\boldsymbol{w}) + \boldsymbol{\Delta}(\boldsymbol{w})^{\mathsf{T}}\tilde{\boldsymbol{\xi}} + \frac{1}{2}\tilde{\boldsymbol{\xi}}^{\mathsf{T}}\boldsymbol{\Gamma}(\boldsymbol{w})\tilde{\boldsymbol{\xi}}, \tag{4.33}$$

where we use the auxiliary functions

$$\theta(\boldsymbol{w}) = \sum_{i=1}^{m} w_i\theta_i, \quad \boldsymbol{\Delta}(\boldsymbol{w}) = \sum_{i=1}^{m} w_i\boldsymbol{\Delta}_i, \quad \text{and} \quad \boldsymbol{\Gamma}(\boldsymbol{w}) = \sum_{i=1}^{m} w_i\boldsymbol{\Gamma}_i,$$

which are all linear in $\boldsymbol{w}$. We emphasize that, in contrast to Section 4.4, we now allow for short-sales of derivatives.

In the remainder of this section we derive a Worst-Case VaR optimization model based on the quadratic approximation (4.33).

4.5.2 Worst-Case Quadratic VaR Model

We define the *Worst-Case Quadratic VaR* (WCQVaR) of a fixed portfolio $\boldsymbol{w} \in \mathcal{W}$ in terms of the Taylor expansion (4.33).

$$\begin{aligned}
\mathrm{WCQVaR}_\epsilon(\boldsymbol{w}) &= \min\left\{\gamma \;:\; \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left\{\gamma \le -\boldsymbol{w}^{\mathsf{T}} f(\tilde{\boldsymbol{\xi}})\right\} \le \epsilon\right\} \\
&= \min\left\{\gamma \;:\; \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left\{\gamma \le -\theta(\boldsymbol{w}) - \boldsymbol{\Delta}(\boldsymbol{w})^{\mathsf{T}}\tilde{\boldsymbol{\xi}} - \frac{1}{2}\tilde{\boldsymbol{\xi}}^{\mathsf{T}}\boldsymbol{\Gamma}(\boldsymbol{w})\tilde{\boldsymbol{\xi}}\right\} \le \epsilon\right\}
\end{aligned} \tag{4.34}$$

Note that the WCQVaR approximates the portfolio return $\boldsymbol{w}^{\mathsf{T}}\tilde{\boldsymbol{r}}$ by a (possibly non-convex) quadratic function of the basic asset returns $\tilde{\boldsymbol{\xi}}$.

Theorem 4.5.1 below shows how the WCQVaR of a portfolio $\boldsymbol{w}$ can be computed by solving a tractable SDP. We first recall the $\mathcal{S}$-lemma, which is a crucial ingredient for the proof of Theorem 4.5.1. We refer to Pólik and Terlaky [PT07] for a derivation and an in-depth survey of its manifold uses.

Lemma 4.5.1 ($\mathcal{S}$-lemma) *Let $f_i(\boldsymbol{\xi}) = \boldsymbol{\xi}^\mathsf{T}\mathbf{A}_i\boldsymbol{\xi}$, $i = 0, \ldots, p$ be quadratic functions of $\boldsymbol{\xi} \in \mathbb{R}^n$. Then, $f_0(\boldsymbol{\xi}) \geq 0$ for all $\boldsymbol{\xi}$ with $f_i(\boldsymbol{\xi}) \leq 0$, $i = 1, \ldots, p$, if there exist constants $\tau_i \geq 0$ such that*

$$\mathbf{A}_0 + \sum_{i=1}^{p} \tau_i \mathbf{A}_i \succcurlyeq \mathbf{0}.$$

For $p = 1$, the converse implication holds if there exists a strictly feasible point $\bar{\boldsymbol{\xi}}$ with $f_1(\bar{\boldsymbol{\xi}}) < 0$.

Theorem 4.5.1 *The* WCQVaR *of a fixed portfolio $\boldsymbol{w} \in \mathcal{W}$ can be computed by solving the following tractable SDP.*

$$\begin{aligned} \mathrm{WCQVaR}_\epsilon(\boldsymbol{w}) = \inf \quad & \gamma \\ \text{s. t.} \quad & \mathbf{M} \in \mathbb{S}^{n+1}, \quad \tau \in \mathbb{R}, \quad \gamma \in \mathbb{R} \\ & \langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq \tau\epsilon, \quad \mathbf{M} \succcurlyeq \mathbf{0}, \quad \tau \geq 0, \\ & \mathbf{M} + \begin{bmatrix} \boldsymbol{\Gamma}(\boldsymbol{w}) & \boldsymbol{\Delta}(\boldsymbol{w}) \\ \boldsymbol{\Delta}(\boldsymbol{w})^\mathsf{T} & -\tau + 2(\gamma + \theta(\boldsymbol{w})) \end{bmatrix} \succcurlyeq \mathbf{0} \end{aligned} \tag{4.35}$$

Proof For the given portfolio $\boldsymbol{w} \in \mathcal{W}$ and for any fixed $\gamma \in \mathbb{R}$, we introduce the set $\mathcal{Q}_\gamma \subseteq \mathbb{R}^n$, defined through

$$\mathcal{Q}_\gamma = \left\{ \boldsymbol{\xi} \in \mathbb{R}^n \ : \ \gamma \leq -\theta(\boldsymbol{w}) - \boldsymbol{\Delta}(\boldsymbol{w})^\mathsf{T}\boldsymbol{\xi} - \frac{1}{2}\boldsymbol{\xi}^\mathsf{T}\boldsymbol{\Gamma}(\boldsymbol{w})\boldsymbol{\xi} \right\}. \tag{4.36}$$

As in Section 4.4, the first step towards a tractable reformulation of WCQVaR is to solve the worst-case probability problem

$$\pi_{\mathrm{wc}} = \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\{\tilde{\boldsymbol{\xi}} \in \mathcal{Q}_\gamma\}, \tag{4.37}$$

which can be identified as the subordinate maximization problem in (4.34). Lemma 4.4.1

implies that (4.37) can equivalently be formulated as

$$\pi_{\mathrm{wc}} = \inf_{\mathbf{M}\in\mathbb{S}^{n+1}} \left\{ \langle \mathbf{\Gamma}, \mathbf{M} \rangle \ : \ \mathbf{M} \succcurlyeq \mathbf{0}, \quad \left[\boldsymbol{\xi}^{\mathsf{T}}\ 1\right] \mathbf{M} \left[\boldsymbol{\xi}^{\mathsf{T}}\ 1\right]^{\mathsf{T}} \geq 1 \quad \forall \boldsymbol{\xi} \in \mathcal{Q}_{\gamma} \right\}. \tag{4.38}$$

By the definition of $\mathcal{Q}$, the semi-infinite constraint in problem (4.38) is equivalent to

$$\left[\boldsymbol{\xi}^{\mathsf{T}}\ 1\right] (\mathbf{M} - \mathrm{diag}(\mathbf{0}, 1)) \left[\boldsymbol{\xi}^{\mathsf{T}}\ 1\right]^{\mathsf{T}} \geq 0 \quad \forall \boldsymbol{\xi} : \left[\boldsymbol{\xi}^{\mathsf{T}}\ 1\right] \begin{bmatrix} \frac{1}{2}\mathbf{\Gamma}(\boldsymbol{w}) & \frac{1}{2}\mathbf{\Delta}(\boldsymbol{w}) \\ \frac{1}{2}\mathbf{\Delta}(\boldsymbol{w})^{\mathsf{T}} & \gamma + \theta(\boldsymbol{w}) \end{bmatrix} \left[\boldsymbol{\xi}^{\mathsf{T}}\ 1\right]^{\mathsf{T}} \leq 0.$$

By using the $\mathcal{S}$-lemma and by analogous reasoning as in Section 4.4.2, we can replace the semi-infinite constraint in problem (4.38) by

$$\exists \tau \geq 0 \ : \ \mathbf{M} + \begin{bmatrix} \tau\mathbf{\Gamma}(\boldsymbol{w}) & \tau\mathbf{\Delta}(\boldsymbol{w}) \\ \tau\mathbf{\Delta}(\boldsymbol{w})^{\mathsf{T}} & -1 + 2\tau(\gamma + \theta(\boldsymbol{w})) \end{bmatrix} \succcurlyeq \mathbf{0}$$

without changing the optimal value of the problem. Thus, the worst-case probability problem (4.37) can be rewritten as

$$\begin{aligned} \pi_{\mathrm{wc}} = \inf \quad & \langle \mathbf{\Omega}, \mathbf{M} \rangle \\ \mathrm{s.\,t.} \quad & \mathbf{M} \in \mathbb{S}^{n+1}, \quad \tau \in \mathbb{R}, \quad \mathbf{M} \succcurlyeq \mathbf{0}, \quad \tau \geq 0 \\ & \mathbf{M} + \begin{bmatrix} \tau\mathbf{\Gamma}(\boldsymbol{w}) & \tau\mathbf{\Delta}(\boldsymbol{w}) \\ \tau\mathbf{\Delta}(\boldsymbol{w})^{\mathsf{T}} & -1 + 2\tau(\gamma + \theta(\boldsymbol{w})) \end{bmatrix} \succcurlyeq \mathbf{0}. \end{aligned} \tag{4.39}$$

The WCQVaR of the portfolio $\boldsymbol{w}$ can therefore be obtained by solving the following non-convex optimization problem.

$$\begin{aligned} \mathrm{WCQVaR}_{\epsilon}(\boldsymbol{w}) = \inf \quad & \gamma \\ \mathrm{s.\,t.} \quad & \mathbf{M} \in \mathbb{S}^{n+1}, \quad \tau \in \mathbb{R}, \quad \gamma \in \mathbb{R} \\ & \langle \mathbf{\Omega}, \mathbf{M} \rangle \leq \epsilon, \quad \mathbf{M} \succcurlyeq \mathbf{0}, \quad \tau \geq 0 \\ & \mathbf{M} + \begin{bmatrix} \tau\mathbf{\Gamma}(\boldsymbol{w}) & \tau\mathbf{\Delta}(\boldsymbol{w}) \\ \tau\mathbf{\Delta}(\boldsymbol{w})^{\mathsf{T}} & -1 + 2\tau(\gamma + \theta(\boldsymbol{w})) \end{bmatrix} \succcurlyeq \mathbf{0} \end{aligned} \tag{4.40}$$

By analogous reasoning as in Section 4.4.2, it can be shown that any feasible solution of problem (4.40) has a strictly positive τ-component. Thus we may divide the matrix inequality in (4.40) by τ. After the variable transformation $\tau \to 1/\tau$ and $\mathbf{M} \to \mathbf{M}/\tau$, we obtain the postulated SDP (4.35). ∎

Remark 4.5.1 *In the absence of derivatives, that is, if the market only contains basic assets, then $m = n$, and the coefficient functions in the delta-gamma approximation* (4.33) *reduce to $\theta(\boldsymbol{w}) = 0$, $\boldsymbol{\Delta}(\boldsymbol{w}) = \boldsymbol{w}$, and $\boldsymbol{\Gamma}(\boldsymbol{w}) = \mathbf{0}$. In this special case, the* WCQVaR *is computed by solving the following SDP.*

$$\begin{aligned} \text{WCQVaR}_\epsilon(\boldsymbol{w}) = \inf \quad & \gamma \\ \text{s.t.} \quad & \mathbf{M} \in \mathbb{S}^{n+1}, \quad \tau \in \mathbb{R}, \quad \gamma \in \mathbb{R} \\ & \langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq \tau\epsilon, \quad \mathbf{M} \succcurlyeq \mathbf{0}, \quad \tau \geq 0 \\ & \mathbf{M} + \begin{bmatrix} \mathbf{0} & \boldsymbol{w} \\ \boldsymbol{w}^\mathsf{T} & -\tau + 2\gamma \end{bmatrix} \succcurlyeq \mathbf{0} \end{aligned}$$

El Ghaoui et al. *[EGOO03] have shown (using similar arguments as in Theorem 4.4.1) that this SDP has the closed form solution*

$$\text{WCVaR}(\boldsymbol{w}) = -\boldsymbol{\mu}^\mathsf{T}\boldsymbol{w} + \kappa(\epsilon)\sqrt{\boldsymbol{w}^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{w}}, \quad \textit{where} \quad \kappa(\epsilon) = \sqrt{\frac{1-\epsilon}{\epsilon}}.$$

Thus, the WCQVaR *model is a direct extension of the* WCVaR *model* (4.7)*.*

Problem (4.35) constitutes a convex SDP that facilitates the efficient computation of the WCQVaR for any fixed portfolio $\boldsymbol{w} \in \mathcal{W}$. Since the matrix inequality in (4.35) is linear in $(\mathbf{M}, \boldsymbol{\tau}, \gamma)$ and $\boldsymbol{w}$, one can reinterpret $\boldsymbol{w}$ as a decision variable without impairing the problem's convexity. This observation reveals that we can efficiently minimize the WCQVaR over all portfolios $\boldsymbol{w} \in \mathcal{W}$ by solving the following SDP.

$$
\begin{array}{ll}
\inf & \gamma \\
\text{s.t.} & \mathbf{M} \in \mathbb{S}^{n+1}, \quad \tau \in \mathbb{R}, \quad \gamma \in \mathbb{R}, \quad \boldsymbol{w} \in \mathcal{W} \\
& \langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq \tau \epsilon, \quad \mathbf{M} \succcurlyeq \mathbf{0}, \quad \tau \geq 0 \\
& \mathbf{M} + \begin{bmatrix} \boldsymbol{\Gamma}(\boldsymbol{w}) & \boldsymbol{\Delta}(\boldsymbol{w}) \\ \boldsymbol{\Delta}(\boldsymbol{w})^\mathsf{T} & -\tau + 2(\gamma + \theta(\boldsymbol{w})) \end{bmatrix} \succcurlyeq \mathbf{0}
\end{array}
\tag{4.41}
$$

Remark 4.5.2 *Unlike in Section 4.4, there seems to be no equivalent SOCP formulation for the SDP* (4.41)*. In particular, there is no simple way to adapt the arguments in the proof of Theorem 4.4.1 to the current setting. The reason for this is a fundamental difference between the corresponding SDP problems* (4.26) *and* (4.41)*. In fact, the top left principal submatrix in the last LMI constraint is independent of* $\boldsymbol{w}$ *in* (4.26) *but not in* (4.41)*.*

4.5.3 Robust Optimization Perspective on WCQVaR

We now highlight the close connection between robust optimization and WCQVaR minimization. In the next theorem we elaborate an equivalence between the WCQVaR minimization problem and a robust optimization problem whose uncertainty set is embedded into a space of positive semidefinite matrices.

Theorem 4.5.2 *The* WCQVaR *minimization problem* (4.41) *is equivalent to the robust optimization problem*

$$
\begin{array}{lll}
\underset{\boldsymbol{w} \in \mathbb{R}^m, \gamma \in \mathbb{R}}{\text{minimize}} & \gamma & \\
\text{subject to} & -\langle \mathbf{Q}(\boldsymbol{w}), \mathbf{Z} \rangle \leq \gamma & \forall \mathbf{Z} \in \mathcal{U}_\epsilon^q \\
& \boldsymbol{w} \in \mathcal{W}, &
\end{array}
\tag{4.42}
$$

where

$$
\mathbf{Q}(\boldsymbol{w}) = \begin{bmatrix} \frac{1}{2}\boldsymbol{\Gamma}(\boldsymbol{w}) & \frac{1}{2}\boldsymbol{\Delta}(\boldsymbol{w}) \\ \frac{1}{2}\boldsymbol{\Delta}(\boldsymbol{w})^\mathsf{T} & \theta(\boldsymbol{w}) \end{bmatrix},
$$

and the uncertainty set $\mathcal{U}_\epsilon^q \subseteq \mathbb{S}^{n+1}$ *is defined as*

$$\mathcal{U}_\epsilon^q = \left\{ \mathbf{Z} = \begin{bmatrix} \mathbf{X} & \boldsymbol{\xi} \\ \boldsymbol{\xi}^\mathsf{T} & 1 \end{bmatrix} \in \mathbb{S}^{n+1} \;:\; \boldsymbol{\Omega} - \epsilon \mathbf{Z} \succcurlyeq \mathbf{0},\; \mathbf{Z} \succcurlyeq \mathbf{0} \right\}. \tag{4.43}$$

Proof For some fixed portfolio $\boldsymbol{w} \in \mathcal{W}$, the WCQVaR can be computed by solving problem (4.35), which involves the LMI constraint

$$\mathbf{M} + \begin{bmatrix} \boldsymbol{\Gamma}(\boldsymbol{w}) & \boldsymbol{\Delta}(\boldsymbol{w}) \\ \boldsymbol{\Delta}(\boldsymbol{w})^\mathsf{T} & -\tau + 2(\gamma + \theta(\boldsymbol{w})) \end{bmatrix} \succcurlyeq \mathbf{0}. \tag{4.44}$$

Without loss of generality, we can rewrite the matrix $\mathbf{M}$ as

$$\mathbf{M} = \begin{bmatrix} \mathbf{V} & \boldsymbol{v} \\ \boldsymbol{v}^\mathsf{T} & u \end{bmatrix}.$$

With this new notation, the LMI constraint (4.44) is representable as

$$\begin{aligned}
& [\boldsymbol{\xi}^\mathsf{T}\ 1] \begin{bmatrix} \mathbf{V} + \boldsymbol{\Gamma}(\boldsymbol{w}) & \boldsymbol{v} + \boldsymbol{\Delta}(\boldsymbol{w}) \\ (\boldsymbol{v} + \boldsymbol{\Delta}(\boldsymbol{w}))^\mathsf{T} & u - \tau + 2(\gamma + \theta(\boldsymbol{w})) \end{bmatrix} [\boldsymbol{\xi}^\mathsf{T}\ 1]^\mathsf{T} \geq 0 && \forall \boldsymbol{\xi} \in \mathbb{R}^n \\
\iff\ & \boldsymbol{\xi}^\mathsf{T}(\mathbf{V} + \boldsymbol{\Gamma}(\boldsymbol{w}))\boldsymbol{\xi} + 2\boldsymbol{\xi}^\mathsf{T}(\boldsymbol{v} + \boldsymbol{\Delta}(\boldsymbol{w})) + u - \tau + 2(\gamma + \theta(\boldsymbol{w})) \geq 0 && \forall \boldsymbol{\xi} \in \mathbb{R}^n \\
\iff\ & \gamma \geq -\frac{1}{2}\boldsymbol{\xi}^\mathsf{T}(\mathbf{V} + \boldsymbol{\Gamma}(\boldsymbol{w}))\boldsymbol{\xi} - \boldsymbol{\xi}^\mathsf{T}(\boldsymbol{v} + \boldsymbol{\Delta}(\boldsymbol{w})) - \theta(\boldsymbol{w}) - \frac{1}{2}(u - \tau) && \forall \boldsymbol{\xi} \in \mathbb{R}^n \\
\iff\ & \gamma \geq \sup_{\boldsymbol{\xi} \in \mathbb{R}^n} \left\{ -\frac{1}{2}\boldsymbol{\xi}^\mathsf{T}(\mathbf{V} + \boldsymbol{\Gamma}(\boldsymbol{w}))\boldsymbol{\xi} - \boldsymbol{\xi}^\mathsf{T}(\boldsymbol{v} + \boldsymbol{\Delta}(\boldsymbol{w})) - \theta(\boldsymbol{w}) - \frac{1}{2}(u - \tau) \right\}.
\end{aligned}$$

Thus, the WCQVaR problem (4.35) can be rewritten as

$$\begin{array}{ll}
\inf & \displaystyle\sup_{\boldsymbol{\xi} \in \mathbb{R}^n}\ -\frac{1}{2}\boldsymbol{\xi}^\mathsf{T}(\mathbf{V} + \boldsymbol{\Gamma}(\boldsymbol{w}))\boldsymbol{\xi} - \boldsymbol{\xi}^\mathsf{T}(\boldsymbol{v} + \boldsymbol{\Delta}(\boldsymbol{w})) - \theta(\boldsymbol{w}) - \frac{1}{2}(u - \tau) \\
\text{s.t.} & \mathbf{V} \in \mathbb{S}^n, \quad \boldsymbol{v} \in \mathbb{R}^n, \quad \tau \in \mathbb{R}, \quad u \in \mathbb{R} \\
& \begin{bmatrix} \mathbf{V} & \boldsymbol{v} \\ \boldsymbol{v}^\mathsf{T} & u \end{bmatrix} \succcurlyeq \mathbf{0}, \quad \tau \geq 0, \quad \langle \mathbf{V}, \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\mathsf{T} \rangle + 2\boldsymbol{v}^\mathsf{T}\boldsymbol{\mu} + u \leq \tau\epsilon.
\end{array} \tag{4.45}$$

Note that if $\mathbf{V} + \boldsymbol{\Gamma}(\boldsymbol{w})$ is not positive semidefinite, the inner maximization problem in (4.45) is unbounded. However, this implies that any $\mathbf{V} \in \mathbb{S}^n$ is infeasible in the outer minimization problem unless $\mathbf{V} + \boldsymbol{\Gamma}(\boldsymbol{w}) \succcurlyeq \mathbf{0}$. Therefore, we can add the constraint $\mathbf{V} + \boldsymbol{\Gamma}(\boldsymbol{w}) \succcurlyeq \mathbf{0}$ to the minimization problem in (4.45) without changing its feasible region. With this constraint appended, the min-max problem (4.45) becomes a saddlepoint problem because its objective is concave in $\boldsymbol{\xi}$ for any fixed $(\mathbf{V}, \boldsymbol{v}, u, \tau)$ and convex in $(\mathbf{V}, \boldsymbol{v}, u, \tau)$ for any fixed $\boldsymbol{\xi}$. Moreover, the feasible sets of the outer and inner problems are convex and independent of each other. Thus, we may interchange the 'inf' and 'sup' operators to obtain the following equivalent problem, see, e.g., [DM74, Theorem 5.1].

$$
\begin{aligned}
\max_{\boldsymbol{\xi} \in \mathbb{R}^n} \ \min \quad & -\frac{1}{2}\boldsymbol{\xi}^\mathsf{T}(\mathbf{V} + \boldsymbol{\Gamma}(\boldsymbol{w}))\boldsymbol{\xi} - \boldsymbol{\xi}^\mathsf{T}(\boldsymbol{v} + \boldsymbol{\Delta}(\boldsymbol{w})) - \theta(\boldsymbol{w}) - \frac{1}{2}(u - \tau) \\
\text{s.t.} \quad & \mathbf{V} \in \mathbb{S}^n, \quad \boldsymbol{v} \in \mathbb{R}^n, \quad \tau \in \mathbb{R}, \quad u \in \mathbb{R} \\
& \begin{bmatrix} \mathbf{V} & \boldsymbol{v} \\ \boldsymbol{v}^\mathsf{T} & u \end{bmatrix} \succcurlyeq \mathbf{0}, \quad \tau \geq 0, \quad \langle \mathbf{V}, \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\mathsf{T} \rangle + 2\boldsymbol{v}^\mathsf{T}\boldsymbol{\mu} + u \leq \tau\epsilon.
\end{aligned}
\tag{4.46}
$$

We proceed by dualizing the inner minimization problem in (4.46). After a few elementary simplification steps, this dual problem reduces to

$$
\begin{aligned}
\max \quad & -\frac{1}{2}\langle \boldsymbol{\Gamma}(\boldsymbol{w}), \boldsymbol{\xi}\boldsymbol{\xi}^\mathsf{T} + \mathbf{Y} \rangle - \boldsymbol{\xi}^\mathsf{T}\boldsymbol{\Delta}(\boldsymbol{w}) - \theta(\boldsymbol{w}) \\
\text{s.t.} \quad & \mathbf{Y} \in \mathbb{S}^n, \quad \alpha \in \mathbb{R}, \quad \mathbf{Y} \succcurlyeq \mathbf{0}, \quad 1 \leq \alpha \leq \frac{1}{\epsilon} \\
& \begin{bmatrix} \alpha(\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\mathsf{T}) - (\boldsymbol{\xi}\boldsymbol{\xi}^\mathsf{T} + \mathbf{Y}) & \alpha\boldsymbol{\mu} - \boldsymbol{\xi} \\ (\alpha\boldsymbol{\mu} - \boldsymbol{\xi})^\mathsf{T} & \alpha - 1 \end{bmatrix} \succcurlyeq \mathbf{0}.
\end{aligned}
\tag{4.47}
$$

Note that strong duality holds because the inner problem in (4.46) is strictly feasible for any $\epsilon > 0$, see [VB96]. This allows us to replace the inner minimization problem in (4.46) by the maximization problem (4.47), which yields the following equivalent formulation for the

WCQVaR problem (4.35).

$$
\begin{aligned}
\max \quad & -\frac{1}{2}\langle \mathbf{\Gamma}(\boldsymbol{w}), \boldsymbol{\xi}\boldsymbol{\xi}^\mathsf{T} + \mathbf{Y}\rangle - \boldsymbol{\xi}^\mathsf{T}\mathbf{\Delta}(\boldsymbol{w}) - \theta(\boldsymbol{w}) \\
\text{s. t.} \quad & \mathbf{Y} \in \mathbb{S}^n, \quad \boldsymbol{\xi} \in \mathbb{R}^n, \quad \alpha \in \mathbb{R}, \quad \mathbf{Y} \succcurlyeq \mathbf{0}, \quad 1 \le \alpha \le \frac{1}{\epsilon} \\
& \begin{bmatrix} \alpha(\mathbf{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\mathsf{T}) - (\boldsymbol{\xi}\boldsymbol{\xi}^\mathsf{T} + \mathbf{Y}) & \alpha\boldsymbol{\mu} - \boldsymbol{\xi} \\ (\alpha\boldsymbol{\mu} - \boldsymbol{\xi})^\mathsf{T} & \alpha - 1 \end{bmatrix} \succcurlyeq \mathbf{0}
\end{aligned}
$$

We now introduce a new decision variable $\mathbf{X} = \boldsymbol{\xi}\boldsymbol{\xi}^\mathsf{T} + \mathbf{Y}$, which allows us to reformulate the above problem as

$$
\begin{aligned}
\max \quad & -\frac{1}{2}\langle \mathbf{\Gamma}(\boldsymbol{w}), \mathbf{X}\rangle - \boldsymbol{\xi}^\mathsf{T}\mathbf{\Delta}(\boldsymbol{w}) - \theta(\boldsymbol{w}) \\
\text{s. t.} \quad & \mathbf{X} \in \mathbb{S}^n, \quad \boldsymbol{\xi} \in \mathbb{R}^n, \quad \alpha \in \mathbb{R}, \quad 1 \le \alpha \le \frac{1}{\epsilon} \\
& \begin{bmatrix} \alpha(\mathbf{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\mathsf{T}) - \mathbf{X} & \alpha\boldsymbol{\mu} - \boldsymbol{\xi} \\ (\alpha\boldsymbol{\mu} - \boldsymbol{\xi})^\mathsf{T} & \alpha - 1 \end{bmatrix} \succcurlyeq \mathbf{0}, \quad \mathbf{X} - \boldsymbol{\xi}\boldsymbol{\xi}^\mathsf{T} \succcurlyeq \mathbf{0}.
\end{aligned}
$$

By definition of $\mathbf{\Omega}$ as the second-order moment matrix of the basic asset returns, see (4.21), the first LMI constraint in the above problem can be rewritten as

$$
\alpha\mathbf{\Omega} - \begin{bmatrix} \mathbf{X} & \boldsymbol{\xi} \\ \boldsymbol{\xi}^\mathsf{T} & 1 \end{bmatrix} \succcurlyeq \mathbf{0}.
$$

Furthermore, by using Schur complements, the following equivalence holds.

$$
\mathbf{X} - \boldsymbol{\xi}\boldsymbol{\xi}^\mathsf{T} \succcurlyeq \mathbf{0} \iff \begin{bmatrix} \mathbf{X} & \boldsymbol{\xi} \\ \boldsymbol{\xi}^\mathsf{T} & 1 \end{bmatrix} \succcurlyeq \mathbf{0}
$$

Therefore, problem (4.47) can be reformulated as

$$\begin{array}{ll} \max & -\left\langle \begin{bmatrix} \frac{1}{2}\mathbf{\Gamma}(\boldsymbol{w}) & \frac{1}{2}\mathbf{\Delta}(\boldsymbol{w}) \\ \frac{1}{2}\mathbf{\Delta}(\boldsymbol{w})^\mathsf{T} & \theta(\boldsymbol{w}) \end{bmatrix}, \begin{bmatrix} \mathbf{X} & \boldsymbol{\xi} \\ \boldsymbol{\xi}^\mathsf{T} & 1 \end{bmatrix} \right\rangle \\ \text{s. t.} & \mathbf{X} \in \mathbb{S}^n, \quad \boldsymbol{\xi} \in \mathbb{R}^n, \quad \alpha \in \mathbb{R}, \quad 1 \leq \alpha \leq \dfrac{1}{\epsilon} \\ & \alpha\mathbf{\Omega} - \begin{bmatrix} \mathbf{X} & \boldsymbol{\xi} \\ \boldsymbol{\xi}^\mathsf{T} & 1 \end{bmatrix} \succcurlyeq \mathbf{0}, \quad \begin{bmatrix} \mathbf{X} & \boldsymbol{\xi} \\ \boldsymbol{\xi}^\mathsf{T} & 1 \end{bmatrix} \succcurlyeq \mathbf{0}. \end{array}$$

Since the objective function is independent of α and $\mathbf{\Omega} \succ \mathbf{0}$, the optimal choice for α is $1/\epsilon$; in fact, this choice of α generates the largest feasible set. We conclude that the WCQVaR for a fixed portfolio $\boldsymbol{w}$ can be computed by solving the following problem.

$$\begin{array}{ll} \max & -\left\langle \begin{bmatrix} \frac{1}{2}\mathbf{\Gamma}(\boldsymbol{w}) & \frac{1}{2}\mathbf{\Delta}(\boldsymbol{w}) \\ \frac{1}{2}\mathbf{\Delta}(\boldsymbol{w})^\mathsf{T} & \theta(\boldsymbol{w}) \end{bmatrix}, \begin{bmatrix} \mathbf{X} & \boldsymbol{\xi} \\ \boldsymbol{\xi}^\mathsf{T} & 1 \end{bmatrix} \right\rangle \\ \text{s. t.} & \mathbf{X} \in \mathbb{S}^n, \quad \boldsymbol{\xi} \in \mathbb{R}^n, \quad \mathbf{\Omega} - \epsilon \begin{bmatrix} \mathbf{X} & \boldsymbol{\xi} \\ \boldsymbol{\xi}^\mathsf{T} & 1 \end{bmatrix} \succcurlyeq \mathbf{0}, \quad \begin{bmatrix} \mathbf{X} & \boldsymbol{\xi} \\ \boldsymbol{\xi}^\mathsf{T} & 1 \end{bmatrix} \succcurlyeq \mathbf{0} \end{array}$$

The WCQVaR minimization problem (4.41) can therefore be expressed as the min-max problem

$$\min_{\boldsymbol{w} \in \mathcal{W}} \max_{\mathbf{Z} \in \mathcal{U}_\epsilon^q} -\langle \mathbf{Q}(\boldsymbol{w}), \mathbf{Z} \rangle, \tag{4.48}$$

which is manifestly equivalent to the postulated semi-infinite program (4.42). ∎

It may not be evident how the uncertainty set $\mathcal{U}_\epsilon^q$ (defined in (4.43)) associated with the WCQVaR formulation is related to the ellipsoidal uncertainty set $\mathcal{U}_\epsilon$ defined in Section 4.2.2. We now demonstrate that there exists a strong connection between these two uncertainty sets, even though they are embedded in spaces of different dimensions.

Corollary 4.5.1 *If the constraint $\mathbf{\Gamma}(\boldsymbol{w}) \succcurlyeq \mathbf{0}$ is appended to the definition of the set $\mathcal{W}$ of* *admissible portfolios, then the robust optimization problem* (4.42) *reduces to*

$$\begin{aligned} \underset{\boldsymbol{w}\in\mathbb{R}^m,\gamma\in\mathbb{R}}{\text{minimize}} \quad & \gamma \\ \text{subject to} \quad & -\theta(\boldsymbol{w}) - \boldsymbol{\Delta}(\boldsymbol{w})^\mathsf{T}\boldsymbol{\xi} - \frac{1}{2}\boldsymbol{\xi}^\mathsf{T}\boldsymbol{\Gamma}(\boldsymbol{w})\boldsymbol{\xi} \le \gamma \qquad \forall \boldsymbol{\xi} \in \mathcal{U}_\epsilon \\ & \boldsymbol{w} \in \mathcal{W}, \end{aligned} \tag{4.49}$$

where $\mathcal{U}_\epsilon$ *is the ellipsoidal uncertainty set defined in Section 4.2.2.*

Proof The inner maximization problem in (4.48) can be written as

$$\begin{aligned} \max \quad & -\theta(\boldsymbol{w}) - \boldsymbol{\Delta}(\boldsymbol{w})^\mathsf{T}\boldsymbol{\xi} - \frac{1}{2}\langle \boldsymbol{\Gamma}(\boldsymbol{w}), \mathbf{X}\rangle \\ \text{s. t.} \quad & \mathbf{X} \in \mathbb{S}^n, \quad \boldsymbol{\xi} \in \mathbb{R}^n, \quad \mathbf{X} - \boldsymbol{\xi}\boldsymbol{\xi}^\mathsf{T} \succcurlyeq \mathbf{0} \\ & \begin{bmatrix} (\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\mathsf{T}) - \epsilon\mathbf{X} & \boldsymbol{\mu} - \epsilon\boldsymbol{\xi} \\ (\boldsymbol{\mu} - \epsilon\boldsymbol{\xi})^\mathsf{T} & 1 - \epsilon \end{bmatrix} \succcurlyeq \mathbf{0}. \end{aligned}$$

By introducing the decision variable $\mathbf{Y} = \mathbf{X} - \boldsymbol{\xi}\boldsymbol{\xi}^\mathsf{T}$ as in the proof of Theorem 4.5.2, the above problem can be reformulated as

$$\begin{aligned} \max \quad & -\theta(\boldsymbol{w}) - \boldsymbol{\Delta}(\boldsymbol{w})^\mathsf{T}\boldsymbol{\xi} - \frac{1}{2}\boldsymbol{\xi}^\mathsf{T}\boldsymbol{\Gamma}(\boldsymbol{w})\boldsymbol{\xi} - \frac{1}{2}\langle \boldsymbol{\Gamma}(\boldsymbol{w}), \mathbf{Y}\rangle \\ \text{s. t.} \quad & \mathbf{Y} \in \mathbb{S}^n, \quad \boldsymbol{\xi} \in \mathbb{R}^n, \quad \mathbf{Y} \succcurlyeq \mathbf{0} \\ & \begin{bmatrix} (\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\mathsf{T}) - \epsilon(\mathbf{Y} + \boldsymbol{\xi}\boldsymbol{\xi}^\mathsf{T}) & \boldsymbol{\mu} - \epsilon\boldsymbol{\xi} \\ (\boldsymbol{\mu} - \epsilon\boldsymbol{\xi})^\mathsf{T} & 1 - \epsilon \end{bmatrix} \succcurlyeq \mathbf{0}. \end{aligned} \tag{4.50}$$

We will now argue that $\mathbf{Y} = \mathbf{0}$ at optimality. This holds due to the following two facts: (i) for $\mathbf{Y} = \mathbf{0}$ we obtain the largest feasible set, and (ii) we have $\langle \boldsymbol{\Gamma}(\boldsymbol{w}), \mathbf{Y}\rangle \ge 0$ for all $\mathbf{Y} \succcurlyeq \mathbf{0}$ because $\boldsymbol{\Gamma}(\boldsymbol{w}) \succcurlyeq \mathbf{0}$ by assumption. Thus problem (4.50) reduces to

$$\begin{aligned} \max_{\boldsymbol{\xi}\in\mathbb{R}^n} \quad & -\theta(\boldsymbol{w}) - \boldsymbol{\Delta}(\boldsymbol{w})^\mathsf{T}\boldsymbol{\xi} - \frac{1}{2}\boldsymbol{\xi}^\mathsf{T}\boldsymbol{\Gamma}(\boldsymbol{w})\boldsymbol{\xi} \\ \text{s. t.} \quad & \begin{bmatrix} (\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\mathsf{T}) - \epsilon\boldsymbol{\xi}\boldsymbol{\xi}^\mathsf{T} & \boldsymbol{\mu} - \epsilon\boldsymbol{\xi} \\ (\boldsymbol{\mu} - \epsilon\boldsymbol{\xi})^\mathsf{T} & 1 - \epsilon \end{bmatrix} \succcurlyeq \mathbf{0}. \end{aligned}$$

Using similar arguments as in Theorem 4.4.1 (in particular, see (4.8.2)), we can show that the semidefinite constraint in the above problem is equivalent to

$$\begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{\xi}-\boldsymbol{\mu} \\ (\boldsymbol{\xi}-\boldsymbol{\mu})^\mathsf{T} & \kappa(\epsilon)^2 \end{bmatrix} \succcurlyeq \mathbf{0} \iff (\boldsymbol{\xi}-\boldsymbol{\mu})^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\xi}-\boldsymbol{\mu}) \leq \kappa(\epsilon)^2.$$

Thus the original min-max formulation (4.48) can be reexpressed as

$$\min_{\boldsymbol{w}\in\mathcal{W}} \max_{\boldsymbol{\xi}\in\mathcal{U}_\epsilon} -\theta(\boldsymbol{w}) - \boldsymbol{\Delta}(\boldsymbol{w})^\mathsf{T}\boldsymbol{\xi} - \frac{1}{2}\boldsymbol{\xi}^\mathsf{T}\boldsymbol{\Gamma}(\boldsymbol{w})\boldsymbol{\xi},$$

which is equivalent to the postulated robust optimization problem. ∎

Remark 4.5.3 *Note that the robust optimization problem* (4.49) *can be reformulated as*

$$\begin{aligned} &\underset{\boldsymbol{w}\in\mathbb{R}^m,\gamma\in\mathbb{R}}{\text{minimize}} && \gamma \\ &\text{subject to} && -\boldsymbol{w}^\mathsf{T}\boldsymbol{r} \leq \gamma \qquad \forall \boldsymbol{r}\in\mathcal{U}_\epsilon^{q2} \\ &&& \boldsymbol{w}\in\mathcal{W}, \end{aligned} \tag{4.51}$$

where the uncertainty set $\mathcal{U}_\epsilon^{q2}$ *is defined as*

$$\mathcal{U}_\epsilon^{q2} = \left\{ \boldsymbol{r}\in\mathbb{R}^m : \begin{array}{l} \exists\boldsymbol{\xi}\in\mathbb{R}^n \quad \textit{such that} \\ (\boldsymbol{\xi}-\boldsymbol{\mu})^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\xi}-\boldsymbol{\mu}) \leq \kappa(\epsilon)^2 \quad \textit{and} \\ r_i = \theta_i + \boldsymbol{\xi}^\mathsf{T}\boldsymbol{\Delta}_i + \frac{1}{2}\boldsymbol{\xi}^\mathsf{T}\boldsymbol{\Gamma}_i\boldsymbol{\xi} \quad \forall i=1,\dots,m \end{array} \right\}$$

In contrast to the simple ellipsoidal set $\mathcal{U}_\epsilon$*, the set* $\mathcal{U}_\epsilon^{q2}$ *is asymmetrically oriented around* $\boldsymbol{\mu}$*. This asymmetry is caused by the quadratic functions that map the basic asset returns* $\boldsymbol{\xi}$ *to the asset returns* $\boldsymbol{r}$*, see Figure 4.3. As a result, the* WCQVaR *model may provide a tighter approximation of the actual VaR of a portfolio containing derivatives than the* WCVaR *model.*

It seems that a min-max formulation (4.51) with an uncertainty set embedded into $\mathbb{R}^m$ is only available if $\boldsymbol{\Gamma}(\boldsymbol{w}) \succcurlyeq \mathbf{0}$, that is, if the portfolio return is a convex quadratic function of the basic assets returns. In general, however, one needs to resort to the more general formulation (4.42),

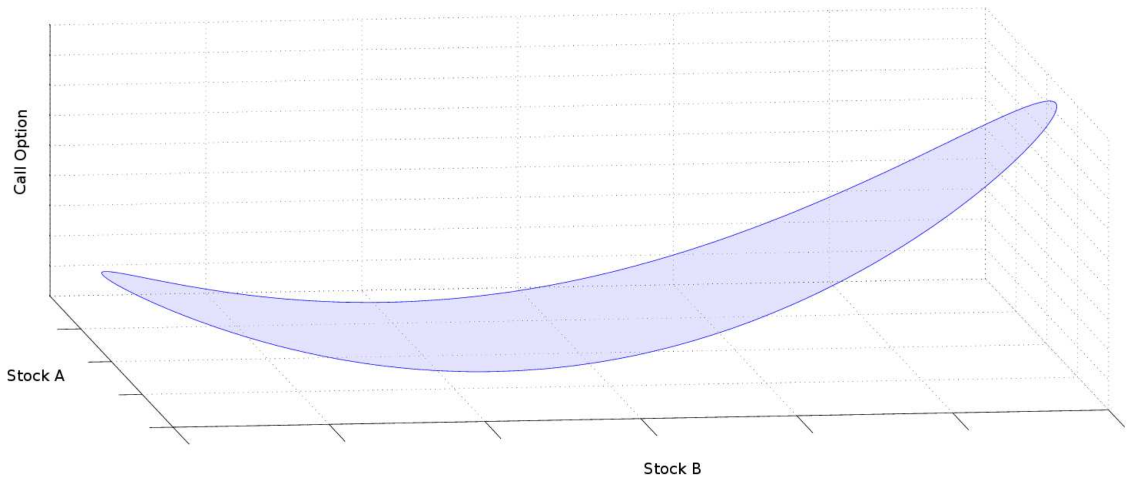

Figure 4.3: Illustration of the $\mathcal{U}_\epsilon^{q2}$ uncertainty set: the classical ellipsoidal uncertainty set has been transformed by the quadratic approximation of the return of the call option written on stock B.

in which the uncertainty set is embedded into $\mathbb{S}^{n+1}$; the dimension increase can compensate for the non-convexity of the portfolio return function.

Example 4.5.1 *We repeat the same experiment as in Example 4.4.1 but estimate the portfolio VaR after 2 days instead of 21 days. Since the VaR is no longer evaluated at the maturity time of the options, we use the* WCQVaR *model instead of the* WCPVaR *model. We use an analogous Monte-Carlo simulation as in Example 4.4.1 to generate the stock and option returns over a 2-day investment period as well as the corresponding sample means* $\boldsymbol{\mu_r}$ *and covariance matrix* $\boldsymbol{\Sigma_r}$*. We determine their values to be*

$$\boldsymbol{\mu_r} = \begin{bmatrix} 0.0009 \\ 0.0006 \\ 0.0106 \\ -0.0083 \end{bmatrix}, \quad \textit{and} \quad \boldsymbol{\Sigma_r} = \begin{bmatrix} 0.0007 & 0.0001 & 0.0107 & -0.0020 \\ 0.0001 & 0.0003 & 0.0014 & -0.0068 \\ 0.0107 & 0.0014 & 0.1636 & -0.0302 \\ -0.0020 & -0.0068 & -0.0302 & 0.1514 \end{bmatrix},$$

where the first two entries in $\boldsymbol{\mu_r}$ *belong to the stock returns, followed by the call and put option returns. The entries for the covariance matrix obey this ordering.*

Let us now compute WCVaR *at* $\epsilon = 10\%$*. We have* $\kappa(0.1) = \sqrt{(1-0.1)/0.1} = 3$*. We now* *insert the above parameter values into equation* (4.7)*, and compute* $\mathrm{WCVaR}_{0.1}(\boldsymbol{w})$ *as*

$$-\begin{bmatrix}0.0009\\0.0006\\0.0106\\-0.0083\end{bmatrix}^{\mathsf{T}}\begin{bmatrix}0.25\\0.25\\0.25\\0.25\end{bmatrix}+3\sqrt{\begin{bmatrix}0.25\\0.25\\0.25\\0.25\end{bmatrix}^{\mathsf{T}}\begin{bmatrix}0.0007 & 0.0001 & 0.0107 & -0.0020\\0.0001 & 0.0003 & 0.0014 & -0.0068\\0.0107 & 0.0014 & 0.1636 & -0.0302\\-0.0020 & -0.0068 & -0.0302 & 0.1514\end{bmatrix}\begin{bmatrix}0.25\\0.25\\0.25\\0.25\end{bmatrix}},$$

which is equal to 0.3831*.*

Next, we evaluate WCQVaR *at* $\epsilon = 10\%$*. The coefficients of the quadratic approximation function* (4.33) *are calculated using the standard Black-Scholes greek formulas (see, e.g., [Mac92]). Thus, we have*

$$\begin{aligned}\theta(\boldsymbol{w}) &= -0.00019513\times 0.25 - 0.00017798\times 0.25 = -9.3276\times 10^{-5}\\ \boldsymbol{\Delta}(\boldsymbol{w}) &= 0.25\begin{bmatrix}1\\0\end{bmatrix}+0.25\begin{bmatrix}0\\1\end{bmatrix}+0.25\begin{bmatrix}14.7872\\0\end{bmatrix}-0.25\begin{bmatrix}0\\21.6419\end{bmatrix}=\begin{bmatrix}3.9468\\-5.1605\end{bmatrix}\\ \boldsymbol{\Gamma}(\boldsymbol{w}) &= 0.25\begin{bmatrix}128.4907 & 0\\0 & 0\end{bmatrix}+0.25\begin{bmatrix}0 & 0\\0 & 316.5187\end{bmatrix}=\begin{bmatrix}32.1227 & 0\\0 & 79.1297\end{bmatrix}.\end{aligned}$$

Furthermore, WCQVaR *requires the second-order moment matrix* $\boldsymbol{\Omega}$ *of the stock returns, which we compute using* (4.21)*. We have*

$$\boldsymbol{\Omega}=\begin{bmatrix}0.0007 & 0.0001 & 0.0009\\0.0001 & 0.0003 & 0.0006\\0.0009 & 0.0006 & 1.0000\end{bmatrix}.$$

We now compute $\mathrm{WCQVaR}_{0.1}(\boldsymbol{w})$ *by inserting the above parameter values into problem* (4.35)*,* *by which we obtain*

$$\begin{aligned}
\inf \quad & \gamma \\
\text{s.t.} \quad & \mathbf{M} \in \mathbb{S}^3, \quad \tau \in \mathbb{R}, \quad \gamma \in \mathbb{R} \\
& \left\langle \begin{bmatrix} 0.0007 & 0.0001 & 0.0009 \\ 0.0001 & 0.0003 & 0.0006 \\ 0.0009 & 0.0006 & 1.0000 \end{bmatrix}, \mathbf{M} \right\rangle \leq 0.1\tau, \quad \mathbf{M} \succcurlyeq \mathbf{0}, \quad \tau \geq 0, \\
& \mathbf{M} + \begin{bmatrix} 32.1227 & 0 & 3.9468 \\ 0 & 79.1297 & -5.1605 \\ 3.9468 & -5.1605 & -\tau + 2(\gamma - 9.3276 \times 10^{-5}) \end{bmatrix} \succcurlyeq \mathbf{0}.
\end{aligned}$$

The above problem is solved using the SDPT3 optimization toolkit [TTT03] and we find an optimal objective value $\gamma^* = 0.2899$.

Figure 4.4 (left) displays the sampled portfolio loss distribution, which is still skewed, although considerably less than in Example 4.4.1. In Figure 4.4 (right) we compare Monte-Carlo VaR*,* WCVaR*, and* WCQVaR *for different confidence levels. Even for the short horizon time under consideration, the* WCVaR *model still fails to give a realistic VaR estimate. At* $\epsilon = 1\%$, WCVaR *is more than 3 times as large as the corresponding* WCQVaR *value. This example demonstrates that the* WCQVaR *can offer significantly better VaR estimates than* WCVaR *when the portfolio contains options.*

4.6 Computational Results

In Section 4.6.1 we compare the out-of-sample performance of the WCQVaR in the context of an index tracking application and analyze the benefits of including options in the investment strategy. We refer to Chapter 3 for an in-depth analysis of the in- and out-of-sample performance of the robust optimization problem (4.29), whose equivalence to our novel WCPVaR model was established in Theorem 4.4.2. All computations are performed within Matlab 2008b

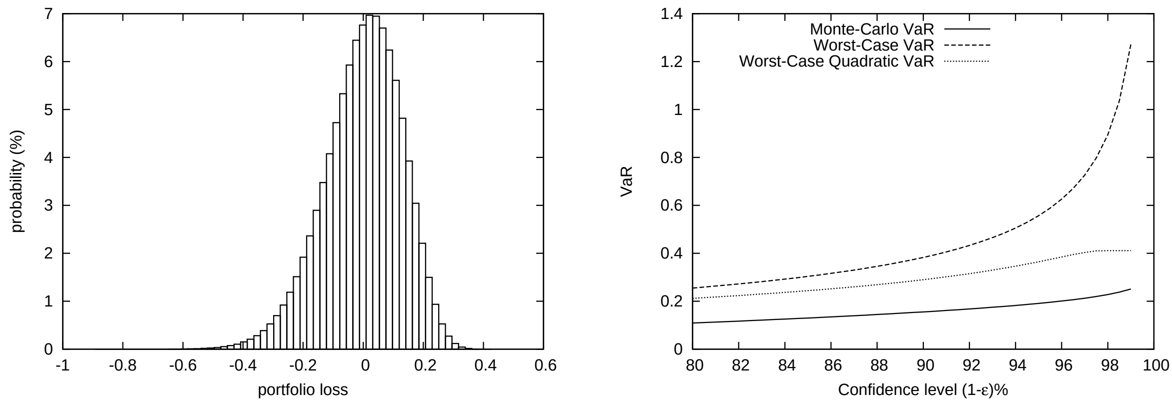

Figure 4.4: Left: The portfolio loss distribution obtained via Monte-Carlo simulation. Note that negative values represent gains. Right: The VaR estimates at different confidence levels obtained via Monte-Carlo sampling, WCVaR, and WCQVaR.

and by using the YALMIP interface [LÖ4] of the SDPT3 optimization toolkit [TTT03]. We use a 2.0 GHz Core 2 Duo machine running Linux Ubuntu 9.04.

4.6.1 Index Tracking using Worst-Case VaR

Index tracking is a common and important problem in portfolio management. The aim is to *replicate* the behavior of a given stock market index, sometimes referred to as the *benchmark*, with a given set of other assets not containing the index itself.

We let $\tilde{r}_1$ denote the random return of the benchmark over the investment interval $[0, T]$. In order to replicate this benchmark, we are given $m-1$ assets, whose vector of returns is denoted by $\tilde{\boldsymbol{r}}_{-1}$. This set of assets includes $n-1$ basic assets as well as $m-n$ options derived from the basic assets. We denote by $\boldsymbol{w}_{-1} \in \mathbb{R}^{m-1}$ the asset weights in the replicating portfolio.

Typically, the level of discrepancy between the benchmark and the portfolio is quantified by the *tracking-error* $\mathbb{E}(|\boldsymbol{w}_{-1}^\mathsf{T}\tilde{\boldsymbol{r}}_{-1} - \tilde{r}_1|)$. Note that minimizing the tracking-error penalizes both under- and over-performance of the portfolio relative to the benchmark.

We adopt a slightly different approach. Instead of minimizing the tracking-error, we are only concerned about the portfolio falling short of the benchmark. The *excess-return* of a portfolio $\boldsymbol{w}_{-1}$ relative to the benchmark is computed as $\boldsymbol{w}^\mathsf{T}\tilde{\boldsymbol{r}}$ where $\boldsymbol{w} = [-1 \;\; \boldsymbol{w}_{-1}^\mathsf{T}]^\mathsf{T}$ and $\tilde{\boldsymbol{r}} = [\tilde{r}_1 \;\; \tilde{\boldsymbol{r}}_{-1}^\mathsf{T}]^\mathsf{T}$. In order to measure the risk of the replicating portfolio falling below the benchmark, we can use the VaR at confidence $\epsilon = 5\%$.[4] The optimal replicating portfolio is found by minimizing $\mathrm{VaR}_\epsilon(\boldsymbol{w})$ over all admissible portfolios $\boldsymbol{w} \in \mathcal{W}$ with

$$\mathcal{W} = \left\{ \boldsymbol{w} \in \mathbb{R}^m \ : \ \boldsymbol{w}^+ - \boldsymbol{w}^- = \boldsymbol{w}, \ \ \boldsymbol{e}^\mathsf{T} \boldsymbol{w}^- \leq \alpha + 1, \ \ \boldsymbol{w}^+ \geq \boldsymbol{0}, \ \ \boldsymbol{w}^- \geq \boldsymbol{0}, \ \ \boldsymbol{e}^\mathsf{T} \boldsymbol{w} = 0 \right\}. \quad (4.52)$$

The inclusion $\boldsymbol{w} \in \mathcal{W}$ implies that the portfolio weights $\boldsymbol{w}_{-1}$ sum up to 1 and that the total amount of shortsales in the replicating portfolio is limited to $\alpha = 4\%$.

Since we include options in the replicating portfolio, we use $\mathrm{WCQVaR}_\epsilon(\boldsymbol{w})$ to approximate the VaR objective. The optimal portfolios are found by solving problem (4.41).

We now compare the out-of-sample performance of the optimal portfolios containing options with those where investment in options is prohibited. Recall that in the absence of options WCQVaR reduces to WCVaR, see Remark 4.5.1.

We assess the out-of-sample behavior of the WCQVaR model using a rolling-horizon backtest procedure. The aim is to minimize the under-performance of the replicating portfolio relative to the S&P 500 index, which is often taken as a proxy for the market portfolio. The replicating portfolio is based on the 30 stock constituents of the Dow Jones Industrial Average, as well as some options written on these. We only include options that expire between 30 and 60 days after the investment dates. This ensures that the option payoffs are differentiable and accurately representable by the delta-gamma approximation. Moreover, longer term options tend to be more illiquid and are therefore not included.

Daily stock and option data are obtained from the Optionmetrics IvyDB database, which is one of the most complete sources of historical option data available. We consider a historical data range from January 2nd, 2004 to October 10th, 2008, containing a total of 1181 trading days. We use the following rolling-horizon backtest procedure. At every investment date we estimate the mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ of the stock returns using the daily returns of the previous 600 trading days. Thus, our backtest starts on the 600th trading day in the historical data set. We compute the out-of-sample returns of the optimal replicating portfolios using the

[4] We ran the backtests in this section with different values of ϵ. Although we only report results for $\epsilon = 5\%$, the general conclusions are independent of the choice of ϵ.

stock and option prices on the next available trading day. This process is repeated for all but the first 600 trading days in our data set.

For simplicity, we use the mid-prices of the assets to calculate the returns. Furthermore, the WCQVaR model requires information about the options' delta and gamma sensitivities. These are obtained from the implied volatilities reported in the Optionmetrics database and are calculated using the Black-Scholes formula.[5] We disregard transaction costs and income taxes on option returns, which are beyond the scope of this thesis.

The same rolling-horizon procedure is used to obtain the out-of-sample returns of the optimal replicating portfolios with and without options. On average the optimal stock-only portfolios are found in 2.1 seconds, whereas the portfolios with options are found in 7.4 seconds. In total we obtain two sequences of $L = 581$ out-of-sample portfolio returns, corresponding to the strategies with and without options, which are denote by $\{r_l^o\}_{l=1}^L$ and $\{r_l^s\}_{l=1}^L$, respectively. The returns of the benchmark are denoted by $\{r_{1,l}\}_{l=1}^L$.

Since the portfolios minimize the under-performance with respect to the benchmark, it is of interest to analyze how much wealth the robust strategies generate relative to the benchmark. By assuming an initial capital of 1 dollar, we calculate the relative wealth ω_l^k at the end of period l for portfolio strategy $k = o, s$ as

$$\omega_l^k = \frac{\prod_{m=1}^l (1 + r_m^k)}{\prod_{m=1}^l (1 + r_{1,m})}.$$

Figure 4.5 displays the relative wealth generated over time by the robust strategies. Both strategies outperform the benchmark over the entire test period. However, the inclusion of options improves the performance considerably. Over the test period, the strategy with options outperforms the benchmark by 56%, whereas the stock-only strategy only outperforms the benchmark by 12%. The annualized average excess-return of the stock-only strategy is 4.9% and that of the option strategy amounts to 19%.

[5]In order to avoid the use of erroneous option data, we only selected those options for which the implied volatility was supplied in the database and which had a bid and ask price greater than 0. We found that this procedure allowed us to filter out incorrect entries.

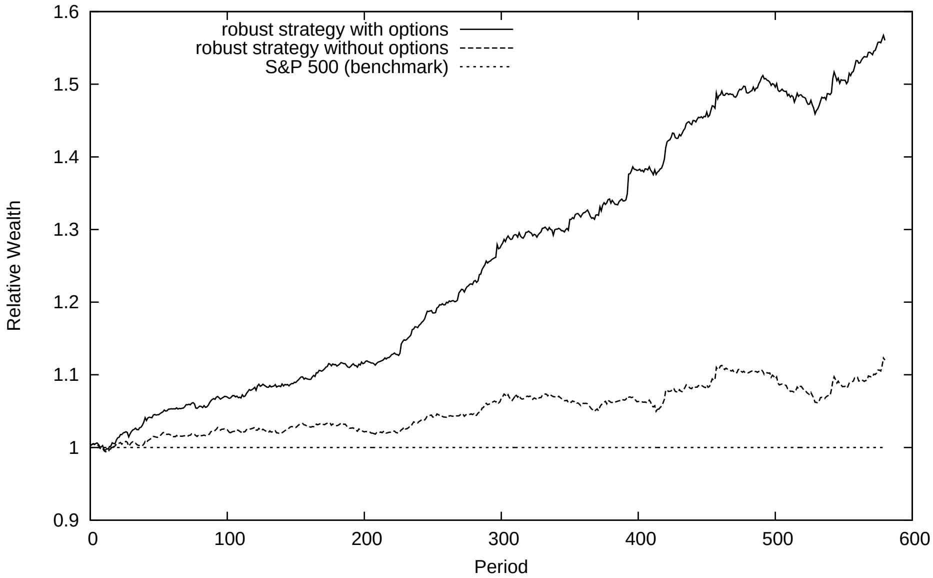

Figure 4.5: Cumulative relative wealth over time of the robust strategies using daily rebalancing between 22/05/2006 and 10/10/2008.

The Sharpe ratio [Sha66] is frequently used to assess the performance of an investment strategy. It is calculated as $(\hat{\mu} - r_f)/\hat{\sigma}$, where $\hat{\mu}$ and $\hat{\sigma}$ represent the annualized estimated mean and standard deviation of the out-of-sample returns, respectively, and $r_f = 3\%$ is the risk-free rate per annum. The stock-only strategy has a Sharpe ratio of 0.13, while the option strategy achieves a value of 0.97. These results clearly demonstrate the benefits of including options in the replicating portfolio.

We observe that all optimal portfolios $\boldsymbol{w}$ satisfy $\boldsymbol{\Gamma}(\boldsymbol{w}) \succcurlyeq \mathbf{0}$, although this was not imposed as a constraint. This implies that the delta-gamma approximation (4.33) of the optimal portfolio return is convex in the returns of the underlying assets. Alexander has observed this phenomenon in a simulation experiment and argues that it is a natural consequence of the risk minimization process. In fact, a portfolio with a convex payoff loses less from downward price moves and benefits disproportionately from upward price moves of the underlying assets [Ale08].

We further observe that the optimal portfolios hold both long and short positions in options on the same underlying asset. It is known that short-sales of options can generate high expected returns (see, e.g., [CS02]) but they also carry considerable risk. Thus, optimal portfolios always cover the short-sale of an option by a long position in another option on the same underlying asset. On average the optimal portfolios allocate 11% of wealth in options and 89% in stocks. This implies that the high expected returns generated by the option strategy are not due to risky positions in options, but rather result from a balanced investment in a mixture of both stocks and options.

Next, we assess the *realized* VaRs of the stock-only and option strategies. These are obtained by first computing the ϵ-quantiles of all out-of-sample excess-returns of both strategies and then multiplying these values by -1 (recall that VaR measures the degree of under-performance). For $\epsilon = 5\%$ the realized VaR of the stock-only strategy amounts to 0.29%, while that of the option strategy is 0.33%. For $\epsilon = 1\%$, the realized VaR values are 0.49% and 0.54%, respectively. These results indicate that the option strategy has a slightly higher out-of-sample VaR than the stock-only strategy. However, since the option strategy achieves much higher excess-returns on average, the differences in VaR are negligible. Interestingly, the worst-case daily under-performance of the stock-only strategy is 0.78%, whilst that of the option strategy is 0.61%. Thus, the option strategy performs better in terms of worst-case under-performance relative to the benchmark.

The WCQVaR model described in Section 4.5 assumes the underlying asset returns to be the only sources of uncertainty in the market. It is known, however, that implied volatilities constitute important risk factors for portfolios containing options. In particular, long dated options are highly sensitive to fluctuations in the volatilities of the underlying assets. The sensitivity of the portfolio return with respect to the volatilities is commonly referred to as *vega risk*. The WCQVaR model can easily be modified to include implied volatilities as additional risk factors. The arising *delta-gamma-vega-approximation* of the portfolio return is still a quadratic function of the risk factors. Thus, the theoretical derivations in Section 4.5 remain valid in this generalized setting. However, estimating first- and second-order moments of the implied volatilities requires the modeling and calibration of the implied volatility surface over time, which is beyond the scope of this thesis. We conjecture that extending the WCQVaR model to account for vega risk can further improve the realized VaR of the option strategy.

4.7 Conclusions

Derivatives depend non-linearly on their underlying assets. In this chapter we generalized the WCVaR model by explicitly incorporating this non-linear relationship into the problem formulation. To this end, we developed two new models.

The WCPVaR model is suited for portfolios containing European options maturing at the investment horizon. WCPVaR expresses the option returns as convex-piecewise linear functions of the underlying assets. A benefit of this model is that it does not require knowledge of the pricing models of the options in the portfolio. However, in order to be tractably solvable, the WCPVaR model precludes short-sales of options.

The WCQVaR model can handle portfolios containing general option types and does not rely on short-sales restrictions. It exploits the popular delta-gamma approximation to model the portfolio return. In contrast to WCPVaR, WCQVaR does require knowledge of the option pricing models to determine the quadratic approximation. Through numerical experiments we demonstrate that the WCPVaR and WCQVaR models can provide much tighter VaR estimates of a portfolio containing options than the WCVaR model which does not explicitly account for non-linear dependencies between the asset returns.

We analyzed the performance of the WCQVaR model in the context of an index tracking application and find that including options in the investment strategy significantly improves the out-of-sample performance. Although options are typically seen as a risky investment, our numerical results indicate that their use in a robust optimization framework can offer substantial benefits.

4.8 Appendix

4.8.1 Proof of Lemma 4.4.1

Define the indicator function of the set $\mathcal{S}$ as

$$\mathbb{I}_{\mathcal{S}}(\boldsymbol{\xi}) = \begin{cases} 1 & \text{if } \boldsymbol{\xi} \in \mathcal{S}, \\ 0 & \text{otherwise.} \end{cases}$$

The worst-case probability problem (4.19) can equivalently be expressed as

$$\begin{aligned} \pi_{\text{wc}} \quad = \quad \sup_{\mu \in \mathcal{M}_+} \quad & \int_{\mathbb{R}^n} \mathbb{I}_{\mathcal{S}}(\boldsymbol{\xi}) \mu(\mathrm{d}\boldsymbol{\xi}) \\ \text{s.t.} \quad & \int_{\mathbb{R}^n} \mu(\mathrm{d}\boldsymbol{\xi}) = 1 \\ & \int_{\mathbb{R}^n} \boldsymbol{\xi} \mu(\mathrm{d}\boldsymbol{\xi}) = \boldsymbol{\mu} \\ & \int_{\mathbb{R}^n} \boldsymbol{\xi}\boldsymbol{\xi}^\intercal \mu(\mathrm{d}\boldsymbol{\xi}) = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\intercal, \end{aligned} \tag{4.53}$$

where $\mathcal{M}_+$ represents the cone of nonnegative Borel measures on $\mathbb{R}^n$. The optimization variable of the semi-infinite linear program (4.53) is the nonnegative measure μ. As can be seen, the first constraint forces μ to be a probability measure. The following constraints enforce consistency with the given first- and second-order moments, respectively.

We now assign dual variables $y_0 \in \mathbb{R}$, $\boldsymbol{y} \in \mathbb{R}^n$, and $\mathbf{Y} \in \mathbb{S}^n$ to the equality constraints in (4.53), respectively, and introduce the following dual problem (see, e.g., [Sha01]).

$$\begin{aligned} \inf \quad & y_0 + \boldsymbol{y}^\intercal \boldsymbol{\mu} + \langle \mathbf{Y}, \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\intercal \rangle \\ \text{s.t.} \quad & y_0 \in \mathbb{R}, \quad \boldsymbol{y} \in \mathbb{R}^n, \quad \mathbf{Y} \in \mathbb{S}^n \\ & y_0 + \boldsymbol{y}^\intercal \boldsymbol{\xi} + \langle \mathbf{Y}, \boldsymbol{\xi}\boldsymbol{\xi}^\intercal \rangle \geq \mathbb{I}_{\mathcal{S}}(\boldsymbol{\xi}) \quad \forall \boldsymbol{\xi} \in \mathbb{R}^n \end{aligned} \tag{4.54}$$

Because $\boldsymbol{\Sigma} \succ \mathbf{0}$, it can be shown that strong duality holds [Isi60]. Therefore the worst-case probability π_{wc} coincides with the optimal value of the dual problem (4.54).

By defining

$$\mathbf{M} = \begin{bmatrix} \mathbf{Y} & \frac{1}{2}\boldsymbol{y} \\ \frac{1}{2}\boldsymbol{y}^\top & y_0 \end{bmatrix},$$

problem (4.54) can be reformulated as

$$\begin{aligned} \inf_{\mathbf{M}\in\mathbb{S}^{n+1}} \quad & \langle \boldsymbol{\Omega}, \mathbf{M} \rangle \\ \text{s. t.} \quad & \left[\boldsymbol{\xi}^\top \ 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\top \ 1\right]^\top \geq \mathbb{I}_{\mathcal{S}}(\boldsymbol{\xi}) \quad \forall \boldsymbol{\xi} \in \mathbb{R}^n. \end{aligned} \tag{4.55}$$

By definition of $\mathbb{I}_{\mathcal{S}}(\boldsymbol{\xi})$, the constraint in (4.55) can be expanded in terms of two semi-infinite constraints.

$$\left[\boldsymbol{\xi}^\top \ 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\top \ 1\right]^\top \geq 0 \quad \forall \boldsymbol{\xi} \in \mathbb{R}^n \tag{4.56a}$$

$$\left[\boldsymbol{\xi}^\top \ 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\top \ 1\right]^\top \geq 1 \quad \forall \boldsymbol{\xi} \in \mathcal{S} \tag{4.56b}$$

Since (4.56a) is equivalent to $\mathbf{M} \succcurlyeq \mathbf{0}$, the claim follows. ∎

4.8.2 Proof of Theorem 4.4.1

In order to obtain the postulated SOCP reformulation, we calculate the dual associated with problem (4.26), which, after some simplification steps, reduces to

$$\begin{aligned} \text{WCPVaR}_\epsilon(\boldsymbol{w}) = \max \quad & (\boldsymbol{e} - \boldsymbol{\delta})^\top \boldsymbol{w}^{\boldsymbol{\eta}} - 2\boldsymbol{m}^\top \boldsymbol{w}^{\boldsymbol{\xi}} \\ \text{s. t.} \quad & \alpha \in \mathbb{R}, \quad \boldsymbol{\delta} \in \mathbb{R}^{m-n}, \quad \boldsymbol{m} \in \mathbb{R}^n, \quad \mathbf{Z} \in \mathbb{S}^n \\ & 0 \leq \alpha \leq \frac{1}{2\epsilon}, \quad \alpha\boldsymbol{\Omega} \succcurlyeq \mathbf{Y} = \begin{bmatrix} \mathbf{Z} & \boldsymbol{m} \\ \boldsymbol{m}^\top & 1/2 \end{bmatrix} \succcurlyeq \mathbf{0}, \\ & \boldsymbol{\delta} - 2\mathbf{B}\boldsymbol{m} - \boldsymbol{a} \geq \mathbf{0}, \quad \boldsymbol{\delta} \geq \mathbf{0}. \end{aligned} \tag{4.57}$$

Note that problem (4.57) is strictly feasible, which implies that strong conic duality holds [VB96]. This confirms that the optimal value of the dual problem (4.57) exactly matches the WCPVaR.

By the definition of $\mathbf{\Omega}$ in (4.21), we may conclude that

$$\alpha\mathbf{\Omega} \succcurlyeq \mathbf{Y} \iff \begin{bmatrix} \alpha(\mathbf{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\mathsf{T}) - \mathbf{Z} & \alpha\boldsymbol{\mu} - \boldsymbol{m} \\ (\alpha\boldsymbol{\mu} - \boldsymbol{m})^\mathsf{T} & \alpha - 1/2 \end{bmatrix} \succcurlyeq \mathbf{0} \implies \alpha \geq 1/2.$$

This allows us to divide the matrix inequality in problem (4.57) by α. Subsequently, we apply the variable substitution $(\mathbf{Z}, \boldsymbol{m}, \alpha) \rightarrow (\mathbf{V}, \boldsymbol{v}, y)$ with $\mathbf{V} = \mathbf{Z}/\alpha$, $\boldsymbol{v} = \boldsymbol{m}/\alpha$, and $y = \frac{1}{2\alpha} \in [\epsilon, 1]$. We thus obtain the following problem reformulation.

$$\begin{aligned} \text{WCPVaR}_\epsilon(\boldsymbol{w}) = \max \quad & (\boldsymbol{e} - \boldsymbol{\delta})^\mathsf{T}\boldsymbol{w}^{\boldsymbol{\eta}} - \frac{\boldsymbol{v}^\mathsf{T}\boldsymbol{w}^{\boldsymbol{\xi}}}{y} \\ \text{s.t.} \quad & y \in \mathbb{R}, \quad \boldsymbol{\delta} \in \mathbb{R}^{m-n}, \quad \boldsymbol{v} \in \mathbb{R}^n, \quad \mathbf{V} \in \mathbb{S}^n \\ & \epsilon \leq y \leq 1, \quad \mathbf{\Omega} \succcurlyeq \begin{bmatrix} \mathbf{V} & \boldsymbol{v} \\ \boldsymbol{v}^\mathsf{T} & y \end{bmatrix} \succcurlyeq \mathbf{0} \\ & \boldsymbol{\delta} \geq \frac{\mathbf{B}\boldsymbol{v}}{y} + \boldsymbol{a}, \quad \boldsymbol{\delta} \geq \mathbf{0} \end{aligned} \tag{4.58}$$

Assume first that $y = 1$ at optimality. Then, by the definition of $\mathbf{\Omega}$ and the linear matrix inequality in problem (4.58), we find $\boldsymbol{v} = \boldsymbol{\mu}$, while (4.58) reduces to

$$\begin{aligned} & \max_{\boldsymbol{\delta} \in \mathbb{R}^{m-n}} \left\{ (\boldsymbol{e} - \boldsymbol{\delta})^\mathsf{T}\boldsymbol{w}^{\boldsymbol{\eta}} - \boldsymbol{\mu}^\mathsf{T}\boldsymbol{w}^{\boldsymbol{\xi}} \ : \ \boldsymbol{\delta} \geq \boldsymbol{a} + \mathbf{B}\boldsymbol{\mu}, \ \boldsymbol{\delta} \geq \mathbf{0} \right\} \\ = & - \boldsymbol{\mu}^\mathsf{T}\boldsymbol{w}^{\boldsymbol{\xi}} - (\max\{-\boldsymbol{e}, \boldsymbol{a} + \mathbf{B}\boldsymbol{\mu} - \boldsymbol{e}\})^\mathsf{T}\boldsymbol{w}^{\boldsymbol{\eta}} \\ = & - f(\boldsymbol{\mu})^\mathsf{T}\boldsymbol{w}. \end{aligned} \tag{4.59}$$

Assume now that $y < 1$ at optimality. By the definition of $\mathbf{\Omega}$ and by using Schur complements, we find

$$\begin{aligned} \mathbf{\Omega} \succcurlyeq \begin{bmatrix} \mathbf{V} & \boldsymbol{v} \\ \boldsymbol{v}^\mathsf{T} & y \end{bmatrix} & \iff \begin{bmatrix} \mathbf{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\mathsf{T} - \mathbf{V} & \boldsymbol{\mu} - \boldsymbol{v} \\ (\boldsymbol{\mu} - \boldsymbol{v})^\mathsf{T} & 1 - y \end{bmatrix} \succcurlyeq \mathbf{0} \\ & \iff \mathbf{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\mathsf{T} - \mathbf{V} - \frac{1}{1-y}(\boldsymbol{\mu} - \boldsymbol{v})(\boldsymbol{\mu} - \boldsymbol{v})^\mathsf{T} \succcurlyeq \mathbf{0}. \end{aligned} \tag{4.60a}$$

A similar argument yields the equivalence

$$\begin{bmatrix} \mathbf{V} & \boldsymbol{v} \\ \boldsymbol{v}^\mathsf{T} & y \end{bmatrix} \succcurlyeq \mathbf{0} \iff \mathbf{V} - \frac{1}{y}\boldsymbol{v}\boldsymbol{v}^\mathsf{T} \succcurlyeq \mathbf{0}. \tag{4.60b}$$

By combining (4.60a) and (4.60b), the linear matrix inequality constraints in problem (4.58) are equivalent to

$$\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\mathsf{T} - \frac{1}{1-y}(\boldsymbol{\mu} - \boldsymbol{v})(\boldsymbol{\mu} - \boldsymbol{v})^\mathsf{T} \succcurlyeq \mathbf{V} \succcurlyeq \frac{1}{y}\boldsymbol{v}\boldsymbol{v}^\mathsf{T}.$$

The decision variable $\mathbf{V}$ can now be eliminated from the problem, while the linear matrix inequality constraints in (4.58) can be replaced by

$$\begin{aligned} & \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\mathsf{T} \succcurlyeq \frac{1}{1-y}(\boldsymbol{\mu} - \boldsymbol{v})(\boldsymbol{\mu} - \boldsymbol{v})^\mathsf{T} + \frac{1}{y}\boldsymbol{v}\boldsymbol{v}^\mathsf{T} \\ \iff\quad & \boldsymbol{\Sigma} \succcurlyeq \frac{1}{y(1-y)}(\boldsymbol{v} - y\boldsymbol{\mu})(\boldsymbol{v} - y\boldsymbol{\mu})^\mathsf{T}. \end{aligned} \tag{4.61}$$

The above arguments imply that problem (4.58) can be reformulated as

$$\mathrm{WCPVaR}_\epsilon(\boldsymbol{w}) = \max\{\phi(y) \ : \ y \in [\epsilon, 1]\},$$

where

$$\begin{aligned} \phi(y) = \max \quad & (\boldsymbol{e} - \boldsymbol{\delta})^\mathsf{T}\boldsymbol{w}^{\boldsymbol{\eta}} - \frac{\boldsymbol{v}^\mathsf{T}\boldsymbol{w}^{\boldsymbol{\xi}}}{y} \\ \text{s.t.} \quad & \boldsymbol{\delta} \in \mathbb{R}^{m-n}, \quad \boldsymbol{v} \in \mathbb{R}^n \\ & \boldsymbol{\Sigma} \succcurlyeq \frac{1}{y(1-y)}(\boldsymbol{v} - y\boldsymbol{\mu})(\boldsymbol{v} - y\boldsymbol{\mu})^\mathsf{T}, \\ & \boldsymbol{\delta} \geq \frac{\mathbf{B}\boldsymbol{v}}{y} + \boldsymbol{a}, \quad \boldsymbol{\delta} \geq \mathbf{0}. \end{aligned} \tag{4.62}$$

For any fixed $y \in [\epsilon, 1)$, we have that $y^{-1}(1-y)^{-1} > 0$, and the linear matrix inequality in (4.62) can be rewritten as

$$\begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{v} - y\boldsymbol{\mu} \\ (\boldsymbol{v} - y\boldsymbol{\mu})^\mathsf{T} & y(1-y) \end{bmatrix} \succcurlyeq \mathbf{0}.$$

Since $\mathbf{\Sigma} \succ \mathbf{0}$, this linear matrix inequality holds if and only if

$$(\boldsymbol{v} - y\boldsymbol{\mu})^\mathsf{T}\mathbf{\Sigma}^{-1}(\boldsymbol{v} - y\boldsymbol{\mu}) \leq y(1-y),$$

which is equivalent to the second-order cone constraint

$$\left\|\mathbf{\Sigma}^{-1/2}(\boldsymbol{v} - y\boldsymbol{\mu})\right\|_2 \leq \sqrt{y(1-y)}.$$

For $y \in [\epsilon, 1)$, the value of $\phi(y)$ can thus be found by solving the following SOCP.

$$\begin{aligned} \phi(y) = \max \quad & (\boldsymbol{e} - \boldsymbol{\delta})^\mathsf{T}\boldsymbol{w}^{\boldsymbol{\eta}} - \frac{\boldsymbol{v}^\mathsf{T}\boldsymbol{w}^{\boldsymbol{\xi}}}{y} \\ \text{s.t.} \quad & \boldsymbol{\delta} \in \mathbb{R}^{m-n}, \quad \boldsymbol{v} \in \mathbb{R}^n \\ & \left\|\mathbf{\Sigma}^{-1/2}(\boldsymbol{v} - y\boldsymbol{\mu})\right\|_2 \leq \sqrt{y(1-y)} \\ & \boldsymbol{\delta} \geq \frac{\mathbf{B}\boldsymbol{v}}{y} + \boldsymbol{a}, \quad \boldsymbol{\delta} \geq \mathbf{0} \end{aligned} \tag{4.63}$$

Note that the above problem is strictly feasible for $y \in [\epsilon, 1)$. By strong conic duality the associated dual problem has the same optimal value [AG03]. We thus obtain that $\phi(y) = \phi'(y)$ for $y \in [\epsilon, 1)$, where

$$\phi'(y) = \min_{\mathbf{0} \leq \boldsymbol{g} \leq \boldsymbol{w}^{\boldsymbol{\eta}}} \; -\boldsymbol{\mu}^\mathsf{T}(\boldsymbol{w}^{\boldsymbol{\xi}} + \mathbf{B}^\mathsf{T}\boldsymbol{g}) + \sqrt{\frac{1-y}{y}}\left\|\mathbf{\Sigma}^{1/2}(\boldsymbol{w}^{\boldsymbol{\xi}} + \mathbf{B}^\mathsf{T}\boldsymbol{g})\right\|_2 - \boldsymbol{a}^\mathsf{T}\boldsymbol{g} + \boldsymbol{e}^\mathsf{T}\boldsymbol{w}^{\boldsymbol{\eta}}.$$

Note that for $y = 1$, we also have $\phi(1) = \phi'(1)$ since

$$\begin{aligned} \phi'(1) &= \min_{\mathbf{0} \leq \boldsymbol{g} \leq \boldsymbol{w}^{\boldsymbol{\eta}}} \; -\boldsymbol{\mu}^\mathsf{T}(\boldsymbol{w}^{\boldsymbol{\xi}} + \mathbf{B}^\mathsf{T}\boldsymbol{g}) - \boldsymbol{a}^\mathsf{T}\boldsymbol{g} + \boldsymbol{e}^\mathsf{T}\boldsymbol{w}^{\boldsymbol{\eta}} \\ &= -\boldsymbol{\mu}^\mathsf{T}\boldsymbol{w}^{\boldsymbol{\xi}} - (\max\{-\boldsymbol{e}, \boldsymbol{a} + \mathbf{B}\boldsymbol{\mu} - \boldsymbol{e}\})^\mathsf{T}\boldsymbol{w}^{\boldsymbol{\eta}} \\ &= \phi(1), \end{aligned}$$

where the second equality follows from (4.59). Maximizing $\phi(y)$ over y yields the desired WCPVaR value. Since $\sqrt{(1-y)/y}$ is monotonically decreasing in y, we have $y = \epsilon$ at optimality. This results in the following optimization problem

$$\mathrm{WCPVaR}_\epsilon(\boldsymbol{w}) = \min_{\boldsymbol{0}\leq\boldsymbol{g}\leq\boldsymbol{w}^{\boldsymbol{\eta}}} \; -\boldsymbol{\mu}^\mathsf{T}(\boldsymbol{w}^{\boldsymbol{\xi}} + \mathbf{B}^\mathsf{T}\boldsymbol{g}) + \sqrt{\frac{1-\epsilon}{\epsilon}}\left\|\boldsymbol{\Sigma}^{1/2}(\boldsymbol{w}^{\boldsymbol{\xi}} + \mathbf{B}^\mathsf{T}\boldsymbol{g})\right\|_2 - \boldsymbol{a}^\mathsf{T}\boldsymbol{g} + \boldsymbol{e}^\mathsf{T}\boldsymbol{w}^{\boldsymbol{\eta}},$$

which is the postulated reformulation of WCPVaR as the optimal value of a SOCP. ∎

Chapter 5

Distributionally Robust Joint Chance Constraints

In Chapter 4, we derived tractable reformulations of Worst-Case Value-at-Risk measures by solving moment problems. It is important to note that the Worst-Case Value-at-Risk is equivalent to a distributionally robust chance constrained program. In this chapter, we leverage on the reformulation techniques developed in Chapter 4 to further investigate general distributionally robust chance constrained programs. In particular, we focus on *joint* chance constraints, which require a system of uncertainty-affected constraints to be jointly satisfied with a given probability. Problems involving such constraints are typically difficult to solve. In fact, even finding a feasible solution for such problems can be intractable. However, we will show that distributionally robust joint chance constrained programs can be approximated using tractable semidefinite programs and that the arising approximations can be reformulated as straight robust optimization problems.

5.1 Introduction

A large class of decision problems in engineering and finance can be formulated as *chance constrained programs* of the form

$$\begin{array}{ll} \underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}} & \boldsymbol{c}^\mathsf{T}\boldsymbol{x} \\ \text{subject to} & \mathbb{Q}\left(\boldsymbol{a}_i(\tilde{\boldsymbol{\xi}})^\mathsf{T}\boldsymbol{x} \leq \boldsymbol{b}_i(\tilde{\boldsymbol{\xi}}) \quad \forall i=1,\ldots,m\right) \geq 1-\epsilon \\ & \boldsymbol{x}\in\mathcal{X}, \end{array} \tag{5.1}$$

where $\boldsymbol{x} \in \mathbb{R}^n$ is the decision vector, $\mathcal{X} \subseteq \mathbb{R}^n$ is a convex closed set that can be represented by semidefinite constraints, and $\boldsymbol{c} \in \mathbb{R}^n$ is a cost vector. Without much loss of generality, we assume that $\boldsymbol{c}$ is deterministic. The *chance constraint* in (5.1) requires a set of m uncertainty-affected inequalities to be jointly satisfied with a probability of at least $1-\epsilon$, where $\epsilon \in (0,1)$ is a desired safety factor specified by the modeler. The uncertain constraint coefficients $\boldsymbol{a}_i(\tilde{\boldsymbol{\xi}}) \in \mathbb{R}^n$ and $\boldsymbol{b}_i(\tilde{\boldsymbol{\xi}}) \in \mathbb{R}$, $i=1,\ldots,m$, depend affinely on a random vector $\tilde{\boldsymbol{\xi}} \in \mathbb{R}^k$, whose distribution $\mathbb{Q}$ is assumed to be known. We thus have

$$\boldsymbol{a}_i(\tilde{\boldsymbol{\xi}}) = \boldsymbol{a}_i^0 + \sum_{j=1}^k \boldsymbol{a}_i^j \tilde{\xi}_j \quad \text{and} \quad \boldsymbol{b}_i(\tilde{\boldsymbol{\xi}}) = b_i^0 + \sum_{j=1}^k b_i^j \tilde{\xi}_j.$$

For ease of notation we introduce auxiliary functions $y_i^j : \mathbb{R}^n \to \mathbb{R}$, which are defined through

$$y_i^j(\boldsymbol{x}) = (\boldsymbol{a}_i^j)^\mathsf{T}\boldsymbol{x} - b_i^j, \quad i=1,\ldots,n,\ j=0,\ldots,k.$$

These functions enable us to rewrite the chance constraint in problem (5.1) as

$$\mathbb{Q}\left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T}\tilde{\boldsymbol{\xi}} \leq 0 \quad \forall i=1,\ldots,m\right) \geq 1-\epsilon, \tag{5.2}$$

where $\boldsymbol{y}_i(\boldsymbol{x}) = [y_i^1(\boldsymbol{x}),\ldots,y_i^k(\boldsymbol{x})]^\mathsf{T}$ for $i=1,\ldots,m$. By convention, (5.2) is referred to as an *individual* or *joint* chance constraint if $m=1$ or $m>1$, respectively. Chance constrained programs were first discussed by Charnes *et al.* [CCS58], Miller and Wagner [MW65] and Prékopa [Pre70]. Although they have been studied for a long time, they have not found wide application in practice due to the following reasons.

Firstly, computing the optimal solution of a chance constrained program is notoriously difficult. In fact, even checking the feasibility of a fixed decision $\boldsymbol{x}$ requires the computation of a multi-dimensional integral, which becomes increasingly difficult as the dimension k of the random vector $\tilde{\boldsymbol{\xi}}$ increases. Furthermore, even though the inequalities in the chance constraint (5.2) are biaffine in $\boldsymbol{x}$ and $\tilde{\boldsymbol{\xi}}$, the feasible set of problem (5.1) is typically nonconvex and sometimes even disconnected.

Secondly, in order to evaluate the chance constraint (5.2), full and accurate information about the probability distribution $\mathbb{Q}$ of the random vector $\tilde{\boldsymbol{\xi}}$ is required. However, in many practical situations $\mathbb{Q}$ must be estimated from historical data and is therefore itself uncertain. Typically, one has only partial information about $\mathbb{Q}$, e.g. about its moments or its support. Replacing the unknown distribution $\mathbb{Q}$ in (5.1) by an estimate $\hat{\mathbb{Q}}$ corrupted by measurement errors may lead to over-optimistic solutions which often fail to satisfy the chance constraint under the true distribution $\mathbb{Q}$.

In a few special cases chance constraints can be reformulated as tractable convex constraints. For example, it is known that if the random vector $\tilde{\boldsymbol{\xi}}$ follows a Gaussian distribution, then an individual chance constraint can be equivalently expressed as a single second-order cone constraint. In this case, the chance constrained problem becomes a tractable second-order cone program (SOCP), which can be solved in polynomial time, see Alizadeh and Goldfarb [AG03]. More generally, Calafiore and El Ghaoui [CEG06] have shown that individual chance constraints can be converted to second-order cone constraints whenever the random vector $\tilde{\boldsymbol{\xi}}$ is governed by a radial distribution. Tractability results for joint chance constraints are even more scarce. In a seminal paper, Prékopa [Pre70] has shown that joint chance constraints are convex when only the right-hand side coefficients $\boldsymbol{b}_i(\tilde{\boldsymbol{\xi}})$ are uncertain and follow a log-concave distribution. However, under generic distributions, chance constrained programs are computationally intractable. Indeed, Shapiro and Nemirovski [NS06] point out that computing the probability of a weighted sum of uniformly distributed variables being nonpositive is already $\mathcal{NP}$-hard.

Recently, Calafiore and Campi [CC06] as well as Luedtke and Ahmed [LA08] have proposed to replace the chance constraint (5.2) by a pointwise constraint that must hold at a finite number of sample points drawn randomly from the distribution $\mathbb{Q}$. A similar approach was advised by Erdoğan and Iyengar [EI06]. The advantage of this Monte Carlo approach is that no structural assumptions about $\mathbb{Q}$ are needed and that the resulting approximate problem is convex. Calafiore and Campi [CC06] showed that one requires $\mathcal{O}(n/\epsilon)$ samples to guarantee that a solution of the approximate problem is feasible in the original chance constrained program. However, this implies that it may be computationally prohibitive to solve large problems or to solve problems for which a small violation probability ϵ is required.

A natural way to *immunize* the chance constraint (5.2) against uncertainty in the probability distribution is to adopt a distributionally robust approach. To this end, let $\mathcal{P}$ denote the set of all probability distributions on $\mathbb{R}^k$ that are consistent with the known properties of $\mathbb{Q}$, such as its first and second moments and/or its support. Consider now the following *ambiguous* or *distributionally robust chance constraint.*

$$\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T}\tilde{\boldsymbol{\xi}} \leq 0 \quad \forall i = 1, \ldots, m\right) \geq 1 - \epsilon \tag{5.3}$$

It is easily verified that whenever $\boldsymbol{x}$ satisfies (5.3) and $\mathbb{Q} \in \mathcal{P}$, then $\boldsymbol{x}$ also satisfies the chance constraint (5.2) under the true probability distribution $\mathbb{Q}$. Replacing the chance constraint (5.2) with its distributionally robust counterpart (5.3) yields the following distributionally robust chance constrained program

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}} \quad & \boldsymbol{c}^\mathsf{T}\boldsymbol{x} \\ \text{subject to} \quad & \inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T}\tilde{\boldsymbol{\xi}} \leq 0 \quad \forall i = 1, \ldots, m\right) \geq 1 - \epsilon \\ & \boldsymbol{x} \in \mathcal{X}, \end{aligned} \tag{5.4}$$

which constitutes a conservative approximation for problem (5.1) in the sense that it has the same objective function but a smaller feasible set.

A common method to simplify the distributionally robust joint chance constraint (5.3), which looks even less tractable than (5.2), is to decompose it into m individual chance constraints by using Bonferroni's inequality. Indeed, by ensuring that the total sum of violation probabilities of the individual chance constraints does not exceed ϵ, the feasibility of the joint chance constraint is guaranteed. Nemirovski and Shapiro [NS06] propose to divide the overall violation probability ϵ equally among the m individual chance constraints. However, the Bonferroni inequality is not necessarily tight, and the corresponding decomposition could therefore be over-conservative. In fact, for positively correlated constraint functions, the quality of the approximation is known to decrease as m increases [CSSC09]. Consequently, the Bonferroni method may result in a poor approximation for problems with joint chance constraints that involve many inequalities.

A recent attempt to improve on the Bonferroni approximation is due to Chen *et al.* [CSSC09]. They first elaborate a convex conservative approximation for a joint chance constraint in terms of a Worst-Case Conditional Value-at-Risk (CVaR) constraint. Then, they rely on a classical inequality in order statistics to determine a tractable conservative approximation for the Worst-Case CVaR and show that the resulting approximation for the joint chance constraint necessarily outperforms the Bonferroni approximation. An attractive feature of this method is that the arising approximate constraints are second-order conic representable. However, the employed probabilistic inequality is not necessarily tight, which may again render the approximation over-conservative.

The principal aim of this chapter is to develop new tools and models for approximating joint chance constraints under the assumption that only the first- and second-order moments as well as the support of the random vector $\tilde{\boldsymbol{\xi}}$ are known. We embrace the modern approach to approximate robust chance constraints by Worst-Case CVaR constraints, but in contrast to the state-of-the-art methods described above, we find an exact semidefinite programming (SDP) reformulation of the Worst-Case CVaR which does not rely on potentially loose probabilistic inequalities. This exact reformulation is facilitated by the theory of moment problems as well as conic duality arguments. We also propose an efficient sequential SDP algorithm to solve the distributionally robust chance constrained program (5.4).

Our secondary objective is to gain deeper insights into the relationship between robust chance constrained programming and straight robust optimization. While it is well known that there is a close connection between ambiguous individual chance constraints and robust semi-infinite constraints (see, e.g., Ben-Tal *et al.* [BTEGN09]), the representability of ambiguous joint chance constraints as straight robust constraints has not been thoroughly investigated. We show that robust joint chance constraints can indeed be reformulated as robust semi-infinite constraints and thereby develop a natural extension of the theory of ambiguous individual chance constraints. The main contributions in this chapter can be summarized as follows:

(1) We prove that a distributionally robust individual chance constraint is equivalent to a Worst-Case CVaR constraint if the underlying constraint function is either concave or (possibly nonconcave) quadratic in $\boldsymbol{\xi}$. We also demonstrate that this equivalence can fail to hold even if the constraint function is convex and piecewise linear in $\boldsymbol{\xi}$.

(2) We show that a robust individual chance constraint can be reformulated as a robust semi-infinite constraint involving a new type of uncertainty set embedded in the space of positive semidefinite matrices. This uncertainty set can be interpreted as a lifted version of an ellipsoid in the $\boldsymbol{\xi}$-space.

(3) We develop an SDP-based approximation for robust joint chance constraints and prove that this approximation consistently outperforms the state-of-the-art methods described above. We show that the approximation quality is controlled by a set of scaling parameters and that the approximation becomes exact if the scaling parameters are chosen optimally.

(4) We present an intuitive dual interpretation for the joint chance constraint approximation and prove that a distributionally robust joint chance constraint can be reformulated as a robust semi-infinite constraint. The corresponding uncertainty set is intimately tied to the lifted ellipsoidal uncertainty sets discovered in the context of robust individual chance constraints.

(5) We analyze numerically the performance of the new joint chance constraint approximation when applied to a dynamic water reservoir control problem.

The remainder of this chapter is organized as follows. In Section 5.2 we review and extend existing approximations for distributionally robust individual chance constraints and investigate the relation between individual chance constraints and Worst-Case CVaR constraints. In

Section 5.3 we elaborate a new approximation for joint chance constraints. We show that the approximation quality is controlled by a set of scaling parameters and prove that the approximation becomes exact if the scaling parameters are chosen optimally. We also show that the arising approximate constraints can be reformulated as robust semi-infinite constraints. In Section 5.4 we analyze the performance of our joint chance constraint approximation in the context of a dynamic water reservoir control problem.

Notation. We use lower-case bold face letters to denote vectors and upper-case bold face letters to denote matrices. The space of symmetric matrices of dimension n is denoted by $\mathbb{S}^n$. For any two matrices $\mathbf{X}, \mathbf{Y} \in \mathbb{S}^n$, we let $\langle \mathbf{X}, \mathbf{Y} \rangle = \mathrm{Tr}(\mathbf{X}\mathbf{Y})$ be the trace scalar product, while the relation $\mathbf{X} \succcurlyeq \mathbf{Y}$ ($\mathbf{X} \succ \mathbf{Y}$) implies that $\mathbf{X} - \mathbf{Y}$ is positive semidefinite (positive definite). Random variables are always represented by symbols with tildes, while their realizations are denoted by the same symbols without tildes. For $x \in \mathbb{R}$, we define $x^+ = \max\{x, 0\}$.

5.2 Distributionally Robust Individual Chance Constraints

It is known that robust individual chance constraints can be conservatively approximated by Worst-Case CVaR constraints. In this section, we first show how the theory of moment problems can be used to reformulate these Worst-Case CVaR constraints in terms of tractable semidefinite constraints. Subsequently, we prove that the Worst-Case CVaR constraints are in fact equivalent to the underlying robust chance constraints for a large class of constraint functions. Finally, we illuminate the relation between robust chance constrained programming and classical robust optimization.

Distributional Assumptions. In the remainder of this chapter we let $\boldsymbol{\mu} \in \mathbb{R}^k$ be the mean vector and $\boldsymbol{\Sigma} \in \mathbb{S}^k$ be the covariance matrix of the random vector $\tilde{\boldsymbol{\xi}}$ under the true distribution $\mathbb{Q}$. Thus, we implicitly assume that $\mathbb{Q}$ has finite second-order moments. Without loss of generality we also assume that $\boldsymbol{\Sigma} \succ \mathbf{0}$. Furthermore, we let $\mathcal{P}$ denote the set of all probability distributions on $\mathbb{R}^k$ that have the same first- and second-order moments as $\mathbb{Q}$.

5.2.1 The Worst-Case CVaR Approximation

For $m = 1$, (5.3) reduces to a distributionally robust *individual* chance constraint

$$\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\intercal\tilde{\boldsymbol{\xi}} \leq 0\right) \geq 1-\epsilon, \tag{5.5}$$

whose feasible set is denoted by

$$\mathcal{X}^{\mathrm{ICC}} = \left\{\boldsymbol{x}\in\mathbb{R}^n \;:\; \inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\intercal\tilde{\boldsymbol{\xi}} \leq 0\right) \geq 1-\epsilon\right\}.$$

In the remainder of this section we will demonstrate that $\mathcal{X}^{\mathrm{ICC}}$ has a manifestly tractable representation in terms of Linear Matrix Inequalities (LMIs). To this end, we first recall the definition of CVaR due to Rockafellar and Uryasev [RU02]. For a given measurable loss function $L : \mathbb{R}^k \to \mathbb{R}$, probability distribution $\mathbb{P}$ on $\mathbb{R}^k$, and tolerance $\epsilon \in (0,1)$, the CVaR at level ϵ with respect to $\mathbb{P}$ is defined as

$$\mathbb{P}\text{-CVaR}_\epsilon(L(\tilde{\boldsymbol{\xi}})) = \inf_{\beta\in\mathbb{R}} \left\{\beta + \frac{1}{\epsilon}\mathbb{E}_{\mathbb{P}}\left((L(\tilde{\boldsymbol{\xi}}) - \beta)^+\right)\right\}, \tag{5.6}$$

where $\mathbb{E}_{\mathbb{P}}(\cdot)$ denotes expectation with respect to $\mathbb{P}$. CVaR essentially evaluates the conditional expectation of loss above above the $(1-\epsilon)$-quantile of the loss distribution. It can be shown that CVaR represents a convex functional of the random variable $L(\tilde{\boldsymbol{\xi}})$.

CVaR can be used to construct convex approximations for chance constraints. Indeed, it is well known that

$$\mathbb{P}\left(L(\tilde{\boldsymbol{\xi}}) \leq \mathbb{P}\text{-CVaR}_\epsilon(L(\tilde{\boldsymbol{\xi}}))\right) \geq 1-\epsilon,$$

for any measurable loss function L, see, e.g., Ben-Tal *et al.* [BTEGN09, §4.3.3]. Thus, $\mathbb{P}\text{-CVaR}_\epsilon(L(\tilde{\boldsymbol{\xi}})) \leq 0$ is sufficient to imply $\mathbb{P}(L(\tilde{\boldsymbol{\xi}}) \leq 0) \geq 1-\epsilon$. As this implication holds for any probability distribution and loss function, we conclude that

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-CVaR}_\epsilon\left(y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\intercal\tilde{\boldsymbol{\xi}}\right) \leq 0 \implies \inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\intercal\tilde{\boldsymbol{\xi}} \leq 0\right) \geq 1-\epsilon. \tag{5.7}$$

Thus, the *Worst-Case CVaR* constraint on the left hand side constitutes a conservative approximation for the distributionally robust chance constraint on the right hand side of (5.7). The above discussion motivates us to define the feasible set

$$\mathcal{Z}^{\text{ICC}} = \left\{ \boldsymbol{x} \in \mathbb{R}^n \ : \ \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-CVaR}_\epsilon \left(y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\intercal \tilde{\boldsymbol{\xi}} \right) \leq 0 \right\}, \tag{5.8}$$

and the implication (5.7) gives rise to the following elementary result.

Proposition 5.2.1 *The feasible set $\mathcal{Z}^{\text{ICC}}$ constitutes a conservative approximation for $\mathcal{X}^{\text{ICC}}$, that is, $\mathcal{Z}^{\text{ICC}} \subseteq \mathcal{X}^{\text{ICC}}$.*

We will now show that $\mathcal{Z}^{\text{ICC}}$ has a tractable representation in terms of LMIs.

Theorem 5.2.1 *The feasible set $\mathcal{Z}^{\text{ICC}}$ can be written as*

$$\mathcal{Z}^{\text{ICC}} = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \begin{array}{l} \exists (\beta, \mathbf{M}) \in \mathbb{R} \times \mathbb{S}^{k+1}, \\ \mathbf{M} \succcurlyeq \mathbf{0}, \quad \beta + \frac{1}{\epsilon} \langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq 0, \\ \mathbf{M} - \begin{bmatrix} \mathbf{0} & \frac{1}{2}\boldsymbol{y}(\boldsymbol{x}) \\ \frac{1}{2}\boldsymbol{y}(\boldsymbol{x})^\intercal & y^0(\boldsymbol{x}) - \beta \end{bmatrix} \succcurlyeq \mathbf{0} \end{array} \right\}.$$

The proof of Theorem 5.2.1 relies on the following well-known result about *worst-case expectation problems*, which will play a key role in many of the subsequent derivations. We relegate its proof to Appendix 5.6.1.

Lemma 5.2.1 *Let $f : \mathbb{R}^k \to \mathbb{R}$ be a measurable function, and define the worst-case expectation θ_{wc} as*

$$\theta_{\text{wc}} = \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left((f(\tilde{\boldsymbol{\xi}}))^+ \right), \tag{5.9}$$

where $\mathcal{P}$ represents the usual set of all probability distributions on $\mathbb{R}^k$ with given mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma} \succ \mathbf{0}$. Then,

$$\theta_{\text{wc}} = \inf_{\mathbf{M} \in \mathbb{S}^{k+1}} \left\{ \langle \boldsymbol{\Omega}, \mathbf{M} \rangle \ : \ \mathbf{M} \succcurlyeq \mathbf{0}, \quad \left[\boldsymbol{\xi}^\intercal \ 1 \right] \mathbf{M} \left[\boldsymbol{\xi}^\intercal \ 1 \right]^\intercal \geq f(\boldsymbol{\xi}) \quad \forall \boldsymbol{\xi} \in \mathbb{R}^k \right\},$$

where

$$\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\top & \boldsymbol{\mu} \\ \boldsymbol{\mu}^\top & 1 \end{bmatrix}$$

is the second-order moment matrix of $\tilde{\boldsymbol{\xi}}$.

Proof of Theorem 5.2.1: By using (5.6), the Worst-Case CVaR in (5.8) can be expressed as

$$\begin{aligned} \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-CVaR}_\epsilon \left(y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\top \tilde{\boldsymbol{\xi}} \right) &= \sup_{\mathbb{P}\in\mathcal{P}} \inf_{\beta\in\mathbb{R}} \left\{ \beta + \frac{1}{\epsilon} \mathbb{E}_\mathbb{P} \left((y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\top \tilde{\boldsymbol{\xi}} - \beta)^+ \right) \right\} \\ &= \inf_{\beta\in\mathbb{R}} \left\{ \beta + \frac{1}{\epsilon} \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_\mathbb{P} \left((y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\top \tilde{\boldsymbol{\xi}} - \beta)^+ \right) \right\}, \end{aligned} \tag{5.10}$$

where the interchange of the maximization and minimization operations is justified by a stochastic saddle point theorem due to Shapiro and Kleywegt [SK02], see also Delage and Ye [DY10] or Natarajan *et al.* [NPS09]. We now show that the Worst-Case CVaR (5.10) of some fixed decision $\boldsymbol{x} \in \mathbb{R}^n$ can be computed by solving a tractable SDP. To this end, we first derive an SDP reformulation of the worst-case expectation problem

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_\mathbb{P} \left((y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\top \tilde{\boldsymbol{\xi}} - \beta)^+ \right),$$

which can be identified as the subordinate maximization problem in (5.10). Lemma 5.2.1 enables us to reformulate this worst-case expectation problem as

$$\begin{aligned} \inf_{\mathbf{M}\in\mathbb{S}^{k+1}} \quad & \langle \boldsymbol{\Omega}, \mathbf{M} \rangle \\ \text{s.t.} \quad & \mathbf{M} \succcurlyeq \mathbf{0}, \quad \left[\boldsymbol{\xi}^\top \ 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\top \ 1\right]^\top \geq y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\top \boldsymbol{\xi} - \beta \quad \forall \boldsymbol{\xi} \in \mathbb{R}^k. \end{aligned} \tag{5.11}$$

Note that the semi-infinite constraint in (5.11) can be written as the following LMI.

$$\begin{bmatrix} \boldsymbol{\xi} \\ 1 \end{bmatrix}^\top \left(\mathbf{M} - \begin{bmatrix} \mathbf{0} & \frac{1}{2}\boldsymbol{y}(\boldsymbol{x}) \\ \frac{1}{2}\boldsymbol{y}(\boldsymbol{x})^\top & y^0(\boldsymbol{x}) - \beta \end{bmatrix} \right) \begin{bmatrix} \boldsymbol{\xi} \\ 1 \end{bmatrix} \geq 0 \ \forall \boldsymbol{\xi} \in \mathbb{R}^k \iff \mathbf{M} - \begin{bmatrix} \mathbf{0} & \frac{1}{2}\boldsymbol{y}(\boldsymbol{x}) \\ \frac{1}{2}\boldsymbol{y}(\boldsymbol{x})^\top & y^0(\boldsymbol{x}) - \beta \end{bmatrix} \succcurlyeq \mathbf{0}$$

This in turn allows us to reformulate the worst-case expectation problem as

$$\begin{aligned} \inf_{\mathbf{M}\in\mathbb{S}^{k+1}} \quad & \langle \mathbf{\Omega}, \mathbf{M}\rangle \\ \text{s.t.} \quad & \mathbf{M} \succcurlyeq \mathbf{0}, \quad \mathbf{M} - \begin{bmatrix} \mathbf{0} & \frac{1}{2}\boldsymbol{y}(\boldsymbol{x}) \\ \frac{1}{2}\boldsymbol{y}(\boldsymbol{x})^\mathsf{T} & y^0(\boldsymbol{x}) - \beta \end{bmatrix} \succcurlyeq \mathbf{0}. \end{aligned} \tag{5.12}$$

By replacing the subordinate worst-case expectation problem in (5.10) by (5.12), we obtain

$$\begin{aligned} \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-CVaR}_\epsilon\left(y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\mathsf{T}\tilde{\boldsymbol{\xi}}\right) \quad = \quad \inf \quad & \beta + \tfrac{1}{\epsilon}\langle \mathbf{\Omega}, \mathbf{M}\rangle \\ \text{s.t.} \quad & \mathbf{M} \in \mathbb{S}^{k+1}, \quad \beta \in \mathbb{R} \\ & \mathbf{M} \succcurlyeq \mathbf{0}, \quad \mathbf{M} - \begin{bmatrix} \mathbf{0} & \frac{1}{2}\boldsymbol{y}(\boldsymbol{x}) \\ \frac{1}{2}\boldsymbol{y}(\boldsymbol{x})^\mathsf{T} & y^0(\boldsymbol{x}) - \beta \end{bmatrix} \succcurlyeq \mathbf{0}, \end{aligned} \tag{5.13}$$

and thus the claim follows. ∎

5.2.2 The Exactness of the Worst-Case CVaR Approximation

So far we have shown that the feasible set $\mathcal{Z}^{\text{ICC}}$ defined in terms of a Worst-Case CVaR constraint constitutes a tractable conservative approximation for $\mathcal{X}^{\text{ICC}}$. We now demonstrate that this approximation is in fact *exact*, that is, we show that the implication (5.7) is in fact an equivalence. We first recall the nonlinear Farkas Lemma as well as the $\mathcal{S}$-lemma, which are crucial ingredients for the proof of this result. We refer to Pólik and Terlaky [PT07] for a derivation and an in-depth survey of the $\mathcal{S}$-lemma as well as a review of the Farkas Lemma.

Lemma 5.2.2 (Farkas Lemma) *Let $f_0, \ldots, f_p : \mathbb{R}^k \to \mathbb{R}$ be convex functions, and assume that there exists a strictly feasible point $\bar{\boldsymbol{\xi}}$ with $f_i(\bar{\boldsymbol{\xi}}) < 0$, $i = 1, \ldots, p$. Then, $f_0(\boldsymbol{\xi}) \geq 0$ for all $\boldsymbol{\xi}$ with $f_i(\boldsymbol{\xi}) \leq 0$, $i = 1, \ldots, p$, if and only if there exist constants $\tau_i \geq 0$ such that*

$$f_0(\boldsymbol{\xi}) + \sum_{i=1}^{p} \tau_i f_i(\boldsymbol{\xi}) \geq 0 \quad \forall \boldsymbol{\xi} \in \mathbb{R}^k.$$

Lemma 5.2.3 ($\mathcal{S}$-lemma) *Let $f_i(\boldsymbol{\xi}) = \boldsymbol{\xi}^\mathsf{T}\mathbf{A}_i\boldsymbol{\xi}$ with $\mathbf{A}_i \in \mathbb{S}^n$ be quadratic functions of $\boldsymbol{\xi} \in \mathbb{R}^n$*

for $i = 0, \ldots, p$. Then, $f_0(\boldsymbol{\xi}) \geq 0$ for all $\boldsymbol{\xi}$ with $f_i(\boldsymbol{\xi}) \leq 0$, $i = 1, \ldots, p$, if there exist constants $\tau_i \geq 0$ such that

$$\mathbf{A}_0 + \sum_{i=1}^{p} \tau_i \mathbf{A}_i \succcurlyeq \mathbf{0}.$$

For $p = 1$, the converse implication holds if there exists a strictly feasible point $\bar{\boldsymbol{\xi}}$ with $f_1(\bar{\boldsymbol{\xi}}) < 0$.

Theorem 5.2.2 *Let $L : \mathbb{R}^k \to \mathbb{R}$ be a continuous loss function that is either*

(i) concave in $\boldsymbol{\xi}$, or

(ii) (possibly nonconcave) quadratic in $\boldsymbol{\xi}$.

Then, the following equivalence holds.

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-CVaR}_\epsilon \left(L(\tilde{\boldsymbol{\xi}}) \right) \leq 0 \iff \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left(L(\tilde{\boldsymbol{\xi}}) \leq 0 \right) \geq 1 - \epsilon \tag{5.14}$$

Proof Consider the *Worst-Case Value-at-Risk* of the loss function L, which is defined as

$$\text{WC-VaR}_\epsilon(L(\tilde{\boldsymbol{\xi}})) = \inf_{\gamma \in \mathbb{R}} \left\{ \gamma \ : \ \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left(L(\tilde{\boldsymbol{\xi}}) \leq \gamma \right) \geq 1 - \epsilon \right\}. \tag{5.15}$$

By definition, the WC-VaR is indeed equal to the $(1-\epsilon)$-quantile of $L(\tilde{\boldsymbol{\xi}})$ evaluated under some worst-case distribution in $\mathcal{P}$. We first show that the following equivalence holds.

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left(L(\tilde{\boldsymbol{\xi}}) \leq 0 \right) \geq 1 - \epsilon \iff \text{WC-VaR}_\epsilon \left(L(\tilde{\boldsymbol{\xi}}) \right) \leq 0 \tag{5.16}$$

Indeed, if the left hand side of (5.16) is satisfied, then $\gamma = 0$ is feasible in (5.15), which implies that $\text{WC-VaR}_\epsilon(L(\tilde{\boldsymbol{\xi}})) \leq 0$. To see that the converse implication holds as well, we note that for any fixed $\mathbb{P} \in \mathcal{P}$, the mapping $\gamma \mapsto \mathbb{P}(L(\tilde{\boldsymbol{\xi}}) \leq \gamma)$ is upper semi-continuous, see [PAS09]. Thus, the related mapping $\gamma \mapsto \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}(L(\tilde{\boldsymbol{\xi}}) \leq \gamma)$ is also upper semi-continuous. If $\text{WC-VaR}_\epsilon(L(\tilde{\boldsymbol{\xi}})) \leq 0$, there exists a sequence $\{\gamma_n\}_{n \in \mathbb{N}}$ that converges to zero and is feasible in (5.15), which implies

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left(L(\tilde{\boldsymbol{\xi}}) \leq 0 \right) \geq \limsup_{n \to \infty} \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left(L(\tilde{\boldsymbol{\xi}}) \leq \gamma_n \right) \geq 1 - \epsilon.$$

Thus, (5.16) follows.

To prove the postulated equivalence (5.14), it is now sufficient to show that

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-CVaR}_\epsilon \left(L(\tilde{\boldsymbol{\xi}})\right) = \text{WC-VaR}_\epsilon \left(L(\tilde{\boldsymbol{\xi}})\right).$$

Note that (5.15) can be rewritten as

$$\text{WC-VaR}_\epsilon(L(\tilde{\boldsymbol{\xi}})) \; = \; \inf_{\gamma\in\mathbb{R}} \left\{\gamma \; : \; \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(L(\tilde{\boldsymbol{\xi}}) > \gamma\right) \leq \epsilon\right\}. \tag{5.17}$$

We proceed by simplifying the subordinate worst-case probability problem $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}(L(\tilde{\boldsymbol{\xi}}) > \gamma)$, which can be expressed as

$$\inf_{\mathbf{M}\in\mathbb{S}^{k+1}} \left\{\langle \boldsymbol{\Omega}, \mathbf{M}\rangle \; : \; \mathbf{M} \succcurlyeq \mathbf{0}, \quad \left[\boldsymbol{\xi}^\mathsf{T} \; 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\mathsf{T} \; 1\right]^\mathsf{T} \geq 1 \quad \forall \boldsymbol{\xi} \; : \; \gamma - L(\boldsymbol{\xi}) < 0\right\}, \tag{5.18}$$

see Lemma 4.4.1.

We will now argue that for all but one value of γ problem (5.18) is equivalent to

$$\begin{array}{ll} \inf & \langle \boldsymbol{\Omega}, \mathbf{M}\rangle \\ \text{s.t.} & \mathbf{M}\in\mathbb{S}^{k+1}, \quad \tau\in\mathbb{R}, \quad \mathbf{M}\succcurlyeq\mathbf{0}, \quad \tau\geq 0 \\ & \left[\boldsymbol{\xi}^\mathsf{T} \; 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\mathsf{T} \; 1\right]^\mathsf{T} - 1 + \tau\left(\gamma - L(\boldsymbol{\xi})\right) \geq 0 \quad \forall \boldsymbol{\xi}\in\mathbb{R}^k. \end{array} \tag{5.19}$$

For ease of exposition, we define $h = \inf_{\boldsymbol{\xi}\in\mathbb{R}^k} \gamma - L(\boldsymbol{\xi})$. The equivalence of (5.18) and (5.19) is proved case by case. Assume first that $h < 0$. Then, the strict inequality in the parameter range of the semi-infinite constraint in (5.18) can be replaced by a weak inequality without affecting its optimal value. The equivalence then follows from the Farkas Lemma (when $L(\boldsymbol{\xi})$ is concave in $\boldsymbol{\xi}$) or from the $\mathcal{S}$-lemma (when $L(\boldsymbol{\xi})$ is quadratic in $\boldsymbol{\xi}$). Assume next that $h > 0$. Then, the semi-infinite constraint in (5.18) becomes redundant and, since $\boldsymbol{\Omega} \succ \mathbf{0}$, the optimal solution of (5.18) is given by $\mathbf{M} = \mathbf{0}$ with a corresponding optimal value of 0. The optimal value of problem (5.19) is also equal to 0. Indeed, by choosing $\tau = 1/h$, the semi-infinite constraint in (5.19) is satisfied for any $\mathbf{M} \succcurlyeq \mathbf{0}$. Finally, note that (5.18) and (5.19) may be different for

$h = 0$.

Since (5.18) and (5.19) are equivalent for all but one value of γ and since their optimal values are nonincreasing in γ, we can express WC-VaR$_\epsilon(L(\tilde{\boldsymbol{\xi}}))$ in (5.17) as

$$
\begin{aligned}
\text{WC-VaR}_\epsilon(L(\tilde{\boldsymbol{\xi}})) \;=\; \inf \quad & \gamma \\
\text{s.t.} \quad & \mathbf{M} \in \mathbb{S}^{k+1}, \quad \tau \in \mathbb{R}, \quad \gamma \in \mathbb{R} \\
& \langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq \epsilon, \quad \mathbf{M} \succcurlyeq \mathbf{0}, \quad \tau \geq 0 \\
& \left[\boldsymbol{\xi}^\intercal \; 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\intercal \; 1\right]^\intercal - 1 + \tau\left(\gamma - L(\boldsymbol{\xi})\right) \geq 0 \quad \forall \boldsymbol{\xi} \in \mathbb{R}^k.
\end{aligned}
\tag{5.20}
$$

It can easily be shown that $\langle \boldsymbol{\Omega}, \mathbf{M} \rangle \geq 1$ for any feasible solution of (5.20) with vanishing τ-component. However, since $\epsilon < 1$, this is in conflict with the constraint $\langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq \epsilon$. We thus conclude that no feasible point can have a vanishing τ-component. This allows us to divide the semi-infinite constraint in problem (5.20) by τ. Subsequently we perform variable substitutions in which we replace τ by $1/\tau$ and $\mathbf{M}$ by $\mathbf{M}/\tau$. This yields the following reformulation of problem (5.20).

$$
\begin{aligned}
\text{WC-VaR}_\epsilon(L(\tilde{\boldsymbol{\xi}})) \;=\; \inf \quad & \gamma \\
\text{s.t.} \quad & \mathbf{M} \in \mathbb{S}^{k+1}, \quad \tau \in \mathbb{R}, \quad \gamma \in \mathbb{R} \\
& \tfrac{1}{\epsilon}\langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq \tau, \quad \mathbf{M} \succcurlyeq \mathbf{0}, \quad \tau \geq 0 \\
& \left[\boldsymbol{\xi}^\intercal \; 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\intercal \; 1\right]^\intercal - \tau + \gamma - L(\boldsymbol{\xi}) \geq 0 \quad \forall \boldsymbol{\xi} \in \mathbb{R}^k
\end{aligned}
$$

Note that, since $\boldsymbol{\Omega} \succ \mathbf{0}$ and $\mathbf{M} \succcurlyeq \mathbf{0}$, we have $\frac{1}{\epsilon}\langle \boldsymbol{\Omega}, \mathbf{M} \rangle \geq 0$. This allows us to remove the redundant nonnegativity constraint on τ. We now introduce a new decision variable $\beta = \gamma - \tau$, which allows us to eliminate γ.

$$
\begin{aligned}
\text{WC-VaR}_\epsilon(L(\tilde{\boldsymbol{\xi}})) \;=\; \inf \quad & \beta + \tau \\
\text{s.t.} \quad & \mathbf{M} \in \mathbb{S}^{k+1}, \quad \tau \in \mathbb{R}, \quad \beta \in \mathbb{R} \\
& \tfrac{1}{\epsilon}\langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq \tau, \quad \mathbf{M} \succcurlyeq \mathbf{0} \\
& \left[\boldsymbol{\xi}^\intercal \; 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\intercal \; 1\right]^\intercal + \beta - L(\boldsymbol{\xi}) \geq 0 \quad \forall \boldsymbol{\xi} \in \mathbb{R}^k
\end{aligned}
$$

Note that at optimality $\tau = \frac{1}{\epsilon}\langle \mathbf{\Omega}, \mathbf{M}\rangle$, which finally allows us to express WC-VaR$_\epsilon(L(\tilde{\boldsymbol{\xi}}))$ as

$$\begin{aligned} \text{WC-VaR}_\epsilon(L(\tilde{\boldsymbol{\xi}})) \;=\; \inf \quad & \beta + \tfrac{1}{\epsilon}\langle \mathbf{\Omega}, \mathbf{M}\rangle \\ \text{s.t.} \quad & \mathbf{M} \in \mathbb{S}^{k+1}, \quad \beta \in \mathbb{R}, \quad \mathbf{M} \succcurlyeq \mathbf{0} \\ & \left[\boldsymbol{\xi}^\intercal \; 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\intercal \; 1\right]^\intercal + \beta - L(\boldsymbol{\xi}) \geq 0 \quad \forall \boldsymbol{\xi} \in \mathbb{R}^k. \end{aligned} \tag{5.21}$$

Recall now that by Lemma 5.2.1 we have

$$\begin{aligned} \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-CVaR}_\epsilon\left(L(\tilde{\boldsymbol{\xi}})\right) &= \inf_{\beta\in\mathbb{R}} \quad \left\{\beta + \frac{1}{\epsilon}\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left((L(\tilde{\boldsymbol{\xi}}) - \beta)^+\right)\right\} \\ &= \inf \quad \beta + \tfrac{1}{\epsilon}\langle \mathbf{\Omega}, \mathbf{M}\rangle \\ &\quad\ \text{s.t.} \quad \mathbf{M} \in \mathbb{S}^{k+1}, \quad \beta \in \mathbb{R}, \quad \mathbf{M} \succcurlyeq \mathbf{0} \\ &\qquad\qquad \left[\boldsymbol{\xi}^\intercal \; 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\intercal \; 1\right]^\intercal + \beta - L(\boldsymbol{\xi}) \geq 0 \quad \forall \boldsymbol{\xi} \in \mathbb{R}^k, \end{aligned}$$

which is clearly equivalent to (5.21). This observation completes the proof. ∎

Corollary 5.2.1 *The following equivalence holds*

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-CVaR}_\epsilon\left(y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\intercal \tilde{\boldsymbol{\xi}}\right) \leq 0 \iff \inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\intercal \tilde{\boldsymbol{\xi}} \leq 0\right) \geq 1 - \epsilon,$$

which implies that $\mathcal{Z}^{\text{ICC}} = \mathcal{X}^{\text{ICC}}$.

Proof The claim follows immediately from Theorem 5.2.2 by observing that $L(\boldsymbol{\xi}) = y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\intercal \boldsymbol{\xi}$ is linear (and therefore concave) in $\boldsymbol{\xi}$. ∎

In the following example we demonstrate that the equivalence (5.14) can fail to hold even if the loss function L is convex and piecewise linear in $\boldsymbol{\xi}$.

Example 5.2.1 *Let $\tilde{\xi}$ be a scalar random variable with mean $\mu = 0$ and standard deviation $\sigma = 1$. Moreover, let $\mathcal{P}$ be the set of all probability distributions on $\mathbb{R}$ consistent with the given mean and standard deviation. Consider now the loss function $L(\xi) = \max\{\xi - 1, 4\xi - 4\}$, and note that L is strictly increasing and convex in ξ. In particular, L is neither concave* *nor quadratic and thus falls outside the scope of Theorem 5.2.2. We now show that for this particular L the Worst-Case CVaR constraint* $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-CVaR}_{\frac{1}{2}}(L(\tilde{\xi})) \leq 0$ *is violated even though the distributionally robust individual chance constraint* $\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}(L(\tilde{\xi}) \leq 0) \geq 1/2$ *is satisfied. To this end, we note that the Chebychev inequality* $\mathbb{P}(\tilde{\xi} - \mu \geq \kappa\sigma) \leq 1/(1+\kappa^2)$ *for* $\kappa = 1$ *implies*

$$\begin{aligned}
\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(\tilde{\xi} \geq 1\right) \leq \frac{1}{2} \iff & \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(L(\tilde{\xi}) \geq L(1) = 0\right) \leq \frac{1}{2} \\
\implies & \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(L(\tilde{\xi}) > 0\right) \leq \frac{1}{2} \\
\iff & \inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(L(\tilde{\xi}) \leq 0\right) \geq \frac{1}{2},
\end{aligned}$$

where the first equivalence follows from the monotonicity of L. Assume now that the true distribution $\mathbb{Q}$ *of* $\tilde{\xi}$ *is discrete and defined through* $\mathbb{Q}(\tilde{\xi} = -2) = 1/8$, $\mathbb{Q}(\tilde{\xi} = 0) = 3/4$, *and* $\mathbb{Q}(\tilde{\xi} = 2) = 1/8$. *It is easy to verify that* $\mathbb{Q} \in \mathcal{P}$ *and that* $\mathbb{Q}\text{-CVaR}_{\frac{1}{2}}(L(\tilde{\xi})) = 0.25$. *Thus,* $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-CVaR}_{\frac{1}{2}}(L(\tilde{\xi})) \geq 0.25 > 0$. *We therefore conclude that the Worst-Case CVaR constraint is not equivalent to the robust chance constraint.*

5.2.3 Robust Optimization Perspective on Individual Chance Constraints

There exists a close relationship between distributionally robust chance constrained programming and straight robust optimization, a powerful modeling paradigm for decision problems subject to non-stochastic data uncertainty, see, e.g., [BTEGN09, BTN98, BTN99]. In order to elicit this connection, we consider the following semi-infinite constraint.

$$y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\xi} \leq 0 \quad \forall \boldsymbol{\xi} \in \mathcal{U}_\epsilon \tag{5.22}$$

In contrast to the chance constraint (5.5), which requires the underlying linear inequality to be satisfied with a certain probability, the above semi-infinite constraint forces the inequality to be satisfied for all realizations of $\tilde{\boldsymbol{\xi}}$ within a prescribed *uncertainty set* $\mathcal{U}_\epsilon$. The shape of the uncertainty set should reflect the modeler's knowledge about the distribution of the random vector $\tilde{\boldsymbol{\xi}}$, e.g., full or partial information about its support and its moments. Recall that in our current setting only the first- and second-order moments of $\tilde{\boldsymbol{\xi}}$ are assumed to be known. Moreover, the size of $\mathcal{U}_\epsilon$ should be chosen in such a way as to guarantee that the set of all $\boldsymbol{x} \in \mathbb{R}^n$ feasible in (5.22) approximates $\mathcal{X}^{\text{ICC}}$.

Chen *et al.* [CSSC09] have derived an uncertainty set with these desirable properties. Inspired by their approach, we derive an uncertainty set with the strong property that (5.5) and (5.22) are indeed equivalent. The following theorem constitutes a first step towards this goal.

Theorem 5.2.3 *For any fixed* $\boldsymbol{x} \in \mathcal{X}$*, the robust chance constraint* (5.5) *is equivalent to the semi-infinite constraint*

$$\left\langle \begin{bmatrix} \mathbf{0} & \frac{1}{2}\boldsymbol{y}(\boldsymbol{x}) \\ \frac{1}{2}\boldsymbol{y}(\boldsymbol{x})^\mathsf{T} & y^0(\boldsymbol{x}) \end{bmatrix}, \boldsymbol{\Lambda} \right\rangle \leq 0 \quad \forall \boldsymbol{\Lambda} \in \mathcal{U}_\epsilon^1, \tag{5.23}$$

where the uncertainty set $\mathcal{U}_\epsilon^1 \subseteq \mathbb{S}^{k+1}$ *is defined as*

$$\mathcal{U}_\epsilon^1 = \left\{ \boldsymbol{\Lambda} \in \mathbb{S}^{k+1} \;:\; \boldsymbol{\Lambda} \succcurlyeq \mathbf{0}, \quad \frac{1}{\epsilon}\boldsymbol{\Omega} - \boldsymbol{\Lambda} \succcurlyeq \mathbf{0}, \quad \Lambda_{k+1,k+1} = 1 \right\}. \tag{5.24}$$

Proof: The equivalence between robust individual chance constraints and Worst-Case CVaR constraints holds for any fixed $\boldsymbol{x} \in \mathcal{X}$. The Worst-Case CVaR can be computed by solving the SDP (5.13). After a few elementary simplification steps the dual of problem (5.13) reduces to

$$\begin{aligned} \sup_{\boldsymbol{\Lambda} \in \mathbb{S}^{k+1}} \quad & \left\langle \begin{bmatrix} \mathbf{0} & \frac{1}{2}\boldsymbol{y}(\boldsymbol{x}) \\ \frac{1}{2}\boldsymbol{y}(\boldsymbol{x})^\mathsf{T} & y^0(\boldsymbol{x}) \end{bmatrix}, \boldsymbol{\Lambda} \right\rangle \\ \text{s.t.} \quad & \boldsymbol{\Lambda} \succcurlyeq \mathbf{0}, \quad \frac{1}{\epsilon}\boldsymbol{\Omega} - \boldsymbol{\Lambda} \succcurlyeq \mathbf{0}, \quad \Lambda_{k+1,k+1} = 1. \end{aligned} \tag{5.25}$$

Note that strong duality holds because the primal problem (5.13) is convex and the dual problem (5.25) is strictly feasible for any $\epsilon \in (0,1)$ since $\boldsymbol{\Omega} \succ \mathbf{0}$ (as a result of $\boldsymbol{\Sigma} \succ \mathbf{0}$). Constraining the Worst-Case CVaR to be nonpositive is therefore equivalent to requiring that the optimal objective value of problem (5.25) is nonpositive, which is manifestly equivalent to the semi-infinite constraint (5.23). ∎

Next, we demonstrate that the uncertainty set $\mathcal{U}_\epsilon^1$, which is embedded in the space of positive semidefinite matrices $\mathbb{S}^{k+1}$, is closely related to an *ellipsoidal uncertainty set* in the $\boldsymbol{\xi}$-space.

Corollary 5.2.2 *For any fixed* $\boldsymbol{x} \in \mathcal{X}$*, the robust chance constraint* (5.5) *is equivalent to the semi-infinite constraint*

$$y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\xi} \leq 0 \quad \forall \boldsymbol{\xi} \in \mathcal{U}_\epsilon^{\text{ell}}, \tag{5.26}$$

where

$$\mathcal{U}_\epsilon^{\text{ell}} = \left\{ \boldsymbol{\xi} \in \mathbb{R}^k \ : \ (\boldsymbol{\xi} - \boldsymbol{\mu})^\mathsf{T} \boldsymbol{\Sigma}^{-1} (\boldsymbol{\xi} - \boldsymbol{\mu}) \leq \frac{1-\epsilon}{\epsilon} \right\}.$$

Proof: By expanding the trace in (5.23), we can reformulate this semi-infinite constraint as

$$y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\xi} \leq 0 \quad \forall \boldsymbol{\xi} \in \text{proj}_{\boldsymbol{\xi}}(\mathcal{U}_\epsilon^1),$$

where $\text{proj}_{\boldsymbol{\xi}}(\mathcal{U}_\epsilon^1)$ denotes the projection of the uncertainty set $\mathcal{U}_\epsilon^1$ to the space $\mathbb{R}^k$ and is defined as

$$\text{proj}_{\boldsymbol{\xi}}(\mathcal{U}_\epsilon^1) = \left\{ \boldsymbol{\xi} \in \mathbb{R}^k \ : \ \exists \mathbf{X} \in \mathbb{S}^k, \quad \begin{bmatrix} \mathbf{X} & \boldsymbol{\xi} \\ \boldsymbol{\xi}^\mathsf{T} & 1 \end{bmatrix} \succcurlyeq \mathbf{0}, \quad \frac{1}{\epsilon}\boldsymbol{\Omega} - \begin{bmatrix} \mathbf{X} & \boldsymbol{\xi} \\ \boldsymbol{\xi}^\mathsf{T} & 1 \end{bmatrix} \succcurlyeq \mathbf{0} \right\}.$$

The equivalence of $\text{proj}_{\boldsymbol{\xi}}(\mathcal{U}_\epsilon^1)$ and $\mathcal{U}_\epsilon^{ell}$ follows directly from Corollary 4.5.1. ∎

Lemma 5.2.2 allows us to interpret the uncertainty set $\mathcal{U}_\epsilon^1$ defined in (5.24) as a version of the ellipsoidal uncertainty set $\mathcal{U}_\epsilon^{\text{ell}}$ lifted to the space of positive semidefinite matrices $\mathbb{S}^{k+1}$. We emphasize that the simplification of (5.23) to (5.26) in Corollary 5.2.2 is only possible because the inequality in the chance constraint (5.5) is linear in the random vector $\tilde{\boldsymbol{\xi}}$. We further remark that the semi-infinite constraint (5.26) can be expressed as a single SOCP constraint, see, e.g., El Ghaoui *et al.* [EGOO03]. The importance of the lifted uncertainty set $\mathcal{U}_\epsilon^1$ will become evident in the next section, where we will show that the uncertainty set associated with a distributionally robust *joint* chance constraint of type (5.3) is not ellipsoidal but can be interpreted as a generalization of $\mathcal{U}_\epsilon^1$.

5.3 Distributionally Robust Joint Chance Constraints

We define the feasible set $\mathcal{X}^{\mathrm{JCC}}$ of the distributionally robust *joint* chance constraint (5.3) as

$$\mathcal{X}^{\mathrm{JCC}} = \left\{ \boldsymbol{x} \in \mathbb{R}^n \ : \ \inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T}\tilde{\boldsymbol{\xi}} \le 0 \quad \forall i = 1,\dots,m \right) \ge 1-\epsilon \right\}.$$

The aim of this section is to investigate the structure of $\mathcal{X}^{\mathrm{JCC}}$ and to elaborate tractable conservative approximations. We first review two existing approximations and discuss their benefits and shortcomings.

5.3.1 The Bonferroni Approximation

A popular approximation for $\mathcal{X}^{\mathrm{JCC}}$ is based on Bonferroni's inequality. Note that the robust joint chance constraint (5.3) is equivalent to

$$\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(\bigcap_{i=1}^m \left\{ y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T}\tilde{\boldsymbol{\xi}} \le 0 \right\} \right) \ge 1-\epsilon \iff \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(\bigcup_{i=1}^m \left\{ y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T}\tilde{\boldsymbol{\xi}} > 0 \right\} \right) \le \epsilon.$$

Furthermore, Bonferroni's inequality implies that

$$\mathbb{P}\left(\bigcup_{i=1}^m \left\{ y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T}\tilde{\boldsymbol{\xi}} > 0 \right\} \right) \le \sum_{i=1}^m \mathbb{P}\left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T}\tilde{\boldsymbol{\xi}} > 0 \right) \quad \forall \mathbb{P} \in \mathcal{P}.$$

For any vector of safety factors $\boldsymbol{\epsilon} \in \mathcal{E} = \{\boldsymbol{\epsilon} \in \mathbb{R}^m_+ \ : \ \sum_{i=1}^m \epsilon_i \le \epsilon\}$, the system of distributionally robust *individual* chance constraints

$$\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T}\tilde{\boldsymbol{\xi}} \le 0 \right) \ge 1-\epsilon_i \quad \forall i = 1,\dots,m \tag{5.27}$$

represents a conservative approximation for the distributionally robust *joint* chance constraint (5.3). By Theorem 5.2.1, we can reformulate each of the individual chance constraints in (5.27) in terms of tractable LMIs. In fact, as explained at the end of Section 5.2.3, we can further reduce these LMIs to SOCP constraints, but this further simplification is irrelevant for our purposes.

Thus, for any $\boldsymbol{\epsilon} \in \mathcal{E}$, the assertion that $\boldsymbol{x} \in \mathcal{Z}_{\mathrm{B}}^{\mathrm{JCC}}(\boldsymbol{\epsilon})$, where

$$\mathcal{Z}_{\mathrm{B}}^{\mathrm{JCC}}(\boldsymbol{\epsilon}) = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \begin{array}{ll} \exists (\beta_i, \mathbf{M}_i) \in \mathbb{R} \times \mathbb{S}^{k+1} & \forall i = 1, \ldots, m, \\ \mathbf{M}_i \succcurlyeq \mathbf{0}, \quad \beta_i + \frac{1}{\epsilon_i} \langle \boldsymbol{\Omega}, \mathbf{M}_i \rangle \leq 0 & \forall i = 1, \ldots, m, \\ \mathbf{M}_i - \begin{bmatrix} \mathbf{0} & \frac{1}{2}\boldsymbol{y}_i(\boldsymbol{x}) \\ \frac{1}{2}\boldsymbol{y}_i(\boldsymbol{x})^\intercal & y_i^0(\boldsymbol{x}) - \beta_i \end{bmatrix} \succcurlyeq \mathbf{0} & \forall i = 1, \ldots, m \end{array} \right\},$$

is a sufficient condition to guarantee that $\boldsymbol{x}$ satisfies the original distributionally robust joint chance constraint (5.3). The above arguments culminate in the following result.

Theorem 5.3.1 (Bonferroni Approximation) *For any $\boldsymbol{\epsilon} \in \mathcal{E}$ we have $\mathcal{Z}_{\mathrm{B}}^{\mathrm{JCC}}(\boldsymbol{\epsilon}) \subseteq \mathcal{X}^{\mathrm{JCC}}$.*

A major shortcoming of the Bonferroni approximation is that the approximation quality depends critically on the choice of $\boldsymbol{\epsilon} \in \mathcal{E}$. Unfortunately, the problem of finding the best $\boldsymbol{\epsilon} \in \mathcal{E}$ for a generic chance constrained problem of type (5.4) is nonconvex and believed to be intractable [NS06]. As a result, in most applications of Bonferroni's inequality the "risk budget" ϵ is equally divided among the m individual chance constraints in (5.27) by setting $\epsilon_i = \epsilon/m$ for $i = 1, \ldots, m$. This approach was first advocated by Nemirovski and Shapiro [NS06].

The Bonferroni approximation can be overly conservative even if $\boldsymbol{\epsilon} \in \mathcal{E}$ is chosen optimally. The following example, which is adapted from Chen *et al.* [CSSC09], highlights this shortcoming.

Example 5.3.1 *Assume that the inequalities in the chance constraint* (5.3) *are perfectly positively correlated in the sense that*

$$y_i^0(\boldsymbol{x}) = \delta_i \hat{y}^0(\boldsymbol{x}) \quad \textit{and} \quad y_i(\boldsymbol{x}) = \delta_i \hat{y}(\boldsymbol{x})$$

for some affine functions $\hat{y}^0 : \mathbb{R}^n \to \mathbb{R}$ and $\hat{y} : \mathbb{R}^n \to \mathbb{R}^k$ and for some fixed constants $\delta_i > 0$ for $i = 1, \ldots, m$. In this case, it can readily be seen that the joint chance constraint (5.3) *is equivalent to the robust individual chance constraint*

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\left(y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\intercal \tilde{\boldsymbol{\xi}} \leq 0 \right) \geq 1 - \epsilon. \tag{5.28}$$

Thus, the least conservative choice for ϵ_i *which guarantees that* (5.27) *implies* (5.3) *is* $\epsilon_i = \epsilon$ *for* $i = 1, \ldots, m$. *However, this means that the* ϵ_i *sum to* $m\epsilon$ *instead of* ϵ *as required by the Bonferroni approximation. In fact, the optimal choice for* $\boldsymbol{\epsilon} \in \mathcal{E}$ *is* $\epsilon_i = \epsilon/m$ *for* $i = 1, \ldots, m$. *This example demonstrates that the quality of the Bonferroni approximation diminishes as* m *increases if the inequalities in the joint chance constraint are positively correlated.*

5.3.2 Approximation by Chen, Sim, Sun and Teo

In order to mitigate the potential over-conservatism of the Bonferroni approximation, Chen *et al.* [CSSC09] proposed an approximation based on a different inequality from probability theory. The starting point is the observation that the joint chance constraint (5.3) can be reformulated as

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\left(\max_{i=1,\ldots,m} \left\{ \alpha_i \left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T} \tilde{\boldsymbol{\xi}} \right) \right\} \leq 0 \right) \geq 1 - \epsilon \tag{5.29}$$

for any vector of strictly positive *scaling parameters* $\boldsymbol{\alpha} \in \mathcal{A} = \{\boldsymbol{\alpha} \in \mathbb{R}^m \ : \ \boldsymbol{\alpha} > \mathbf{0}\}$. Note that the choice of $\boldsymbol{\alpha} \in \mathcal{A}$ does not affect the feasible region of the chance constraint (5.29). Although these scaling parameters are seemingly unnecessary, it turns out that they can be tuned to improve the approximation to be developed below. We emphasize that (5.29) represents a distributionally robust *individual* chance constraint, which can be conservatively approximated by a Worst-Case CVaR constraint. Thus, for any $\boldsymbol{\alpha} \in \mathcal{A}$, the requirement

$$\boldsymbol{x} \in \mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha}) = \left\{ \boldsymbol{x} \in \mathbb{R}^n \ : \ \sup_{\mathbb{P} \in \mathcal{P}} \mathrm{CVaR}_\epsilon \left(\max_{i=1,\ldots,m} \left\{ \alpha_i \left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T} \tilde{\boldsymbol{\xi}} \right) \right\} \right) \leq 0 \right\} \tag{5.30}$$

implies that $\boldsymbol{x} \in \mathcal{X}^{\mathrm{JCC}}$, see Proposition 5.2.1. It is important to note that, in contrast to the chance constraint (5.29), the Worst-Case CVaR constraint $\boldsymbol{x} \in \mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ *does* depend on the choice of $\boldsymbol{\alpha} \in \mathcal{A}$. Thus, the Worst-Case CVaR constraint in (5.30) is not equivalent to the robust chance constraint (5.29) since the max function in (5.29) is convex piecewise linear, see also Theorem 5.2.2 and Example 5.2.1.

The following theorem due to Chen *et al.* [CSSC09] provides a tractable conservative approximation for $\mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$.

Theorem 5.3.2 (Approximation by Chen *et al.*) *For any* $\boldsymbol{\alpha} \in \mathcal{A}$ *we have* $\mathcal{Z}_{\mathrm{U}}^{\mathrm{JCC}}(\boldsymbol{\alpha}) \subseteq \mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha}) \subseteq \mathcal{X}^{\mathrm{JCC}}$ *where*

$$\mathcal{Z}_{\mathrm{U}}^{\mathrm{JCC}}(\boldsymbol{\alpha}) = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \begin{array}{ll} \exists \beta \in \mathbb{R}, \quad w^0 \in \mathbb{R}, \quad \boldsymbol{w} \in \mathbb{R}^k, \quad \{\mathbf{M}_i\}_{i=0}^m \in \mathbb{S}^{k+1}, & \\ \beta + \frac{1}{\epsilon}\left(\sum_{i=0}^m \langle \boldsymbol{\Omega}, \mathbf{M}_i \rangle\right) \leq 0, \ \mathbf{M}_0 - \begin{bmatrix} \mathbf{0} & \frac{1}{2}\boldsymbol{w} \\ \frac{1}{2}\boldsymbol{w}^\mathsf{T} & w^0 - \beta \end{bmatrix} \succcurlyeq \mathbf{0} & \\ \mathbf{M}_i - \begin{bmatrix} \mathbf{0} & \frac{1}{2}(\alpha_i \boldsymbol{y}_i(\boldsymbol{x}) - \boldsymbol{w}) \\ \frac{1}{2}(\alpha_i \boldsymbol{y}_i(\boldsymbol{x}) - \boldsymbol{w})^\mathsf{T} & \alpha_i y_i^0(\boldsymbol{x}) - w^0 \end{bmatrix} \succcurlyeq \mathbf{0} & \forall i = 1, \ldots, m \\ \mathbf{M}_i \succcurlyeq \mathbf{0} & \forall i = 0, \ldots, m, \end{array} \right\}, \tag{5.31}$$

Proof The inclusion $\mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha}) \subseteq \mathcal{X}^{\mathrm{JCC}}$ follows from Proposition 5.2.1. To prove $\mathcal{Z}_{\mathrm{U}}^{\mathrm{JCC}}(\boldsymbol{\alpha}) \subseteq \mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$, we note that the constraint $\boldsymbol{x} \in \mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ is equivalent to $\mathcal{J}(\boldsymbol{x}, \boldsymbol{\alpha}) \leq 0$, where

$$\begin{aligned} \mathcal{J}(\boldsymbol{x}, \boldsymbol{\alpha}) &= \sup_{\mathbb{P} \in \mathcal{P}} \mathrm{CVaR}_\epsilon \left(\max_{i=1,\ldots,m} \left\{ \alpha_i \left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T} \tilde{\boldsymbol{\xi}} \right) \right\} \right) \\ &= \inf_{\beta \in \mathbb{R}} \left\{ \beta + \frac{1}{\epsilon} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_\mathbb{P} \left(\left[\max_{i=1,\ldots,m} \left\{ \alpha_i \left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T} \tilde{\boldsymbol{\xi}} \right) \right\} - \beta \right]^+ \right) \right\} \ \leq \ 0. \end{aligned} \tag{5.32}$$

Due to a classical result in order statistics by Meilijson and Nadas [MN79], we have

$$\begin{aligned} & \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_\mathbb{P} \left(\left[\max_{i=1,\ldots,m} \left\{ \alpha_i \left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T} \tilde{\boldsymbol{\xi}} \right) \right\} - \beta \right]^+ \right) \\ \leq \ & \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_\mathbb{P} \left((w^0 + \boldsymbol{w}^\mathsf{T} \tilde{\boldsymbol{\xi}} - \beta)^+ \right) + \sum_{i=1}^m \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_\mathbb{P} \left(\left[\alpha_i y_i^0(\boldsymbol{x}) - w^0 + (\alpha_i \boldsymbol{y}_i(\boldsymbol{x}) - \boldsymbol{w})^\mathsf{T} \tilde{\boldsymbol{\xi}} \right]^+ \right) \\ = \ & G(w^0, \boldsymbol{w}, \beta, \boldsymbol{\alpha}, \boldsymbol{x}) \end{aligned}$$

for any fixed $(w^0, \boldsymbol{w}) \in \mathbb{R} \times \mathbb{R}^k$. This estimate provides the following upper bound on the Worst-Case CVaR.

$$\mathcal{J}(\boldsymbol{x}, \boldsymbol{\alpha}) \ \leq \ \hat{\mathcal{J}}(\boldsymbol{x}, \boldsymbol{\alpha}) \ = \inf_{\beta \in \mathbb{R}, \ w^0 \in \mathbb{R}, \ \boldsymbol{w} \in \mathbb{R}^k} \left\{ \beta + \frac{1}{\epsilon} G(w^0, \boldsymbol{w}, \beta, \boldsymbol{\alpha}, \boldsymbol{x}) \right\} \tag{5.33}$$

The evaluation of $\hat{\mathcal{J}}(\boldsymbol{x}, \boldsymbol{\alpha})$ involves the solution of $m+1$ subordinate worst-case expectation problems, all of which have equivalent tractable SDP formulations of the type (5.12). This enables us to reformulate $\hat{\mathcal{J}}(\boldsymbol{x},\boldsymbol{\alpha})$ as

$$\begin{aligned} \hat{\mathcal{J}}(\boldsymbol{x},\boldsymbol{\alpha}) \quad = \quad \inf \quad & \beta + \frac{1}{\epsilon}\left(\sum_{i=0}^{m}\langle \boldsymbol{\Omega}, \mathbf{M}_i\rangle\right) \\ \text{s.t.} \quad & \boldsymbol{w}\in\mathbb{R}^k, \quad w^0\in\mathbb{R}, \quad \beta\in\mathbb{R}, \quad \mathbf{M}_i\in\mathbb{S}^{k+1} \quad \forall i=0,\dots,m \\ & \mathbf{M}_i - \begin{bmatrix} \mathbf{0} & \frac{1}{2}(\alpha_i\boldsymbol{y}_i(\boldsymbol{x})-\boldsymbol{w}) \\ \frac{1}{2}(\alpha_i\boldsymbol{y}_i(\boldsymbol{x})-\boldsymbol{w})^\intercal & \alpha_i y_i^0(\boldsymbol{x})-w^0 \end{bmatrix} \succcurlyeq \mathbf{0} \quad \forall i=1,\dots,m \\ & \mathbf{M}_0 - \begin{bmatrix} \mathbf{0} & \frac{1}{2}\boldsymbol{w} \\ \frac{1}{2}\boldsymbol{w}^\intercal & w^0-\beta \end{bmatrix} \succcurlyeq \mathbf{0}, \quad \mathbf{M}_i\succcurlyeq\mathbf{0} \quad \forall i=0,\dots,m. \end{aligned} \tag{5.34}$$

Thus, the assertion $\boldsymbol{x}\in\mathcal{Z}_{\mathrm{U}}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ implies $\boldsymbol{x}\in\mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$, and we conclude that $\mathcal{Z}_{\mathrm{U}}^{\mathrm{JCC}}(\boldsymbol{\alpha})\subseteq\mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ for all $\boldsymbol{\alpha}\in\mathcal{A}$. ∎

Note that, since the feasible set $\mathcal{Z}_{\mathrm{U}}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ constitutes a tractable conservative approximation for $\mathcal{X}^{\mathrm{JCC}}$ for any $\boldsymbol{\alpha}\in\mathcal{A}$, the union $\bigcup_{\boldsymbol{\alpha}\in\mathcal{A}}\mathcal{Z}_{\mathrm{U}}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ still constitutes a conservative approximation for $\mathcal{X}^{\mathrm{JCC}}$. Chen *et al.* [CSSC09] prove also that their approximation is tighter than the Bonferroni approximation by showing that $\mathcal{Z}_{\mathrm{B}}^{\mathrm{JCC}}(\boldsymbol{\epsilon})\subseteq\bigcup_{\boldsymbol{\alpha}\in\mathcal{A}}\mathcal{Z}_{\mathrm{U}}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ for all $\boldsymbol{\epsilon}\in\mathcal{E}$. Unfortunately, similar to the Bonferroni approach, the approximation by Chen *et al.* depends critically on the choice of $\boldsymbol{\alpha}$, while the problem of finding the best $\boldsymbol{\alpha}\in\mathcal{A}$ for a generic chance constrained program of the type (5.4) is nonconvex and therefore believed to be intractable.

5.3.3 The Worst-Case CVaR Approximation

Both approximations discussed so far rely on inequalities from probability theory, which are not necessarily tight. In this section we show that the set $\mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ has in fact an exact tractable representation in terms of LMIs. This result relies on the solution of a moment problem that allows us to determine an exact reformulation of the Worst-Case CVaR in (5.32), and it leads to a tighter convex approximation for the feasible set $\mathcal{X}^{\mathrm{JCC}}$.

Theorem 5.3.3 *For any fixed* $\boldsymbol{x} \in \mathbb{R}^n$ *and* $\boldsymbol{\alpha} \in \mathcal{A}$, *we have*

$$\mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha}) = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \begin{array}{l} \exists (\beta, \mathbf{M}) \in \mathbb{R} \times \mathbb{S}^{k+1}, \\ \beta + \frac{1}{\epsilon}\langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq 0, \quad \mathbf{M} \succcurlyeq \mathbf{0}, \\ \mathbf{M} - \begin{bmatrix} \mathbf{0} & \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x}) \\ \frac{1}{2}\alpha_i \boldsymbol{y}_i^\intercal & \alpha_i y_i^0(\boldsymbol{x}) - \beta \end{bmatrix} \succcurlyeq \mathbf{0} \quad \forall i = 1, \ldots, m \end{array} \right\}. \tag{5.35}$$

Proof As in Section 5.2, the first step towards a tractable reformulation of the Worst-Case CVaR in the definition of $\mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ is to solve the worst-case expectation problem

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left(\left[\max_{i=1,\ldots,m} \left\{ \alpha_i \left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\intercal \tilde{\boldsymbol{\xi}} \right) \right\} - \beta \right]^+ \right). \tag{5.36}$$

For any fixed $\boldsymbol{x} \in \mathcal{X}$, $\beta \in \mathbb{R}$, and $\boldsymbol{\alpha} \in \mathcal{A}$, Lemma 5.2.1 enables us to reformulate (5.36) as

$$\begin{array}{lll} \inf\limits_{\mathbf{M} \in \mathbb{S}^{k+1}} & \langle \boldsymbol{\Omega}, \mathbf{M} \rangle \\ \text{s.t.} & \mathbf{M} \succcurlyeq \mathbf{0}, \quad \left[\boldsymbol{\xi}^\intercal \ 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\intercal \ 1\right]^\intercal \geq \max\limits_{i=1,\ldots,m} \left\{ \alpha_i \left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\intercal \tilde{\boldsymbol{\xi}} \right) \right\} - \beta \quad \forall \boldsymbol{\xi} \in \mathbb{R}^k. \end{array} \tag{5.37}$$

We emphasize that (5.37) represents a lossless reformulation of the worst-case expectation problem (5.36). The semi-infinite constraint in (5.37) can be expanded into m simpler semi-infinite constraints of the form

$$\left[\boldsymbol{\xi}^\intercal \ 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\intercal \ 1\right]^\intercal \geq \alpha_i \left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\intercal \boldsymbol{\xi} \right) - \beta \quad \forall \boldsymbol{\xi} \in \mathbb{R}^k, \ i = 1, \ldots, m.$$

Using similar arguments as in Section 5.2, these semi-infinite constraints can be equivalently expressed as the following system of LMIs.

$$\mathbf{M} - \begin{bmatrix} \mathbf{0} & \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x}) \\ \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x})^\intercal & \alpha_i y_i^0(\boldsymbol{x}) - \beta \end{bmatrix} \succcurlyeq \mathbf{0} \quad \forall i = 1, \ldots, m$$

We can therefore reformulate the worst-case expectation problem (5.36) as

$$\begin{array}{ll} \inf\limits_{\mathbf{M}\in\mathbb{S}^{k+1}} & \langle \mathbf{\Omega}, \mathbf{M}\rangle \\ \text{s. t.} & \mathbf{M} \succcurlyeq \mathbf{0}, \quad \mathbf{M} - \begin{bmatrix} \mathbf{0} & \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x}) \\ \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T} & \alpha_i y_i^0(\boldsymbol{x}) - \beta \end{bmatrix} \succcurlyeq \mathbf{0} \quad \forall i = 1, \dots, m. \end{array} \tag{5.38}$$

Substituting (5.38) into (5.32) yields

$$\begin{array}{rcll} \mathcal{J}(\boldsymbol{x}, \boldsymbol{\alpha}) & = & \inf & \beta + \frac{1}{\epsilon}\langle \mathbf{\Omega}, \mathbf{M}\rangle \\ & & \text{s. t.} & \mathbf{M} \in \mathbb{S}^{k+1}, \quad \beta \in \mathbb{R} \\ & & & \mathbf{M} \succcurlyeq \mathbf{0}, \quad \mathbf{M} - \begin{bmatrix} \mathbf{0} & \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x}) \\ \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T} & \alpha_i y_i^0(\boldsymbol{x}) - \beta \end{bmatrix} \succcurlyeq \mathbf{0} \quad \forall i = 1, \dots, m, \end{array} \tag{5.39}$$

where $\mathcal{J}(\boldsymbol{x}, \boldsymbol{\alpha})$ is defined as in the proof of Theorem 5.3.2. Since $\mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha}) = \{\boldsymbol{x} \in \mathbb{R}^n : \mathcal{J}(\boldsymbol{x}, \boldsymbol{\alpha}) \leq 0\}$, the claim follows. ∎

Theorem 5.3.3 establishes that $\mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ has an exact representation in terms of LMIs. We have already seen in Section 5.3.2 that $\mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha}) \subseteq \mathcal{X}^{\mathrm{JCC}}$ for all $\boldsymbol{\alpha} \in \mathcal{A}$ and that $\mathcal{Z}_{\mathrm{U}}^{\mathrm{JCC}}(\boldsymbol{\alpha}) \subseteq \mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$, see Theorem 5.3.2. Thus, $\mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ constitutes a tractable conservative approximation for $\mathcal{X}^{\mathrm{JCC}}$ which is at least as tight as $\mathcal{Z}_{\mathrm{U}}^{\mathrm{JCC}}(\boldsymbol{\alpha})$.

Remark 5.3.1 *As a consistency check, we can verify that the constraints in* (5.31) *imply the constraint in* (5.35)*. To this end, assume first that* $\boldsymbol{x} \in \mathcal{Z}_{\mathrm{U}}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ *for some given* $\boldsymbol{\alpha} \in \mathcal{A}$*. This implies that there exists some* $(\beta, w^0, \boldsymbol{w}, \{\mathbf{M}_i\}_{i=0}^m)$ *that satisfy*

$$\mathbf{M}_0 + \mathbf{M}_i - \begin{bmatrix} \mathbf{0} & \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x}) \\ \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T} & \alpha_i y_i^0(\boldsymbol{x}) - \beta \end{bmatrix} \succcurlyeq \mathbf{0} \quad \forall i = 1, \dots, m, \tag{5.40a}$$

and

$$\beta + \frac{1}{\epsilon}\left(\sum_{i=0}^{m} \langle \mathbf{\Omega}, \mathbf{M}_i\rangle\right) \leq 0, \tag{5.40b}$$

see (5.31)*. We now prove that* $\boldsymbol{x} \in \mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ *by showing that* $\boldsymbol{\alpha}$, β, $\mathbf{M} = \sum_{i=0}^m \mathbf{M}_i$, *and* $\boldsymbol{x}$ *are* *feasible in* (5.35). *Firstly, since* $\mathbf{M}_i \succcurlyeq \mathbf{0}$ *for* $i = 0, \ldots, m$ *we have that* $\mathbf{M} \succcurlyeq \mathbf{0}$. *Secondly, it is easy to see that the inequality* (5.40b) *is equivalent to*

$$\beta + \frac{1}{\epsilon} \langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq 0.$$

Finally, since $\mathbf{M} = \sum_{i=0}^{m} \mathbf{M}_i \succcurlyeq \mathbf{M}_0 + \mathbf{M}_j$ *for all* $j = 1, \ldots, m$, *the inequalities in* (5.40a) *imply that*

$$\mathbf{M} - \begin{bmatrix} \mathbf{0} & \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x}) \\ \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x})^\intercal & \alpha_i y_i^0(\boldsymbol{x}) - \beta \end{bmatrix} \succcurlyeq \mathbf{0} \quad \forall i = 1, \ldots, m.$$

Thus, all the constraints in (5.35) *are satisfied.* ■

Recall from Section 5.3.2 that $\mathcal{Z}_{\mathrm{B}}^{\mathrm{JCC}}(\boldsymbol{\epsilon}) \subseteq \bigcup_{\boldsymbol{\alpha} \in \mathcal{A}} \mathcal{Z}_{\mathrm{U}}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ for all $\boldsymbol{\epsilon} \in \mathcal{E}$. Moreover, we have $\mathcal{Z}_{\mathrm{U}}^{\mathrm{JCC}}(\boldsymbol{\alpha}) \subseteq \mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha}) \subset \mathcal{X}^{\mathrm{JCC}}$ for all $\boldsymbol{\alpha} \in \mathcal{A}$. This allows us to conclude that our new approximation is at least as tight as the two state-of-the-art approximations discussed above.

Remark 5.3.2 *In contrast to the classical Bonferroni approximation, the new approximation behaves reasonably in situations in which the* m *inequalities in the chance constraint* (5.3) *are positively correlated. Indeed, by choosing* $\alpha_i := 1/\delta_i > 0$ *for all* $i = 1, \ldots, m$ *in Example 5.3.1, the constraint* $\boldsymbol{x} \in \mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ *is equivalent to*

$$\exists \beta \in \mathbb{R},\ \mathbf{M} \in \mathbb{S}^{k+1} \ : \ \beta + \frac{1}{\epsilon}\langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq 0, \quad \mathbf{M} \succcurlyeq \mathbf{0}, \quad \mathbf{M} - \begin{bmatrix} \mathbf{0} & \frac{1}{2}\boldsymbol{y}(\boldsymbol{x}) \\ \frac{1}{2}\boldsymbol{y}(\boldsymbol{x})^\intercal & y^0(\boldsymbol{x}) - \beta \end{bmatrix} \succcurlyeq \mathbf{0},$$

which can easily be identified as the SDP reformulation of the individual chance constraint (5.28). *This implies that* $\mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha}) = \mathcal{X}^{\mathrm{ICC}}$ *for all* $\boldsymbol{\alpha} \in \mathcal{A}$ *in Example 5.3.1, see also Theorem 5.2.1. Thus, by choosing* $\boldsymbol{\alpha}$ *appropriately, our method can provide tight approximations for distributionally robust joint chance constraints, even in situations when the* m *inequalities are positively correlated.*

5.3.4 Dual Interpretation of the Worst-Case CVaR Approximation

In this section we explore a different way to find a tractable conservative approximation for the chance constraint (5.3). Subsequently, we will prove that this approximation is equivalent to the Worst-Case CVaR approximation.

Consider again the robust *individual* chance constraint (5.29) which is equivalent to the robust *joint* chance constraint (5.3) for any fixed $\boldsymbol{\alpha} \in \mathcal{A}$. Instead of approximating (5.29) by a Worst-Case CVaR constraint, we can approximate the max-function in the chance constraint (5.29) by a quadratic majorant of the form $q(\boldsymbol{\xi}) = \boldsymbol{\xi}^\mathsf{T}\mathbf{Q}\boldsymbol{\xi} + \boldsymbol{\xi}^\mathsf{T}\boldsymbol{q} + q^0$ that satisfies

$$\begin{aligned} & q(\boldsymbol{\xi}) \geq \max_{i=1,\dots,m} \left\{\alpha_i \left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T}\boldsymbol{\xi}\right)\right\} \quad \forall \boldsymbol{\xi} \in \mathbb{R}^k, \\ \iff \quad & q(\boldsymbol{\xi}) \geq \alpha_i \left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T}\boldsymbol{\xi}\right) \quad \forall \boldsymbol{\xi} \in \mathbb{R}^k,\ i = 1, \dots, m. \end{aligned} \tag{5.41}$$

Replacing the max function in (5.29) by $q(\boldsymbol{\xi})$ yields the distributionally robust (individual) *quadratic* chance constraint

$$\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(\tilde{\boldsymbol{\xi}}^\mathsf{T}\mathbf{Q}\tilde{\boldsymbol{\xi}} + \tilde{\boldsymbol{\xi}}^\mathsf{T}\boldsymbol{q} + q^0 \leq 0\right) \geq 1 - \epsilon. \tag{5.42}$$

For further argumentation, we define

$$\mathcal{Z}_{\mathrm{Q}}^{\mathrm{JCC}}(\boldsymbol{\alpha}) = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \begin{array}{l} \exists \mathbf{Q} \in \mathbb{S}^k,\ \boldsymbol{q} \in \mathbb{R}^k,\ q^0 \in \mathbb{R} \quad \text{such that} \\ q(\boldsymbol{\xi}) = \boldsymbol{\xi}^\mathsf{T}\mathbf{Q}\boldsymbol{\xi} + \boldsymbol{\xi}^\mathsf{T}\boldsymbol{q} + q^0 \text{ satisfies (5.41) and (5.42)} \end{array} \right\}. \tag{5.43}$$

Proposition 5.3.1 *For any fixed $\boldsymbol{\alpha} \in \mathcal{A}$ the feasible set $\mathcal{Z}_{\mathrm{Q}}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ constitutes a conservative approximation for $\mathcal{X}^{\mathrm{JCC}}$, that is, $\mathcal{Z}_{\mathrm{Q}}^{\mathrm{JCC}}(\boldsymbol{\alpha}) \subseteq \mathcal{X}^{\mathrm{JCC}}$.*

Proof Note that any $\boldsymbol{x}$ feasible in (5.29) is also feasible in (5.43) since

$$\mathbb{P}\left(\tilde{\boldsymbol{\xi}}^\mathsf{T}\mathbf{Q}\tilde{\boldsymbol{\xi}} + \tilde{\boldsymbol{\xi}}^\mathsf{T}\boldsymbol{q} + q^0 \leq 0\right) \ \leq\ \mathbb{P}\left(\max_{i=1,\dots,m} \left\{\alpha_i(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T}\tilde{\boldsymbol{\xi}})\right\} \leq 0\right) \quad \forall \mathbb{P} \in \mathcal{P}.$$

Since $\boldsymbol{x}$ is feasible in (5.29) if and only if $\boldsymbol{x} \in \mathcal{X}^{\mathrm{JCC}}$, the claim follows. ∎

Theorem 5.3.4 *For any fixed $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{\alpha} \in \mathcal{A}$ we have*

$$\mathcal{Z}_{\mathrm{Q}}^{\mathrm{JCC}}(\boldsymbol{\alpha}) = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \begin{array}{l} \exists \mathbf{Q} \in \mathbb{S}^k, \quad \boldsymbol{q} \in \mathbb{R}^k, \quad q^0 \in \mathbb{R}, \quad \beta \in \mathbb{R}, \quad \mathbf{M} \in \mathbb{S}^{k+1}, \\ \beta + \frac{1}{\epsilon}\langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq 0, \quad \mathbf{M} \succcurlyeq \mathbf{0}, \quad \mathbf{M} - \begin{bmatrix} \mathbf{Q} & \frac{1}{2}\boldsymbol{q} \\ \frac{1}{2}\boldsymbol{q}^\mathsf{T} & q^0 - \beta \end{bmatrix} \succcurlyeq \mathbf{0}, \\ \begin{bmatrix} \mathbf{Q} & \frac{1}{2}(\boldsymbol{q} - \alpha_i \boldsymbol{y}_i(\boldsymbol{x})) \\ \frac{1}{2}(\boldsymbol{q} - \alpha_i \boldsymbol{y}_i(\boldsymbol{x}))^\mathsf{T} & q^0 - \alpha_i y_i^0(\boldsymbol{x}) \end{bmatrix} \succcurlyeq \mathbf{0} \quad \forall i = 1, \ldots, m \end{array} \right\}.$$

Proof Note that the constraints in (5.41) are equivalent to

$$\begin{bmatrix} \mathbf{Q} & \frac{1}{2}(\boldsymbol{q} - \alpha_i \boldsymbol{y}_i(\boldsymbol{x})) \\ \frac{1}{2}(\boldsymbol{q} - \alpha_i \boldsymbol{y}_i(\boldsymbol{x}))^\mathsf{T} & q^0 - \alpha_i y_i^0(\boldsymbol{x}) \end{bmatrix} \succcurlyeq \mathbf{0} \quad \forall i = 1, \ldots, m.$$

Moreover, by Theorem 5.2.2, the robust quadratic chance constraint (5.42) is equivalent to the Worst-Case CVaR constraint

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-CVaR}\left(\tilde{\boldsymbol{\xi}}^\mathsf{T}\mathbf{Q}\tilde{\boldsymbol{\xi}} + \tilde{\boldsymbol{\xi}}^\mathsf{T}\boldsymbol{q} + q^0\right) \;=\; \inf_{\beta \in \mathbb{R}} \left\{ \beta + \frac{1}{\epsilon} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}\left(\left[\tilde{\boldsymbol{\xi}}^\mathsf{T}\mathbf{Q}\tilde{\boldsymbol{\xi}} + \tilde{\boldsymbol{\xi}}^\mathsf{T}\boldsymbol{q} + q^0 - \beta\right]^+\right)\right\} \;\leq\; 0. \tag{5.44}$$

As usual, we first find an SDP reformulation of the subordinate worst-case expectation problem

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}\left(\left[\tilde{\boldsymbol{\xi}}^\mathsf{T}\mathbf{Q}\tilde{\boldsymbol{\xi}} + \tilde{\boldsymbol{\xi}}^\mathsf{T}\boldsymbol{q} + q^0 - \beta\right]^+\right).$$

By Lemma 5.2.1, this problem can be rewritten as

$$\begin{array}{ll} \inf\limits_{\mathbf{M} \in \mathbb{S}^{k+1}} & \langle \boldsymbol{\Omega}, \mathbf{M} \rangle \\ \text{s.t.} & \mathbf{M} \succcurlyeq \mathbf{0}, \quad \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right]^\mathsf{T} \;\geq\; \boldsymbol{\xi}^\mathsf{T}\mathbf{Q}\boldsymbol{\xi} + \boldsymbol{\xi}^\mathsf{T}\boldsymbol{q} + q^0 - \beta \quad \forall \boldsymbol{\xi} \in \mathbb{R}^k. \end{array} \tag{5.45}$$

Note that the semi-infinite constraint in (5.45) is equivalent to

$$\begin{bmatrix} \boldsymbol{\xi} \\ 1 \end{bmatrix}^\mathsf{T} \left(\mathbf{M} - \begin{bmatrix} \mathbf{Q} & \frac{1}{2}\boldsymbol{q} \\ \frac{1}{2}\boldsymbol{q}^\mathsf{T} & q^0 - \beta \end{bmatrix} \right) \begin{bmatrix} \boldsymbol{\xi} \\ 1 \end{bmatrix} \geq 0 \quad \forall \boldsymbol{\xi} \in \mathbb{R}^k \quad \iff \quad \mathbf{M} - \begin{bmatrix} \mathbf{Q} & \frac{1}{2}\boldsymbol{q} \\ \frac{1}{2}\boldsymbol{q}^\mathsf{T} & q^0 - \beta \end{bmatrix} \succcurlyeq \mathbf{0},$$

which enables us to rewrite (5.45) as

$$\begin{aligned} \inf_{\mathbf{M}\in\mathbb{S}^{k+1}} \quad & \langle \boldsymbol{\Omega}, \mathbf{M}\rangle \\ \text{s.t.} \quad & \mathbf{M} \succcurlyeq \mathbf{0}, \quad \mathbf{M} - \begin{bmatrix} \mathbf{Q} & \frac{1}{2}\boldsymbol{q} \\ \frac{1}{2}\boldsymbol{q}^\intercal & q^0 - \beta \end{bmatrix} \succcurlyeq \mathbf{0}. \end{aligned} \tag{5.46}$$

Substituting (5.46) into (5.44) shows that the robust quadratic chance constraint (5.42) is equivalent to

$$\begin{aligned} 0 \;\geq\; \inf \quad & \beta + \tfrac{1}{\epsilon}\langle \boldsymbol{\Omega}, \mathbf{M}\rangle \\ \text{s.t.} \quad & \mathbf{M} \in \mathbb{S}^{k+1}, \quad \beta \in \mathbb{R} \\ & \mathbf{M} \succcurlyeq \mathbf{0}, \quad \mathbf{M} - \begin{bmatrix} \mathbf{Q} & \frac{1}{2}\boldsymbol{q} \\ \frac{1}{2}\boldsymbol{q}^\intercal & q^0 - \beta \end{bmatrix} \succcurlyeq \mathbf{0}. \end{aligned}$$

Thus, the claim follows. ∎

In the following theorem we show that the approximate feasible set $\mathcal{Z}_{\mathrm{Q}}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ is equivalent to the set $\mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ found in Section 5.3.3. This implies that the approximation of a distributionally robust joint chance constraint by a Worst-Case CVaR constraint is equivalent to the approximation of the max function implied by the joint chance constraint by a quadratic majorant. Note that both approximations depend of the choice of the scaling parameters $\boldsymbol{\alpha}$.

Theorem 5.3.5 *For any $\boldsymbol{\alpha} \in \mathcal{A}$ we have $\mathcal{Z}_{\mathrm{Q}}^{\mathrm{JCC}}(\boldsymbol{\alpha}) = \mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$.*

Proof By defining the combined variable

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Q} & \frac{1}{2}\boldsymbol{q} \\ \frac{1}{2}\boldsymbol{q}^\intercal & q^0 \end{bmatrix},$$

the set $\mathcal{Z}_{\mathrm{Q}}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ can be rewritten as

$$\mathcal{Z}_{\mathrm{Q}}^{\mathrm{JCC}}(\boldsymbol{\alpha}) = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \begin{array}{ll} \exists \mathbf{Y} \in \mathbb{S}^k, \quad \beta \in \mathbb{R}, \quad \mathbf{M} \in \mathbb{S}^{k+1}, & \\ \beta + \frac{1}{\epsilon}\langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq 0, \quad \mathbf{M} \succcurlyeq \mathbf{0} & \\ \mathbf{M} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0}^\intercal & \beta \end{bmatrix} \succcurlyeq \mathbf{Y} \succcurlyeq \begin{bmatrix} \mathbf{0} & \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x}) \\ \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x})^\intercal & \alpha_i y_i^0(\boldsymbol{x}) \end{bmatrix} & \forall i = 1, \ldots, m \end{array} \right\},$$

It is easy to see that $\mathbf{Y}$ may be eliminated from the above representation of $\mathcal{Z}_{\mathrm{Q}}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ by rewriting the last constraint group as

$$\mathbf{M} - \begin{bmatrix} \mathbf{0} & \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x}) \\ \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x})^\intercal & \alpha_i y_i^0(\boldsymbol{x}) - \beta \end{bmatrix} \succcurlyeq \mathbf{0} \quad \forall i = 1, \ldots, m.$$

This observation establishes the postulated equivalence. ∎

5.3.5 The Exactness of the Worst-Case CVaR Approximation

So far we have shown that, for any fixed $\boldsymbol{\alpha} \in \mathcal{A}$, the feasible set $\mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ constitutes a tractable conservative approximation for $\mathcal{X}^{\mathrm{JCC}}$. This implies that the union of all sets of the type $\mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ for $\boldsymbol{\alpha} \in \mathcal{A}$ still constitutes a conservative approximation for $\mathcal{X}^{\mathrm{JCC}}$. We now demonstrate that this improved approximation is in fact exact.

Theorem 5.3.6 *The Worst-Case CVaR approximation is exact if $\boldsymbol{\alpha}$ is treated as a decision variable. Formally, we have*

$$\mathcal{X}^{\mathrm{JCC}} = \bigcup_{\boldsymbol{\alpha} \in \mathcal{A}} \mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha}).$$

Proof Recall from Section 5.3.1 that the feasible set $\mathcal{X}^{\mathrm{JCC}}$ can be written as

$$\mathcal{X}^{\mathrm{JCC}} = \left\{ \boldsymbol{x} \in \mathbb{R}^n \ : \ \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\left(\bigcup_{i=1}^{m} \left\{ y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\intercal \tilde{\boldsymbol{\xi}} > 0 \right\} \right) \leq \epsilon \right\}.$$

By Lemma 4.4.1 we may thus conclude that

$$\mathcal{X}^{\mathrm{JCC}} = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \begin{array}{l} \exists \mathbf{M} \in \mathbb{S}^{k+1}, \quad \langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq \epsilon, \quad \mathbf{M} \succcurlyeq \mathbf{0}, \\ \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right]^\mathsf{T} \geq 1 \quad \forall \boldsymbol{\xi} \in \bigcup_{i=1}^m \left\{ y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T} \boldsymbol{\xi} > 0 \right\} \end{array} \right\}.$$

The semi-infinite constraint in the above representation of $\mathcal{X}^{\mathrm{JCC}}$ can be reexpressed as

$$\left[\boldsymbol{\xi}^\mathsf{T}\ 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right]^\mathsf{T} \geq 1 \quad \forall \boldsymbol{\xi} \ : \ y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T} \boldsymbol{\xi} > 0, \quad \forall i = 1, \ldots, m,$$

which, by the $\mathcal{S}$-lemma, is equivalent to

$$\exists \boldsymbol{\alpha} \geq \mathbf{0}, \quad \mathbf{M} - \begin{bmatrix} \mathbf{0} & \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x}) \\ \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T} & \alpha_i y_i^0(\boldsymbol{x}) + 1 \end{bmatrix} \succcurlyeq \mathbf{0} \quad \forall i = 1, \ldots, m.$$

Thus, the feasible set $\mathcal{X}^{\mathrm{JCC}}$ can be written as

$$\mathcal{X}^{\mathrm{JCC}} = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \begin{array}{l} \exists \mathbf{M} \in \mathbb{S}^{k+1}, \quad \boldsymbol{\alpha} \in \mathbb{R}^m, \\ \langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq \epsilon, \quad \mathbf{M} \succcurlyeq \mathbf{0}, \quad \boldsymbol{\alpha} > \mathbf{0}, \\ \mathbf{M} - \begin{bmatrix} \mathbf{0} & \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x}) \\ \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T} & \alpha_i y_i^0(\boldsymbol{x}) + 1 \end{bmatrix} \succcurlyeq \mathbf{0} \quad \forall i = 1, \ldots, m \end{array} \right\}. \tag{5.47}$$

Note that we require here without loss of generality that $\boldsymbol{\alpha}$ is strictly positive. Indeed, it can be shown that no feasible $\boldsymbol{\alpha}$ has any vanishing components. By Theorem 5.3.3, we have

$$\bigcup_{\boldsymbol{\alpha} \in \mathcal{A}} \mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha}) = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \begin{array}{l} \exists \beta \in \mathbb{R}, \quad \mathbf{M} \in \mathbb{S}^{k+1}, \quad \boldsymbol{\alpha} \in \mathcal{A} \\ \beta + \frac{1}{\epsilon} \langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq 0, \quad \mathbf{M} \succcurlyeq \mathbf{0}, \\ \mathbf{M} - \begin{bmatrix} \mathbf{0} & \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x}) \\ \frac{1}{2}\alpha_i \boldsymbol{y}_i^\mathsf{T} & \alpha_i y_i^0(\boldsymbol{x}) - \beta \end{bmatrix} \succcurlyeq \mathbf{0} \quad \forall i = 1, \ldots, m \end{array} \right\}. \tag{5.48}$$

Note that any feasible β in $\mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ must be nonpositive since $\beta + \frac{1}{\epsilon}\langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq 0$, $\mathbf{M} \succcurlyeq \mathbf{0}$, and $\boldsymbol{\Omega} \succ \mathbf{0}$.

It is clear that $\mathcal{X}^{\mathrm{JCC}} \subseteq \bigcup_{\boldsymbol{\alpha} \in \mathcal{A}} \mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ since we are free to set $\beta = -1$ in (5.48) and since

$-1 + \frac{1}{\epsilon}\langle \mathbf{\Omega}, \mathbf{M} \rangle \leq 0$ is equivalent to $\langle \mathbf{\Omega}, \mathbf{M} \rangle \leq \epsilon$. To prove the converse inclusion, we select an arbitrary $\boldsymbol{x} \in \bigcup_{\boldsymbol{\alpha} \in \mathcal{A}} \mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ and a corresponding $(\beta, \mathbf{M}, \boldsymbol{\alpha})$ satisfying the constraints in (5.48). Assume first that $\beta < 0$. Since all constraints in (5.48) are homogeneous of degree 1 in $(\beta, \mathbf{M}, \boldsymbol{\alpha})$, the scaled variables $(\beta', \mathbf{M}', \boldsymbol{\alpha}') = (-1, -\mathbf{M}/\beta, \boldsymbol{\alpha}/\beta)$ are also feasible in (5.48). Moreover, $(\boldsymbol{x}, \beta', \mathbf{M}', \boldsymbol{\alpha}')$ is feasible in (5.47), and thus $\boldsymbol{x} \in \mathcal{X}^{\mathrm{JCC}}$. Assume now that $\beta = 0$. Then, the constraints in (5.48) imply $\mathbf{M} = \mathbf{0}$. Since $\boldsymbol{\alpha} > \mathbf{0}$, this in turn implies that

$$
\begin{aligned}
& y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T} \boldsymbol{\xi} \leq 0 \quad \forall \boldsymbol{\xi} \in \mathbb{R}^k, \quad \forall i = 1, \ldots, m \\
\Longrightarrow\ & \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T} \tilde{\boldsymbol{\xi}} \leq 0 \quad \forall i = 1, \ldots, m \right) = 1.
\end{aligned}
$$

These observations make it clear that $\boldsymbol{x} \in \mathcal{X}^{\mathrm{JCC}}$, which concludes the proof. ∎

Theorem 5.3.6 implies that the original joint chance constrained program

$$
\underset{\boldsymbol{x} \in \mathcal{X} \cap \mathcal{X}^{\mathrm{JCC}}}{\text{minimize}} \quad \boldsymbol{c}^\mathsf{T} \boldsymbol{x}
$$

is equivalent to

$$
\underset{\substack{\boldsymbol{x} \in \mathcal{X} \cap \mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha}) \\ \boldsymbol{\alpha} \in \mathcal{A}}}{\text{minimize}} \quad \boldsymbol{c}^\mathsf{T} \boldsymbol{x} \tag{5.49}
$$

Unfortunately, optimizing jointly over $\boldsymbol{x} \in \mathcal{X} \cap \mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ and $\boldsymbol{\alpha} \in \mathcal{A}$ in (5.49) involves Bilinear Matrix Inequalities (BMIs). It is known that generic BMI constrained problems are $\mathcal{NP}$-hard, see [TO95]. Similar nonconvexities arise also in the approximations discussed in Sections 5.3.1 and 5.3.2, which underlines the general perception that problems with distributionally robust joint chance constraints are hard to solve.

Recall, however, that for any fixed $\boldsymbol{\alpha} \in \mathcal{A}$, the set $\mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ is representable in terms of tractable LMI constraints involving the auxiliary variables β and $\mathbf{M}$. In particular, the constraints in (5.48) are convex in $\beta, \mathbf{M}$, and $\boldsymbol{x}$ for any fixed $\boldsymbol{\alpha}$, and convex in $\boldsymbol{\alpha}$ for any fixed $\beta, \mathbf{M}$, and $\boldsymbol{x}$. In Section 5.3.8 we will use this property to propose an algorithm for solving (5.49) approximately.

5.3.6 Robust Optimization Perspective on Joint Chance Constraints

In Section 5.2.3 we have shown that robust individual chance constraints are equivalent to robust semi-infinite constraints involving an uncertainty set that can be interpreted as an ellipsoid lifted to the space of positive semidefinite matrices. In this section we show that there exists also a close relationship between robust *joint* chance constraints and robust semi-infinite optimization.

We first show that one can construct a robust counterpart for the constraint $\boldsymbol{x} \in \mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$.

Theorem 5.3.7 *For any fixed $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{\alpha} \in \mathcal{A}$ we have $\boldsymbol{x} \in \mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ if and only if*

$$\sum_{i=1}^{m} \alpha_i \left\langle \begin{bmatrix} \mathbf{0} & \frac{1}{2}\boldsymbol{y}_i(\boldsymbol{x}) \\ \frac{1}{2}\boldsymbol{y}_i(\boldsymbol{x})^\intercal & y_i^0(\boldsymbol{x}) \end{bmatrix}, \boldsymbol{\Lambda}_i \right\rangle \leq 0 \quad \forall (\boldsymbol{\Lambda}_1, \ldots, \boldsymbol{\Lambda}_m) \in \mathcal{U}_\epsilon^m, \tag{5.50}$$

where the uncertainty set $\mathcal{U}_\epsilon^m \subseteq (\mathbb{S}^{k+1})^m$ is defined as

$$\mathcal{U}_\epsilon^m = \left\{ (\boldsymbol{\Lambda}_1, \ldots, \boldsymbol{\Lambda}_m) \in (\mathbb{S}^{k+1})^m : \begin{array}{l} \boldsymbol{\Lambda}_i \succcurlyeq \mathbf{0} \quad \forall i = 1, \ldots, m \\ \frac{1}{\epsilon}\boldsymbol{\Omega} - \sum_{i=1}^m \boldsymbol{\Lambda}_i \succcurlyeq \mathbf{0} \\ \sum_{i=1}^m \Lambda_{k+1,k+1}^i = 1 \end{array} \right\}. \tag{5.51}$$

Proof In the proof of Theorem 5.3.3 we have seen that $\boldsymbol{x} \in \mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ if and only if the optimal value of the SDP in (5.39) is nonpositive. After a few elementary simplification steps, the dual of this SDP reduces to

$$\begin{array}{ll} \sup & \displaystyle\sum_{i=1}^{m} \alpha_i \left\langle \begin{bmatrix} \mathbf{0} & \frac{1}{2}\boldsymbol{y}_i(\boldsymbol{x}) \\ \frac{1}{2}\boldsymbol{y}_i(\boldsymbol{x})^\intercal & y_i^0(\boldsymbol{x}) \end{bmatrix}, \boldsymbol{\Lambda}_i \right\rangle \\ \text{s. t.} & \boldsymbol{\Lambda}_1, \ldots, \boldsymbol{\Lambda}_m \in \mathbb{S}^{k+1} \\ & \displaystyle\frac{1}{\epsilon}\boldsymbol{\Omega} - \sum_{i=1}^{m} \boldsymbol{\Lambda}_i \succcurlyeq \mathbf{0}, \quad \sum_{i=1}^{m} \Lambda_{k+1,k+1}^i = 1 \\ & \boldsymbol{\Lambda}_i \succcurlyeq \mathbf{0} \quad \forall i = 1, \ldots, m. \end{array} \tag{5.52}$$

Using similar arguments as in Theorem 5.2.3, it can be shown that strong duality holds. Thus,

$\boldsymbol{x} \in \mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ if and only if the optimal value of problem (5.52) is nonpositive, which is manifestly equivalent to the postulated semi-infinite constraint (5.50). ■

Remark 5.3.3 *Note that for $m = 1$, the uncertainty set $\mathcal{U}_\epsilon^m$ defined in* (5.51) *reduces to $\mathcal{U}_\epsilon^1$ defined in* (5.24)*, which is the uncertainty set associated with a distributionally robust individual chance constraint. In this case, the robust counterpart* (5.50) *adopts the form*

$$\alpha \left\langle \begin{bmatrix} \mathbf{0} & \frac{1}{2}\boldsymbol{y} \\ \frac{1}{2}\boldsymbol{y}^\intercal & y^0 \end{bmatrix}, \boldsymbol{\Lambda} \right\rangle \leq 0 \quad \forall \boldsymbol{\Lambda} \in \mathcal{U}_\epsilon^1.$$

The scalar $\alpha > 0$ can be divided away by which we recover the semi-infinite constraint (5.23)*. This shows that the robust counterpart* (5.50) *of the joint chance constraint* (5.3) *encapsulates that of the individual chance constraint* (5.5) *as a special case.*

In the next theorem we show that one can also construct a robust counterpart for the constraint $\boldsymbol{x} \in \mathcal{Z}_{\mathrm{U}}^{\mathrm{JCC}}(\boldsymbol{\alpha})$, see Section 5.3.2.

Theorem 5.3.8 *For any fixed $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{\alpha} \in \mathcal{A}$, we have $\boldsymbol{x} \in \mathcal{Z}_U^{JCC}(\boldsymbol{\alpha})$ if and only if*

$$\sum_{i=1}^m \alpha_i \left\langle \begin{bmatrix} \mathbf{0} & \frac{1}{2}\boldsymbol{y}_i(\boldsymbol{x}) \\ \frac{1}{2}\boldsymbol{y}_i(\boldsymbol{x})^\intercal & y_i^0(\boldsymbol{x}) \end{bmatrix}, \boldsymbol{\Lambda}_i \right\rangle \leq 0 \quad \forall (\boldsymbol{\Lambda}_1, \ldots, \boldsymbol{\Lambda}_m) \in \hat{\mathcal{U}}_\epsilon^m, \tag{5.53}$$

where the uncertainty set $\hat{\mathcal{U}}_\epsilon^m \subseteq (\mathbb{S}^{k+1})^m$ is defined as

$$\hat{\mathcal{U}}_\epsilon^m = \left\{ (\boldsymbol{\Lambda}_1, \ldots, \boldsymbol{\Lambda}_m) \in (\mathbb{S}^{k+1})^m : \begin{array}{ll} \boldsymbol{\Lambda}_i \succcurlyeq \mathbf{0} & \forall i = 1, \ldots, m \\ \frac{1}{\epsilon}\boldsymbol{\Omega} - \boldsymbol{\Lambda}_i \succcurlyeq \mathbf{0} & \forall i = 1, \ldots, m \\ \sum_{i=1}^m \Lambda_{k+1,k+1}^i = 1 & \end{array} \right\}$$

Proof The proof relies on deriving the dual of the SDP (5.34) and then follows the argumentation in the proof of Theorem 5.3.7. ■

The semi-infinite constraints (5.50) and (5.53) both represent (conservative) robust counterparts for the distributionally robust joint chance constraint (5.3). Note that (5.53) with the uncertainty set $\hat{\mathcal{U}}_\epsilon^m$ is generally more conservative than (5.50) with the uncertainty set $\mathcal{U}_\epsilon^m$. Indeed, it can easily be seen that $\mathcal{U}_\epsilon^m \subseteq \hat{\mathcal{U}}_\epsilon^m$ since the inequalities $\mathbf{\Lambda}_i \succcurlyeq \mathbf{0}$ for $i = 1, \ldots, m$ and $\frac{1}{\epsilon}\mathbf{\Omega} - \sum_{i=1}^m \mathbf{\Lambda}_i \succcurlyeq \mathbf{0}$ imply that $\frac{1}{\epsilon}\mathbf{\Omega} - \mathbf{\Lambda}_i \succcurlyeq \mathbf{0}$ for $i = 1, \ldots, m$. Note that $\mathcal{U}_\epsilon^m$ is generically a strict subset of $\hat{\mathcal{U}}_\epsilon^m$, which is consistent with Remark 5.3.1.

5.3.7 Injecting Support Information

In many practical applications the support of the (true) distribution $\mathbb{Q}$ of $\tilde{\boldsymbol{\xi}}$ is known to be a strict subset of $\mathbb{R}^k$. Disregarding this information in the definition of $\mathcal{P}$ can result in unnecessarily conservative robust chance constraints. In this section we briefly outline how support information can be used to tighten robust chance constraints and their approximations developed in Section 5.3. To this end, we first revise our distributional assumptions.

Distributional Assumptions. The random vector $\tilde{\boldsymbol{\xi}}$ has a distribution $\mathbb{Q}$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma} \succ \mathbf{0}$. We assume that $\mathbb{Q}$ is supported on $\Xi = \{\boldsymbol{\xi} \in \mathbb{R}^k \ : \ [\boldsymbol{\xi}^\mathsf{T}\ 1]\mathbf{W}_i[\boldsymbol{\xi}^\mathsf{T}\ 1]^\mathsf{T} \leq 0 \ \forall i = 1, \ldots, l\}$, where $\mathbf{W}_i \in \mathbb{S}^{k+1}$ for all $i = 1, \ldots, l$.[1] Thus, we have $\mathbb{Q}(\tilde{\boldsymbol{\xi}} \in \Xi) = 1$. We define $\mathcal{P}_\Xi$ as the set of all probability distributions supported on Ξ that have the same first- and second-order moments as $\mathbb{Q}$.

In this section we are interested in tractable conservative approximations for the feasible set

$$\mathcal{X}_\Xi^{\mathrm{JCC}} = \left\{ \boldsymbol{x} \in \mathbb{R}^n \ : \ \inf_{\mathbb{P} \in \mathcal{P}_\Xi} \mathbb{P}\left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T}\tilde{\boldsymbol{\xi}} \leq 0 \quad \forall i = 1, \ldots, m \right) \geq 1 - \epsilon \right\}.$$

As before, we study approximate feasible sets of the form

$$\mathcal{Z}_\Xi^{\mathrm{JCC}}(\boldsymbol{\alpha}) = \left\{ \boldsymbol{x} \in \mathbb{R}^n \ : \ \sup_{\mathbb{P} \in \mathcal{P}_\Xi} \mathrm{CVaR}_\epsilon \left(\max_{i=1,\ldots,m} \left\{ \alpha_i \left(y_i^0(\boldsymbol{x}) + \boldsymbol{y}_i(\boldsymbol{x})^\mathsf{T}\tilde{\boldsymbol{\xi}} \right) \right\} \right) \leq 0 \right\}$$

for $\boldsymbol{\alpha} \in \mathcal{A}$. By using similar arguments as in Section 5.2.1, one can show that $\mathcal{Z}_\Xi^{\mathrm{JCC}}(\boldsymbol{\alpha}) \subseteq \mathcal{X}_\Xi^{\mathrm{JCC}}$ for all $\boldsymbol{\alpha} \in \mathcal{A}$. However, the sets $\mathcal{Z}_\Xi^{\mathrm{JCC}}(\boldsymbol{\alpha})$ have no longer an exact representation in terms of

[1] Note that every finite intersection of half-spaces and ellipsoids in $\mathbb{R}^k$ is representable as a set of the form Ξ.

LMIs. Instead, they need to be conservatively approximated.

Theorem 5.3.9 *For any fixed $\boldsymbol{\alpha} \in \mathcal{A}$, we have $\mathcal{Y}_{\Xi}^{\mathrm{JCC}}(\boldsymbol{\alpha}) \subseteq \mathcal{Z}_{\Xi}^{\mathrm{JCC}}(\boldsymbol{\alpha}) \subseteq \mathcal{X}_{\Xi}^{\mathrm{JCC}}$, where $\mathcal{Y}_{\Xi}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ has the following tractable reformulation in terms of LMIs.*

$$\mathcal{Y}_{\Xi}^{\mathrm{JCC}}(\boldsymbol{\alpha}) = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \begin{array}{ll} \exists \mathbf{M} \in \mathbb{S}^{k+1}, \quad \beta \in \mathbb{R}, \quad \boldsymbol{\tau}_i \in \mathbb{R}^l, & \\ \beta + \frac{1}{\epsilon}\langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq 0, \quad \boldsymbol{\tau}_i \geq \mathbf{0} & \forall i = 0, \dots, m \\ \mathbf{M} + \sum_{j=1}^{l} \tau_{0,j} \mathbf{W}_j \succcurlyeq \mathbf{0} & \\ \mathbf{M} + \sum_{j=1}^{l} \tau_{i,j} \mathbf{W}_j - \begin{bmatrix} \mathbf{0} & \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x}) \\ \frac{1}{2}\alpha_i \boldsymbol{y}_i(\boldsymbol{x})^\top & \alpha_i y_i^0(\boldsymbol{x}) - \beta \end{bmatrix} \succcurlyeq \mathbf{0} & \forall i = 1, \dots, m \end{array} \right\}. \tag{5.54}$$

Furthermore, for $l = 1$, we have $\mathcal{Y}_{\Xi}^{\mathrm{JCC}}(\boldsymbol{\alpha}) = \mathcal{Z}_{\Xi}^{\mathrm{JCC}}(\boldsymbol{\alpha})$.

Proof The proof widely parallels the proof of Theorem 5.3.1. The only difference is that $\mathbb{R}^k$ is replaced by Ξ and that we use the $\mathcal{S}$-lemma to approximate (for $l > 1$) or reformulate (for $l = 1$) the semi-infinite constraints over Ξ by LMI constraints.

Remark 5.3.4 *While $\mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ is exactly representable in terms of LMIs in the absence of support information, Theorem 5.3.9 only provides a conservative LMI approximation for $\mathcal{Z}_{\Xi}^{\mathrm{JCC}}(\boldsymbol{\alpha})$. Nevertheless, it is easily verified that $\mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha}) \subseteq \mathcal{Y}_{\Xi}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ and therefore $\mathcal{Y}_{\Xi}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ constitutes a better approximation for $\mathcal{Z}_{\Xi}^{\mathrm{JCC}}(\boldsymbol{\alpha})$ than $\mathcal{Z}^{\mathrm{JCC}}(\boldsymbol{\alpha})$. In fact, by setting $\boldsymbol{\tau}_i = \mathbf{0}$ for all $i = 0, \dots, m$,* (5.54) *reduces to* (5.39).

Remark 5.3.5 *Support information can also be used in a straightforward way to tighten the approximations discussed in Sections 5.3.1 and 5.3.2.*

5.3.8 Optimizing over the Scaling Parameters

By Theorem 5.3.6, the original distributionally robust chance constrained program (5.4) can be written as

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{R}^n,\boldsymbol{\alpha}\in\mathcal{A}}{\text{minimize}} \quad & \boldsymbol{c}^{\mathsf{T}}\boldsymbol{x} \\ \text{subject to} \quad & \mathcal{J}(\boldsymbol{x},\boldsymbol{\alpha}) \leq 0 \\ & \boldsymbol{x} \in \mathcal{X}, \end{aligned} \tag{5.55}$$

where the Worst-Case CVaR functional $\mathcal{J}(\boldsymbol{x},\boldsymbol{\alpha})$ is defined as in (5.32). Unfortunately, as discussed in Section 5.3.3, $\mathcal{J}(\boldsymbol{x},\boldsymbol{\alpha})$ is merely biconvex, but not jointly convex in $\boldsymbol{x}$ and $\boldsymbol{\alpha}$. Thus, optimization problem (5.55) is nonconvex. By Theorem 5.3.3, however, the problem becomes convex and tractable when the values of the scaling parameters $\boldsymbol{\alpha}$ are frozen.

For the further argumentation we define the set $\bar{\mathcal{A}} = \{\boldsymbol{\alpha} : \boldsymbol{\alpha} \geq \delta \boldsymbol{e}\}$, where $\boldsymbol{e}$ denotes the vector of ones and $\delta > 0$ represents a small tolerance, which we set to 10^{-7}. Note that, unlike $\mathcal{A}$, the set $\bar{\mathcal{A}}$ is closed. Consider now the following optimization model where $\boldsymbol{\alpha} \in \bar{\mathcal{A}}$ is fixed.

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n} \quad & \boldsymbol{c}^{\mathsf{T}}\boldsymbol{x} \\ \text{s.t.} \quad & \mathcal{J}(\boldsymbol{x},\boldsymbol{\alpha}) \leq 0 \\ & \boldsymbol{x} \in \mathcal{X} \end{aligned} \tag{5.56}$$

We emphasize again that by Theorem 5.3.3 (5.56) is equivalent to a tractable SDP and that any $\boldsymbol{x}$ feasible in (5.56) is also feasible in the original chance constrained problem (5.4). In the remainder of this section we develop an algorithm that repeatedly solves (5.56) while systematically improving the scaling parameters $\boldsymbol{\alpha}$.

The main idea of this approach, which is inspired by [CSSC09], is to minimize $\mathcal{J}(\boldsymbol{x},\boldsymbol{\alpha})$ over $\boldsymbol{\alpha} \in \bar{\mathcal{A}}$ with the aim of enlarging the feasible region of problem (5.56) and thereby improving the objective value. To this end, we introduce the following optimization model which depends parametrically on $\boldsymbol{x} \in \mathcal{X}$.

$$\begin{aligned} \min_{\boldsymbol{\alpha}\in\mathbb{R}^m} \quad & \mathcal{J}(\boldsymbol{x},\boldsymbol{\alpha}) \\ \text{s.t.} \quad & \boldsymbol{\alpha} \in \bar{\mathcal{A}} \end{aligned} \tag{5.57}$$

Theorem 5.3.3 implies that (5.57) can also be expressed as a tractable SDP.

Assume that $\boldsymbol{x}^*$ is an optimal solution of problem (5.56) for a given $\boldsymbol{\alpha} \in \bar{\mathcal{A}}$. By the feasibility of $\boldsymbol{x}^*$ in (5.56) we know that $\mathcal{J}(\boldsymbol{x}^*, \boldsymbol{\alpha}) \leq 0$. Keeping $\boldsymbol{x}^*$ fixed, we then solve problem (5.57) to obtain the optimal scaling parameters $\boldsymbol{\alpha}^*$ corresponding to $\boldsymbol{x}^*$. By construction, we find

$$\mathcal{J}(\boldsymbol{x}^*, \boldsymbol{\alpha}^*) \ \leq \ \mathcal{J}(\boldsymbol{x}^*, \boldsymbol{\alpha}) \ \leq 0. \tag{5.58}$$

The above inequalities imply that the optimal objective value of problem (5.56) with input $\boldsymbol{\alpha}^*$ must not exceed $\boldsymbol{c}^\mathsf{T}\boldsymbol{x}^*$. Therefore, by solving the problems (5.56) and (5.57) in alternation, we obtain a sequence of monotonically decreasing objective values. This motivates the following algorithm, which relies on the availability of an initial feasible solution $\boldsymbol{x}_{\text{init}}$ for problem (5.56).

Algorithm 5.3.1 *Sequential Convex Optimization Procedure*

1. ***Initialization***. *Let* $\boldsymbol{x}_{\text{init}}$ *be some feasible solution of problem* (5.56). *Set the current solution to* $\boldsymbol{x}^0 \leftarrow \boldsymbol{x}_{\text{init}}$, *the current objective value to* $f^0 \leftarrow \boldsymbol{c}^\mathsf{T}\boldsymbol{x}^0$, *and the iteration counter to* $t \leftarrow 1$.

2. ***Scaling Parameter Optimization***. *Solve problem* (5.57) *with input* $\boldsymbol{x}^{t-1}$ *and let* $\boldsymbol{\alpha}^*$ *denote an optimal set of scaling parameters. Set* $\boldsymbol{\alpha}^t \leftarrow \boldsymbol{\alpha}^*$.

3. ***Decision Optimization***. *Solve problem* (5.56) *with input* $\boldsymbol{\alpha}^t$ *and let* $\boldsymbol{x}^*$ *denote an optimal solution. Set* $\boldsymbol{x}^t \leftarrow \boldsymbol{x}^*$ *and* $f^t \leftarrow \boldsymbol{c}^\mathsf{T}\boldsymbol{x}^t$.

4. ***Termination***. *If* $(f^t - f^{t-1})/|f^{t-1}| \leq \gamma$ *(where* γ *is a given small tolerance), output* $\boldsymbol{x}^t$ *and stop. Otherwise, set* $t \leftarrow t + 1$ *and go back to Step 2.*

Theorem 5.3.10 *Assume that* $\boldsymbol{x}_{\text{init}}$ *is feasible in problem* (5.56) *for some* $\boldsymbol{\alpha} \in \bar{\mathcal{A}}$. *Then, the sequence of objective values* $\{f^t\}$ *generated by Algorithm 5.3.1 is monotonically decreasing. If the set* $\mathcal{X}$ *is bounded, then the sequence* $\{\boldsymbol{x}^t\}$ *is also bounded, while the sequence* $\{f^t\}$ *converges to a finite limit.*

Proof By the inequality (5.58), an update of the scaling parameters from $\boldsymbol{\alpha}^{t-1}$ to $\boldsymbol{\alpha}^{t}$ in Step 2 of the algorithm preserves the feasibility of $\boldsymbol{x}^{t-1}$ in problem (5.56). This guarantees that the sequence of objective values $\{f^t\}$ is monotonically decreasing. Furthermore, it is readily seen that the solution sequence $\{\boldsymbol{x}^t\}$ is bounded if the feasible set $\mathcal{X}$ is bounded. Since (5.56) has a continuous objective function, the monotonicity of the objective value sequence implies that $\{f^t\}$ has a finite limit. ■

Remark 5.3.6 *Algorithm 5.3.1 can also be used in the presence of support information as discussed in Section 5.3.7. In this case, the Worst-Case CVaR functional $\mathcal{J}(\boldsymbol{x}, \boldsymbol{\alpha})$ has to be redefined in the obvious way. Algorithm 5.3.1 can further be used in the context of the approximation by Chen* et al.*, see Section 5.3.2. In this case, $\mathcal{J}(\boldsymbol{x}, \boldsymbol{\alpha})$ is replaced by its conservative approximation $\hat{\mathcal{J}}(\boldsymbol{x}, \boldsymbol{\alpha})$ defined in* (5.33)*. Details are omitted for brevity of exposition.*

We emphasize that Algorithm 5.3.1 does not necessarily find the global optimum of problem (5.55). Nevertheless, as confirmed by the numerical results in the next section, the method can perform well in practice.

5.4 Numerical Results

We consider a dynamic water reservoir control problem for hydro power generation, which is inspired by a model due to Andrieu *et al.* [AHR10]. Let $\tilde{\boldsymbol{\xi}} = (\tilde{\xi}_1, \ \tilde{\xi}_2, \ldots, \ \tilde{\xi}_T)$ denote the sequence of stochastic inflows (precipitation) into the reservoir at time instances $t = 1, \ldots, T$. The history of inflows up to time t is denoted by $\tilde{\boldsymbol{\xi}}^t = (\tilde{\xi}_1, \ldots, \tilde{\xi}_t)$, where $\tilde{\boldsymbol{\xi}}^T = \tilde{\boldsymbol{\xi}}$. We let $\boldsymbol{\mu} \in \mathbb{R}^T$ and $\boldsymbol{\Sigma} \in \mathbb{S}^T$ denote the mean vector and covariance matrix of $\tilde{\boldsymbol{\xi}}$, respectively. Furthermore, $\tilde{\boldsymbol{\xi}}$ is supported on a rectangle of the form $\Xi = [\boldsymbol{l}, \ \boldsymbol{u}]$. However, we assume that no further information about the true distribution of $\tilde{\boldsymbol{\xi}}$ is available. As usual, we let $\mathcal{P}_\Xi$ denote the set of all distributions supported on Ξ with matching first- and second-order moments. We denote by $x_t(\tilde{\boldsymbol{\xi}}^t)$ the amount of water released from the reservoir in period t. Note that the decision $x_t(\tilde{\boldsymbol{\xi}}^t)$ is selected at time t after $\tilde{\boldsymbol{\xi}}^t$ has been observed and is therefore a function of the observation history. We require $x_t(\tilde{\boldsymbol{\xi}}^t) \geq 0$ almost surely for all $\mathbb{P} \in \mathcal{P}_\Xi$ and $t = 1, \ldots, T$. The water level at time t is computed as the sum of the initial level l_0 and the cumulative inflows minus the cumulative releases up to time t, that is,

$$l_0 + \sum_{i=1}^{t} \tilde{\xi}_i - \sum_{i=1}^{t} x_t(\tilde{\boldsymbol{\xi}}^t).$$

We require that the water level remains between some upper threshold l_{high} (flood reserve) and some lower threshold l_{low} (dead storage) over all time periods $t = 1, \ldots, T$ with probability $1 - \epsilon$, where $\epsilon \in (0, 1)$. The water released in any period t is used to produce electric energy which is sold at a periodic price

$$c_t = 10 + 5 \sin \left[\frac{\pi(1-t)}{3} \right] \quad \forall t = 1, \ldots, T.$$

The worst-case expected profit over all time periods is computed as

$$\inf_{\mathbb{P} \in \mathcal{P}_\Xi} \mathbb{E}_\mathbb{P} \left(\sum_{t=1}^{T} c_t x_t(\tilde{\boldsymbol{\xi}}^t) \right).$$

In order to determine an admissible control strategy that maximizes the worst-case profit, we must solve the following distributionally robust joint chance constrained problem.

$$\begin{array}{ll} \underset{x_1(\cdot), \ldots, x_T(\cdot)}{\text{maximize}} & \inf_{\mathbb{P} \in \mathcal{P}_\Xi} \mathbb{E}_\mathbb{P} \left(\sum_{t=1}^{T} c_t x_t(\tilde{\boldsymbol{\xi}}^t) \right) \\ \text{subject to} & \inf_{\mathbb{P} \in \mathcal{P}_\Xi} \mathbb{P} \left(l_{\text{low}} \leq l_0 + \sum_{i=1}^{t} \tilde{\xi}_i - \sum_{i=1}^{t} x_t(\tilde{\boldsymbol{\xi}}^t) \leq l_{\text{high}} \quad \forall t = 1, \ldots, T \right) \geq 1 - \epsilon \\ & x_t(\tilde{\boldsymbol{\xi}}^t) \geq 0 \quad \mathbb{P}\text{-a.s.} \quad \forall \mathbb{P} \in \mathcal{P}_\Xi, \; t = 1, \ldots, T \end{array} \tag{5.59}$$

Note that (5.59) is an infinite dimensional problem since the control decisions $x_t(\cdot)$ are generic measurable functionals of the uncertain inflows. To reduce the problem complexity, we focus on policies that are affine functions of $\tilde{\boldsymbol{\xi}}$. Thus, we optimize over affine disturbance feedback policies of the form

$$x_t(\tilde{\boldsymbol{\xi}}^t) = x_t^0 + \boldsymbol{x}_t^\mathsf{T} \mathbf{P}_t \tilde{\boldsymbol{\xi}} \quad \forall t = 1, \ldots, T, \tag{5.60}$$

where $x_t^0 \in \mathbb{R}$, $\boldsymbol{x}_t \in \mathbb{R}^t$ and $\mathbf{P}_t : \mathbb{R}^T \to \mathbb{R}^t$ is a truncation operator that maps $\tilde{\boldsymbol{\xi}}$ to $\tilde{\boldsymbol{\xi}}^t$. By focusing on affine control policies we conservatively approximate the infinite dimensional dynamic problem (5.59) by a problem with a polynomial number of variables, namely, the coefficients $\{x_t^0, \boldsymbol{x}_t\}_{t=1}^T$. For more details on the use of affine control policies in robust control and stochastic programming, see, e.g., Ben-Tal *et al.* [BTEGN09], Chen *et al.* [CSS], and Kuhn *et al.* [KWG09].

By applying now standard robust optimization techniques [BTEGN09], the requirement that $x_t(\tilde{\boldsymbol{\xi}}^t) \geq 0$ holds almost surely can be expressed as

$$
\begin{aligned}
& x_t^0 + \boldsymbol{x}_t^\mathsf{T} \mathbf{P}_t \boldsymbol{\xi} \geq 0 \quad \forall \boldsymbol{\xi} \in \Xi \\
\iff \quad & 0 \leq \min_{\boldsymbol{\xi} \in \mathbb{R}^T} \left\{ x_t^0 + \boldsymbol{x}_t^\mathsf{T} \mathbf{P}_t \boldsymbol{\xi} \ : \ \boldsymbol{l} \leq \boldsymbol{\xi} \leq \boldsymbol{u} \right\} \\
\iff \quad & 0 \leq \max_{\boldsymbol{\lambda}_t \in \mathbb{R}^T} \left\{ x_t^0 + \boldsymbol{x}_t^\mathsf{T} \mathbf{P}_t \boldsymbol{u} + \boldsymbol{\lambda}_t^\mathsf{T} (\boldsymbol{l} - \boldsymbol{u}) \ : \ \boldsymbol{\lambda}_t \geq \mathbf{P}_t^\mathsf{T} \boldsymbol{x}_t, \ \boldsymbol{\lambda}_t \geq \mathbf{0} \right\} \\
\iff \quad & \exists \boldsymbol{\lambda}_t \in \mathbb{R}^T \ : \ x_t^0 + \boldsymbol{x}_t^\mathsf{T} \mathbf{P}_t \boldsymbol{u} + \boldsymbol{\lambda}_t^\mathsf{T} (\boldsymbol{l} - \boldsymbol{u}) \geq 0, \ \boldsymbol{\lambda}_t \geq \mathbf{P}_t^\mathsf{T} \boldsymbol{x}_t, \ \boldsymbol{\lambda}_t \geq \mathbf{0}.
\end{aligned}
$$

By substituting (5.60) into (5.59) we thus obtain the following conservative approximation for (5.59).

$$
\begin{array}{ll}
\text{maximize} & \displaystyle\sum_{t=1}^T c_t \left(x_t^0 + \boldsymbol{x}_t^\mathsf{T} \mathbf{P}_t \boldsymbol{\mu} \right) \\
\text{subject to} & \boldsymbol{\lambda}_t \in \mathbb{R}^T, \quad \boldsymbol{x}_t \in \mathbb{R}^t \quad \forall t = 1, \ldots, T \\
& \displaystyle\inf_{\mathbb{P} \in \mathcal{P}_\Xi} \mathbb{P} \left(\begin{array}{l} l_0 - l_{\text{high}} + \displaystyle\sum_{i=1}^t \tilde{\xi}_i - \left(\sum_{i=1}^t x_i^0 + \boldsymbol{x}_i^\mathsf{T} \mathbf{P}_i \tilde{\boldsymbol{\xi}} \right) \leq 0 \quad \forall t = 1, \ldots, T \\ l_{\text{low}} - l_0 - \displaystyle\sum_{i=1}^t \tilde{\xi}_i + \left(\sum_{i=1}^t x_i^0 + \boldsymbol{x}_i^\mathsf{T} \mathbf{P}_i \tilde{\boldsymbol{\xi}} \right) \leq 0 \quad \forall t = 1, \ldots, T \end{array} \right) \geq 1 - \epsilon \\
& \left. \begin{array}{l} x_t^0 + \boldsymbol{x}_t^\mathsf{T} \mathbf{P}_t \boldsymbol{u} + \boldsymbol{\lambda}_t^\mathsf{T} (\boldsymbol{l} - \boldsymbol{u}) \geq 0 \\ \boldsymbol{\lambda}_t \geq \mathbf{P}_t^\mathsf{T} \boldsymbol{x}_t, \ \boldsymbol{\lambda}_t \geq \mathbf{0} \end{array} \right\} \quad \forall t = 1, \ldots, T
\end{array}
\tag{5.61}
$$

Note that the joint chance constraint in (5.61) involves $2T$ inequalities that are bilinear in the decisions $\{\boldsymbol{x}_t\}_{t=1}^T$ and the random vector $\tilde{\boldsymbol{\xi}}$. Problem (5.61) can therefore be identified as a special instance of problem (5.4) and is amenable to the approximation methods described in

Section 5.3. In the remainder of this section, we compare the performance of these approximation methods.

In the subsequent tests, we set $T = 5$, $l_0 = 1$, $l_{\text{low}} = 1$, and $l_{\text{high}} = 5$. The mean value of $\tilde{\xi}_t$ is assumed to be 1, while its standard deviation is set to 10%, over all time periods. Furthermore, we set the correlation of different stochastic inflows to 25% for adjacent time periods and 0% otherwise. Finally, we assume that $\Xi = [0, 2]^T$. All tests are run for a range of reliability levels ϵ between 1% and 10% in steps of 1%.

We first solve problem (5.61) using the Bonferroni approximation by decomposing the joint chance constraint into $2T$ individual chance constraints with reliability factors $\epsilon_i = \epsilon/(2T)$ for $i = 1, \ldots, 2T$. The resulting optimal objective value is denoted by V^B, and the associated optimal solution is used to initialize Algorithm 5.3.1. We run the algorithm using our new approximation as well as the approximation by Chen *et al.* described in Section 5.3.2. We denote the resulting optimal objective values by V^M and V^U, respectively. In both cases the algorithm's convergence threshold is set to $\gamma = 10^{-6}$.

In Table 5.1 we report the optimal objective values and the improvement of V^M relative to V^U and V^B. As expected, all three methods yield optimal objective values that increase with ϵ because the joint chance constraint becomes less restrictive as ϵ grows. At $\epsilon = 1\%$ the objective values of the different approximations coincide. However, V^M exceed V^U and V^B for all the other values of ϵ. In this particular example, our method outperforms the Bonferroni approximation by up to 24% and the approximation by Chen *et al.* by up to 11%, see Table 5.1.

5.5 Conclusions

In this chapter we developed tractable SDP-based approximations for distributionally robust individual and joint chance constraints. We first showed that distributionally robust individual chance constraints are equivalent to Worst-Case CVaR constraints if the underlying constraint functions are concave or (possibly non-concave) quadratic in the uncertain parameters. We also showed that individual chance constraints can be reformulated as robust semi-infinite

ϵ	V^M	V^U	V^B	$(V^M - V^U)/V^U$	$(V^M - V^B)/V^B$
1%	44.3	44.3	44.3	0.0%	0.0%
2%	44.9	44.3	44.3	1.3%	1.3%
3%	49.4	46.4	44.3	6.4%	11.4%
4%	52.4	48.6	44.5	7.8%	17.6%
5%	54.5	48.8	45.2	11.7%	20.5%
6%	56.3	52.0	46.0	8.3%	22.5%
7%	57.8	54.2	46.7	6.5%	23.6%
8%	58.9	55.8	47.3	5.5%	24.5%
9%	59.9	57.1	47.8	4.8%	25.2%
10%	60.7	58.2	48.8	4.3%	24.5%

Table 5.1: Optimal objective values of the water reservoir control problem when using our new approximation (V^M), the approximation by Chen *et al.* (V^U), and the Bonferroni approximation (V^B). We also report the relative differences between V^M and V^U as well as V^M and V^B.

constraints involving uncertainty sets that can be interpreted as ellipsoids lifted to the space of positive semidefinite matrices.

Subsequently, we used the theory of moment problems to obtain a new approximation for joint chance constraints. We prove that this approximation is tighter that the classical Bonferroni approximation as well as a more recent approximation suggested by Chen *et al.* [CSSC09]. The approximation quality is controlled by a set of scaling parameters. We also showed that the approximation becomes exact if the scaling parameters are chosen optimally. Unfortunately, however, optimizing jointly over the decision variables of the original problem and the scaling parameters leads to a BMI constrained problem and is therefore nonconvex. We therefore proposed a sequential convex optimization algorithm that guides the choice of the scaling parameters. The new approximation also enables us to reformulate joint chance constraints as robust semi-infinite constraints whose uncertainty sets are reminiscent of the lifted ellipsoidal uncertainty sets characteristic for individual chance constraints.

We evaluated the new joint chance constraint approximation in the context of a dynamic water reservoir control problem and demonstrate numerically its superiority over the state-of-the-art approximations.

5.6 Appendix

5.6.1 Proof of Lemma 5.2.1

Proof The worst-case expectation problem (5.9) can equivalently be expressed as

$$\begin{aligned} \theta_{\mathrm{wc}} = \sup_{\mu \in \mathcal{M}_+} \quad & \int_{\mathbb{R}^k} \max\{0, f(\boldsymbol{\xi})\}\mu(\mathrm{d}\boldsymbol{\xi}) \\ \text{s. t.} \quad & \int_{\mathbb{R}^k} \mu(\mathrm{d}\boldsymbol{\xi}) = 1 \\ & \int_{\mathbb{R}^k} \boldsymbol{\xi}\mu(\mathrm{d}\boldsymbol{\xi}) = \boldsymbol{\mu} \\ & \int_{\mathbb{R}^k} \boldsymbol{\xi}\boldsymbol{\xi}^\intercal \mu(\mathrm{d}\boldsymbol{\xi}) = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\intercal, \end{aligned} \tag{5.62}$$

where $\mathcal{M}_+$ represents the cone of nonnegative Borel measures on $\mathbb{R}^k$. The optimization variable of the semi-infinite linear program (5.62) is the nonnegative measure μ. Note that the first constraint forces μ to be a probability measure. The other two constraints enforce consistency with the given first- and second-order moments, respectively.

We now assign dual variables $y_0 \in \mathbb{R}$, $\boldsymbol{y} \in \mathbb{R}^k$, and $\mathbf{Y} \in \mathbb{S}^k$ to the equality constraints in (5.62), respectively, and introduce the following dual problem (see, e.g., [Sha01]).

$$\begin{aligned} \inf \quad & y_0 + \boldsymbol{y}^\intercal \boldsymbol{\mu} + \langle \mathbf{Y}, \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\intercal \rangle \\ \text{s. t.} \quad & y_0 \in \mathbb{R}, \quad \boldsymbol{y} \in \mathbb{R}^k, \quad \mathbf{Y} \in \mathbb{S}^k \\ & y_0 + \boldsymbol{y}^\intercal \boldsymbol{\xi} + \langle \mathbf{Y}, \boldsymbol{\xi}\boldsymbol{\xi}^\intercal \rangle \geq \max\{0, f(\boldsymbol{\xi})\} \quad \forall \boldsymbol{\xi} \in \mathbb{R}^k \end{aligned} \tag{5.63}$$

Because $\boldsymbol{\Sigma} \succ \mathbf{0}$, it can be shown that strong duality holds [Isi60]. Therefore the worst-case probability θ_{wc} coincides with the optimal value of the dual problem (5.63).

By defining the combined variable

$$\mathbf{M} = \begin{bmatrix} \mathbf{Y} & \frac{1}{2}\boldsymbol{y} \\ \frac{1}{2}\boldsymbol{y}^\intercal & y_0 \end{bmatrix},$$

problem (5.63) reduces to

$$\begin{aligned}
\inf_{\mathbf{M}\in\mathbb{S}^{k+1}} \quad & \langle \boldsymbol{\Omega}, \mathbf{M} \rangle \\
\text{s. t.} \quad & \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right]^\mathsf{T} \geq \max\{0, f(\boldsymbol{\xi})\} \quad \forall \boldsymbol{\xi} \in \mathbb{R}^k.
\end{aligned} \tag{5.64}$$

Note that the semi-infinite constraint in (5.64) can be expanded in terms of two equivalent semi-infinite constraints.

$$\left[\boldsymbol{\xi}^\mathsf{T}\ 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right]^\mathsf{T} \geq 0 \qquad \forall \boldsymbol{\xi} \in \mathbb{R}^k \tag{5.65a}$$

$$\left[\boldsymbol{\xi}^\mathsf{T}\ 1\right] \mathbf{M} \left[\boldsymbol{\xi}^\mathsf{T}\ 1\right]^\mathsf{T} \geq f(\boldsymbol{\xi}) \quad \forall \boldsymbol{\xi} \in \mathbb{R}^k \tag{5.65b}$$

Since (5.65a) is equivalent to $\mathbf{M} \succcurlyeq \mathbf{0}$, the claim follows. ∎

Chapter 6

Conclusion

6.1 Summary of Thesis Achievements

The main achievements of this thesis can be summarized as follows.

In Chapter 3, we extended robust portfolio optimization to accommodate options. Moreover, we showed how the options can be used to provide strong insurance guarantees, which also hold when the stock returns are realized outside of the prescribed uncertainty set. The arising model can be interpretted as a fusion between robust portfolio optimization and classical portfolio insurance. Using conic and linear duality, we reformulated the model as a convex second-order cone program, which is scalable in the amount of stocks and options and can be solved efficiently with standard optimization packages. The proposed methodology can be applied to a wide range of uncertainty sets and can therefore be seen as a generic extension to the robust portfolio optimization framework. Through extensive numerical backtesting, we observed that on average the non-insured portfolios achieve higher expected returns than the insured portfolios, whereas the insured portfolios obtain higher Sharpe ratios. The results also indicate that the performance of the insured portfolios is highly dependent on the required level of insurance. When the insurance level is set too high, the cost of insurance causes the performance to deteriorate. Therefore, the level of insurance should be tuned to the market; to preserve wealth, higher insurance levels can benefit the portfolio when the market is volatile and experiences jumps. Lower insurance levels are preferable in less volatile periods since unnecessary insurance costs are avoided.

In Chapter 4, we generalized the WCVaR model by explicitly incorporating the non-linear relationships between derivatives and their underlying assets. To this end, we developed two new models. The WCPVaR model is suited for portfolios containing European options maturing at the investment horizon. WCPVaR expresses the option returns as convex-piecewise linear functions of the underlying assets. A benefit of this model is that it does not require knowledge of the pricing models of the options in the portfolio. However, in order to be tractably solvable, the WCPVaR model precludes short-sales of options. The WCQVaR model can handle portfolios containing general option types and does not rely on short-sales restrictions. It exploits the popular delta-gamma approximation to model the portfolio return. In contrast to WCPVaR, WCQVaR does require knowledge of the option pricing models to determine the quadratic approximation. Through numerical experiments we demonstrate that the WCPVaR and WCQVaR models can provide much tighter VaR estimates of a portfolio containing options than the WCVaR model which does not explicitly account for non-linear dependencies between the asset returns. Using historical backtesting, we analyzed the performance of the WCQVaR model in the context of an index tracking application and find that including options in the investment strategy significantly improves the out-of-sample performance. Although options are typically seen as a risky investments, our numerical results indicate that their use in a robust optimization framework can offer substantial benefits.

In Chapter 5, we developed tractable SDP-based approximations for distributionally robust individual and joint chance constraints. We first showed that distributionally robust individual chance constraints are equivalent to Worst-Case CVaR constraints if the underlying constraint functions are concave or (possibly non-concave) quadratic in the uncertain parameters. We also showed that individual chance constraints can be reformulated as robust semi-infinite constraints involving uncertainty sets that can be interpreted as ellipsoids lifted to the space of positive semidefinite matrices. Subsequently, we used the theory of moment problems to obtain a new approximation for joint chance constraints. We prove that this approximation is tighter that the classical Bonferroni approximation as well as a more recent approximation suggested by Chen *et al.* [CSSC09]. The approximation quality is controlled by a set of scaling parameters. We also showed that the approximation becomes exact if the scaling parameters are chosen optimally. Unfortunately, however, optimizing jointly over the decision variables of the original problem and the scaling parameters leads to a BMI constrained problem and is therefore nonconvex. We therefore proposed a sequential convex optimization algorithm that guides the choice of the scaling parameters. The new approximation also enables us to reformulate joint chance constraints as robust semi-infinite constraints whose uncertainty sets are reminiscent of the lifted ellipsoidal uncertainty sets characteristic for individual chance constraints. We evaluated the new joint chance constraint approximation in the context of a dynamic water reservoir control problem and demonstrated numerically its superiority over the state-of-the-art approximations.

6.2 Directions for Future Research

Possible directions for future research are outlined as follows.

Stochastic Volatility. The WCQVaR model described in Chapter 4 assumes the underlying asset returns to be the only sources of uncertainty in the market. It is known, however, that implied volatilities constitute important risk factors for portfolios containing options. In particular, long dated options are highly sensitive to fluctuations in the volatilities of the underlying assets. The sensitivity of the portfolio return with respect to the volatilities is commonly referred to as *vega risk*. The WCQVaR model can easily be modified to include implied volatilities as additional risk factors. The arising *delta-gamma-vega-approximation* of the portfolio return is still a quadratic function of the risk factors. Thus, the theoretical derivations in Chapter 4 remain valid in this generalized setting. It would be interesting to investigate if extending the WCQVaR model to account for vega risk can further improve the performance of the model in historical backtests.

Reducing the Conservatism of Distributionally Robust Optimization . In this thesis, we only assumed that the first- and second-order moments as well as the support of the true probability distribution were known. It would be worthwhile to investigate if more distributional information can be injected into the distributionally robust optimization framework without impairing the tractability of the models. As more distributional information is included, the set of compatible probability distributions is reduced and the distributionally robust optimization problems become less conservative.

Modeling Nonlinear Dependencies through Copulas. In this thesis, we modeled the nonlinear relationships between derivative and their underlying assets. Copulas are a different approach to model nonlinear dependencies. It would be interesting to investigate whether copula information can be incorporated within the distributionally robust optimization framework in a tractable manner.

Practical Applications. It would be beneficial to further apply the developed theory of robust chance constraints to real-world applications.

Bibliography

[ABRW99] D. Ahn, J. Boudoukh, M. Richardson, and R. F. Whitelaw. Optimal risk management using options. *The Journal of Finance*, 54(1):359–375, 1999.

[ADEH99] P. Artzner, F. Delbaen, J. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.

[AG03] F. Alizadeh and D. Goldfarb. Second-order cone programming. *Mathematical Programming*, 95(1):3–51, 2003.

[AHR10] L. Andrieu, R. Henrion, and J. Römisch. A model for dynamic chance constraints in hydro power reservoir management. *European Journal of Operational Research*, In Press, Corrected Proof, 2010.

[Ale08] C. Alexander. *Market Risk Analysis Volume IV: Value-at-Risk Models.* John Wiley & Sons, 2008.

[BB08] D. Bertsimas and D.B. Brown. Constructing uncertainty sets for robust linear optimization. To appear in Operations Research, 2008.

[BL91] F. Black and R. Litterman. Asset allocation: Combining investor views with market equilibrium. *The Journal of Fixed Income*, 1(3):7–18, 1991.

[BP02] D. Bertsimas and I. Popescu. On the relation between option and stock prices: A convex optimization approach. *Operations Research*, 50(2):358–374, 2002.

[BP08] D. Bertsimas and D. Pachamanova. Robust multiperiod portfolio management in the presence of transaction costs. *Computers & Operations Research*, 35(1):3–17, 2008.

[Bro93] M. Broadie. Computing efficient frontiers using estimated parameters. *Annals of Operations Research*, 45(1):21–58, 1993.

[BS73] F. Black and M. S. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–54, 1973.

[BTBN06] A. Ben-Tal, S. Boyd, and A. Nemirovski. Extending scope of robust optimization: Comprehensive robust counterparts of uncertain problems. *Math. Program., Ser. B*, 107:63–89, 2006.

[BTEGN09] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust Optimization.* Princeton University Press, 2009.

[BTMN00] A. Ben-Tal, T. Margalit, and A. Nemirovski. Robust modeling of multi-stage portfolio problems. *In: Frenk, H., Roos, K., Terlaky, T., Zhang, S. (Eds.), High performance optimization*, pages 303–328, 2000.

[BTN98] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805, 1998.

[BTN99] A. Ben-Tal and A. Nemirovski. Robust solutions of uncertain linear programs. *Operations Research Letters*, 25(1):1–13, 1999.

[BV04] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004.

[CC06] G. Calafiore and M. C. Campi. The scenario approach to robust control design. *IEEE Transactions on Automatic Control*, 51(5):742–753, 2006.

[CCS58] A. Charnes, W.W. Cooper, and G.H. Symonds. Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil. *Management Science*, 4(3):235–263, 1958.

[CEG06] G. Calafiore and L. El Ghaoui. Distributionally robust chance-constrained linear programs with applications. *Journal of Optimization Theory and Applications*, 130(1):1–22, 2006.

[Cho93] V. K. Chopra. Improving optimization. *Journal of Investing*, 2(3):51–59, 1993.

[CHT93] V. K. Chopra, C. R. Hensel, and A. L. Turner. Massaging mean-variance inputs: Returns from alternative investment strategies in the 1980s. *Management Science*, 39(7):845–855, 1993.

[CS02] J. Coval and T. Shumway. Expected option returns. *The Journal of Finance*, 56(3):983 – 1009, 2002.

[CS06] S. Ceria and R. Stubbs. Incorporating estimation errors into portfolio selection: Robust portfolio construction. *Journal of Asset Management*, 7(2):109–127, 2006.

[CSS] X. Chen, M. Sim, and P. Sun. A robust optimization perspective on stochastic programming. *Operations Research*, 55(6):1058–1071.

[CSSC09] W. Chen, M. Sim, J. Sun, and Teo C.P. From CVaR to uncertainty set: Implications in joint chance-constrained optimization. *Operations Research*, 2009.

[CTEG09] G. Calafiore, U. Topcu, and L. El Ghaoui. Parameter estimation with expected and residual-at-risk criteria. *Systems & Control Letters*, 58(1):39–46, 2009.

[CZ93] V. K. Chopra and W. T. Ziemba. The effect of errors in means, variances and covariances on optimal portfolio choice. *Journal of Portfolio Management*, 19(2):6–11, 1993.

[DM74] V. F. Dem'Yanov and V. N. Malozemov. *Introduction to minimax*. Keter Publishing House, Jerusalem, 1974.

[DN09] V. DeMiguel and F. J. Nogales. Portfolio selection with robust estimation. *Operations Research*, 57(3):560–577, 2009.

[DO00] C. Dert and B. Oldenkamp. Optimal guaranteed return portfolios and the casino effect. *Operations Research*, 48(5):768–775, 2000.

[DY10] E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 2010.

[EGOO03] L. El Ghaoui, M. Oks, and F. Outstry. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research*, 51(4):543–556, 2003.

[EI06] E. Erdoğan and G. Iyengar. Ambiguous chance constrained problems and robust optimization. *Mathematical Programming, Series B*, 107:37–61, 2006.

[FKD07] D. Fabozzi, P. Kolm, and Pachamanova D. *Robust Portfolio Optimization and Management*. Wiley, 2007.

[GI03] D. Goldfarb and G. Iyengar. Robust portfolio selection problems. *Mathematics of Operations Research*, 28(1):1–38, 2003.

[Har91] W.V. Harlow. Asset allocation in a downside-risk framework. *Financial analysts journal*, 47(5):28–40, 1991.

[HRS94] M. Howe, B. Rustem, and M. Selby. Minimax hedging strategy. *Computational Economics*, 7(4):245–275, 1994.

[Idz02] T. Idzorek. A step-by-step guide to the black-litterman model. Technical report, Duke University, 2002.

[Isi60] K. Isii. The extrema of probability determined by generalized moments (i) bounded random variables. *Annals of the Institute of Statistical Mathematics*, 12(2):119–134, 1960.

[Jas02] S. R. Jaschke. The cornish-fisher expansion in the context of delta-gamma-normal approximations. *Journal of Risk*, 4(4):33–53, 2002.

[JM03] R. Jagannathan and T. Ma. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance*, 58(4):1651–1684, 2003.

[Jor01] P. Jorion. *Value-at-Risk: The New Benchmark for Managing Financial Risk*. McGraw-Hill, 2001.

[KWG09] D. Kuhn, W. Wiesemann, and A. Georghiou. Primal and dual linear decision rules in stochastic and robust optimization. *Mathematical Programming*, 2009.

[LÖ4] J. Löfberg. YALMIP : A toolbox for modeling and optimization in MATLAB. In *Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004.

[LA08] J. Luedtke and S. Ahmed. A sample approximation approach for optimization with probabilistic constraints. *SIAM Journal on Optimization*, 19(2):674–699, 2008.

[Lev92] L. Levy. Stochastic dominance and expected utility: Survey and analysis. *Management Science*, 38(4):555–593, 1992.

[LS08] A. Lucas and A. Siegmann. The effect of shortfall as a risk measure for portfolios with hedge funds. *Journal of Business Finance & Accounting*, 35(1-2):200–226, 2008.

[LSK06] F. Lutgens, S. Sturm, and A. Kolen. Robust one-period option hedging. *Operations Research*, 54(6):1051–1062, 2006.

[Lue98] D. G. Luenberger. *Investment Science*. Oxford University Press, Madison Avenue, New York, 1998.

[LVBL98] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret. Applications of second-order cone programming. *Linear Algebra and its Applications*, 284(1):193–228, 1998.

[Mac92] L. G. MacMillan. *Options as a strategic investment*. Prentice Hall, 1992.

[Mar52] H. Markowitz. Portfolio selection. *Journal of Finance*, 7(1):77–91, 1952.

[Mer76] R. C. Merton. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3(1-2):125–144, 1976.

[Mer80] R. C. Merton. On estimating the expected return on the market: an exploratory investigation. *Journal of Financial Economics*, 8(4):323–361, 1980.

[Meu05] A. Meucci. *Risk and Asset Allocation*. Springer, Berlin, 2005.

[Mic01] R. O. Michaud. *Efficient Asset Management: A Practical Guide to Stock Portfolio Management and Asset Allocation*. Oxford University Press, 2001.

[MN79] I. Meilijson and A. Nadas. Convex majorization with an application to the length of critical path. *Journal of Applied Probability*, (16):671–677, 1979.

[MU99] J. Mina and A. Ulmer. Delta-gamma four ways. Technical report, RiskMetrics Group, 1999.

[MW65] L.B. Miller and H. Wagner. Chance-constrained programming with joint constraints. *Operations Research*, 13(6):930–945, 1965.

[NPS08] K. Natarajan, D. Pachamanova, and M. Sim. Incorporating asymmetric distributional information in robust value-at-risk optimization. *Management Science*, 54(3):573–585, 2008.

[NPS09] K. Natarajan, D. Pachamanova, and M. Sim. Constructing risk measures from uncertainty sets. *Operations Research*, 57(5):1129–1141, 2009.

[NS06] A. Nemirovski and A. Shapiro. Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, 17(4):969–996, 2006.

[PAS09] B. K. Pagnoncelli, S. Ahmed, and A. Shapiro. Sample average approximation method for chance constrained programming: Theory and applications. *Journal of Optimization Theory and Applications*, 142(2), 2009.

[Pre70] A. Prekopa. On probabilistic constrained programming. In *Proceedings of the Princeton Symposium on Mathematical Programming*, pages 113–138, Princeton University Press, Princeton, 1970.

[PT07] I. Polik and T. Terlaky. A survey of the S-lemma. *SIAM Review*, 49(3):371–481, 2007.

[RBM00] B. Rustem, R. G. Becker, and W. Marty. Robust min-max portfolio strategies for rival forecast and risk scenarios. *Journal of Economic Dynamics and Control*, 24(11-12):1591–1621, 2000.

[RH02] B. Rustem and M. Howe. *Algorithms for Worst-Case Design and Applications to Risk Management*. Princeton University Press, 2002.

[RU02] R.T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–41, 2002.

[Sca58] H. Scarf. A min-max solution of an inventory problem. studies in the mathematical theory of inventory and production. pages 201–209, 1958.

[Sha66] W. F. Sharpe. Mutual fund performance. *The Journal of Business*, 39(1):119–138, 1966.

[Sha01] A. Shapiro. *On duality theory of conic linear problems*. Semi-infinite programming: recent advances (M.A.Goberna and M.A. Lopez, eds.), Kluwer Academic Publishers, 2001.

[SK02] A. Shapiro and A.J. Kleywegt. Minimax analysis of stochastic problems. *Optimization Methods and Software*, (17):523–542, 2002.

[TK04] R. H. Tütüncü and M. Koenig. Robust asset allocation. *Annals of Operations Research*, 132(1-4):157–187, 2004.

[TO95] O. Toker and H. Ozbay. On the NP-hardness of solving bilinear matrix inequalities and simultaneous stabilization with static output feedback. In *Proceedings of the American Control Conference*, pages 2525–2526, Seatle, Washington, 1995.

[TTT03] R. H. Tütüncü, K. C. Toh, and M. J. Todd. Solving semidefinite-quadratic-linear programs using SDPT3. *Mathematical Programming Ser. B*, 95:189–217, 2003.

[VB96] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Review*, 38(1):49–95, 1996.