Johannes Moosbrugger

Distributionally Robust Portfolio Optimization

MASTER THESIS

written to obtain the academic degree of a

Master of Science (MSc)

Master programme financial and actuarial mathematics

submitted at the

Graz University of Technology

Advisor:
Ao.Univ.-Prof. Dipl.-Ing. Dr.techn. Eranda Dragoti-Cela

Institute for Optimization and Discrete Mathematics

Graz, April 2015

EIDESSTATTLICHE ERKLÄRUNG

AFFIDAVIT

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Das in TUGRAZonline hochgeladene Textdokument ist mit der vorliegenden Masterarbeit identisch.

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present master‘s thesis.

Datum/Date

Unterschrift/Signature

ABSTRACT

The Markowitz mean-variance efficiency framework states probably the most popular and widely known approach in modern portfolio optimization. It allows the specification of a so-called risk aversion parameter which intends to govern the risk taken by investing into the resulting portfolio. A major shortcoming of this approach is its assumption of the asset returns being (jointly) normally distributed random variables. We avoid this assumption by introducing a new method to a posteriori measure the performance of a portfolio. A priori, this measure states a one dimensional random variable whose probability function only depends on the unknown asset return distribution and the length of the investment horizon. We then use well known methods from distributionally robust optimization to reformulate the resulting portfolio optimization problem as a tractable conic program whose size does not depend on the length of the investment horizon. Thereby, we only assume the first- and second-order moments of the asset return distribution to be known. This approach allows for an easy robustification against the ambiguity that arises from estimating these two moments and also for a choice of a risk aversion parameter. For a certain choice of this risk aversion parameter, a robustified approximation of the growth optimal portfolio is attained. The empirical backtests show that the robust portfolios offer a more moderate performance, i.e. the performances follow their mean-variance efficient counterparts but are less extreme.

ZUSAMMENFASSUNG

Der Ansatz der effizienten Portfolios nach Markowitz gilt als Grundlage der modernen Portfoliotherie. Dabei ist es möglich einen sogenannten Risikoaversionsparameter zu wählen. Dieser soll das Risiko, dem man durch ein Investment in das resultierende Portfolio ausgesetzt ist, steuern. Ein großer Kritikpunkt an diesem Ansatz ist die Annahme, dass die Renditen der betrachteten Vermögenswerte normalverteilte Zufallsvariablen sind. Um dieses Problem zu umgehen führen wir ein neues Maß zur a posteriori Bewertung von Portfoliorenditen ein. A priori stellt dieses Maß eine eindimensionale Zufallsvariable dar dessen Verteilungsfunktion nur von der Verteilung der Renditen und der Länge des Betrachtungszeitraumes abhängt. Mit der Hilfe von bekannten Methoden der verteilungsrobusten Optimierung können wir das Portfolio-Optimierungsproblem als ein konisches Programm formulieren, dessen Größe nicht mehr von der Länge des Betrachtungszeitraumes abhängt. Dabei nehmen wir lediglich an, dass die ersten beiden Momente der Rendite-Verteilung bekannt sind. Dieser Ansatz erlaubt uns eine einfache Robustifizierung gegenüber den Schätzfehlern dieser ersten beiden Momente und ebenso die Festlegung der Risikoaversion durch einen Parameter. Für eine bestimmte Wahl dieses Risikoaversionsparameters erhalten wir eine robustifizierte Approximation des sogenannten wachstumsoptimalen Portfolios. Die empirischen Tests belegen, dass diese robusten Portfolios ein gemäßigteres Verhalten aufweisen, das heißt dass die beobachteten Renditen zwar jenen der klassischen effizienten Portfolios entsprechen, jedoch nicht so extrem sind.

Contents

List of Figures

List of Tables

List of Algorithms

1 Introduction

In this thesis, we investigate the problem of choosing an optimal combination of various risky assets to invest in. A very intuitive and popular approach was presented by [Markowitz, 1952]. He introduced the notion of an efficient trade-off between the expected return and the risk of a portfolio, which he quantified by the mean and the variance of the portfolio returns. Beside its simplicity and intuitive approach, the main reason for the huge popularity of the Markowitz model is that the so-called mean-variance efficient portfolios can be easily computed as the optimal solution of a quadratic optimization problem. The trade-off between return and risk can be specified by a risk aversion parameter. The most risk averse portfolio in this framework is called minimum-variance portfolio and simply minimizes the portfolio variance.

On the other hand, the Markowitz approach is burdened by some major disadvantages. Since it is only considered for one rebalancing period, and its consecutive application will lead to almost sure ruin in the long run (see [Roll, 1973]), its usage for long term investing is not advisable. Also, the true mean and variance of the asset-returns are assumed to be known. As these moments can only be estimated and the Markowitz approach does not account for the implied ambiguity, this constitutes a likely source of errors.

Of course, since Markowitz first published his work, many other approaches have been made to tackle these shortcomings. The so-called Kelly strategy, which maximizes the expected portfolio growth rate ("growth-optimal portfolios", see [Luenberger, 1998]), is one of these other approaches which gained wide popularity. It can be shown that, in the long run, the growth-optimal portfolio accumulates more wealth than any other portfolio with probability one. Like the mean-variance efficient portfolio, the Kelly strategy can be easily computed by solving a single-period convex optimization problem.

A drawback of the Kelly strategy is that the time needed to assure that it outperforms any other strategy can be impracticably long. Also, the Kelly strategy cannot be tailored to a specific time horizon and ignores moment ambiguity.

In their recently drafted paper [Rujeerapaiboon et al., 2014] introduced the robust growth-optimal portfolio that *"offers similar performance guarantees as the classical growth-optimal portfolio but for finite investment horizons and ambiguous return distribution"*. This behaviour is achieved by maximizing the worst-case value-at-risk at level ϵ of a quadratic approximation of the portfolio growth rate, where the worst case is taken over all distributions in a predefined ambiguity set. The robust growth-optimal portfolio can also be efficiently computed as the solution of a second-order cone program (shortly SOCP) whose size does not depend on the length of the investment horizon and allows for easy robustification against moment ambiguity.

We introduce a method to a posteriori evaluate the performance of a portfolio by a quadratic polynomial function of the realized portfolio returns. By definition, this quadratic polynomial has a positive curvature and a parametrized minimum at γ. This evaluation states a random variable a priori which we aim to "minimize". We then use the idea and conceptional derivation presented in [Rujeerapaiboon et al., 2014] to reformulate our portfolio optimization problem as a SOCP. As will be clear by the definition of γ, this parameter will influence the risk aversion of the resulting portfolio. For certain choices of γ, the robust growth-optimal portfolio or the Markowitz minimum-variance portfolio can be obtained. We will also derive a technical lower bound γ^* for the possible values of γ. If γ is chosen as γ^*, the resulting portfolio marks the most risk averse robust portfolio in the sense of minimal expected variance. We will see that this robust risk averse portfolio is in some sense the distributionally robust counterpart to the classical minimum-variance portfolio. Furthermore, as all of our robust portfolios are mean-variance efficient in the classical sense, we interpret our approach as an distributionally robust extension to the mean-variance efficiency framework. The parameter γ defines the exact position of a robust portfolio on the efficient frontier. This "robust efficient frontier", i.e. the set of all robust portfolios which arises from different choices of γ, will thereby only cover a small fraction of the classical efficient frontier. Similar to the approach of Rujeerapaiboon et al., we can further robustify our portfolio optimization against moment ambiguity. The resulting problem will also be a SOCP. It will be clear from the problem formulation that these portfolios are also mean-variance efficient in the classical sense. Therefore, the set of all robust portfolios with moment ambiguity will lie on the efficient frontier and we will see that it moved "towards" the minimum variance portfolio.

The remainder of this thesis is structured as follows. As our resulting portfolio optimization problem will be formulated as a second-order cone program, we give a brief introduction of conic programming in Section 2. In Section 3, we will recap the classical portfolio selection model of Markowitz and the Kelly growth-optimal strategy. Our method of evaluating portfolios, inspired by the approach of Rujeerapaiboon et al., with the introduction of the new risk-aversion parameter γ is presented in Section 4. The preliminaries and the actual formulation of the resulting optimization problem as a second-order cone program are introduced in Section 5. There, we also address the implication of particular choices of γ. In Section 6 we present our methods of choice for estimating moments of asset-returns and the corresponding uncertainty cones. From our empirical backtests in Section 7 we see that the robust risk averse portfolios show more moderate performances than the minimum-variance portfolio. This means that in scenarios where the minimum-variance portfolio performed good, so did its robust counterpart, but not as good. Vice versa, in scenarios where the minimum-variance portfolio performed bad, so did the robust risk averse portfolio, but not as bad. We will also see that the robust portfolios outperformed the naive equally weighted portfolio when applied to the components of the Dow Jones Industrial Average over the time period December 2005 to December 2010, which covers the outbreak of the global financial crisis. We conclude in Section 8.

1.1 Notation

The following abbreviations/notations are used:

(QC)LP…(Quadratically Constrained) Linear Program

(QC)QP…(Quadratically Constrained) Quadratic Program

SOCP…Second-order Cone Program

SDP…Semidefinite Program

$\mathbb{I}$…The identity matrix in the appropriate dimension

$\mathbf{1}$…The vector of ones in the appropriate dimension

δ_{st}…The Kronecker-Delta, i.e. $\delta_{st} = \begin{cases} 1 & if \quad s = t \\ 0 & else \end{cases}$

P_0^n…The set of all not degenerated probability distributions on $\mathbb{R}^n$

$\mathbf{X}'$…The transpose of $\mathbf{X} \in \mathbb{R}^{m \times n}$

$\mathbb{S}^n$ ($\mathbb{S}^n_+$)…The space of symmetric (symmetric positive semidefinite) matrices in $\mathbb{R}^{n \times n}$

$Tr(\mathbf{X})$…The trace of a matrix $\mathbf{X} = \{x_{ij}\}_{i,j=1,\ldots,n} \in \mathbb{R}^{n \times n}$ given by $Tr(\mathbf{X}) = \sum_{i=1}^{n} x_{ii}$

$\langle \mathbf{X}, \mathbf{Y} \rangle = Tr(\mathbf{XY})$…The trace scalar product for any $\mathbf{X}, \mathbf{Y} \in \mathbb{S}^n$

$\mathbf{X} \succeq \mathbf{Y} (\mathbf{X} \succ \mathbf{Y})$…Indicates that $\mathbf{X} - \mathbf{Y}$ is positive semidefinite (positive definite)

$\lambda_{min}(\mathbf{X})$…The smallest eigenvalue of a matrix $\mathbf{X} \in \mathbb{R}^{n \times n}$

$\lambda_{max}(\mathbf{X})$…The largest eigenvalue of a matrix $\mathbf{X} \in \mathbb{R}^{n \times n}$

$\mathbf{X}^{1/2}$…The "square root" of a matrix $\mathbf{X} \in \mathbb{S}^n_+$, i.e. $\mathbf{X}^{1/2} \cdot \mathbf{X}^{1/2} = \mathbf{X}$

2 Second-Order Cone Programming

Since we are going to reformulate the distributionally robust portfolio optimization problem as a second-order cone program in Section 5.2, we want to recall some basic notions about cone optimization in general and second-order cone programming in particular. We will define (second-order) cones in Section 2.1, where we also state the general formulation of a conic optimization problem. Second-order cone programs and their general formulation are reviewed in Section 2.2. In Section 2.3 we will present some techniques on how to solve such problems. These also represent the reasons for our interest in SOCPs, since they enable us to numerically solve our portfolio optimization problem with little computational effort.
This short introduction is based mainly on [Lobo et al., 1998] and [Alizadeh and Goldfarb, 2002].

2.1 Cones in $\mathbb{R}^n$

In order to solve SOCPs, we should of course remember the definition of a cone in the n-dimensional space of real numbers denoted by $\mathbb{R}^n$, where n is a natural number.

Definition 2.1 (Cone)**.** A n-dimensional cone is a subset of $\mathbb{R}^n$ which is closed under multiplication with a non-negative scalar, i.e.

$$C \subseteq \mathbb{R}^n \text{ is a cone} \Leftrightarrow \forall x \in C\, \forall \lambda \in \mathbb{R}_+ : \quad x \cdot \lambda \in C.$$

A pointed cone is a cone which does not contain any line, or equivalently $C \cap (-C) = \{0\}$, where $-C := \{-x | x \in C\}$.

A pointed and convex cone C with non-empty interior $int(C)$ induces a partial order on $\mathbb{R}^n$ with

$$\begin{aligned} x \succeq_C y &\Leftrightarrow x - y \in C \\ x \succ_C y &\Leftrightarrow x - y \in int(C), \end{aligned}$$

where $x, y \in \mathbb{R}^n$ and $int(C)$ denotes the interior of C.
The above definition includes of course many different types of cones. One example in the two-dimensional Euclidean space can be seen in Figure 1. Note that also the set $\mathbb{S}^n_+$ of symmetric and positive semidefinite $n \times n$-matrices is a cone. For our particular application though, we can restrict our attention to so-called second-order cones.

Definition 2.2 (Second-order cone)**.** The $(n+1)$-dimensional second-order cone (also called Lawrence or ice-cream cone) in $\mathbb{R}^{n+1}$ is defined as

$$C_2^{n+1} := \left\{ (x_0, x_1, \ldots, x_n) \in \mathbb{R}^{n+1} : x_0 \geq \|(x_1, x_2, \ldots, x_n)\|_2 \right\},$$

where $\| \cdot \|_2$ denotes the Euclidean vector norm, i.e. $\|(x_1, x_2, \ldots, x_n)\|_2 = \sqrt{x_1^2 + \ldots + x_n^2}$.

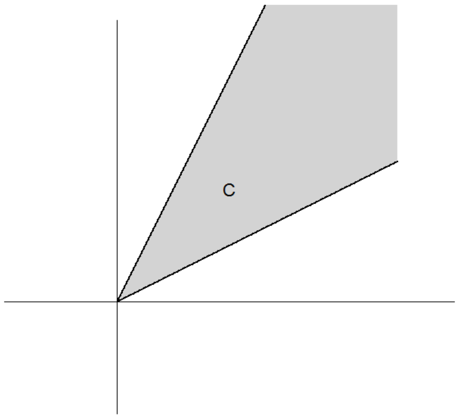

Figure 1: Cone in the two-dimensional Euclidean space

The reason for calling second-order cones "ice-cream cones" becomes clear if we visualize it in three dimensions, which is plotted in Figure 2. Note that this cone is a pointed, convex and closed cone.

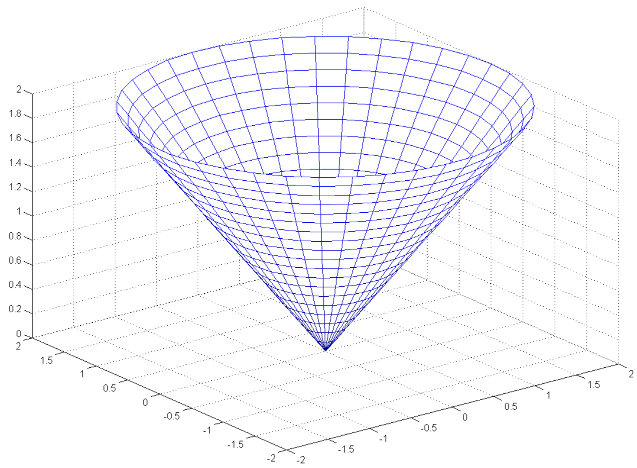

Figure 2: The three dimensional second-order cone

For every cone, we can define its so-called dual cone.

Definition 2.3. Let C be a cone. Its dual cone C^* is defined as

$$C^* := \{x \in \mathbb{R}^n | \forall y \in C : \langle x, y \rangle \geq 0\},$$

where $\langle \cdot, \cdot \rangle$ denotes an inner product on $\mathbb{R}^n$. Here we use $\langle x, y \rangle = x_1 y_1 + \ldots + x_n y_n$.

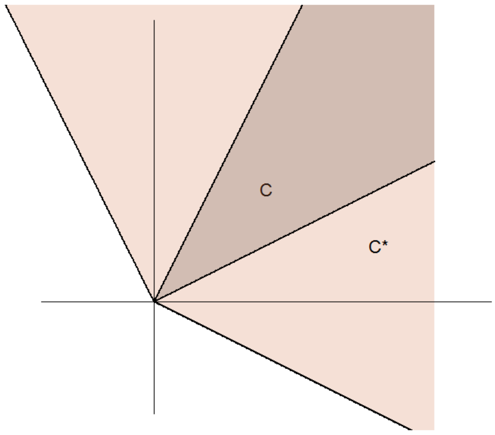

Figure 3: Cone and dual cone in the two-dimensional Euclidean space

The dual cone of our exemplary cone in the two-dimensional Euclidean space can be seen in Figure 3. If a cone coincides with its dual, it is called a self-dual cone.

We can now formulate a conic optimization problem. Let $\mathbf{A} \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ and $e \in \mathbb{R}^n$. For C a closed and convex cone in $\mathbb{R}^n$, a conic optimization problem is given by

$$\begin{aligned} &\min_x e'x \\ &\text{s.t. } \mathbf{A}x = b \quad \text{and} \quad x \in C. \end{aligned} \tag{1}$$

Note that for $C = \mathbb{R}^n_+$ the conic optimization Problem (1) reduces to a linear optimization problem.

2.2 General Second-Order Cone Program-Formulation

Similar to the conic optimization Problem (1), in a second-order cone program we minimize a linear objective function where the set of all feasible points is now given as the intersection of an affine set and finitely many transformed second-order cones.

Let $\mathbf{A}_i \in \mathbb{R}^{(n_i-1)\times n}$, $\mathbf{F} \in \mathbb{R}^{p\times n}$, $g \in \mathbb{R}^p$, $b_i \in \mathbb{R}^{n-1}$, $c_i \in \mathbb{R}^n$, $d_i \in \mathbb{R}$ and $e \in \mathbb{R}^n$ where $n_i \in \mathbb{N}$ for $i = 1, \ldots, N$ and $n, p \in \mathbb{N}$. A second-order cone program is then given by

$$\begin{aligned} &\min_x e'x \\ &\text{s.t. } \|\mathbf{A}_i x + b_i\|_2 \leq c_i' x + d_i,\ i = 1, \ldots, N \\ &\mathbf{F}x = g, \end{aligned} \tag{2}$$

where $\|\cdot\|_2$ again denotes the Euclidean norm.

Note that the constraints $\|\mathbf{A}_i x + b_i\|_2 \leq c_i' x + d_i$ are called second-order cone constraints since

$$\|\mathbf{A}_i x + b_i\|_2 \leq c_i' x + d_i \Leftrightarrow \begin{bmatrix} c_i' \\ \mathbf{A_i} \end{bmatrix} x + \begin{bmatrix} d_i \\ b_i \end{bmatrix} \in C_2^{n_i}.$$

As its objective is a convex function and the set of all feasible solutions is a convex set, the second-order cone Problem (2) states a convex optimization problem.
Many common convex constraints can be represented as second-order cone constraints. Therefore, a lot of basic optimization problems can be formulated as SOCPs. Some of the most prominent examples are listed below, see [Alizadeh and Goldfarb, 2002] for a more exhaustive overview.

- **Linear Programs (LPs):**
  If $\mathbf{A}_i = 0$ and $b_i = 0$ (or equivalently $n_i = 1$) for all $i = 1, \ldots, N$, the second-order cone Program (2) reduces to a linear program of the form

$$\begin{aligned} &\min_x e' x \\ &\text{s.t. } 0 \leq c_i' x + d_i,\ i = 1, \ldots, N \\ &\mathbf{F} x = g. \end{aligned}$$

- **Quadratically Constrained Linear Programs (QCLPs):**
  For $c_i = 0$, the i-th constraint of Program (2) reduces to $\|\mathbf{A}_i x + b_i\|_2 \leq d_i$. If we assume $d_i \geq 0$, this is equivalent to $\|\mathbf{A}_i x + b_i\|_2^2 \leq d_i^2$. Therefore, if for all $i = 1, \ldots, N$ the parameters c_i are equal to zero, the second-order cone program reduces to a quadratically constrained linear program given by

$$\begin{aligned} &\min_x e' x \\ &\text{s.t. } \|\mathbf{A}_i x + b_i\|_2^2 \leq d_i^2,\ i = 1, \ldots, N \\ &\mathbf{F} x = g. \end{aligned}$$

- **(Convex) Quadratic Programs (QPs):**
  For $\mathbf{P_0} \in \mathbb{R}^{n \times n}$ symmetric, $q_0, \in \mathbb{R}^n$, $a_i \in \mathbb{R}^n$, $b_i \in \mathbb{R}$ and $r_0 \in \mathbb{R}$, where $i = 1, \ldots, N$, a convex quadratic program is given by

$$\begin{aligned} &\min_x x' \mathbf{P}_0 x + 2 q_0' x + r_0 \\ &\text{s.t. } a_i' x \leq b_i,\ i = 1, \ldots, N. \end{aligned}$$

  If we assume $\mathbf{P}_0$ to be positive definite (i.e. $\mathbf{P}_0 \succ 0$), the above problem can be reformulated as a second-order cone program of the form

$$\begin{aligned} &\min_{t,x} \quad t \\ &\text{s.t. } \|\mathbf{P}_0^{1/2} x + (\mathbf{P}_0^{1/2})^{-1} q_0\|_2 \leq t \\ &a_i' x \leq b_i,\ i = 1, \ldots, N, \end{aligned}$$

since every solution that minimizes $\|\mathbf{P}_0^{1/2}x + (\mathbf{P}_0^{1/2})^{-1}q_0\|_2^2 = x'\mathbf{P}_0x + 2q_0'x + q_0'\mathbf{P}^{-1}q_0$ also minimizes the objective of the original convex quadratic program. Note that QPs can be reformulated as SOCPs in general and therefore also for $\mathbf{P}_0$ only positive semidefinite. We neglect this case since the general reformulation would require additional effort and is irrelevant for the introductory character of this section.

- **(Convex) Quadratically Constrained Quadratic Programs (QCQPs):**
  A general convex quadratically constrained quadratic program is given by

$$\begin{aligned} &\min_x x'\mathbf{P}_0x + 2q_0'x + r_0 \\ \text{s.t. } &x'\mathbf{P}_ix + 2q_i'x + r_i \leq 0,\ i = 1, \dots, N, \end{aligned}$$

  where the matrices $\mathbf{P}_i \in \mathbb{R}^{n\times n}$ are symmetric and positive semidefinite and $q_i \in \mathbb{R}^n$, $r_i \in \mathbb{R}$ for $i = 0, \dots, N$.
  In the special case where the matrices $\mathbf{P}_i$ are positive definite, the above QCQP can be written as

$$\begin{aligned} &\min_{t,x} t \\ \text{s.t. } &\|\mathbf{P}_0^{1/2}x + (\mathbf{P}_0^{1/2})^{-1}q_0\|_2 \leq t \\ &\|\mathbf{P}_i^{1/2}x + (\mathbf{P}_i^{1/2})^{-1}q_i\|_2 \leq (q_i'\mathbf{P}_i^{-1}q_i - r_i)^{1/2},\ i = 1, \dots, N, \end{aligned}$$

  which states a SOCP with $N + 1$ constraints.
  Note that the constraints $\|\mathbf{P}_i^{1/2}x + (\mathbf{P}_i^{1/2})^{-1}q_0\|_2 \leq (q_i'\mathbf{P}_i^{-1}q_i - r_i)^{1/2}$ are equivalent to $x'\mathbf{P}_ix + 2q_i'x + q_i'\mathbf{P}_i^{-1}q_i \leq q_i'\mathbf{P}_i^{-1}q_i - r_i$ which coincides with the original QCQP constraints.

In summary, many common optimization problems can be recast as second-order cone programs. On the other side, SOCPs are a special case of semidefinite programs (SDPs). For SDPs, the set of all feasible solutions is given by the cone of positive semidefinite matrices. For further information about SDPs, we refer to [Laurent and Rendl, 2005].

- **Semidefinite Programs (SDPs):**
  For $\mathbf{F}_i, \mathbf{C} \in \mathbb{S}^n$ with $i = 1, \dots, n$ and $e \in \mathbb{R}^n$ a general semidefinite program is given by

$$\begin{aligned} &\min_X tr(\mathbf{C} \cdot \mathbf{X}) \\ \text{s.t. } &(tr(\mathbf{F}_1 \cdot \mathbf{X}), \dots, , tr(\mathbf{F}_n \cdot \mathbf{X}))' = e, \\ &\mathbf{X} \succeq 0. \end{aligned}$$

The (Lagrangian-) dual of an optimization problem is obtained by using non-negative multipliers to add the constraints to the objective function. The resulting function is called Lagrangian and the multipliers are considered as the dual variables. We can then solve for the primal variable values (the original optimization variables) that minimize the Lagrangian as a function of the dual variables. The dual problem is to maximize this function with respect to the dual variables. One can show that the dual of the above SDP is given by

$$\max_x e'x$$
$$\text{s.t.} \sum_{i=1}^{n} x_i \mathbf{F}_i \preceq \mathbf{C},$$

where $x = (x_1, \ldots, x_n)' \in \mathbb{R}^n$ is the optimization variable.

If we define p^* to be the optimal value of the primal and d^* to be the optimal values of the dual problem, the so-called duality gap is given by $p^* - d^*$. It can be shown that the duality gap is non-negative, i.e. the value of the primal SDP is at least the value of the dual SDP. The primal and the dual are called strictly feasible if there exists a feasible $\mathbf{X} \succ 0$ for the dual and a $x \in \mathbb{R}^n$ such that $\sum_{i=1}^{n} x_i \mathbf{F}_i \prec \mathbf{C}$. In this case, the duality gap is equal to zero. See [Vandenberghe and Boyd, 1996] for more details on the duality of semidefinite problems.

We can reformulate the Second-Order Cone Program (2) as the dual of an SDP. For this purpose we observe that a second-order cone constraint is equivalent to a linear matrix inequality, i.e. for $u \in \mathbb{R}^n$ and $t \in \mathbb{R}$

$$\|u\|_2 \leq t \Leftrightarrow \begin{bmatrix} t\mathbb{I}_n & u \\ u' & t \end{bmatrix} \succeq 0.$$

The above equivalence can be easily verified by using Sylvester's criterion[1]. Therefore, if $t \geq 0$, for the above equivalence to hold it is sufficient that the determinant of the right hand matrix is non-negative. This determinant is given by

$$det\left(\begin{bmatrix} t\mathbb{I}_n & u \\ u' & t \end{bmatrix}\right) = t^n(t - \frac{1}{t}u'u),$$

and obviously $t^n(t - \frac{1}{t}u'u) \geq 0 \Leftrightarrow t \geq \|u\|_2$.
This result can be used to reformulate a constraint $\|\mathbf{A}x + b\|_2 \leq c'x + d$ of a SOCP as

$$\begin{bmatrix} (c'x+d)\mathbb{I} & \mathbf{A}x + b \\ (\mathbf{A}x+b)' & (c'x+d) \end{bmatrix} \succeq 0. \tag{3}$$

If we define

$$\mathbf{F}_i := \begin{bmatrix} c_i & 0 & \ldots & 0 & \mathbf{A}_{1i} \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \ldots & 0 & c_i & \mathbf{A}_{ni} \\ \mathbf{A}_{1i} & \ldots & \ldots & \mathbf{A}_{ni} & c_i \end{bmatrix} \quad \text{and} \quad \mathbf{C} := \begin{bmatrix} -d\mathbb{I} & -b \\ -b' & -d \end{bmatrix},$$

[1] Sylvester's criterion states that a quadratic matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is positive semidefinite if and only if all matrices determined by the upper left $k \times k$ corner of $\mathbf{A}$ ($k = 1, \ldots, n$) have a non-negative determinant

it follows that Equation (3) is equivalent to $\sum_{i=1}^{n} x_i \mathbf{F}_i \succeq \mathbf{C}$. We conclude that in some sense, SOCPs lie somewhere in-between LPs and SDPs, while they also comprise QPs.

2.3 Solving Second-Order Cone Programs

In this section, we will focus our attention on methods to solve SOCPs. We will therefore only briefly review the theoretical background and give a quick overview on available software packages that can handle second-order cone programs.

Similar to LPs, QPs, and SDPs, optimal solutions to SOCPs can be approximated with any given accuracy in polynomial time by using interior point methods. Note that since SOCPs can be recast as SDPs, it would be possible to solve them as SDPs. Nevertheless, this is not advisable since the computational effort to solve SDPs is in general higher than in the case where algorithms specifically designed to solve SOCPs are involved. Since many empirical studies have shown that primal-dual interior-point algorithms often show more appealing properties than primal only, we will now have a look at the dual of the second-order cone Problem (2). To do so, we will first recall the formulation of the general second-order cone Problem (2) which was given as

$$\begin{aligned} &\min_x e'x \\ &\text{s.t. } \|\mathbf{A}_i x + b_i\|_2 \leq c_i' x + d_i,\ i = 1, \ldots, N, \end{aligned} \tag{4}$$

where we neglect the affine restriction for simplicity.

In order to obtain the (Lagrangian-) dual of the above problem we again solve for the primal variable values that minimize the Lagrangian as a function of the dual variables. The dual is then given by maximizing this function with respect to the dual variables. For this purpose, we introduce new variables, and rewrite the second-order cone Problem (4) as

$$\begin{aligned} &\min_x e'x \\ &\text{s.t. } \|y_i\|_2 \leq t_i,\ i = 1, \ldots, N \\ &\quad\ y_i = \mathbf{A}_i x + b_i,\ t_i = c_i' x + d_i,\ i = 1, \ldots, N. \end{aligned}$$

The Lagrangian $L(x, y, t, w, z, \mu) =: L$ is then given by

$$\begin{aligned} L &= e'x + \sum_{i=1}^{N} w_i(\|y_i\|_2 - t_i) + \sum_{i=1}^{N} z_i'(y_i - \mathbf{A}_i x - b_i) + \sum_{i=1}^{N} \mu_i(t_i - c_i' x - d_i) \\ &= (e - \sum_{i=1}^{N} (\mathbf{A}_i' z_i + \mu_i c_i))' x + \sum_{i=1}^{N} (w_i \|y_i\|_2 + z_i' y_i - w_i t_i + \mu_i t_i - b_i' z_i - d_i \mu_i), \end{aligned}$$

which, as a function in x, is bounded from below if and only if $e = \sum_{i=1}^{N}(\mathbf{A}_i' z_i - \mu_i c_i)$. Similarly, we observe that as a function in t_i, the above Lagrangian is bounded from below if and only if $w_i = \mu_i$. We also note that

$$\inf_{y_i}(w_i\|y_i\|_2 + z_i' y_i) = \begin{cases} 0 & \text{if } \|z_i\|_2 \le w_i \\ -\infty & \text{else.} \end{cases}$$

Hence, if we minimize the above Lagrangian with respect to the primal variables, the optimal value is $-\sum_{i=1}^{N}(b_i' z_i + d_i w_i)$. As mentioned earlier, this optimal value serves as the objective function for the dual problem, which we aim to maximize. By considering the restrictions we derived above, the SOCP-dual is thus given as

$$\begin{aligned} &\max_{z_i, w_i} - \sum_{i=1}^{N}(b_i' z_i + d_i w_i) \\ \text{s.t. } &\sum_{i=1}^{N}(\mathbf{A}_i' z_i + c_i w_i) = e,\ i = 1, \dots, N \\ &\|z_i\|_2 \le w_i,\ i = 1, \dots, N, \end{aligned} \tag{5}$$

where $z_i \in \mathbb{R}^{n_i - 1}$ and $w \in \mathbb{R}^N$ are the dual optimization variables.

We observe that the above SOCP-dual (5) again states a convex second-order cone program. Similar to SDPs, the duality gap for SOCPs is always non-negative. If both, the primal and the dual SOCP, are strictly feasible, the duality gap is equal to zero. See [Alizadeh and Goldfarb, 2002, §5] for more details on the duality theory of SOCPs.

Many interior-point methods which were initially developed for linear programming can be extended to solve SOCPs. Likewise, the majority of interior-points methods which were developed for semidefinite programming can be specialized for SOCPs.
Linear primal-dual interior-point methods were initially introduced by [Kojima et al., 1989] and [Monteiro and Adler, 1989]. The basic idea of primal-dual interior-point methods for SOCPs is to use the path-following paradigm. In every iteration, a so-called predictor search direction is computed which aims to minimize the duality gap, which is nothing but the difference between the primal and the dual objectives at the current iterate. The step is then corrected in order to stay close to the so-called central path, an analytic curve in the interior of the set of all feasible solutions which eventually converges to the optimal solution, see [Alizadeh and Goldfarb, 2002, §7] for more details.

For infeasible primal-dual path-following algorithm the initial iterates do not have to be feasible. These algorithms try to achieve feasibility and optimality of their iterates simultaneously. See [Toh et al., 2006] for more details on these type of algorithms.

There exist several solvers that can handle SOCPs. Some of them are listed in Table 1, see for example NEOS[2] (Network Enabled Optimization Server) for a more comprehensive overview of the available software solutions.

Solver	Description
MOSEK	Commercial software package for solving large optimization problems.
SeDuMi	MATLAB toolbox for solving optimization problems over self-dual homogeneous cones
AMPL	Algebraic modelling language with SOCP support
CPLEX	Optimization software package developed by IBM
SDPT3	MATLAB implementation of infeasible path-following algorithms for solving conic programming problems
PENSDP	Stand-alone program for solving general optimization problems

Table 1: SOCP solvers

For our empirical backtests in Section 7.2 we used SDPT3 in combination with the modelling package CVX[3], a MATLAB-based modelling system for convex optimization which would also support the SeDuMi solver.

[2]http://neos.mcs.anl.gov
[3]http://cvxr.com/cvx

3 Portfolio Optimization

Here, we want to lay out the basic ideas and the most important results of the theory of portfolio optimization. We consider an investor who has the choice of $n \in \mathbb{N}$ different assets to invest his capital in. In our framework, it is assumed that the only information available are the historic asset-returns observed in the past. Based on this information, the aim is to derive a portfolio or an investment strategy that somehow "complies best" with the interests of the investor or is optimal in some sense.
In Section 3.1, we will further explain our assumptions and the problem setting. This is followed by a short review of the classical approach of Markowitz and the Kelly strategy in Sections 3.2 and 3.3, respectively. Finally, in Section 3.4, we compare these two basic approaches in a simple example.

3.1 Preliminaries

Let us consider $n \geq 2$ different assets $S^1, S^2, \ldots, S^n$. We denote the price process of asset $i \in \{1, \ldots, n\}$ by P^i, saying that P^i_t is the actual price of asset i at time t. The (absolute) return r^i_t of asset i in the time period $[t-1, t]$ is given by $r^i_t := \frac{P^i_t - P^i_{t-1}}{P^i_{t-1}}$. Let μ_i and σ_i ($i \in \{1, \ldots, n\}$) be the expected value and the standard deviation of the future return $\tilde{r}^i$ of asset S^i on a given time interval. We assume that $\tilde{r}^i \geq -1$, i.e. the worst case scenario is a complete default of asset S^i, which implies a total (100%) loss of our investment. As in general, asset-returns cannot be assumed to be mutually independent, we denote by ρ_{ij} ($i \neq j$) the correlation coefficient of the returns of assets S^i and S^j.

A vector of portfolio weights $w \in \mathbb{R}^n$, where we assume that $\sum_{i=1}^n w^i = 1$ holds, describes the distribution of the capital invested in the assets, meaning a portion w^i of the initial capital is invested in asset S^i, $i = 1, \ldots, n$. We further assume $w^i \geq 0$, i.e. we exclude portfolios with short sales. Obviously, a vector of portfolio weights fully describes a portfolio over the assets $\{S^1, \ldots, S^n\}$. In addition, we denote by $\mathbb{W} \subseteq \mathbb{R}^n$ the set of all so-called admissible portfolios that comply with the above and maybe other possible linear restrictions that arise due to regulatory or institutional reasons.
If we describe the expected value μ and the covariance matrix $\mathbf{\Sigma}$ of the asset-returns $\tilde{r} = (\tilde{r}^1, \ldots, \tilde{r}^n)'$ by

$$\mu := \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} \quad \text{and} \quad \mathbf{\Sigma} := \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \ldots & \rho_{1n}\sigma_1\sigma_n \\ \rho_{21}\sigma_2\sigma_1 & \sigma_2^2 & \ldots & \rho_{2n}\sigma_2\sigma_n \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1}\sigma_n\sigma_1 & \rho_{n2}\sigma_n\sigma_2 & \ldots & \sigma_n^2 \end{bmatrix},$$

we can conveniently write the expected return and variance of a portfolio w hold unchanged for the considered time interval as

$$\mu_w := w'\mu \quad \text{and} \quad \sigma_w^2 := w'\mathbf{\Sigma}w.$$

As $\mathbf{\Sigma}$ states a covariance matrix, it is positive semidefinite, i.e. $\mathbf{\Sigma} \succeq 0$ holds. By assumption $\mathbf{\Sigma} \succ 0$ holds. Note that this is not very restrictive since it only means that there are no redundant assets within $\{S^1, \ldots, S^n\}$.

Moving to a multi-period scenario, we now assume that for a given time horizon $T \in \mathbb{N}$, the portfolio weights $w_t \in \mathbb{R}^n$ may be adjusted at predetermined dates $t = 0, \ldots, T-1$ and define by $\tilde{r}_t^i$ the return of asset S_i in time period $[t-1, t]$.
We also assume the asset-returns $\tilde{r}_t := (\tilde{r}_t^1, \ldots, \tilde{r}_t^n)'$ to follow a weak sense white noise process defined as follows.

Definition 3.1 (Weak Sense White Noise Process)**.** The random vectors $(\tilde{r}_t)_{t=1}^T$ form a weak sense white noise process if they are mutually uncorrelated and share the same mean values $\mathbb{E}_{\mathbb{P}}(\tilde{r}_t) = \mu$ and second-order moments $\mathbb{E}_{\mathbb{P}}(\tilde{r}_t \tilde{r}_t') = \mathbf{\Sigma} + \mu\mu' \quad \forall 1 \leq t \leq T$, where $\mathbb{P}$ describes the (unknown) asset-return distribution.

We call random vectors, which are not only uncorrelated but independent and identically distributed, white noise processes in the strong sense.

Definition 3.2 (Strong Sense White Noise Process)**.** The random vectors $(\tilde{r}_t)_{t=1}^T$ form a strong sense white noise process if they are mutually independent and identically distributed.

A family of vectors $(w_t)_{t=1}^T$ ($w_t \in \mathbb{W}$) describes an investment strategy, meaning portfolio w_t is held in period $(t-1, t]$. A subset of the class of all possible investment strategies is called fixed-mix strategies and the subset is described by the following definition.

Definition 3.3 (Fixed-Mix Strategy)**.** A portfolio strategy $(w_t)_{t=1}^T$ is called a fixed-mix strategy if there is a $w \in \mathbb{W}$ such that $w_t = w$ for all $t = 1, \ldots, T$.

Thus, fixed-mix strategies keep the vector of portfolio weights constant over time.

To maintain tractability, we restrict our attention to fixed-mix strategies due to their simplicity and attractive theoretical properties. As a fixed-mix strategy is defined by a single $w \in \mathbb{W}$, we will describe both the investment strategy and the single portfolio by w.

So far, we have outlined the preliminaries of portfolio theory. We defined what an admissible portfolio is and on what information and assumptions a portfolio-selection method should be based. In the remainder of this section, we will recall two of the most fundamental approaches of portfolio theory, namely the so-called "mean-variance efficient" and "growth-optimal" portfolios.

3.2 Mean-Variance Efficient Portfolios

[Markowitz, 1952] introduced the so-called *"Modern Portfolio Theory"* in his famous 1952 article. The foundation of this theory is the basic assumption that investors are risk averse, i.e. given two portfolios with the same expected return, an investor will choose the one with the lesser risk. Markowitz identified risk by the variance of the portfolio-returns.
The aim is now to minimize the variance of a portfolio under the restriction that its expected return is larger than or equal to a given threshold. Such variance minimizing portfolios are called efficient portfolios.
In the following, we assume the set of all admissible portfolios $\mathbb{W}$ to be of the form $\mathbb{W} = \{w \in \mathbb{R}^n : \mathbf{A}w = b, \mathbf{C}w \geq d, w \geq 0\}$, where $\mathbf{A}, \mathbf{C} \in \mathbb{R}^{n \times n}$ and $b, d \in \mathbb{R}^n$ are given parameters.

Definition 3.4 (Efficient Portfolios)**.** A portfolio $w \in \mathbb{W}$ is called efficient if one of the following holds:

a) $\nexists\, \overline{w} \in \mathbb{W}$ with $\overline{w}'\mu > w'\mu$ and $\overline{w}'\mathbf{\Sigma}\overline{w} \leq w'\mathbf{\Sigma}w$

b) $\nexists\, \overline{w} \in \mathbb{W}$ with $\overline{w}'\mu \geq w'\mu$ and $\overline{w}'\mathbf{\Sigma}\overline{w} < w'\mathbf{\Sigma}w$.

The above definition simply means that a portfolio is efficient if there exists no other portfolio with

a) strictly higher expected return and smaller variance or

b) higher expected return and strictly smaller variance.

It can be shown that efficient portfolios can be determined by solving one of the following parametrised optimization problems (see Theorem 3.1), where R, S and δ are given parameters.

- Minimize the variance subject to a lower return threshold R:

$$\begin{aligned} \min \quad & w'\mathbf{\Sigma}w \\ \text{s.t.} \quad & \mu'w \geq R \\ & w \in \mathbb{W} \end{aligned} \tag{6}$$

- Maximize the return subject to an upper variance threshold S:

$$\begin{aligned} \max \quad & \mu'w \\ \text{s.t.} \quad & w'\mathbf{\Sigma}w \leq S \\ & w \in \mathbb{W} \end{aligned} \tag{7}$$

- Maximize the return which is penalized by the variance multiplied with the so-called risk aversion parameter δ:

$$\begin{aligned} \max \quad & \mu'w - \delta w'\mathbf{\Sigma}w \\ \text{s.t.} \quad & w \in \mathbb{W} \end{aligned} \tag{8}$$

Let R_{min} (R_{max}) describe the minimal (maximal) expected portfolio-return of all admissible portfolios. Obviously, R_{min} and R_{max} correspond to the minimal and maximal expected asset return. For all $r \in [R_{min}, R_{max}]$ let σ_r^2 be the optimal value of the Optimization Problem (6) where $R = r$ is chosen. A commonly used visualization of efficient portfolios is to plot the expected return threshold r against σ_r^2.

Definition 3.5 (Efficient-Frontier)**.** The two-dimensional set $\{(r, \sigma_r^2) : r \in [R_{min}, R_{max}]\}$ is called efficient-frontier (or Pareto-frontier) to the corresponding portfolio optimization problem.

Note that the efficient portfolios, and hence the efficient-frontier, do not depend on the particular formulation (Problem (6), (7) or (8)) of our portfolio optimization problem. This result is summarized in the following easy to verify theorem.

Theorem 3.1. *Problems (6), (7) and (8) are equivalent.*
This means that for all $R \in [R_{min}, R_{max}]$ there exist parameters $S > 0$ and $\delta > 0$ such that an optimal solution w^* of Problem (6) is an optimal solution of Problems (7) and (8), where the corresponding parameters are R, S and δ, respectively.
Analogously, for all $S > 0$ ($\delta > 0$) there exists a $R \in [R_{min}, R_{max}]$ and a $\delta > 0$ ($S > 0$) such that an optimal solution w^* of Problem (7) ((8)) is also an optimal solution of Problem (6) and (8) ((7)), where the parameters are chosen accordingly.

Proof. See [Krokhmal et al., 2002, Appendix A]. □

The big advantages of the mean-variance efficient portfolios are their appealing theoretical derivation and the simplicity of the resulting optimization problem which has to be solved. In fact, if we assume the covariance matrix $\mathbf{\Sigma}$ to be positive definite, i.e. $\mathbf{\Sigma} \succ 0$, and since the corresponding optimization problems state quadratic problems, there exists a unique solution and the so-called Karush-Kuhn-Tucker conditions are necessary and sufficient (see among others [Luenberger and Ye, 2008]). As the mean-variance efficient portfolio optimization problem is a quadratic program, a numerical solution can be found with reasonable computational effort.

Of course, this theoretical and computational simplicity comes with the cost of many assumptions which compromise the mean-variance efficient framework to some degree. The following listing only displays some of these assumptions, where we mainly focus on technical issues, see e.g. [Mandelbrot and Hudson, 2004] or [Elton and Gruber, 1997] for more comprehensive accounts on the drawbacks of the mean-variance efficiency framework.

- Asset-return distribution:
  By assuming that all of the risk is described by the variance of the asset-returns, one implicitly assumes these returns are jointly normally distributed variables.

- Investment-horizon:
  The notion of a particular investment-horizon is completely ignored in this framework. The correlations are assumed to stay the same for whatever time horizon the investor intends to hold his portfolio.

- Asset-return moments:
  It is assumed that the true first- and second-order moments are known and constant over time.

- Availability of assets:
  As we only restrict the optimal solution of the above optimization problem to linear equalities and inequalities, we cannot assure the availability of the asset fraction implied by the resulting portfolio.

Some of these assumptions and the overall conceptual derivation result in some undesirable properties of the obtained efficient portfolios. For example, one can observe that often the resulting portfolios are not very well diversified. Also, since we assume μ and $\mathbf{\Sigma}$ to be the true moments of the asset-returns but in general have to rely on estimates, we somehow "optimize" inevitable estimation errors.

We now want to state some basic attempts that account for some of these shortcomings.

1. Upper Investment Bounds:
   This simple idea is targeted at the often bad diversification of the mean-variance efficient portfolios. Some possible implementations are asset-wise upper bounds $x^i \leq m^i \quad i \in \{1, \ldots, n\}$ or group-wise upper bounds $\sum\limits_{i \in G} x^i \leq m^G$, where m^i and m^G are given constants.

2. Combining Estimates:
   We determine $k \in \mathbb{N}$ different moment estimation $\hat{\mu}_i$, $\hat{\mathbf{\Sigma}}_i$ with $1 \leq i \leq k$ from k different data samples. The optimal portfolios w_i (corresponding to $\hat{\mu}_i$ and $\hat{\mathbf{\Sigma}}_i$) are combined to an aggregated portfolio by a predetermined linear combination. See [Michaud, 1998] for further information.

3. Robust Optimization:
   The so-called uncertainty-sets $\{A^i\}_{i=1}^n$ and B for the true moments $\{\mu^i\}_{i=1}^n$ and $\mathbf{\Sigma}$ are determined. It is assumed that the true moments lie in these uncertainty-sets. The optimization problem is then adapted to comply with all possible moments, i.e.

   $$\begin{aligned} &\min_{w \in \mathbb{W}} \max_{\mathbf{\Sigma} \in B} w' \mathbf{\Sigma} w \\ &\text{s.t.} \quad \mu' w \geq R \quad \forall \mu \in A^1 \times \ldots \times A^n \end{aligned}$$

   Obviously, this accounts for the moment ambiguity of the true asset-return distribution.

Whereas the above models are based on the mean-variance efficiency framework, other approaches have been made to determine "optimal" portfolios. One of them is the growth-optimal framework, which we will present in the following section.

3.3 Growth-Optimal Portfolios

In contrast to the mean-variance efficiency framework, we now take the length T of the investment horizon into account, where naturally T is chosen as a natural number, i.e. $T \in \mathbb{N}$. We recall that the portfolio weights w_t may be adjusted at predetermined dates $t = 0, \dots, T-1$, where w_t describes the portfolio held in time period $(t-1, t]$.
Let $\tilde{V}_T$ define the aggregate return generated by an investment strategy $(w_t)_{t=1}^T$, hence

$$\tilde{V}_T := \prod_{t=1}^{T} [1 + w_t' \tilde{r}_t].$$

Therefore, $\tilde{V}_T$ describes the overall growth of our wealth over the considered investment period. As the time horizon is split in several intervals, it is nearby to have a look at the average growth over these periods.

Definition 3.6 (Growth Rate)**.** We define the portfolio growth rate $\tilde{\varphi}_T$ over an investment horizon of length $T \in \mathbb{N}$ as the natural logarithm of the geometric mean of the returns, i.e.

$$\tilde{\varphi}_T = \ln \left(\sqrt[T]{\prod_{t=1}^{T} [1 + w_t' \tilde{r}_t]} \right) = \frac{1}{T} \sum_{t=1}^{T} \ln(1 + w_t' \tilde{r}_t). \tag{9}$$

Since $\tilde{V}_T$ describes the generated wealth of our investment strategy and by the above definition $\tilde{V}_T = \exp(T \cdot \tilde{\varphi}_T)$ holds, the aim of maximizing terminal wealth is equivalent to maximizing $\tilde{\varphi}_T$ per se.
As the growth rate $\tilde{\varphi}_T$ is a random variable, we can not just "maximize" it. Fortunately, when considering a fixed-mix strategy and if the asset-returns follow a strong sense white noise process, the asymptotic growth rate $\lim\limits_{T \to \infty} \tilde{\varphi}_T$ turns out to be deterministic. This can be shown by simply applying the strong law of large numbers to the right hand side of Equation (9) as T tends to infinity.

Proposition 3.1 (Asymptotic Growth Rate)**.** *Let $(w_t)_{t=1}^T$ be a fixed-mix strategy, hence $w_t = w$ for some $w \in \mathbb{W}$. If the asset-returns $(\tilde{r}_t)_{t=1}^T$ follow a strong sense white noise process, then the following limit equation almost surely holds (with probability 1)*

$$\lim_{T \to \infty} \tilde{\varphi}_T = \mathbb{E}(\ln(1 + w' \tilde{r}_1)). \tag{10}$$

As we were looking for an investment strategy with maximal portfolio growth rate, a nearby candidate for such a portfolio is the one which maximizes the asymptotic growth rate of the implied fixed-mix strategy. The resulting investment strategy is called *"Kelly strategy"* and is simply obtained by maximizing the right hand side of equation (10).

Definition 3.7 (Growth-Optimal Portfolio)**.** An admissible portfolio $w^* \in \mathbb{W}$ is called growth-optimal portfolio if

$$w^* = \operatorname{argmax}_{w \in \mathbb{W}} \mathbb{E}(\ln(1 + w' \tilde{r}_1)). \tag{11}$$

The main reason for which the Kelly strategy is of such interest from a theoretical point of view is that it can be shown that, in the long run, it outperforms any other portfolio strategy. This also includes investment strategies which are not of a fixed mix type. For further informations about this result see [Cover and Thomas, 1991].

Theorem 3.2 (Asymptotic Optimality of the Kelly Strategy)**.** *We denote by $\tilde{\varphi}_T^*$ and $\tilde{\varphi}_T$ the growth rates of the Kelly strategy and some arbitrary other causal portfolio strategy, respectively. Then the following holds.*

$$(\tilde{r}_t)_{t=1}^T \text{ is a strong sense white noise process} \Rightarrow \mathbb{P}\left[\limsup_{T\to\infty}(\tilde{\varphi}_T - \tilde{\varphi}_T^*) \leq 0\right] = 1.$$

Proof. See [Cover and Thomas, 1991, Theorem 15.3.1] □

Theorem 3.2 states that as the length T of the investment horizon tends towards infinity, the probability that the Kelly strategy accumulates more wealth (has a larger growth rate) than any other portfolio strategy is equal to one.

Although this is a quite strong and appealing result, the Kelly strategy is burdened by several major disadvantages. We give a short overview and refer to [MacLean et al., 2010] for a more comprehensive insight in the advantages and disadvantages of the so-called *"Kelly Criterion"*.

- The "Long Run":
  By Theorem 3.2, the dominance of the Kelly strategy holds only asymptotically. Unfortunately, the time until the Kelly strategy has a larger growth rate than any other strategy (with high confidence) may be very large and won't be of any practical interest.

- Ambiguity of the asset-return distribution.
  In order to determine the Kelly strategy, the right hand side of Equation (10) has to be maximized. This expectation value is taken with respect to the true asset-return distribution. As in practice one is obliged to use an estimated distribution, this estimation is a likely source for errors which will significantly affect the performance of the resulting portfolio.

- The Asset-Return Process:
  In order to guarantee the asymptotic dominance of the Kelly strategy, Theorem (10) requires the asset-return process $(\tilde{r}_t)_{t=1}^T$ to be a white noise process in the strong sense. This is too strong of an assumption which generally can not be verified in practice.

In their paper [Rujeerapaiboon et al., 2014] introduced robust growth-optimal portfolios. These portfolios are obtained by maximizing a quadratic approximation of the growth rate, where the asset-return distribution is not assumed to be known but to lie in a predefined ambiguity set of distributions. A big theoretical advantage of this approach is that these portfolios are tailored to finite investment horizons but offer similar performance guarantees as the classical growth-optimal portfolios.

As mentioned earlier, our approach is based on the conceptional ideas of these robust growth-optimal portfolios. In Section 5.4 we will also see that in our framework, these portfolios can be obtained by a certain choice of the risk-aversion parameter.

3.4 Comparison of Mean-Variance Efficient and Growth-Optimal Portfolios

As we have laid out the theoretical backgrounds of the two major historical models of portfolio optimization, i.e. the mean-variance efficient and the growth-optimal framework (see Sections 3.2 and 3.3), the purpose of this section is to get a hold of the differences between these two approaches. For this reason, we present a neat little example originally published by [Hakansson, 1971].

Example 3.1. *We assume that there are two assets S^1 and S^2 in which we can invest in. Short sales are again excluded and all of the capital has to be invested, i.e. for the portfolio weights $w = (w_1, w_2)' \in [0,1] \times [0,1]$ the restriction $w_1 + w_2 = 1$ has to hold. Let the initial ($t = 0$) prices of the assets be given as $S_0^1 = S_0^2 = 1$. We consider the time horizon T to consist of only one time period $[0, T]$ and the random values S_T^1 and S_T^2 of the assets at time T to have marginal distributions of the form*

$$S_T^1 = \begin{cases} 0 & \text{with probability 0.1} \\ 1.5 & \text{with probability 0.9} \end{cases} \qquad \text{and} \qquad S_T^2 = \begin{cases} 1.15 & \text{with probability 0.9} \\ 2.65 & \text{with probability 0.1} \end{cases}.$$

The joint distribution of S_T^1 and S_T^2 is defined by

$$\mathbb{P}(S_T^1 = 0, S_T^2 = 1.15) = 0.1,\ \mathbb{P}(S_T^1 = 0, S_T^2 = 2.65) = 0$$
$$\mathbb{P}(S_T^1 = 1.5, S_T^2 = 1.15) = 0.8 \text{ and } \mathbb{P}(S_T^1 = 1.5, S_T^2 = 2.65) = 0.1.$$

Therefore, the associated returns r_T^1 and r_T^2 are given by

$$r_T^1 = \begin{cases} -1 & \text{with probability 0.1} \\ 0.5 & \text{with probability 0.9} \end{cases} \qquad \text{and} \qquad r_T^2 = \begin{cases} 0.15 & \text{with probability 0.9} \\ 1.65 & \text{with probability 0.1} \end{cases},$$

with the obvious joint distribution.

We now use this exact return-distribution to determine the growth-optimal portfolio. As mentioned in Section 3.3, we obtain this portfolio by maximizing the expected logarithmic return. In our simple example, since $w_2 = 1 - w_1$, this means that we have to maximize

$$\begin{aligned}
\mathbb{E}[\ln(1 + w' r_T)] &= \mathbb{E}[\ln(1 + w_1 r_T^1 + w_2 r_T^2)] \\
&= \mathbb{E}[\ln(1 + w_1 r_T^1 + (1 - w_1) r_T^2)] \\
&= \mathbb{E}[\ln(1 + r_T^2 + (r_T^1 - r_T^2) w_1)] \\
&= 0.1 \ln(1.15 - 1.15 w_1) + 0.8 \ln(1.15 + 0.35 w_1) + 0.1 \ln(2.65 - 1.15 w_1).
\end{aligned}$$

This can be done by setting the derivative (with respect to w_1) of the above expression equal to zero and solving for the weight w_1. This yields $w_1 \in \{0.394, 1.923\}$ with only $w_1 = 0.394$ being feasible for our problem. The growth optimal portfolio is therefore given by

$$w_{GO} = (0.394, 0.606)'.$$

The simplicity of this example allows us to determine the mean-variance efficient portfolios by plotting the return expectation against the variance of each possible portfolio.
Subject to w_1, which uniquely defines one possible portfolio, the expected portfolio-return $\mu_p(w_1)$ and the variance $\sigma_p^2(w_1)$ are given by

$$\begin{aligned}
\mu_p(w_1) &= \mathbb{E}[r_P(w_1)] = w_1 0.35 + (1 - w_1) 0.3 = 0.05 w_1 + 0.3 \\
\sigma_p^2(w_1) &= \mathbb{E}[(r_P(w_1))^2] - \mathbb{E}^2[r_P(w_1)] \\
&= 0.36 w_1^2 - 0.36 w_1 + 0.2025.
\end{aligned}$$

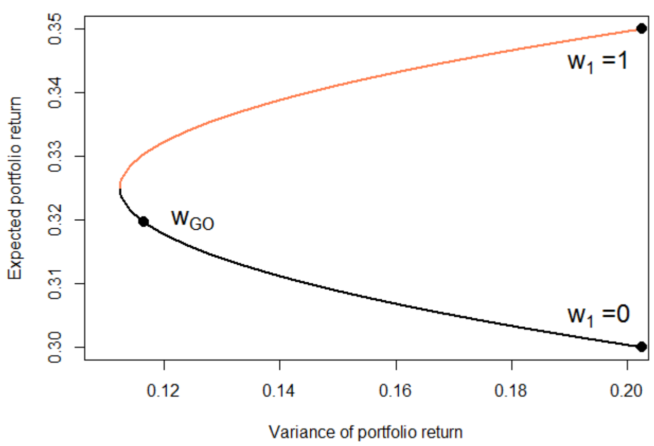

Figure 4: Return-variance contour of Example 3.1

Using Definition 3.4 of efficient portfolios, we identify the portfolios marked orange in Figure 4 as the mean-variance efficient portfolios. This is obvious since for every "black" portfolio there exists another ("orange") portfolio with the same variance but a higher mean. We immediately recognize that the growth-optimal portfolio (marked with w_{GO}*) is not a mean-variance efficient portfolio, so in general the two notions of growth-optimality and mean-variance efficiency lead to different portfolios.*

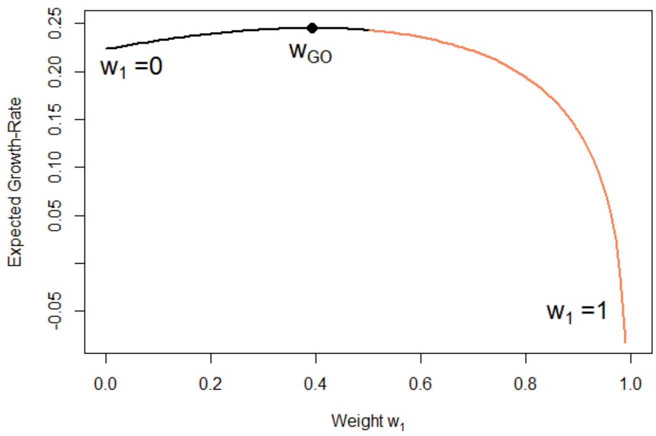

Figure 5: Expected growth rates of Example 3.1

Likewise, we can plot the expected growth rate of every portfolio. In Figure 5 this is visualized by plotting weight w_1 *against the expected growth rate we derived earlier.*

The plot visualizes the defining characteristic of the growth-optimal portfolio, i.e. maximizing the expected growth rate. We can also observe that portfolio $(0, 1)$*, although having the worst properties in the sense of mean-variance (highest variance, lowest mean), has a higher expected growth rate than most of the mean-variance efficient portfolios.*

We conclude that the two notions of mean-variance efficiency and growth-optimality are indeed different approaches to the portfolio optimization problem and will lead to different results. Note that one has to keep in mind that the Markowitz-approach is tailored to a single investment period, while the Kelly-strategy is asymptotically optimal. Nevertheless, as mentioned earlier, the mean-variance efficient portfolios are very popular and often used sequentially for consecutive investment periods.

4 Portfolio Evaluation

There are many ways to a posteriori evaluate the performance of a portfolio or an investment strategy (see Section 7.1). We will introduce our method of "quadratic return penalization" in Section 4.1. It is designed to provide analytic properties which are needed in order to reformulate the resulting portfolio optimization problem as a tractable SOCP. It allows for an intuitive interpretation and the introduction of a risk-aversion parameter γ, similar to the classical approach of Markowitz. This a posteriori evaluation implies a random variable a priori, which represents our objective in terms of portfolio optimization. As the distribution of this random variable is directly linked to the unknown asset-return distribution, we have to account for this ambiguity, which is done in Section 4.2.

4.1 Quadratic Return Penalization

When selecting portfolio weights, one has to specify desirable properties of the portfolio. Following the classical approach, these properties would be high return and low variance. Given a portfolio by its weight-vector $w \in \mathbb{W}$, we need to quantify its performance according to those properties. For this purpose, we define a function of the realized portfolio-returns, which, by definition, penalizes small portfolio-returns with high values and high portfolio-returns with small values. Our choice to do so is a quadratic polynomial function $f_\gamma(x) = (x - \gamma)^2$ with positive curvature and a minimum at $\gamma \in \mathbb{R}$, where γ needs to be chosen sufficiently large such that all realized portfolio-returns are smaller than γ.

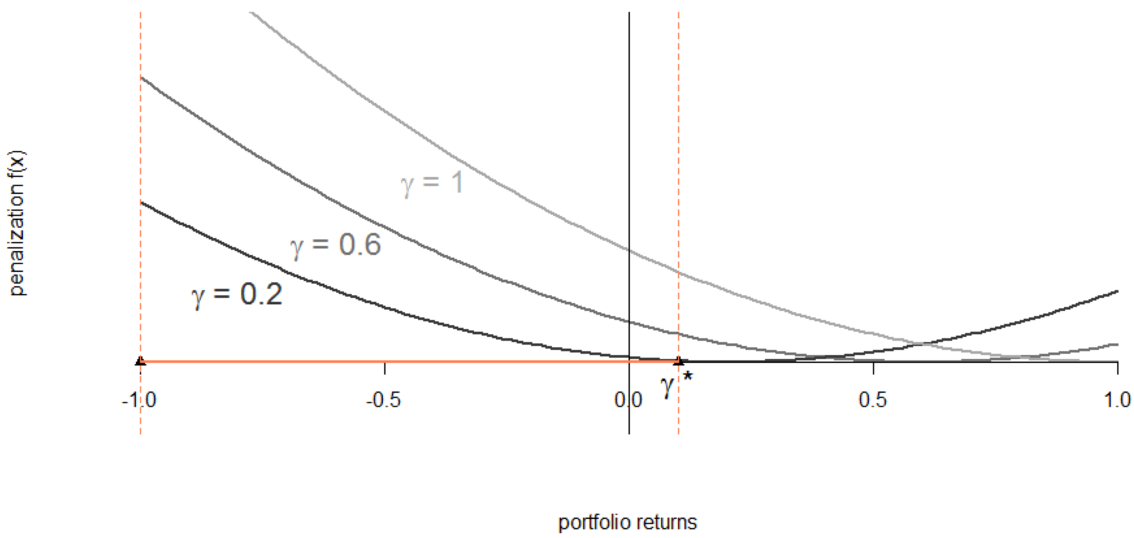

Figure 6: Penalization function $f_\gamma(x)$ for different γ

Note that for a smaller choice of γ, we would penalize high portfolio returns by a high value of $f_\gamma(x)$ too. In the following, we will call the smallest γ that fulfils the above requirement γ^*, so one needs to choose $\gamma \geq \gamma^*$.

Our definition implies that the penalization function is strictly convex. In particular, this holds on the range $[-1, \gamma^*]$, in which lie all realized portfolio returns, according to the definition of γ^*.

This definition also enables us to adjust the curvature of $f_\gamma(x)$ on $[-1, \gamma^*]$ by our choice of γ. For very large γ, the penalization function $f_\gamma(x)$ becomes almost linear on $[-1, \gamma^*]$, whereas for small γ (close to γ^*) we have a distinct curvature of $f_\gamma(x)$ on the relevant interval (see Figure 6 for an exemplary visualization). We observe that an almost linear curvature (big γ) favours portfolios with a high expected return since the relative (with respect to the mean) penalization of a deviation does not depend on the actual location of the mean. On the other side, a distinct curvature (small γ) focuses more on the variance. This is due to the fact that an increase in the mean does not provide as much gain (or rather loss since we aim to minimize the penalization) as it did before. This simple interpretation of curvature is visualized in Example 1. We conclude that γ can be interpreted as a risk-aversion parameter, see Section 5.4.2 for further details on the choice of γ.

We can now measure the a posteriori performance of a given portfolio w over a specific time horizon T by the mean of the penalized portfolio-returns

$$\nu_T^\gamma(w) := \frac{1}{T}\sum_{t=1}^{T} f_\gamma(w'r_t) = \frac{1}{T}\sum_{t=1}^{T}(w'r_t - \gamma)^2. \tag{12}$$

We hence favour portfolios with small values of ν_T^γ.

Example 4.1. *We want to back our interpretation of γ as a risk-aversion parameter by the following simple example. We assume two portfolios named (and visualized) red and blue. The considered investment horizon consists of two periods, so $T = 2$. For each of the portfolios we observed the following portfolio returns in the periods $(0, 1]$ and $(1, 2]$: $r_1^{blue} = -0.2$, $r_2^{blue} = -0.6$ and $r_1^{red} = -0.45$, $r_2^{red} = -0.38$. This returns are marked by the large coloured points in Figure 7.*
We observe that the mean of the blue returns $\hat{\mu}^{blue} = -0.4$ is bigger than the mean of the red returns $\hat{\mu}^{red} = -0.415$ and that obviously the blue variance is bigger than the red.
Since

$$\nu^1(blue) = \frac{1}{2}\left((r_1^{blue} - 1)^2 + (r_2^{blue} - 1)^2)\right) < \frac{1}{2}\left((r_1^{red} - 1)^2 + (r_2^{red} - 1)^2)\right) = \nu^1(red)$$

and

$$\nu^{0.1}(blue) = \frac{1}{2}\left((r_1^{blue} - 0.1)^2 + (r_2^{blue} - 0.1)^2)\right) > \frac{1}{2}\left((r_1^{red} - 1)^2 + (r_2^{red} - 1)^2)\right) = \nu^{0.1}(red)$$

our evaluation method with $\gamma = 1$ would prefer portfolio blue, whereas for $\gamma = 0.1$ we favour portfolio red. Hence, what we have seen is that with γ small, the smaller variance of portfolio red makes it more attractive than portfolio blue, despite the fact that it has a smaller mean return. Vice versa, we see that for γ big, the blue portfolio with its bigger mean return was chosen.

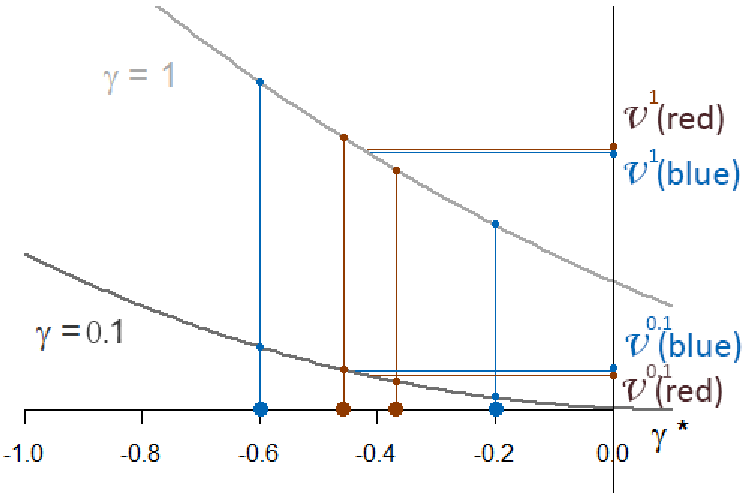

Figure 7: Penalization function with exemplary returns

As the asset-returns r_t are unknown at time $t = 0$, $\nu_T^\gamma(w)$ is a random variable. We denote the fact that we do not know the future returns by writing $\tilde{r}_t$ and consequently

$$\tilde{\nu}_T^\gamma(w) := \frac{1}{T}\sum_{t=1}^{T} f_\gamma(w'\tilde{r}_t) = \frac{1}{T}\sum_{t=1}^{T}(w'\tilde{r}_t - \gamma)^2. \tag{13}$$

By the above definition, we prefer portfolios which tend to have small realizations of $\tilde{\nu}_T^\gamma(w)$, so our task will be the derivation of a portfolio w^* that has a good performance (small values of $\tilde{\nu}_T^\gamma(w)$) with a high level of confidence.

Let $\mathbb{P}$ be the (unknown) probability distribution of the asset-returns $\tilde{r}_t$. Our approach is to choose w^* such that it minimizes the $(1-\epsilon)$-quantile of $\tilde{\nu}_T^\gamma(w)$. Since $\tilde{\nu}_T^\gamma(w)$ is a random variable only dependent on $\{\tilde{r}_1, \ldots, \tilde{r}_T\}$, one can interpret this quantile as the $\mathbb{P}$-Value-at-Risk (VaR) at level ϵ of $\tilde{\nu}_T^\gamma(w)$, so

$$\mathbb{P}\text{-}VaR_\epsilon(\tilde{\nu}_T^\gamma(w)) = \min_{\nu\in\mathbb{R}}\left\{\nu : \mathbb{P}\left(\frac{1}{T}\sum_{t=1}^{T}(w'\tilde{r}_t - \gamma)^2 \le \nu\right) \ge 1-\epsilon\right\}. \tag{14}$$

Note that ϵ is determined by the user and is usually chosen as a small number ≤ 0.1 (e.g. $\epsilon \in [0.01, 0.05, 0.1]$).

We therefore evaluate a single portfolio $w \in \mathbb{W}$ a priori by $\mathbb{P}\text{-}\mathit{VaR}_\epsilon(\tilde{\nu}_T^\gamma(w))$, where we prefer portfolios with small values of $\mathbb{P}\text{-}\mathit{VaR}_\epsilon(\tilde{\nu}_T^\gamma(w))$ since for them, the $(1-\epsilon)$ upper bound of a realization of $\tilde{\nu}_T^\gamma(w)$ is small.
Since we do not know the exact asset-return distribution, we have to robustify our definition of $\mathbb{P}\text{-}\mathit{VaR}_\epsilon(\tilde{\nu}_T^\gamma(w))$ against the ambiguity of $\mathbb{P}$, which we will do in the following section.

4.2 Ambiguity of the Asset-Return Distribution

As mentioned above, in most practical cases the precise asset-return distribution $\mathbb{P}$ is unknown, but we may know some more general properties. We define P as the set of all asset-return distributions with these known properties. Since we assume that the real distribution belongs to P, we have to adapt our portfolio evaluation $\mathbb{P}\text{-}\mathit{VaR}_\epsilon(\tilde{\nu}_T^\gamma(w))$ with respect to this ambiguity set. We will do this by simply requiring the inequality inside the minimum of Equation (14) to hold for all $\mathbb{P} \in P$ and refer to this new evaluation as the worst-case VaR of $\tilde{\nu}_T^\gamma(w)$ at level ϵ, therefore

$$\begin{aligned} \mathit{WVaR}_\epsilon(\tilde{\nu}_T^\gamma(w)) : &= \max_{\mathbb{P}\in P} \mathbb{P}\text{-}\mathit{VaR}_\epsilon(\tilde{\nu}_T^\gamma(w)) \qquad (15) \\ &= \min_{\nu\in\mathbb{R}} \left\{ \nu : \mathbb{P}\left(\frac{1}{T}\sum_{t=1}^{T}(w^t\tilde{r}_t - \gamma)^2 \leq \nu \right) \geq 1-\epsilon \quad \forall \mathbb{P} \in P \right\}. \qquad (16) \end{aligned}$$

By referring to [Roy, 1952], where he states that the first two moments of the asset return distribution *"are the only quantities that can be distilled out of our knowledge of the past"*, we want to motivate our decision to only use the first- and second-order moments of the centered asset-return distribution for the definition of P.
We will therefore assume that the only information we have about the true return-distribution are its (estimated) first- and second-order moments. It follows that P is of the form

$$P_{\mu,\Sigma} = \left\{ \mathbb{P} \in P_0^{nT} : \begin{array}{l} \mathbb{E}[\tilde{r}_t] = \mu \quad \forall t : 1 \leq t \leq T \\ \mathbb{E}[\tilde{r}_s \cdot \tilde{r}_t'] = \delta_{st}\Sigma + \mu\mu' \quad \forall s,t : 1 \leq s \leq t \leq T \end{array} \right\} \qquad (17)$$

where $\mu \in \mathbb{R}^n$ is the known (estimated) mean vector and $\mathbf{\Sigma} \in \mathbb{R}^{n\times n}$ is the known (estimated) covariance matrix of the asset return distribution. See Section 6 for further details on the estimation of μ and $\mathbf{\Sigma}$.

Using all the above considerations, we can finally state our a priori evaluation of a given portfolio w as

$$\mathit{WVaR}_\epsilon(\tilde{\nu}_T^\gamma(w)) := \min_{\nu\in\mathbb{R}} \left\{ \nu : \mathbb{P}\left(\frac{1}{T}\sum_{t=1}^{T}(w^t\tilde{r}_t - \gamma)^2 \leq \nu \right) \geq 1-\epsilon \quad \forall \mathbb{P} \in P_{\mu,\Sigma} \right\}. \qquad (18)$$

5 The Distributionally Robust Portfolio Optimization Problem

As we have seen in Section 4 how to a priori evaluate a given portfolio by the $WVaR_\epsilon(\tilde{\nu}_T^\gamma(w))$ defined in Equation (18), and since we favour portfolios with small values of this measure, we state the resulting portfolio optimization problem as

$$\begin{aligned} \min_{w} \; & WVaR_\epsilon(\tilde{\nu}_T^\gamma(w)) \\ \text{s.t.} \; & w \in \mathbb{W}, \end{aligned} \tag{19}$$

where $\mathbb{W}$ describes the set of all admissible portfolios.

Since we are not able to identify Problem (19) as a classical optimization problem yet, due to the current formulation of the objective function, we cannot compute an optimizer w^* in a straight forward fashion.

In order to determine an optimal solution $w^* \in \mathbb{W}$ to Problem (19), we will first look for an alternative formulation of the objective function $WVaR_\epsilon(\tilde{\nu}_T^\gamma(w))$ in Section 5.1. With some further assumptions on the set of admissible portfolios $\mathbb{W}$, in Section 5.2 we can reformulate the resulting problem as a classic SOCP, which can be solved by well-known algorithms. Due to this reformulation, the computational effort for solving our portfolio optimization problem for w^* is comparable to the classic Markowitz portfolio. Not only will the size of our resulting portfolio optimization problem be independent from the investment horizon $T \in \mathbb{N}$, but also will it admit an easy robustification against the moment-estimations, see Section 5.3. In Section 5.4, we will present some characteristics of w^* with respect to the particular choice of γ.

5.1 The Analytic Representation of the Objective Function

We will now show that Problem (18) admits an analytic solution for every $w \in \mathbb{R}^n$.
To do so, we first define the excess of a portfolio-return at time $t \in [1, \dots, T]$ with respect to γ as

$$\tilde{\eta}_t(w) := w'\tilde{r}_t - \gamma. \tag{20}$$

It follows that $\mathbb{E}[\tilde{\eta}_t(w)] = \mu_w - \gamma$ and $\mathbb{E}[\tilde{\eta}_t(w)\tilde{\eta}_s(w)] = \delta_{ts}\sigma_w^2 + (\mu_w - \gamma)^2$, where again $\mu_w = w'\mu = w'\mathbb{E}[\tilde{r}]$ denotes the expected portfolio return and $\sigma_w^2 = w'\boldsymbol{\Sigma}w$ the portfolio variance.

By using these return-excesses we can reformulate Definition (18) of the $WVaR_\epsilon(\tilde{\nu}_T^\gamma(w))$ as

$$WVaR_\epsilon(\tilde{\nu}_T^\gamma(w)) = \inf_{\nu \in \mathbb{R}} \nu \quad \text{s.t.} \quad \mathbb{P}\left(\frac{1}{T}\sum_{t=1}^{T} \tilde{\eta}_t(w)^2 \leq \nu\right) \geq 1 - \epsilon \quad \forall \mathbb{P} \in P_{\tilde{\eta}(w)}, \tag{21}$$

where $P_{\tilde{\eta}(w)}$ describes the $\tilde{\eta}$-distribution ambiguity set

$$P_{\tilde{\eta}(w)} = \left\{ \mathbb{P} \in P_0^T : \begin{array}{l} \mathbb{E}[\tilde{\eta}_t(w)] = \mu_w - \gamma \quad \forall t : 1 \leq t \leq T \\ \mathbb{E}[\tilde{\eta}_t(w)\tilde{\eta}_s(w)] = \delta_{ts}\sigma_w^2 + (\mu_w - \gamma)^2 \quad \forall s, t : 1 \leq s \leq t \leq T \end{array} \right\}.$$

Note that $(\tilde{\eta}_t(w))_{t=1}^T$ still follows a weak sense white noise process since $(\tilde{r}_t)_{t=1}^T$ does.

In order to convert Problem (21) into a tractable SDP, we use the following theorem (see [Rujeerapaiboon et al., 2014, Theorem A.1]).

Theorem 5.1. *Let P be the set of all probability distributions of a random vector $\tilde{\xi} \in \mathbb{R}^n$ that share the same mean $\mu \in \mathbb{R}^n$ and covariance matrix $\mathbf{\Sigma} \in \mathbb{S}_+^n$, $\mathbf{\Sigma} \succ 0$. Then, for $\mathbf{Q} \in \mathbb{S}_+^n$, $q \in \mathbb{R}^n$ and $q^0 \in \mathbb{R}$ the following holds:*

- *a distributional robust chance constraint given by*

$$\inf_{\mathbb{P} \in P} \mathbb{P}\left(\tilde{\xi}'\mathbf{Q}\tilde{\xi} + \tilde{\xi}'q + q^0 \leq 0\right) \geq 1 - \epsilon,$$

where the first- and second-order moments of all distributions in P equal the given parameters μ and $\mathbf{\Sigma}$, respectively, is equivalent to

$$\exists \mathbf{M} \in \mathbb{S}^{n+1}, \beta \in \mathbb{R} : \quad \beta + \frac{1}{\epsilon}\langle \mathbf{\Omega}, \mathbf{M} \rangle \leq 0, \quad \mathbf{M} \succeq 0$$
$$\text{and} \quad \mathbf{M} \succeq \begin{bmatrix} \mathbf{Q} & \frac{1}{2}q \\ \frac{1}{2}q' & q^0 - \beta \end{bmatrix},$$

where $\mathbf{\Omega}$ is a notational abbreviation for the second-order moment matrix of $\tilde{\xi}$, i.e.

$$\mathbf{\Omega} = \begin{bmatrix} \mathbf{\Sigma} + \mu\mu' & \mu \\ \mu' & 1 \end{bmatrix}.$$

Proof. See [Zymler et al., 2013, Theorem 2.3] □

By applying this theorem to our chance constraint (21) we get

$$\begin{aligned} WVaR_\epsilon(\tilde{\nu}_T^\gamma(w)) = \min \quad & \nu \\ \text{s.t.} \quad & \mathbf{M} \in \mathbb{S}^{T+1}, \beta, \nu \in \mathbb{R} \\ & \beta + \frac{1}{\epsilon}\langle \mathbf{\Omega}(w), \mathbf{M} \rangle \leq 0, \quad \mathbf{M} \succeq 0 \\ & \mathbf{M} - \begin{bmatrix} \mathbb{I} & 0 \\ 0 & -T\nu - \beta \end{bmatrix} \succeq 0, \end{aligned} \tag{22}$$

where $\mathbb{I} \in \mathbb{S}^T$ denotes the identity matrix and $\mathbf{\Omega}(w) \in \mathbb{S}^{T+1}$ describes the first- and second-order moments of $(\tilde{\eta}_1(w), \ldots, \tilde{\eta}_T(w))'$, hence

$$\mathbf{\Omega}(w) = \begin{bmatrix} \mathbb{E}[\tilde{\eta}_t(w)\tilde{\eta}_s(w)]_{t,s=1,\ldots,T} & \mathbb{E}[\tilde{\eta}_t(w)]_{t=1,\ldots,T} \\ (\mathbb{E}[\tilde{\eta}_t(w)]_{t=1,\ldots,T})' & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_w^2 + (\mu_w - \gamma)^2 & (\mu_w - \gamma)^2 & \cdots & (\mu_w - \gamma)^2 & (\mu_w - \gamma) \\ (\mu_w - \gamma)^2 & \sigma_w^2 + (\mu_w - \gamma)^2 & \cdots & (\mu_w - \gamma)^2 & (\mu_w - \gamma) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ (\mu_w - \gamma)^2 & (\mu_w - \gamma)^2 & \cdots & \sigma_w^2 + (\mu_w - \gamma)^2 & (\mu_w - \gamma) \\ (\mu_w - \gamma) & (\mu_w - \gamma) & \cdots & (\mu_w - \gamma) & 1 \end{bmatrix}.$$

In the following, we will omit the reference to the fixed portfolio weights w and write e.g. η_t instead of $\eta_t(w)$.

We will now show that Problem (22) admits an analytical solution. To do so, we will first study the structures of the matrices that appear in the above restrictions. This will allow us to substitute the semidefinite restrictions by simple inequalities.

Definition 5.1 (Compound Symmetry)**.** A matrix $\mathbf{M} \in \mathbb{S}^{T+1}$ is compound symmetric if there exist $\tau_1, \tau_2, \tau_3, \tau_4 \in \mathbb{R}$ with

$$\mathbf{M} = \left[\begin{array}{cccc|c} \tau_1 & \tau_2 & \cdots & \tau_2 & \tau_3 \\ \tau_2 & \tau_1 & \cdots & \tau_2 & \tau_3 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \tau_2 & \tau_2 & \cdots & \tau_1 & \tau_3 \\ \hline \tau_3 & \tau_3 & \cdots & \tau_3 & \tau_4 \end{array}\right]. \tag{23}$$

By the above definition, we can conclude that $\mathbf{\Omega}$ is compound symmetric and state, similar to [Rujeerapaiboon et al., 2014], the following proposition.

Proposition 5.1. *There exists a maximizer* $(\mathbf{M}, \beta, \nu)$ *of (22) with* $\mathbf{M}$ *compound symmetric.*

Proof. Analogous to proof of [Rujeerapaiboon et al., 2014, Proposition 4.2]. □

As we can now restrict our attention to compound symmetric matrices $\mathbf{M}$, we use the following proposition to reformulate the restrictions of (22) (see also [Rujeerapaiboon et al., 2014, Proposition 4.3]).

Proposition 5.2. *For any compound symmetric matrix* $\mathbf{M} \in \mathbb{S}^{T+1}$ *of the form (23), the following equivalence holds*

$$\mathbf{M} \succeq 0 \Leftrightarrow \begin{cases} \tau_1 \geq \tau_2 \\ \tau_4 \geq 0 \\ \tau_1 + (T-1)\tau_2 \geq 0 \\ \tau_4(\tau_1 + (T-1)\tau_2) \geq T\tau_3^2. \end{cases}$$

Proof. [Rujeerapaiboon et al., 2014, Proposition 4.3] □

By applying the above proposition to the restrictions of Problem (22) we obtain the following non-linear program, where the first restriction corresponds to the inner-product restriction of (22), the following four restrictions to $\mathbf{M} \succeq 0$ and the last four restrictions to the last positive semidefinite restriction

$$
\begin{aligned}
WVaR_\epsilon(\tilde{\nu}_T^\gamma(w)) = \min \quad & \nu \\
\text{s.t.} \quad & \tau \in \mathbb{R}^4,\ \beta, \nu \in \mathbb{R} \\
& \beta + \frac{1}{\epsilon}\left[T(\sigma_P^2 + (\mu_w - \gamma)^2)\tau_1 + T(T-1)(\mu_w - \gamma)^2\tau_2 + 2T(\mu_w - \gamma)\tau_3 + \tau_4\right] \leq 0 \\
& \tau_1 \geq \tau_2 \\
& \tau_4 \geq 0 \\
& \tau_1 + (T-1)\tau_2 \geq 0 \\
& \tau_4(\tau_1 + (T-1)\tau_2) \geq T\tau_3^2 \\
& \tau_1 - 1 \geq \tau_2 \\
& \tau_4 + T\nu + \beta \geq 0 \\
& \tau_1 - 1 + (T-1)\tau_2 \geq 0 \\
& (\tau_4 + T\nu + \beta)(\tau_1 - 1 + (T-1)\tau_2) \geq T\tau_3^2.
\end{aligned}
$$

We note that the first restriction is binding in optimality for (τ, ν, β), as $(\tau, \nu - \frac{\Delta}{T}, \beta + \Delta)$ has smaller objective value but stays feasible. Also, there exists an optimal solution for which $\tau_1 = \tau_2 + 1$ since $\left(\frac{\tau_1+(T-1)\tau_2-1}{T} + 1, \frac{\tau_1+(T-1)\tau_2-1}{T}, \tau_3, \tau_4, \nu, \beta\right)$ is feasible if (τ, ν, β) is and has the same objective value.
Therefore, by substituting $\tau_1 = \tau_2 + 1$ and omitting redundant constraints, we can further simplify the above optimization problem to

$$
\begin{aligned}
WVaR_\epsilon(\tilde{\nu}_T^\gamma(w)) = \min \quad & \nu \\
\text{s.t.} \quad & \tau_2, \tau_3, \tau_4, \beta, \nu \in \mathbb{R} \\
& \beta + \frac{1}{\epsilon}\left[T(\sigma_P^2 + (\mu_w - \gamma)^2) + T(\sigma_P^2 + T(\mu_w - \gamma)^2)\tau_2 + 2T(\mu_w - \gamma)\tau_3 + \tau_4\right] = 0 \\
& \tau_4 \geq 0 \\
& \tau_4 + T\nu + \beta \geq 0 \\
& \tau_2 \geq 0 \\
& \tau_4\left(\tau_2 + \frac{1}{T}\right) \geq \tau_3^2 \\
& (\tau_4 + T\nu + \beta)\tau_2 \geq \tau_3^2.
\end{aligned}
\tag{24}
$$

In order to find an analytical solution for Problem (24), we use the transformations

$$r := \frac{\tau_4}{2T}, \quad x := \frac{\tau_4 + T\nu + \beta}{2T}, \quad y := 2T\tau_2 + 1 \quad \text{and} \quad z := \tau_3$$

and achieve a more compact formulation.
For these transformations the following equivalences hold

$$\begin{aligned}
\tau_4 \geq 0 &\Leftrightarrow r \geq 0, \\
\tau_2 \geq 0 &\Leftrightarrow y \geq 1, \\
\tau_4 + T\nu + \beta \geq 0 &\Leftrightarrow x \geq 0, \\
\tau_4 \left(\tau_2 + \frac{1}{T} \right) \geq \tau_3^2 &\Leftrightarrow r(y+1) \geq z^2 \quad \text{and} \\
(\tau_4 + T\nu + \beta)\tau_2 \geq \tau_3^2 &\Leftrightarrow x(y-1) \geq z^2.
\end{aligned}$$

We can also express ν as a linear function of r, x, y and z

$$\begin{aligned}
\nu &= 2(x-r) - \frac{\beta}{T} \\
&= 2(x-r) + \frac{1}{\epsilon T} \left[T(\sigma_w^2 + (\mu_w - \gamma)^2)) + T(\sigma_w^2 + T(\mu_w - \gamma)^2))\tau_2 + 2T(\mu_w - \gamma)\tau_3 + \tau_4 \right] \\
&= 2(x-r) + \frac{1}{\epsilon T} \left[T(\sigma_w^2 + (\mu_w - \gamma)^2)) + T(\sigma_w^2 + T(\mu_w - \gamma)^2))\frac{y-1}{2T} + 2T(\mu_w - \gamma)z + 2Tw \right] \\
&= \left(\frac{2}{\epsilon} - 2 \right) r + 2x + \frac{\sigma_w^2 + T(\mu_w - \gamma)^2}{2\epsilon T} y + \frac{2(\mu_w - \gamma)}{\epsilon} z + \frac{\sigma_w^2 + (\mu_w - \gamma)^2}{\epsilon} - \frac{\sigma_w^2 + T(\mu_w - \gamma)^2}{2\epsilon T} \\
&= \left(\frac{2}{\epsilon} - 2 \right) r + 2x + \frac{\sigma_w^2 + T(\mu_w - \gamma)^2}{2\epsilon T} y + \frac{2(\mu_w - \gamma)}{\epsilon} z + \frac{(\mu_w - \gamma)^2}{2\epsilon} + \frac{\sigma_w^2(2T-1)}{2\epsilon T}.
\end{aligned}$$

Therefore, if we set

$$\begin{aligned}
a := \left(\frac{2}{\epsilon} - 2 \right), \; b := 2, \; c := \frac{\sigma_w^2 + T(\mu_w - \gamma)^2}{2\epsilon T}, \\
d := \frac{2(\mu_w - \gamma)}{\epsilon} \quad \text{and} \quad e := \frac{(\mu_w - \gamma)^2}{2\epsilon} + \frac{\sigma_w^2(2T-1)}{2\epsilon T},
\end{aligned}$$

we can finally express Problem (24), and hence $WVaR_\epsilon(\tilde{\nu}_T^\gamma(w))$, as

$$\begin{aligned}
WVaR_\epsilon(\tilde{\nu}_T^\gamma(w)) = \min \quad & ar + bx + cy + dz + e \\
\text{s.t.} \quad & r, x, y, z \in \mathbb{R} \\
& r \geq 0, \; x \geq 0, \; y \geq 1 \\
& r(y+1) \geq z^2, \; x(y-1) \geq z^2.
\end{aligned} \tag{25}$$

We are now ready to show that $WVaR_\epsilon(\tilde{\nu}_T^\gamma(w))$, respectively Problem (25), admits an analytical solution.

Lemma 5.1. *Consider an optimization problem of the form*

$$\begin{aligned} \min \quad & ar + bx + cy + dz + e \\ s.t. \quad & r, x, y, z \in \mathbb{R} \\ & r \geq 0,\ x \geq 0,\ y \geq 1 \\ & r(y+1) \geq z^2,\ x(y-1) \geq z^2. \end{aligned}$$

For $a, b, c, d, e \in \mathbb{R}$ *and* $\Delta := \sqrt{4(a+b)c - d^2}$ *with*

(i) $a, b, c > 0$

(ii) $4(a+b)c - d^2 > 0$ *and*

(iii) $d + \sqrt{\frac{b}{a}}\Delta < 0$,

the optimal value of the above optimization problem is given by

$$\frac{d^3 + d^2\Delta\left(\sqrt{\frac{b}{a}} - \sqrt{\frac{a}{b}}\right) - d\Delta^2 - 4dc(a+b) + 2c(a+b)\Delta\left(\sqrt{\frac{a}{b}} - \sqrt{\frac{b}{a}}\right)}{2\Delta(a+b)\left(\sqrt{\frac{a}{b}} + \sqrt{\frac{b}{a}}\right)} + e. \tag{26}$$

Proof. As the considered optimization problem states a convex SOCP with two hyperbolic constraints, the Karush-Kuhn-Tucker (KKT) optimality conditions are necessary and sufficient (see [Luenberger and Ye, 2008]).
Hence, we are looking for an candidate solution which satisfies the KKT-conditions.
With

$$p := \frac{-d - \sqrt{\frac{b}{a}}\Delta}{2(a+b)} \quad \text{and} \quad q := \frac{-d + \sqrt{\frac{b}{a}}\Delta}{2(a+b)}$$

we will show that the candidate solution

$$y := \frac{p+q}{q-p},\ z := \frac{2pq}{q-p},\ r := \frac{z^2}{y+1} \quad \text{and} \quad x := \frac{z^2}{y-1}$$

is in fact optimal for our optimization problem.

We observe that by definition $q > p$ and

$$y = \frac{p+q}{q-p} = 1 + 2\frac{p}{q-p}.$$

Hence, $y > 1 \Leftrightarrow p > 0$, which is guaranteed by assumption (*iii*).
Obviously, $r \geq 0$ and $x \geq 0$ hold and the two hyperbolic constraints are binding.
We conclude that the above solution is indeed feasible.

By denoting λ_1, λ_2 and λ_3 the Lagrange multipliers of the three linear inequalities and λ_4 and λ_5 the Lagrange multipliers of the two hyperbolic constraints, it is easy to show that for $\lambda_1 = \lambda_2 = \lambda_3 = 0$, $\lambda_4 = \frac{a}{y+1}$ and $\lambda_5 = \frac{b}{y-1}$ all KKT-conditions are met.

The optimal value is thus given by

$$\begin{aligned} ar + bx + cy + dz + e =& a\frac{z^2}{y+1} + b\frac{z^2}{y-1} + cy + dz + e \\ =& \frac{2ap^2q + 2bpq^2 + c(p+q) + 2dpq}{q-p} + e, \end{aligned}$$

from which we obtain the optimal value (26) by substituting the definitions of p and q. □

We can now use Lemma 5.1 to explicitly state the analytical solution of $WVaR_\epsilon(\tilde{\nu}_T^\gamma(w))$.

Theorem 5.2. *For $\gamma > \mu_w + \sqrt{\frac{\epsilon}{(1-\epsilon)T}}\sigma_w$ the $WVaR_\epsilon(\tilde{\nu}_T^\gamma(w))$ of $w \in \mathbb{R}^n$ is given by*

$$WVaR_\epsilon(\tilde{\nu}_T^\gamma(w)) = \left(\gamma - w'\mu + \sqrt{\frac{1-\epsilon}{\epsilon T}}\sqrt{w'\mathbf{\Sigma}w}\right)^2 + \frac{T-1}{\epsilon T}w'\mathbf{\Sigma}w. \tag{27}$$

Proof. As we have already shown, the $WVaR_\epsilon(\tilde{\nu}_T^\gamma(w))$ is given by the solution of Problem (25), where $a = \left(\frac{2}{\epsilon} - 2\right)$, $b = 2$, $c = \frac{\sigma_w^2 + T(\mu_w - \gamma)^2}{2\epsilon T}$, $d = \frac{2(\mu_w - \gamma)}{\epsilon}$ and $e = \frac{(\mu_w - \gamma)^2}{2\epsilon} + \frac{\sigma_w^2(2T-1)}{2\epsilon T}$. To apply Lemma 5.1 we need to verify its assumptions.

Obviously $a, b, c > 0$ hold, and so does $4(a+b)c - d^2 = \frac{2\sigma_w}{\epsilon\sqrt{T}} > 0$.
It is also easy to verify that assumption (iii) of Lemma 5.1 is equivalent to $\gamma > \mu_w + \sqrt{\frac{\epsilon}{(1-\epsilon)T}}\sigma_w$ and therefore all conditions of Lemma 5.1 are met.
By substituting the definitions of a, b, c, d and e back into the optimal solution (26), one obtains the claimed representation of $WVaR_\epsilon(\tilde{\nu}_T^\gamma(w))$. □

Remark 5.1. Note that the lower bound $\gamma_w^{lb} := \mu_w + \sqrt{\frac{\epsilon}{(1-\epsilon)T}}\sigma_w$ in Theorem 5.2 depends on the actual portfolio weights $w \in \mathbb{R}^n$ we are looking at, so in fact $\gamma_w^{lb} = \gamma^{lb}(w) = w'\mu + \sqrt{\frac{\epsilon}{(1-\epsilon)T}w'\mathbf{\Sigma}w}$.

As we want to use the formulation (27) of $WVaR_\epsilon(\tilde{\nu}_T^\gamma(w))$ as the objective function of our portfolio optimization problem, our chosen γ has to comply with the lower bound restrictions for all considered portfolios, so $\gamma > \gamma_w^{lb} \quad \forall w \in \mathbb{W}$.

By the use of the above result we can finally reformulate our initial Portfolio Optimization Problem (19) as

$$\begin{aligned} &\min_{w} \left(\gamma - w'\mu + \sqrt{\frac{1-\epsilon}{\epsilon T}} \sqrt{w'\mathbf{\Sigma}w} \right)^2 + \frac{T-1}{\epsilon T} w'\mathbf{\Sigma}w \\ &\text{s.t. } w \in \mathbb{W}. \end{aligned} \tag{28}$$

The objective function is now a closed form analytic function of the portfolio weights w. We also observe that the length of the investment time interval does not affect the size of this optimization problem. Naturally, it is also closely related to the portfolio optimization problem derived in [Rujeerapaiboon et al., 2014], see Section 5.4 for this special case.

In order to solve this problem by widely used and accessible methods, we will reformulate it as a second-order cone program (SOCP) in Section 5.2.
The following remark will be used in Section 5.3 where we account for the ambiguity of the true moments μ and $\mathbf{\Sigma}$.

Remark 5.2. As long as $\gamma > \gamma_w^{lb}\,(> w'\mu)$ holds, the worst-case value-at-risk of our portfolio evaluation (27) is decreasing in the portfolio mean return $w'\mu$ and increasing in the portfolio variance $w'\mathbf{\Sigma}w$.

This property simply means that under all portfolios with the same variance, our evaluation method prefers the ones with the highest expected return. Vice versa, under all portfolios with the same expected return, we prefer the ones with the smallest variance. Hence, a portfolio obtained by solving Problem (28) is mean-variance efficient in the classical sense, which we stated in Section 3 Definition 3.4.

We conclude that the distributionally robust portfolios we have just derived are mean-variance efficient in the sense of the classical Markowitz approach presented in Section 3. Therefore, all of this portfolios lie on the classical efficient frontier, where the risk-aversion parameter controls the exact location.

Note that this property does not hold if we take moment ambiguity into account, which will be presented in Section 5.3. Also, not all portfolios on the classical efficient frontier will be “attainable” by certain choices of γ since the process of distributional robustification induces additional restrictions on the selection process.

To provide the possibility of using one of the widely known solvers for SOCPs we will reformulate Problem (28) as a second-order cone program in the following Section 5.2. We have already presented some of the solvers that can be used in Section 2.3.

5.2 Representation as a SOCP

We will now reformulate the Portfolio Optimization Problem (28) as a SOCP.
For this purpose, we recall the notion of cones by the definition of the so-called second-order (or Lorentz) cone which we already stated in Section 2.1.
The second-order cone C_q in $\mathbb{R}^{n+1}$ was defined as

$$C_q := \left\{ (x_0, x_1, \ldots, x_n)' \in \mathbb{R}^{n+1} : x_0 \geq \|(x_1, \ldots, x_n)\|_q \right\}, \tag{29}$$

where $\|(x_1, \ldots, x_n)\|_q := \sqrt[q]{|x_1|^q + \ldots + |x_n|^q}$ for $q \in \mathbb{Z}$ denotes the finite dimensional q-Norm of a vector $x \in \mathbb{R}^n$.

Observation 5.1. *If $\mathbb{W}$ describes the set of admissible portfolios and $\mathbb{W}$ is characterized by a finite number of linear constraints, the Portfolio Optimization Problem (28) reduces to a tractable SOCP whose size is again independent of the investment horizon*

$$\begin{aligned} &\min_{w} \left(\gamma - w'\mu + \sqrt{\frac{1-\epsilon}{\epsilon T}} s \right)^2 + \frac{T-1}{\epsilon T} s^2 \\ &s.t.\, w \in \mathbb{W} \\ &(s, \Sigma^{1/2} w) \in C_2. \end{aligned} \tag{30}$$

Note that by definition of the second-order cone, the last restriction $(s, \Sigma^{1/2} w) \in C_2$ is equivalent to

$$s \geq \|\mathbf{\Sigma}^{1/2} w\|_2 = \sqrt{(\mathbf{\Sigma}^{1/2} w)' (\mathbf{\Sigma}^{1/2} w)} = \sqrt{w' \mathbf{\Sigma} w}. \tag{31}$$

Since we are minimizing and the objective function is increasing in s because $\gamma > w'\mu$ due to the lower bound restriction on γ derived in Theorem 5.2, inequality (31) will be binding in optimality, hence $s = \sqrt{w'\mathbf{\Sigma} w}$. This proves the equivalence of Problem (30) and Problem (28) in terms of optimization.
Program (30) indeed states an SOCP which can be solved by many common program packages (see Section 2.3). This becomes particularly obvious if we rewrite Problem (30) as

$$\begin{aligned} &\min_{w,t} t \\ \text{s.t.}\, &w \in \mathbb{W} \\ &t \geq \left(\gamma - w'\mu + \sqrt{\frac{1-\epsilon}{\epsilon T}} s \right)^2 + \frac{T-1}{\epsilon T} s^2 \\ &(s, \Sigma^{1/2} w) \in C_2, \end{aligned}$$

which is nothing but a quadratically constrained linear program which already has a conic constraint.

5.3 Robustification of the Moment-Estimations

Until now, we assumed that μ and $\mathbf{\Sigma}$ are the exact mean and covariance matrix of the asset-returns. Of course, in practice we are obliged to use estimates $\hat{\mu}$ and $\hat{\mathbf{\Sigma}}$ which are bound to errors. We will account for these errors by following the approach of [Rujeerapaiboon et al., 2014, Section 5.2] and assume that the true moments μ and $\mathbf{\Sigma}$ lie in a convex uncertainty set of the form

$$\mathcal{U} = \left\{(\mu, \mathbf{\Sigma}) \in \mathbb{R}^n \times \mathbb{S}^n : (\mu - \hat{\mu})'\hat{\mathbf{\Sigma}}^{-1}(\mu - \hat{\mu}) \leq \delta_1,\ \delta_3\hat{\mathbf{\Sigma}} \preceq \mathbf{\Sigma} \preceq \delta_2\hat{\mathbf{\Sigma}}\right\},$$

where $\hat{\mu}$ and $\hat{\mathbf{\Sigma}}$ are point estimates we obtain for example by applying the estimation methods presented in Section 6.

Here, $\delta_1 \geq 0$ and $\delta_2 \geq 1 \geq \delta_3 \geq 0$ describe our confidence in the estimations $\hat{\mu}$ and $\hat{\mathbf{\Sigma}}$. See again Section 6 for details on the determination of $\hat{\mu}, \hat{\mathbf{\Sigma}}, \delta_1, \delta_2$ and δ_3.

Taking the moment-ambiguity into account and using our previous result from Equation (27), the worst-case value-at-risk of our portfolio evaluation $\tilde{\nu}_T^\gamma(w)$ is given by

$$WVaR_\epsilon(\tilde{\nu}_T^\gamma(w)) = \max_{(\mu,\mathbf{\Sigma})\in\mathcal{U}} \left(\gamma - w'\mu + \sqrt{\frac{1-\epsilon}{\epsilon T}}\sqrt{w'\mathbf{\Sigma}w}\right)^2 + \frac{T-1}{\epsilon T}w'\mathbf{\Sigma}w.$$

Again, we can reformulate this expression into a closed analytic form, which is done in the following theorem.

Theorem 5.3. *If $\mathbf{\Sigma} \succ 0$ and $\gamma > \gamma^{lb} = w'\mu + \sqrt{\frac{\epsilon}{(1-\epsilon)T}w'\mathbf{\Sigma}w}$ for all $(\mu, \mathbf{\Sigma}) \in \mathcal{U}$, then*

$$WVaR_\epsilon(\tilde{\nu}_T^\gamma(w)) = \left(\gamma - w'\hat{\mu} + \left(\sqrt{\delta_1} + \sqrt{\frac{(1-\epsilon)\delta_2}{\epsilon T}}\right)\sqrt{w'\hat{\mathbf{\Sigma}}w}\right)^2 + \frac{(T-1)\delta_2}{\epsilon T}w'\hat{\mathbf{\Sigma}}w.$$

Proof. As mentioned in Remark 5.2, the $WVaR_\epsilon(\tilde{\nu}_T^\gamma(w))$ is decreasing in the portfolio mean return $w'\mu$ and increasing in the portfolio standard deviation $\sqrt{w'\mathbf{\Sigma}w}$. Hence, an upper bound for the worst-case scenario, which is the highest possible value of $WVaR_\epsilon(\tilde{\nu}_T^\gamma(w))$ for $w \in \mathbb{W}$, can be obtained by substituting the smallest, respectively the highest possible values for the mean return and the standard deviation into the above equation.
The highest possible portfolio variance, given the weight vector w, in our uncertainty set U is obviously given by

$$\max_{(\mu,\Sigma)\in\mathcal{U}} \sqrt{w'\mathbf{\Sigma}w} = \sqrt{\delta_2}\sqrt{w'\hat{\mathbf{\Sigma}}w}.$$

In order to determine the smallest return, i.e. the solution to the optimization problem stated as

$$\min_{(\mu,\Sigma)\in\mathcal{U}} w'\mu \quad \Leftrightarrow \quad \min_{\mu} w'\mu \quad \text{s.t.}\ (\mu - \hat{\mu})'\hat{\mathbf{\Sigma}}^{-1}(\mu - \hat{\mu}) \leq \delta_1,$$

we assume the positive definiteness of $\hat{\mathbf{\Sigma}}$ and use the substitution $z := \hat{\mathbf{\Sigma}}^{-1/2}(\mu - \hat{\mu})$ to reformulate this problem as

$$\min_{z} w'\hat{\mu} + w'\hat{\mathbf{\Sigma}}^{1/2} z \quad \text{s.t.}\, z'z \leq \delta_1.$$

Again, the Karush-Kuhn-Tucker (KKT) conditions are necessary and sufficient (see [Luenberger and Ye, 2008]). As $w \neq 0$ is given, the KKT conditions directly imply that for the optimal solution $\tilde{z}'\tilde{z} = \delta_1$ holds. By solving the remaining KKT-equation, the optimal solution is given as $\tilde{z} = -\frac{\sqrt{\delta_1}}{\sqrt{w'\hat{\mathbf{\Sigma}}w}}\hat{\mathbf{\Sigma}}^{1/2}w'$. We can now conclude that

$$\min_{\mu,\Sigma} w'\mu = w'\hat{\mu} - \sqrt{\delta_1}\sqrt{w'\mathbf{\Sigma}w},$$

which yields the claim. □

We obtain our portfolio optimization problem with robustification against moment ambiguity by updating the objective function of Problem (30) to the above result

$$\begin{aligned} \min_{w} & \left(\gamma_P - w'\hat{\mu} + \left(\sqrt{\delta_1} + \sqrt{\frac{(1-\epsilon)\delta_2}{\epsilon T}}\right) s\right)^2 + \frac{(T-1)\delta_2}{\epsilon T} s^2 \\ \text{s.t.}\, & w \in \mathbb{W} \\ & (s, \Sigma^{1/2} w) \in C_2. \end{aligned} \tag{32}$$

By using the same arguments as in Section 5.2, we immediately formulated the problem with moment robustification as an SOCP. Therefore, the computational effort to compute the optimal solution does not change by introducing moment ambiguity.

Remark 5.3. The requirements $\mathbf{\Sigma} \succ 0$ and $\gamma > \gamma^{lb} = w'\mu + \sqrt{\frac{\epsilon}{(1-\epsilon)T} w'\mathbf{\Sigma}w}$ for all $(\mu, \mathbf{\Sigma}) \in \mathcal{U}$ of Theorem 5.3 are equivalent to

$$\delta_3\hat{\mathbf{\Sigma}} \succ 0 \quad \text{and} \quad \gamma > w'\hat{\mu} + \sqrt{\delta_1}\sqrt{w'\hat{\mathbf{\Sigma}}w} + \sqrt{\frac{\epsilon\delta_2}{(1-\epsilon)T}}\sqrt{w'\hat{\mathbf{\Sigma}}w}.$$

We have now formulated our portfolio optimization problem with and without moment ambiguity in dependence of the risk aversion parameter. In the remainder of this section, we will have a look at the implication of different choices of γ on the resulting portfolios.

5.4 The Choice of the Risk-Aversion Parameter

We have already motivated the interpretation of the parameter γ as some sort of risk-aversion adjustment in Section 4.1. Now we want to have a closer look at the choice of γ and its implications.

In general, there are two lower bound restrictions one needs to consider when choosing the risk-aversion parameter:

- The first one is the technical lower bound restriction which we obtained by the derivation of the analytic formulation of $WVaR_\epsilon(\tilde{\nu}_T^\gamma(w))$ in Section 5.1.
  We have seen that γ has to be bigger than γ_w^{lb} for all admissible portfolios w, so

$$\gamma > w'\mu + \sqrt{\frac{\epsilon}{(1-\epsilon)T} w'\mathbf{\Sigma} w} \quad \forall w \in \mathbb{W}, \tag{33}$$

  where again $\mathbb{W}$ denotes the set of all admissible portfolios.

- We stated the second lower bound restriction for γ in the very beginning of Section 4 (Portfolio Evaluation) where we demanded γ to be *"sufficiently large such that all realized portfolio-returns are smaller than γ"*. The reason for this requirement is the fact that otherwise, due to the definition of our penalization function $f_\gamma(x)$ as a quadratic polynomial with its minimum at γ, we would penalize high (bigger than γ) portfolio-returns like we do small ones. This, of course, is undesirable since we welcome high portfolio returns and therefore should not penalize them more than lower returns.

As the asset-returns, and therefore the portfolio-returns, are modelled as continuous random variables unbounded from above, theoretically we cannot satisfy the second requirement on γ stated above.
Nevertheless, we are now using the penalization function $f_\gamma(x)$ to choose from the set of all feasible portfolios. Hence, for our purpose, γ does not have to comply with the requirement of being bigger than any asset-return realization but actually has to exceed all possible expected portfolio-returns.

Since $\sqrt{\frac{\epsilon}{(1-\epsilon)T} w'\mathbf{\Sigma} w} \geq 0$ holds, a globally chosen γ such that the technical lower bound restriction (33) is fulfilled, is also bigger than the expected portfolio return $w'\mu$ for any portfolio w which is feasible. This implies that such a γ is bigger than the highest portfolio return possible, and therefore the technical lower bound is sufficient to guarantee the subject based second restriction.

In order to comply with the technical restriction, we approximate the exact lower bound $\max_{w\in\mathbb{W}} w'\mu + \sqrt{\frac{\epsilon}{(1-\epsilon)T} w'\mathbf{\Sigma} w}$ by

$$\gamma^* := \max_{w\in\mathbb{W}} w'\mu + \sqrt{\frac{\epsilon}{(1-\epsilon)T}} \max_{w\in\mathbb{W}} \sqrt{w'\mathbf{\Sigma} w}. \tag{34}$$

Since $\gamma^* \geq \max_{w\in\mathbb{W}} w'\mu + \sqrt{\frac{\epsilon}{(1-\epsilon)T}}\sqrt{w'\mathbf{\Sigma} w}$ we can now choose $\gamma \in (\gamma^*, \infty]$ arbitrarily. We observe that γ^* is the mean return plus a multiple of the standard deviation, where the multiplication factor contains the time horizon T in its denominator. For very long investment horizons (T very big), the lower bound is getting smaller and will almost equal the maximum expected return. We will further discuss this observation in Section 5.4.2. We also observe that $0 < \gamma^* \ll 1$ will hold for most practical cases.

Observation 5.2. *If one chooses γ as a function of the portfolio weights w, namely $\gamma(w) = w'\mu$, our methods of reformulating the initial Portfolio Optimization Problem (19) also work fine although the technical lower bound restriction (33) is obviously not satisfied. One can show that with $\tilde{\nu}_T^{\gamma}(w) = \frac{1}{T}\sum_{t=1}^{T}(w'\tilde{r}_t - w'\mu)^2$ instead of (13) and a similar derivation, the objective function of (28) becomes*

$$\min_w \frac{1}{\epsilon} w'\Sigma w.$$

Obviously, this approach is equivalent to minimizing the portfolio variance per se. Thus, we obtain the classical minimum variance portfolio which we will henceforth call MVAR.

In the following, we want to survey some explicit choices of γ. Where in Section 5.4.1 the parameter is chosen equal to one, we will further discuss the role of γ as a risk-aversion parameter in Section 5.4.2.

5.4.1 The Robust Growth-Optimal Portfolio

For $\gamma = 1$ the objective function of our Portfolio Optimization Problem (28) becomes

$$\min_w \left(1 - w'\mu + \sqrt{\frac{1-\epsilon}{\epsilon T}}\sqrt{w'\mathbf{\Sigma} w}\right)^2 + \frac{T-1}{\epsilon T} w'\mathbf{\Sigma} w,$$

which is, in terms of optimization, equivalent to the objective function of the robust growth-optimal portfolio derived in [Rujeerapaiboon et al., 2014, Theorem 4.1], namely

$$\max_w \frac{1}{2}\left(1 - \left(1 - w'\mu + \sqrt{\frac{1-\epsilon}{\epsilon T}}\sqrt{w'\mathbf{\Sigma} w}\right)^2 - \frac{T-1}{\epsilon T} w'\mathbf{\Sigma} w\right).$$

Hence, if we choose $\gamma = 1$, we obtain the robust growth-optimal portfolio which we will denote by RGOP, respectively RGOP+ for the robust growth-optimal portfolio with moment-ambiguity.

Remark 5.4. Analogues to Remark 5.3 the requirements $\mathbf{\Sigma} \succ 0$ and $\gamma > \gamma^{lb} = w'\mu + \sqrt{\frac{\epsilon}{(1-\epsilon)T} w'\mathbf{\Sigma} w}$ for all $(\mu, \mathbf{\Sigma}) \in \mathcal{U}$ of Theorem 5.3 for $\gamma = 1$ are equivalent to

$$\delta_3 \hat{\mathbf{\Sigma}} \succ 0 \quad \text{and} \quad 1 > w'\hat{\mu} + \sqrt{\delta_1}\sqrt{w'\hat{\mathbf{\Sigma}}w} + \sqrt{\frac{\epsilon\delta_2}{(1-\epsilon)T}}\sqrt{w'\hat{\mathbf{\Sigma}}w}.$$

Note that for most practical cases $\gamma = 1 > w'\mu + \sqrt{\frac{\epsilon}{(1-\epsilon)T}}\sqrt{w'\mathbf{\Sigma}w} \quad \forall w \in \mathbb{W}$ holds. Otherwise, this can always be achieved by shortening the rebalancing intervals.
In their work, [Rujeerapaiboon et al., 2014] also state that *"this condition even holds for yearly rebalancing intervals if the means and standard deviations of the asset-returns fall within their typical ranges reported in [Luenberger, 1998, § 8]"*.

The RGOP portfolio is achieved by maximizing the worst-case value-at-risk of a quadratic approximation of the portfolio growth rate. The difference between the growth rate $\tilde{\gamma}_T$, given by

$$\tilde{\gamma}_T = \log \sqrt[T]{\prod_{t=1}^{T}[1 + w_t \tilde{r}_t]} = \frac{1}{T}\sum_{t=1}^{T} log[1 + w_t \tilde{r}_t], \tag{35}$$

and its quadratic approximation

$$\tilde{\gamma}'_T = \frac{1}{T}\sum_{t=1}^{T}\left(w_t\tilde{r}_t - \frac{1}{2}(w_t\tilde{r}_t)^2\right) \tag{36}$$

is thereby reported to be uniformly bounded by 1% under monthly and by 5% under yearly rebalancing (see [Rujeerapaiboon et al., 2014, Section 4]).
We observe that for $\gamma = 1$, our evaluation function becomes

$$\tilde{\nu}_T^1 = \frac{1}{T}\sum_{t=1}^{T}(w_t\tilde{r}_t - 1)^2 = \frac{1}{T}\sum_{t=1}^{T}\left((w_t\tilde{r}_t)^2 - 2w_t\tilde{r}_t + 1\right),$$

which is, in terms of optimization and since in our model we were minimizing, equivalent to the quadratic approximation of the growth rate.

5.4.2 The Risk-Aversion Parameter

In the last part of this section, we again want to motivate the interpretation of γ as a risk-aversion parameter. To better understand the implications of an explicit choice of $\gamma \in (\gamma^*, \infty]$ (with γ^* defined by (34)) to the resulting portfolio, we recall the example figure of our penalization function f_γ with respect to γ in Figure 8.

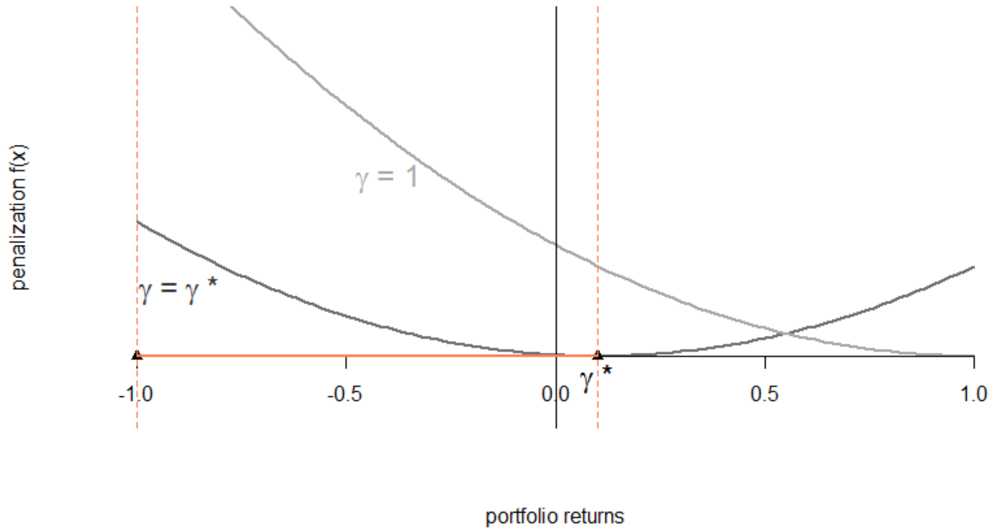

Figure 8: Penalization function $f_\gamma(x)$ for different γ

We observe that as the chosen γ tends towards γ^*, the resulting portfolio will become more risk averse, meaning that both expected return and variance decrease. This is due to the fact that for small γ the curvature of f_γ on $[-1, \gamma^*]$ increases. Then, in order to achieve good performance according to our evaluation $\tilde{\nu}_T^\gamma(w)$ (see (13)), a high mean-return becomes less important compared to small variance. We have already seen that for $\gamma = 1$ we achieve the robust growth optimal portfolio. Since $\gamma = \gamma^*$ is the most risk-averse we can get, we want to name the resulting portfolio Robust Risk Averse Portfolio (RRAP) or likewise RRAP+ under moment ambiguity.
We remark that for $\gamma = \gamma^*$ restriction (33) may not be fulfilled with strict inequality for all $w \in \mathbb{W}$. But since our whole Portfolio Optimization Problem (28) is continuous in γ, a permissible choice $\gamma = \gamma^* + \epsilon_\gamma$ will (numerically) deliver the same resulting portfolio. We also observe that γ^* is decreasing in the length T of the investment horizon. Following the above reasoning, this means that the longer the investment horizon gets, the more risk averse the RRAP (or RRAP+) portfolio will get.

As mentioned earlier, the distributional robust portfolios without moment ambiguity are all mean-variance efficient in the classical sense. Therefore, γ plays the same role as the risk aversion parameter in the Markowitz mean-variance efficiency framework does. This means they both define the location of the portfolio on the efficient frontier.

This completes our derivation of the distributionally robust portfolios. In Section 7.2 we will apply our theory to real life data where we use the moment estimation techniques presented in the following Section 6.

6 Parameter Estimation

In this section, we want to present our method of choice for estimating the first two moments of the asset-returns $\tilde{r} = (\tilde{r}^1, \dots, \tilde{r}^n)$. It is well known that the naive approach of selecting $\hat{\mu}$ and $\hat{\mathbf{\Sigma}}$ as the sample estimators often shows poor out of sample performance. To avoid this problem, we utilize the approach of shrinkage estimators, combining the raw estimate with "other information". In our setting, this is reached by shrinking the sample estimator towards a target estimator. We denote the shrinkage estimator of the expected value μ by μ_{sh} and the shrinkage estimator of the covariance matrix $\mathbf{\Sigma}$ by $\mathbf{\Sigma}_{sh}$.
In Section 6.1 we will present the basic idea of shrinkage estimators and state the particular estimations for the expected value in Section 6.1.1 and for the covariance matrix in Section 6.1.2. Of course the shrinkage estimators are subject to estimation errors. We accounted estimation errors in Section 5.3 by robustifying $WVaR_\epsilon(\tilde{\nu}_T^\gamma(w))$ against all $(\mu, \mathbf{\Sigma}) \in \mathcal{U}$, where $\mathcal{U}$ defined uncertainty cones for both μ and $\mathbf{\Sigma}$. In Section 6.2 we will define those uncertainty cones and state the respective estimations. For later convenience we will denote the sample estimators by

$$\begin{aligned} \mu_{sp} &:= \frac{1}{T}\sum_{t=1}^{T} \tilde{r}_t \quad \text{and} \\ \mathbf{\Sigma}_{sp} &:= \frac{1}{T-1}\sum_{t=1}^{T} (\tilde{r}_t - \mu_{sp})(\tilde{r}_t - \mu_{sp})', \end{aligned}$$

where $T \in \mathbb{N}$ now describes the time horizon for which we observed the asset-returns.

Throughout this section we assume the covariance matrix estimation $\hat{\mathbf{\Sigma}}$ to be positive definite. If this is not the case, one can use various methods to "correct" the estimation. Among others, see [Rebonato and Jäckel, 1999] for further details.

6.1 Shrinkage Estimators of Moments

For our purpose of using the estimators for portfolio optimization, we will focus on the approach of [DeMiguel et al., 2013] on shrinkage estimators. Shrinkage estimators are convex combinations of the sample estimators and a scaled shrinkage target, where the convexity parameter α is called the shrinkage intensity. We will denote the shrinkage targets by μ_{tg} and Σ_{tg} and the scaling parameters by ϕ_μ and ϕ_Σ, respectively.

$$\mu_{sh} := (1 - \alpha_\mu)\mu_{sp} + \alpha_\mu \phi_\mu \mu_{tg} \tag{37}$$

$$\mathbf{\Sigma}_{sh} := (1 - \alpha_\Sigma)\mathbf{\Sigma}_{sp} + \alpha_\Sigma \phi_\Sigma \mathbf{\Sigma}_{tg} \tag{38}$$

The general advantage of shrinkage estimators is that it can be shown that, under general conditions, there exists a shrinkage intensity for which the resulting shrinkage estimator

contains less estimation error than the original sample estimator.
In order to determine μ_{sh} and $\mathbf{\Sigma}_{sh}$, we therefore need to estimate the optimal shrinkage intensities α_μ and α_Σ.

In the case of estimating the mean, DeMiguel et al. derived a closed-form expression for the true optimal shrinkage intensity, assuming that the returns are independent and identically distributed (iid) without any other distributional assumptions.
When estimating the covariance matrix, such a closed-form expression is only given if the returns are assumed to be iid normal. If the returns are iid but not normal, a non parametric procedure to estimate the true optimal shrinkage intensity is presented.

6.1.1 Shrinkage Estimator of Mean Returns

For estimating the mean returns, the idea is to choose the shrinkage intensity such that it minimizes the expected quadratic loss of the estimator. We choose the shrinkage target μ_{tg} as the vector of ones and the scaling factor ϕ_μ to minimize the bias of the shrinkage target, so

$$\mu_{tg} = \mathbf{1} \quad \text{and}$$
$$\phi_\mu = \text{argmin}_\phi \|\phi\mathbf{1} - \mu\|_2^2 = \frac{1}{n}\sum_{i=1}^{n} \mu_i =: \overline{\mu},$$

where μ denotes the true (unknown) mean return vector.
Note that DeMiguel et al. justify the choice of $\mu_{tg} = \mathbf{1}$ by the fact that "in the case where the shrinkage intensity is equal to one, the solution of the estimated mean-variance portfolio would be the minimum-variance portfolio, which is a common benchmark". We follow this approach, and as we mentioned above, we choose the shrinkage intensity to minimize the expected quadratic loss of the estimator. Therefore we select α_μ as the optimal solution of

$$\min_\alpha \mathbb{E}\left[\|\mu_{sh} - \mu\|_2^2\right]$$
$$\text{s.t.} \quad \mu_{sh} = (1-\alpha)\mu_{sp} + \alpha\phi_\mu\mathbf{1}.$$

We are now ready to state the closed-form expression of the true optimal shrinkage intensity in the following proposition as stated in [DeMiguel et al., 2013].

Proposition 6.1. *For $\tilde{r}_t$ iid, the true optimal shrinkage intensity, in the sense of minimal quadratic loss, is given by*

$$\alpha_\mu = \frac{\mathbb{E}\left(\|\mu_{sp} - \mu\|_2^2\right)}{\mathbb{E}\left(\|\mu_{sp} - \mu\|_2^2\right) + \|\phi_\mu\mathbf{1} - \mu\|_2^2} = \frac{(n/T)\overline{\sigma^2}}{(n/T)\overline{\sigma^2} + \|\phi_\mu\mathbf{1} - \mu\|_2^2}, \tag{39}$$

with $\overline{\sigma^2} = trace(\mathbf{\Sigma})/n$.

Proof. See [DeMiguel et al., 2013, Appendix A. Proof of Proposition 1] □

We see that in Equation (39) we are still in need of the true moments μ and $\mathbf{\Sigma}$. [DeMiguel et al., 2013, Section 5] argue in their empirical tests that the use of the sample estimators μ_{sp} and $\mathbf{\Sigma}_{sp}$ instead only bears some reasonable estimation risk.
We can therefore state our estimation for the optimal shrinkage parameter as

$$\hat{\alpha_\mu} = \frac{\text{trace}(\mathbf{\Sigma}_{sp})/T}{\text{trace}(\mathbf{\Sigma}_{sp})/T + \|\phi_\mu \mathbf{1} - \mu_{sp}\|_2^2}. \tag{40}$$

Combining all our above results and plugging them in into equation (39), we obtain the shrinkage estimator for the first moment of $\tilde{r}$ as

$$\mu_{sh} = (1 - \hat{\alpha}_\mu)\mu_{sp} + \hat{\alpha}_\mu \overline{\mu} \mathbf{1}. \tag{41}$$

6.1.2 Shrinkage Estimator of the Covariance-Matrix

Analogous to the approach for the shrinkage estimator of μ, we here choose the shrinkage target $\mathbf{\Sigma}_{tg}$ as the identity matrix and the scaling factor ϕ_Σ to minimize the bias of the shrinkage target, so

$$\begin{aligned} &\mathbf{\Sigma}_{tg} = \mathbb{I} \quad \text{and} \\ &\phi_\Sigma = \text{argmin}_\phi \|\phi \mathbb{I} - \mathbf{\Sigma}\|_F^2 = \frac{1}{n}\sum_{i=1}^{n} \sigma_i^2 =: \overline{\sigma^2}, \end{aligned}$$

where $\|\cdot\|_F$ denotes the Frobenius-Norm of a matrix (for $A \in \mathbb{R}^{m \times n}$ the Frobenius-Norm is defined by $\|A\|_F^2 = \sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2$) and $\mathbf{\Sigma}$ the true (unknown) covariance matrix of the returns with diagonal elements σ_i^2 for $i = 1, \ldots, n$.
We again select the shrinkage intensity α_Σ to minimize the expected quadratic loss, so

$$\begin{aligned} &\min_\alpha \ \mathbb{E}\left[\|\mathbf{\Sigma}_{sh} - \mathbf{\Sigma}\|_F^2\right] \\ &\text{s.t.} \quad \mathbf{\Sigma}_{sh} = (1-\alpha)\mathbf{\Sigma}_{sp} + \alpha \phi_\Sigma \mathbb{I}. \end{aligned}$$

One can now show, by plugging the restriction into the objective function, that this problem is equivalent to

$$\min_\alpha \ \mathbb{E}\left[\|\mathbf{\Sigma}_{sh} - \mathbf{\Sigma}\|_F^2\right] = \min_\alpha (1-\alpha)^2 \mathbb{E}\left[\|\mathbf{\Sigma}_{sp} - \mathbf{\Sigma}\|_F^2\right] + \alpha^2 \|\phi_\Sigma \mathbb{I} - \mathbf{\Sigma}\|_F^2. \tag{42}$$

As the objective function of this problem is nothing but a quadratic function in the optimization variable, we simply set the derivation with respect to α equal to zero and solve the resulting equation. By using Equality (42) the resulting equation is given by

$$\frac{d}{d\alpha}\mathbb{E}\left[\|\mathbf{\Sigma}_{sh} - \mathbf{\Sigma}\|_F^2\right] = -2(1-\alpha)\mathbb{E}\left[\|\mathbf{\Sigma}_{sp} - \mathbf{\Sigma}\|_F^2\right] + 2\alpha\|\phi_\Sigma \mathbb{I} - \mathbf{\Sigma}\|_F^2,$$

which can be easily solved for α and results in

$$\alpha_\Sigma = \frac{\mathbb{E}\left[\|\mathbf{\Sigma}_{sp} - \mathbf{\Sigma}\|_F^2\right]}{\mathbb{E}\left[\|\mathbf{\Sigma}_{sp} - \mathbf{\Sigma}\|_F^2\right] + \|\phi_\Sigma \mathbb{I} - \mathbf{\Sigma}\|_F^2}. \tag{43}$$

By assuming the returns to be iid normal and $T > n + 4$, DeMiguel et al. derive a closed form expression for $\mathbb{E}\left[\|\mathbf{\Sigma}_{sp} - \mathbf{\Sigma}\|_F^2\right]$. As we intend to use this estimator in the context of distributionally robust portfolio optimization, we cannot justify this assumption. Therefore, we will use the technique of smoothed bootstrapping to obtain α_Σ (see [DeMiguel et al., 2013]).

For our purpose of calculating α_Σ, we will use the bootstrap for the approximation of the expected quadratic loss $\mathbb{E}\left[\|\mathbf{\Sigma}_{sp} - \mathbf{\Sigma}\|_F^2\right]$. We then estimate $\|\phi_\Sigma \mathbb{I} - \mathbf{\Sigma}\|_F^2$ by simply using the sample covariance matrix $\mathbf{\Sigma}_{sp}$ instead of the unknown true second moment $\mathbf{\Sigma}$, which we again justify by referring to the numerical results of [DeMiguel et al., 2013], where the same approach is taken.

The "vanilla" version of the bootstrap method for approximating $\mathbb{E}\left[\|\mathbf{\Sigma}_{sp} - \mathbf{\Sigma}\|_F^2\right]$ is to simply generate $B (\in \mathbb{N})$ samples of asset-returns by drawing observations with replacement from the original (observed) sample. For each of these samples $S_b := (r_{t,b}^1, \dots, r_{t,b}^n)_{t=1}^T$ with $b = 1, \dots, B$, we can then easily calculate $\|\mathbf{\Sigma}_{sp} - \mathbf{\Sigma}_{sp,b}\|_F^2$, with $\mathbf{\Sigma}_{sp,b}$ being the sample covariance matrix of S_b. The expected value of the quadratic loss function is then estimated by

$$\hat{\mathbb{E}}\left[\|\mathbf{\Sigma}_{sp} - \mathbf{\Sigma}\|_F^2\right] = \frac{1}{B} \sum_{b=1}^{B} \|\mathbf{\Sigma}_{sp} - \mathbf{\Sigma}_{sp,b}\|_F^2 .$$

When using the multivariate version of the smoothed bootstrap, one simply updates each drawn (sub-) sample $r_{b_t} = (r_{t,b}^1, \dots, r_{t,b}^n)$ of S_b to

$$r_{b_t}^* = \mu_{sp,b} + (\mathbb{I} + \mathbf{\Sigma}_Z)^{-1/2} \left[r_{b_t} - \mu_{sp,b} + \mathbf{\Sigma}_{sp,b}^{1/2} Z_{b_t} \right], \tag{44}$$

where $\mu_{sp,b}$ and $\mathbf{\Sigma}_{sp,b}$ are the sample estimates of the first two moments based on S_b and Z_b is a multivariate normal random variable with zero mean and covariance matrix $\mathbf{\Sigma}_{sp,b}$ (therefore $\mathbf{\Sigma}_Z = \mathbf{\Sigma}_{sp,b}$). This modification implies that we are now sampling observations from a continuous density function and hence, the probability of repeated observations is zero. This result is very appealing, since many repeated observations are likely to lead to singularity of the estimated covariance matrix.
We summarize the "vanilla" bootstrap in Algorithm 1 and the smoothed bootstrap in Algorithm 2, respectively.

Algorithm 1: Bootstrap for Expected Loss of Sample Covariance Matrix

input : $\{r_t\}_{t=1}^T$, $\mathbf{\Sigma}_{sp}$, $B \in \mathbb{N}$
output: Estimation for $\mathbb{E}\left[\|\mathbf{\Sigma}_{sp} - \mathbf{\Sigma}\|_F^2\right]$
for *b=1,...,B* **do**
  $S_b \leftarrow$ sample of T return vectors sampled from $\{r_t\}_{t=1}^T$ with replacement
  $\mathbf{\Sigma}_{sp,b} \leftarrow$ covariance matrix based on sample S_b

return $\frac{1}{B}\sum_{b=1}^{B}\|\mathbf{\Sigma}_{sp} - \mathbf{\Sigma}_{sp,b}\|_F^2$

As mentioned above, in the smoothed bootstrap we additionally update each drawn (sub-) sample.

Algorithm 2: Smoothed Bootstrap for Expected Loss of Sample Covariance Matrix

input : $\{r_t\}_{t=1}^T$, $\mathbf{\Sigma}_{sp}$, $B \in \mathbb{N}$
output: Estimation for $\mathbb{E}\left[\|\mathbf{\Sigma}_{sp} - \mathbf{\Sigma}\|_F^2\right]$
for *b=1,...,B* **do**
  $S_b \leftarrow$ sample of T return vectors sampled from $\{r_t\}_{t=1}^T$ with replacement
  $\mathbf{\Sigma}_{sp,b} \leftarrow$ covariance matrix based on sample S_b
  $Z_{b_t} \leftarrow$ realization of a multivariate random variable with mean zero and covariance matrix $\Sigma_{sp,b}$
  $r^*_{b_t} \leftarrow \mu_{sp,b} + (\mathbb{I} + \mathbf{\Sigma}_Z)^{-1/2}\left[r_{b_t} - \mu_{sp,b} + \mathbf{\Sigma}_{sp,b}^{1/2} Z_{b_t}\right]$
  $\mathbf{\Sigma}^*_{sp,b} \leftarrow$ covariance matrix based on $(r^*_{b_t})_{t=1}^T$

return $\frac{1}{B}\sum_{b=1}^{B}\left\|\mathbf{\Sigma}_{sp} - \mathbf{\Sigma}^*_{sp,b}\right\|_F^2$

We obtain an approximation of the convexity parameter α_Σ by plugging our above (smoothed) results into equation (43), so

$$\hat{\alpha}_\Sigma = \frac{\frac{1}{B}\sum_{b=1}^{B}\left\|\mathbf{\Sigma}_{sp} - \mathbf{\Sigma}^*_{sp,b}\right\|_F^2}{\frac{1}{B}\sum_{b=1}^{B}\left\|\mathbf{\Sigma}_{sp} - \mathbf{\Sigma}^*_{sp,b}\right\|_F^2 + \|\overline{\sigma^2}\mathbb{I} - \mathbf{\Sigma}_{sp}\|_F^2}. \tag{45}$$

Finally, our shrinkage estimator of $\mathbf{\Sigma}$ is given by $\mathbf{\Sigma}_{sh} := (1 - \hat{\alpha}_\Sigma)\mathbf{\Sigma}_{sp} + \hat{\alpha}_\Sigma\overline{\sigma^2}\mathbb{I}$.

6.2 Uncertainty Cones of Moments

In order to robustify our portfolio optimization problem against estimation error of the asset-return moments, in Section 5.3 we assumed the true moments μ and $\mathbf{\Sigma}$ to be in a convex uncertainty set $\mathcal{U}$ of the form

$$\mathcal{U} = \left\{ (\mu, \mathbf{\Sigma}) \in \mathbb{R}^n \times \mathbb{S}^n : (\mu - \hat{\mu})' \hat{\mathbf{\Sigma}}^{-1} (\mu - \hat{\mu}) \leq \delta_1,\ \delta_3 \hat{\mathbf{\Sigma}} \preceq \mathbf{\Sigma} \preceq \delta_2 \hat{\mathbf{\Sigma}} \right\}, \tag{46}$$

with δ_1, δ_2, $\delta_3 \in \mathbb{R}_+$ and $\delta_2 \geq 1 \geq \delta_1$.

Of course, we select the center $(\hat{\mu}, \hat{\mathbf{\Sigma}})$ to be our point estimations μ_{sh} and $\mathbf{\Sigma}_{sh}$, derived in the previous section. Naturally, we cannot find reasonable δ_1, δ_2 and δ_3 for which the true moments lie in $\mathcal{U}$ with probability equal to one. Hence, our task is now to derive such δ_1, δ_2, $\delta_3 \in \mathbb{R}_+$ for which $(\mu, \mathbf{\Sigma}) \in \mathcal{U}$ with a high level of confidence, say $1 - \delta$. In fact, we can immediately choose $\delta_3 = 0$, since in our application we only use the worst case scenario of $\mathbf{\Sigma}$, therefore only the upper bound $\delta_2 \hat{\mathbf{\Sigma}}$ is needed.

Our approach will be based on a bootstrapping technique for estimating thresholds for hypothesis-testing presented by [Bertsimas et al., 2013]. In their paper, Bertsimas et al. interpret the analytical approach of [Delage and Ye, 2010] for deriving such an uncertainty set $\mathcal{U}$, as a hypothesis-test with the null hypothesis being

$$H_0 : \mathbb{E}[\tilde{r}] = \mu_{sh} \quad \text{and} \quad \mathbb{E}[\tilde{r}\tilde{r}'] - \mathbb{E}[\tilde{r}]\mathbb{E}[\tilde{r}'] = \mathbf{\Sigma}_{sh},$$

which is, using our notation, equivalent to

$$H_0 : \mu = \mu_{sh} \quad \text{and} \quad \mathbf{\Sigma} = \mathbf{\Sigma}_{sh}. \tag{47}$$

In order to estimate δ_1 and δ_2, we define two test-statistics T^1 and T^2 by

$$\begin{aligned} T^1(\hat{\mu}, \hat{\mathbf{\Sigma}}) &:= (\hat{\mu} - \mu_{sh})' \hat{\mathbf{\Sigma}}^{-1} (\hat{\mu} - \mu_{sh}) \\ T^2(\hat{\mu}, \hat{\mathbf{\Sigma}}) &:= \max_{\lambda \in \mathbb{R}^n} \frac{\lambda' \mathbf{\Sigma}_{sh} \lambda}{\lambda' \hat{\mathbf{\Sigma}} \lambda}, \end{aligned}$$

where $\hat{\mu}$ and $\hat{\mathbf{\Sigma}}$ denote estimations of μ and $\mathbf{\Sigma}$ based on another sample of observed asset-returns. Of course, we again use the presented shrinkage estimation methods for $\hat{\mu}$ and $\hat{\mathbf{\Sigma}}$ and the "other sample" will be a bootstrap sample which we gain by sampling with replacement from the real observed asset-returns.

The bootstrap method of Bertsimas et al. is then to draw B ($\in \mathbb{N}$) samples of size T with replacement from the observed asset-returns $\{r_t\}_{t=1}^T$. For each of these samples S_b, where $b = 1, \ldots, B$, we then compute $\hat{\mu}_b$, $\hat{\mathbf{\Sigma}}_b$ and T_b^1 or T_b^2 respectively. We then simply estimate δ_i by the $\lceil B(1 - \delta) \rceil$-largest value of $\{T_1^i, \ldots, T_B^i\}$, where $i \in \{1, 2\}$.

For the purpose of readability, we summarize this bootstrap approach in Algorithm 3.

Algorithm 3: Bootstrapping Moment Thresholds

---

input : $\{r_t\}_{t=1}^T$, T^i, $\delta \in (0,1)$, $B \in \mathbb{N}$, $i \in \{1,2\}$
output: Estimation for δ_i
for $b = 1, \ldots, B$ **do**
  $S_b \leftarrow$ sample of T return vectors sampled from $\{r_t\}_{t=1}^T$ with replacement
  $T_b^i \leftarrow T^i(\hat{\mu}_b, \hat{\mathbf{\Sigma}}_b)$ where $\hat{\mu}$ and $\hat{\mathbf{\Sigma}}_j$ are shrinkage estimators of mean and covariance matrix based on S_b
return $\lceil B(1-\delta) \rceil$-*largest value of* $\{T_1^i, \ldots, T_B^i\}$

Note that while $T^1(\hat{\mu}, \hat{\mathbf{\Sigma}})$ can be computed immediately by its definition, $T^2(\hat{\mu}, \hat{\mathbf{\Sigma}})$ is the maximum over all $\lambda \in \mathbb{R}^n$ of a quotient. In order to calculate this maximum, we first compute the square root of the positive definite matrix $\hat{\mathbf{\Sigma}}$, so $\hat{\mathbf{\Sigma}} = \hat{\mathbf{\Sigma}}^{1/2}\hat{\mathbf{\Sigma}}^{1/2}$. By substituting $\hat{\mathbf{\Sigma}}$ by this decomposition and also plugging it in the enumerator, and since all the matrices are symmetric, we get the following expression for our test-statistics.

$$T^2(\hat{\mu}, \hat{\mathbf{\Sigma}}) = \max_{\lambda \in \mathbb{R}^n} \frac{\lambda' \mathbf{\Sigma}_{sh} \lambda}{\lambda \hat{\mathbf{\Sigma}} \lambda} = \max_{\lambda \in \mathbb{R}^n} \frac{(\hat{\mathbf{\Sigma}}^{1/2}\lambda)' \hat{\mathbf{\Sigma}}^{-1/2} \mathbf{\Sigma}_{sh} \hat{\mathbf{\Sigma}}^{-1/2} (\hat{\mathbf{\Sigma}}^{1/2}\lambda)}{(\hat{\mathbf{\Sigma}}^{1/2}\lambda)'(\hat{\mathbf{\Sigma}}^{1/2}\lambda)}$$

This is nothing but the Rayleigh-Quotient of the matrix $\hat{\mathbf{\Sigma}}^{-1/2}\mathbf{\Sigma}_{sh}\hat{\mathbf{\Sigma}}^{-1/2}$. Since this matrix is symmetric and we are maximizing, this equals the biggest eigenvalue. We can therefore compute $T^2(\hat{\mu}, \hat{\mathbf{\Sigma}})$ by

$$T^2(\hat{\mu}, \hat{\mathbf{\Sigma}}) = \lambda_{max}\left(\hat{\mathbf{\Sigma}}^{-1/2}\mathbf{\Sigma}_{sh}\hat{\mathbf{\Sigma}}^{-1/2}\right). \tag{48}$$

Algorithm 3 will deliver estimations $\hat{\delta}_1$ and $\hat{\delta}_2$, which we use to finally determine the convex uncertainty set $\hat{\mathcal{U}}$, in which we assume the true moments μ and $\mathbf{\Sigma}$ lie with probability of at least $(1-\delta)$;

$$\hat{\mathcal{U}} = \left\{(\mu, \mathbf{\Sigma}) \in \mathbb{R}^n \times \mathbb{S}^n : (\mu - \mu_{sh})' \mathbf{\Sigma}_{sh}^{-1} (\mu - \mu_{sh}) \leq \hat{\delta}_1,\ 0 \preceq \mathbf{\Sigma} \preceq \hat{\delta}_2 \mathbf{\Sigma}_{sh}\right\}. \tag{49}$$

7 Numerical Experiments

In this Section we want to apply our derived portfolio theory to real data. We will compare the performance of investment strategies induced by the different portfolio selection methods derived in Section 5.4. In addition, we will compare the results to the equally weighted portfolio, which is known to be hard to outperform (see [DeMiguel et al., 2009]), and the "classical" (Markowitz) minimum variance portfolio we introduced in Section 3.2. We hence compare the following portfolios:

- **(RGOP) Robust Growth Optimal Portfolio:**

  This is the portfolio derived in [Rujeerapaiboon et al., 2014], on which our approach is based. As mentioned in Section 5.4.1, we obtain this portfolio by choosing the risk aversion parameter γ equal to one and solving the corresponding Program (30).

- **(RGOP+) Robust Growth Optimal Portfolio with Moment Uncertainty:**

  This is the portfolio we obtain by choosing γ equal to one and robustifying our moment estimations, i.e. solving Program (32).

- **(RRAP) Robust Risk Averse Portfolio:**

  This is the portfolio we derived in Section 5.4.2. It is the most risk averse portfolio in our framework and is obtained by choosing γ equal to the technical lower bound γ^*, where γ^* is defined by Problem (34), and solving Program (30).

- **(RRAP+) Robust Risk Averse Portfolio with Moment Uncertainty:**

  Similar to the above robust risk averse portfolio but with robustification against moment uncertainty, i.e. the optimal solution of Program (32) with γ equal to γ^*.

- **(MVAR) Minimum Variance Portfolio:**

  This is the "classical" Markowitz minimum variance portfolio which we reviewed in Section 3.2. It is the efficient portfolio with the smallest variance and can be achieved for example by solving Problem (6) for R equal to the smallest expected asset-return.

- **(1/n) Equally Weighted Portfolio:**

  This portfolio simply weights every asset with $1/n$, so $w_1 = \frac{1}{n}\mathbf{1}$, where n is the number of considered assets.

From a theoretical point of view, we would expect the portfolios which consider moment uncertainty, i.e. RRAP+ and RGOP+, to be more conservative than their "vanilla" counterpart. As the RRAPs are by definition the most risk averse under all robust portfolios and the same holds for the MVAR-portfolio in the set of all mean-variance efficient portfolios, we expect a similar but not identical behaviour of these strategies. In contrast, the RGOPs are the only considered portfolios which do not minimize the risk in some sense. Therefore, it is most likely that they differ most from the other considered strategies.

To compare these portfolio selection methods, we define investment strategies by updating the portfolio weights every 12 months and rebalancing the corresponding portfolio accordingly. This means investment strategy RRAP is given by solving Problem (30) calibrated with moment estimates based on the information available at time t=0 and $\gamma = \gamma^*$. This portfolio is then kept for 12 months. After 12 months, or at t=1, Problem (30) is again solved with $\gamma = \gamma^*$, using new estimates based on the available information at t=1. The initial portfolio is then redeployed to match the new portfolio weights and so on. To account this redeployment, we also consider proportional transaction costs of $c = 50$ basis points per Euro traded. The parameter ϵ, which defines the confidence level of the worst-case value-at-risk of our a priori performance measure $\tilde{\nu}_T^{\gamma}(w)$ $(WVaR_{\epsilon}(\tilde{\nu}_T^{\gamma}(w))$, is chosen as $\epsilon = 5\%$.

In Section 7.2 we will compare the performances of the several strategies in different scenarios by using the performance measures presented in Section 7.1. A scenario is thereby defined by the considered set of assets and the time horizon (starting date and number of years) for which the capital is invested. Scenario 1 in Section 7.2.1 consists of the investment period December 2005 to December 2010 and a set of assets which all experienced a price collapse at the outbreak of the global financial crisis in the year 2008. On the other hand, Scenario 2 in Section 7.2.2 also covers the same investment period but includes assets which either did not show any or only modest reaction to the global financial crisis. For the purpose of easy visualization of the portfolio weight assignments we only consider small sets of assets in these two scenarios. In contrast, Scenario 3 in Section 7.2.3 includes all 30 assets of the Dow Jones Industrial Average.

We will also look at the actual influence the risk aversion parameter γ has on the chosen portfolio in practice. As mentioned above, for γ small, i.e. close to the lower bound γ^* defined by Equation (34), we would expect the resulting portfolio to be "near" to the MVAR-portfolio. On the other hand, as γ grows, we expect the resulting asset-weight distribution to differ more from the minimum variance portfolio.

As mentioned above, we consider proportional transaction costs. Hence, excessive redeployment of a portfolio will result in smaller portfolio returns. We will therefore also consider the "evolution" of our portfolios over time in order to determine if a strategy causes a lot of redeployment costs. Naturally we expect the equally weighted portfolio 1/n to cause the least redeployment costs.

For all strategies, the moment and parameter estimations are done with the methods presented in Section 6, using the most recent 60 ($\hat{=}$ five years) monthly observations available at the respective moment in time. As we have seen, for estimating the covariance matrix and the moment uncertainty cone, we use a bootstrapping procedure, for which we choose the number of iterations $B = 500$. The uncertainty cone around the shrinkage point estimations μ_{sh} and $\mathbf{\Sigma}_{sh}$ is determined by the method presented in Section 6.2, where we choose $\delta = 5\%$.

7.1 Performance Measures

To compare the performance of the different investment strategies, we need some measures to quantify the observed behaviours. Let $w_t \in \mathbb{R}^n$ denote a portfolio kept in time period $(t-1, t]$ and $r_t \in \mathbb{R}^n$ the realized asset-returns over this interval, where $t = 1, \ldots, T$. Note that for all periods the portfolio weights w_t sum up to one, which means that all of the capital is invested and we do not consider a cash position. We assume proportional transaction costs of $c = 50$ basis points per Euro traded. Obvious measures are of course the mean return and the standard deviation of the realized portfolio returns which when combined, result in the so-called Sharpe Ratio. We are also interested in the turnover rate and maximum draw-down, describing the relative amount of redeployed capital and the biggest relative decline of aggregate return, respectively.
We summarize the used performance measures in the following list:

1. **Mean Return:**

   We simply take the mean of the realized portfolio returns, where we subtract the costs for the portfolio redeployment at the dates $t = 1, \ldots, T$.

$$\hat{r} := \frac{1}{T} \sum_{t=1}^{T} \left((1 + w_t' r_t) \left(1 - c \sum_{i=1}^{n} |w_t^i - w_{t_-}^i| \right) - 1 \right).$$

   Here, w_{t_-} denotes the portfolio weights of portfolio w_{t-1} at time t, where $w_0 := 0$ accommodates the fact that nothing is invested at time $t = 0$. Obviously, we prefer a high mean return.

2. **Standard Deviation:**

   As this is the most intuitive measure for the volatility of our portfolio returns, we prefer strategies with small standard deviation.

$$\hat{\sigma} := \sqrt{\frac{1}{T-1} \sum_{t=1}^{T} \left((1 + w_t' r_t) \left(1 - c \sum_{i=1}^{n} |w_t^i - w_{t_-}^i| \right) - 1 - \hat{r} \right)^2}.$$

3. **Sharpe Ratio:**

   The Sharpe Ratio combines the two notions of high return and small volatility (variance). It is given by

$$\widehat{SR} := \frac{\hat{r}}{\hat{\sigma}},$$

   where we considered the "risk free" reference interest rate equal to zero.

4. **Turnover Rate:**

   The turnover rate simply accumulates the differences in consecutive portfolios of one strategy. As redeployment is charged by c basis points per Euro traded, we prefer strategies with a small turnover rate. It is given by

$$\widehat{TR} := \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{n} |w_t^i - w_{t_-}^i|.$$

5. **Net Aggregate Return:**

   The net aggregate return describes the worth at the end of the considered time interval (t=T) of one Euro invested at time $t = 0$ in the portfolio strategy. By using

$$\hat{V}_t := \prod_{j=1}^{t} \left[(1 + w_j' r_j) \left(1 - c \sum_{i=1}^{n} |w_t^i - w_{t_-}^i| \right) \right],$$

   where $\hat{V}_t$ describes the value at time t, the net aggregate return is given by

$$\widehat{NR} := \hat{V}_T.$$

6. **Maximum Draw-down:**

   The maximum draw-down describes the biggest relative loss we have experienced over the considered time interval. By using the notation above, it is given by

$$\widehat{MDD} := \max_{1 \leq s < t \leq T} \frac{\hat{V}_s - \hat{V}_t}{\hat{V}_s}.$$

Of course, only in very rare cases will there be one strategy dominating the others in the sense of all the above measures. Therefore, in Section 7.2, we will compare all the performance measures of every strategy with the others.

7.2 Empirical Backtests

7.2.1 Scenario 1: A fragile Market during the Global Financial Crisis

In order to provide traceability and easy visualization, our first scenario consists of seven assets to invest in. The investment horizon is chosen to be December 2005 to December 2010. We therefore use the monthly observations of the asset-returns from December 2000 to December 2005 to get the initial estimates for the first- and second-order moments μ and $\mathbf{\Sigma}$ and the uncertainty cone parameters δ_1 and δ_2. Using these estimates, we can determine the initial portfolios by solving the corresponding optimization problems. These portfolios are held for 12 months, consequently we set the time horizon $T = 12$ accordingly. This routine is repeated every 12 months for five consecutive years using a rolling five year observation window for the estimation. Due to the portfolio recalculation in December, the resulting investment strategies are only fixed-mix strategies for the one year periods in-between the rebalancing dates.

The set of considered assets consists of five American and two European stocks, listed in Table 2. Note that these corporations are chosen to be from very different economic sectors as strongly correlated assets would not facilitate the display of differences in the portfolio strategies.

Corporation	Symbol	Exchange	Description
Coca-Cola Company	KO	New York	Beverage corporation and manufacturer.
Procter & Gamble	PG	New York	Consumer goods company.
Exxon Mobil	XOM	New York	Oil and gas corporation.
Pfizer	PFE	New York	Pharmaceutical corporation.
United Health Group	UNH	New York	Health care company.
L'Oreal	OR.PA	Paris	Cosmetics and beauty company.
Iberdrola	IBE.MC	Madrid	Electric utility company.

Table 2: Scenario 1: List of considered stocks

The adjusted closing prices[4] for the relevant period of time, which are obtained from Yahoo Finance[5], are displayed in Figure 9. The rebalancing dates are marked by the dotted vertical lines. Again, note that for updating the portfolio weights at the rebalancing dates, only the most recent five years to the "left" of the rebalancing date are used.
We immediately observe obvious differences in the performances of the different stock prices. Where the prices of the Spanish Iberdrola stock seem to be fairly stable, the L'Oreal stock varies a lot more. These differences become particularly interesting when we investigate their influence on the considered portfolios.

[4] The actual stock closing price amended to include dividends, stock splits and other corporate actions.
[5] http://finance.yahoo.com

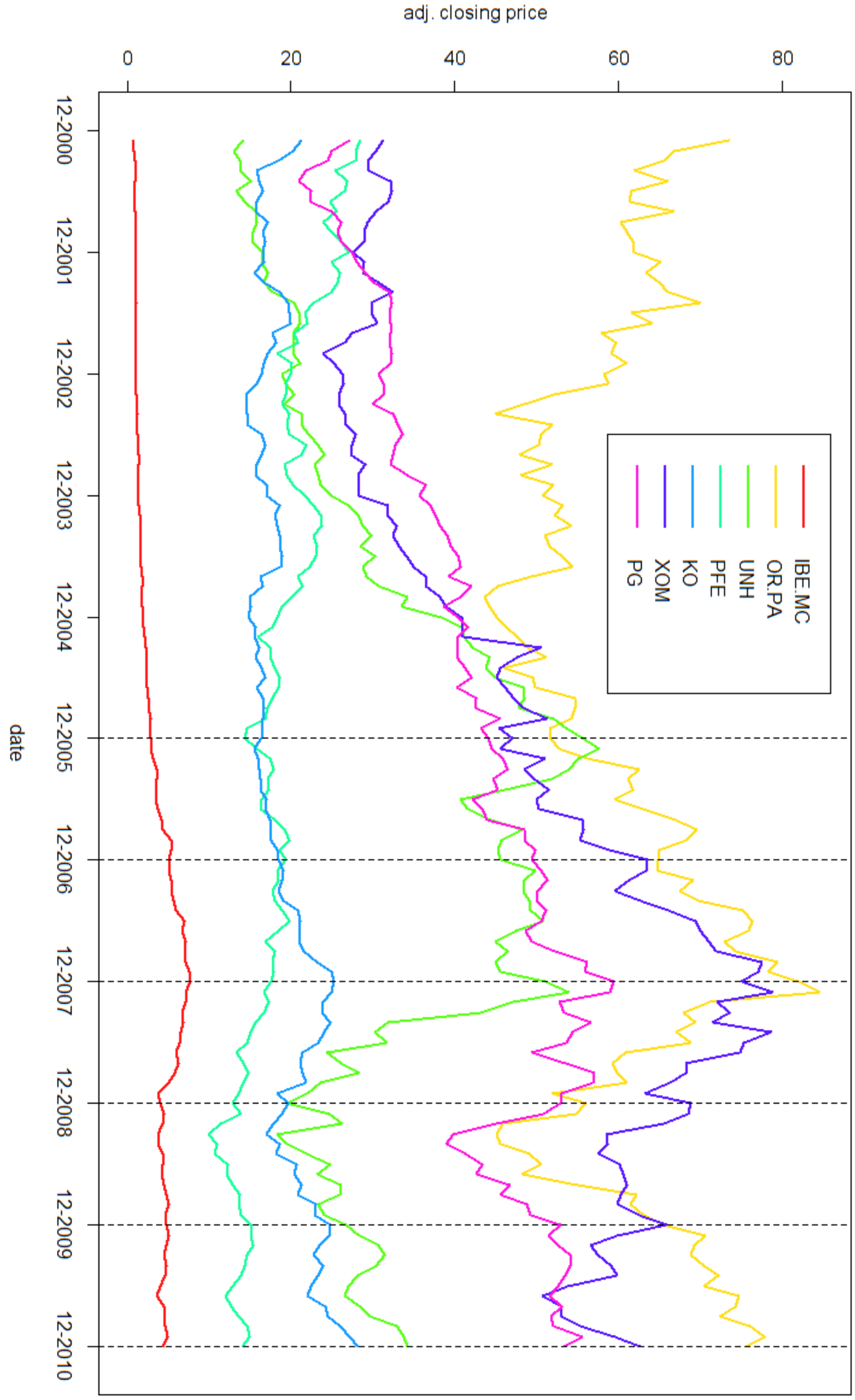

Figure 9: Scenario 1: Adjusted closing prices

First, we want to have a look at the initial portfolios. They are visualized in Figure 10, where we neglected to display the equally weighted portfolio since its weight assignment is obvious and never changes. We observe that of all robust portfolios, the RRAP+ visually (and numerically, in the sense of an arbitrary vector norm) is the closest to the classic (Markowitz) minimum variance portfolio MVAR. In this sense of distance, also the RRAP is closer to the MVAR portfolio than the RGOP and RGOP+. This is an expected pattern, since the RRAP(+) and MVAR portfolios both minimize the risk in some sense, whereas the growth optimal portfolios focus on the growth rate. Note that in this example the RRAP and RRAP+ can only be distinguished numerically.
As moment ambiguity states an additional source of risk, its consideration results in even more risk averse portfolios. This can be particularly observed when looking at the RGOP+, which somehow lies between the RGOP and the risk averse portfolios.

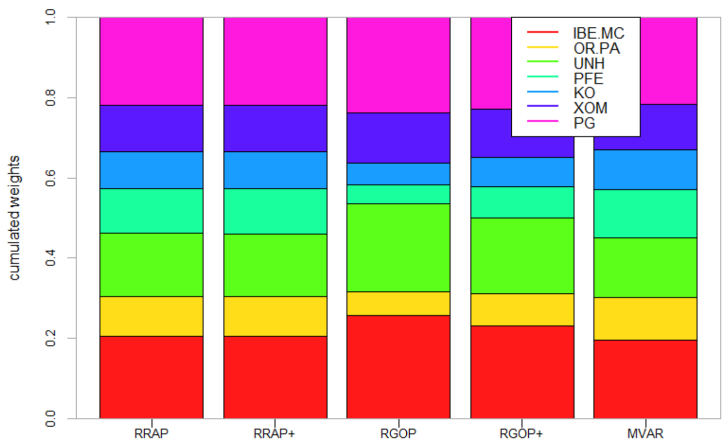

Figure 10: Scenario 1: Initial portfolios

To further interpret these different portfolio weights, in Figure 11 we display the box-plots of the asset-returns observed over the time horizon relevant for the initial portfolios, i.e. December 2000 to December 2005. The median is depicted by the bold horizontal line. We also marked the shrinkage estimators for the mean, as presented in Section 6, by bold black points. The coloured boxes mark the lower and upper quartiles of the returns and the dashed "whiskers" indicate the most extreme observations as long as their distance to the median does not exceed 1.5 times the interquartile length, otherwise these "outliers" are marked by circles.

As expected, we see that for all portfolio strategies the highest weights are observed for the assets with the highest mean return, i.e. the best performing assets from the past. We can see in Figure 10 that the biggest differences between the initial portfolios are observed for the assets IBE.MC and PFE. While the growth optimal portfolios RGOP and RGOP+ favour the higher mean return of IBE.MC, the risk averse portfolios RRAP(+) and MVAR invest more in PFE, although it has a negative expected return. Note that from the boxplot we can not identify the correlation between the assets, and therefore the advantages of choosing PFE instead of IBE.MC in order to minimize the portfolio volatility is not displayed.

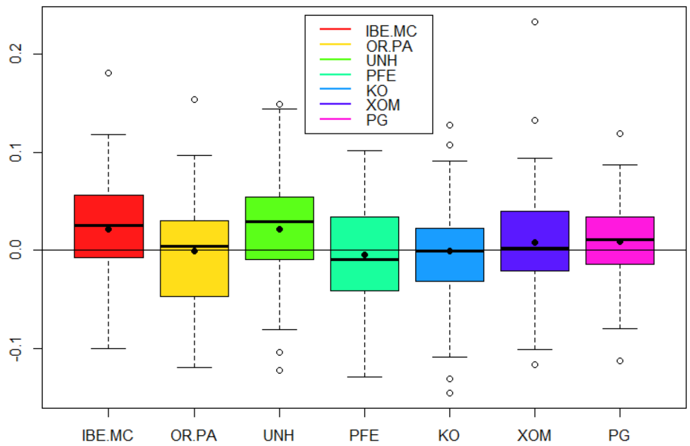

Figure 11: Scenario 1: Boxplots of initial return-history

Next, we want to have a look at the impact of the risk aversion parameter γ on the resulting portfolios. For this purpose, in Figure 12 we display the portfolio weights that we obtain for different choices of γ when solving Problem (30), which is the distributionally robust portfolio optimization problem without moment ambiguity. For $\gamma = \gamma^*$, where in this example $\gamma^* = 0.02477$ as defined by Equation (34), and $\gamma = 1$ we obtain the RRAP and RGOP, respectively.

We observe that for higher, and therefore less risk averse choices of γ, the assets with high expected returns dominate the portfolios. It also seems that for very big γ, the resulting portfolio converges towards some “limit”-portfolio. For γ between γ^* and one, we see the biggest changes in asset weights.

We want to compare these results to the portfolios that take the moment ambiguity into account. In Figure 13 we again display the portfolio weights for different choices of γ, this time for the distributionally robust portfolio optimization problem with moment ambiguity (Problem (32)). We can observe a similar pattern as in Figure 12. For increasing γ, the changes in asset weights decrease. In contrast to the portfolios without moment ambiguity, the changes in asset weights are smaller and the “limit”-portfolio seems to be more diversified. This is due to the fact that all of these portfolios are more risk averse than their “vanilla” counterparts.

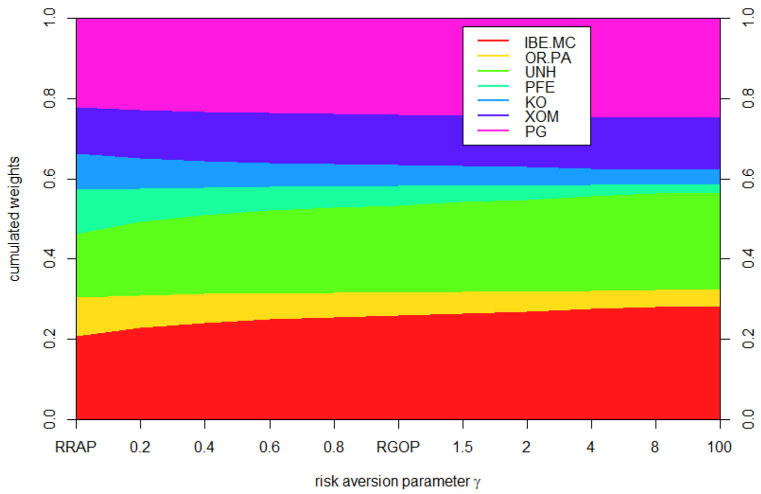

Figure 12: Scenario 1: Initial portfolios without moment ambiguity dependent on γ

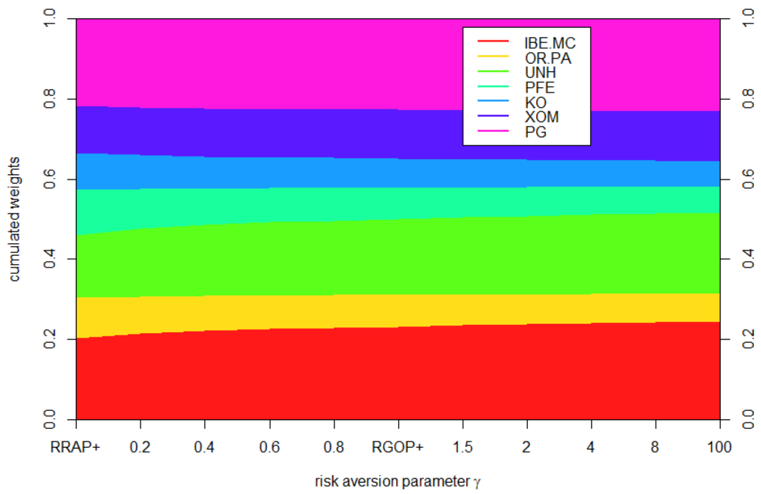

Figure 13: Scenario 1: Initial portfolios with moment ambiguity dependent on γ

As the overall aggregated return is reduced by transaction costs, it is worth having a look at the evolution of the asset-weights over the investment horizon. An excessive

change in portfolio weights will result in high transaction costs at the rebalancing dates in December. Therefore, we would prefer a strategy with less required redeployment. The asset-weights evolution is presented for the robust risk averse and growth optimal strategies in Figures 14 and 15, respectively.
From these figures we see that the slopes of the cumulated RGOP weights are more extreme than those of the RRAP weights. This simply means that, in this scenario, the RGOP strategy causes more transaction costs than the RRAP strategy. It is also worth mentioning that our observations for the initial portfolios, which we stated above, stay true at all rebalancing dates. Under all robust portfolios, the RRAP+ is the closest to the MVAR portfolio. Also, at all rebalancing dates the portfolios with moment ambiguity are closer to the minimum variance portfolio than the ones without.

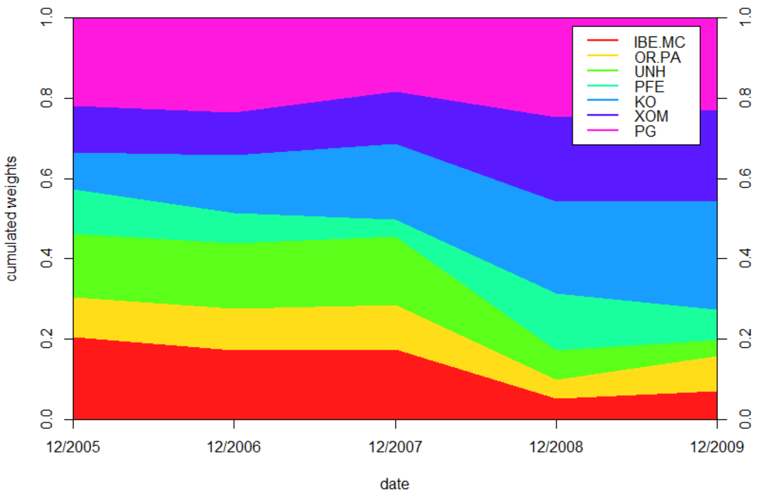

Figure 14: Scenario 1: Asset-weights evolution of the RRAP

Finally, we want to a posteriori compare the investment strategies by using the performance measures introduced in Section 7.1. An investment strategy is said to dominate another strategy if for every considered measure its performance is better.
The results are presented in Table 3, where the figures refer to the monthly portfolio-returns that realized in the considered investment period December 2005 to December 2010.

As we have mentioned above, the equally weighted portfolio strategy is surprisingly hard to outperform, which is also documented in [DeMiguel et al., 2009]. In our example, it outperforms all other strategies in every performance measure.
We also observe that in this scenario, the minimum variance portfolio performed better than the robust portfolios. This is due to the fact that from 2007 to 2009 (the global financial crisis) all of the stock prices experienced a major drawback and consequently risk averse strategies performed better. This behaviour can also be observed if we look at the

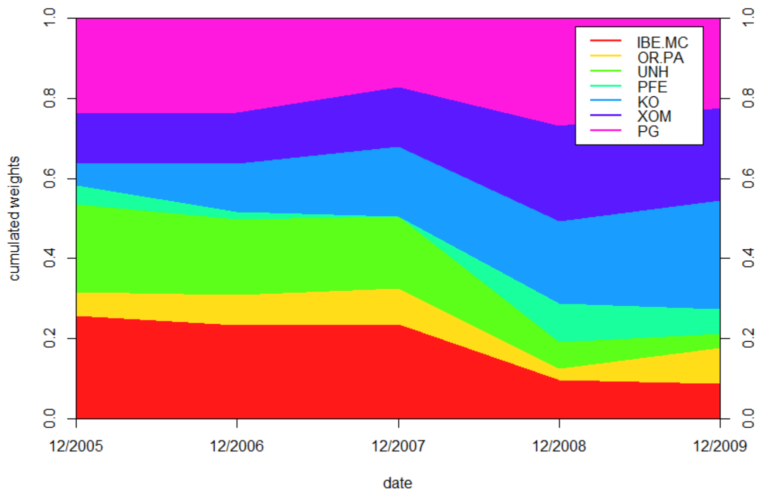

Figure 15: Scenario 1: Asset-weights evolution of the RGOP

	$\hat{r}$	$\hat{\sigma}$	$\widehat{SR}$	$\widehat{TR}$	$\widehat{NR}$	$\widehat{MDD}$
RRAP	0.00549	0.04530	0.12124	0.07064	1.30604	0.42044
RRAP+	0.00550	0.04528	0.12154	0.07052	1.30699	0.42016
RGOP	0.00538	0.04612	0.11670	0.07392	1.29442	0.42611
RGOP+	0.00543	0.04575	0.11868	0.07249	1.29950	0.42443
MVAR	0.00552	0.04517	0.12223	0.06999	1.30881	0.41902
1/n	0.00585	0.04499	0.13004	0.05570	1.33554	0.41135

Table 3: Scenario 1: Performance measures of strategies

robust portfolios. It is remarkable that both strategies with moment ambiguity dominate their "vanilla" counterparts. Also the risk averse strategies RRAP and RRAP+ showed a better performance than the growth optimal portfolios RGOP and RGOP+.
From Figures 14 and 15, which compared the asset-weights evolution, we concluded that the RGOP strategy will cause more transaction costs than the RRAP strategy. This is confirmed by the higher turnover rate that the robust growth optimal strategies produced.

In summary, we have seen a setting in which risk averse strategies performed better than others. This was also observed within the set of robust portfolios, where the risk averse dominated the growth optimal strategies. From all considered strategies, the minimum variance was only outperformed by the equally weighted portfolio. In some sense, the robust strategies "followed" the performance of the minimum variance portfolio.
We will now have a look at another Scenario to see if we can confirm these patterns.

7.2.2 Scenario 2: A robust Market during the Global Financial Crisis

In contrast to Scenario 1, we will now consider a set of assets for which some of the stock prices did not show such a negative reaction to the global financial crisis. The investment horizon is again chosen to be December 2005 to December 2010. We also use exactly the same investment strategies, e.g. the parameters for the respective portfolio optimization problems are estimated using a rolling five year observation window and the portfolios are rebalanced each December for consecutive five years.

This time, the set of considered assets consists of British stocks, which are again chosen to be from different economic sectors and are listed in Table 4.

Corporation	Symbol	Exchange	Description
Babcock International Group	BAB.L	London	Multinational corporation which specialises in support service managing complex assets and infrastructure in safety-critical environments.
British American Tobacco	BATS.L	London	One of the worlds five largest tobacco companies.
GlaxoSmithKline	GSK.L	London	A pharmaceutical, biologics, vaccines and consumer healthcare company.
Kingfisher	KGF.L	London	The largest home improvement retailer in Europe.
United Utilities Group	UU.L	London	The United Kingdom's largest listed water company.
Vodafone Group	VOD.L	London	Multinational telecommunications company.

Table 4: Scenario 2: List of considered stocks

We obtain the adjusted closing prices[6] from Yahoo Finance[7], where the listed symbols in Table 4 state the corresponding ticker symbols.
From Figure 16 we can see that the price movements show another behaviour than in the previous Scenario 1. There are no major drawbacks, and all of the prices at the end of 2010 are almost the same or above the prices from December 2005. We also recognise that the BATS.L stock realized by far the best performance in the sense of aggregated return, while other stock prices, especially VOD.L, relatively showed not as much movement.
The rebalancing dates are again marked by the dotted lines.

[6]The actual stock closing price amended to include dividends, stock splits and other corporate actions.
[7]http://finance.yahoo.com

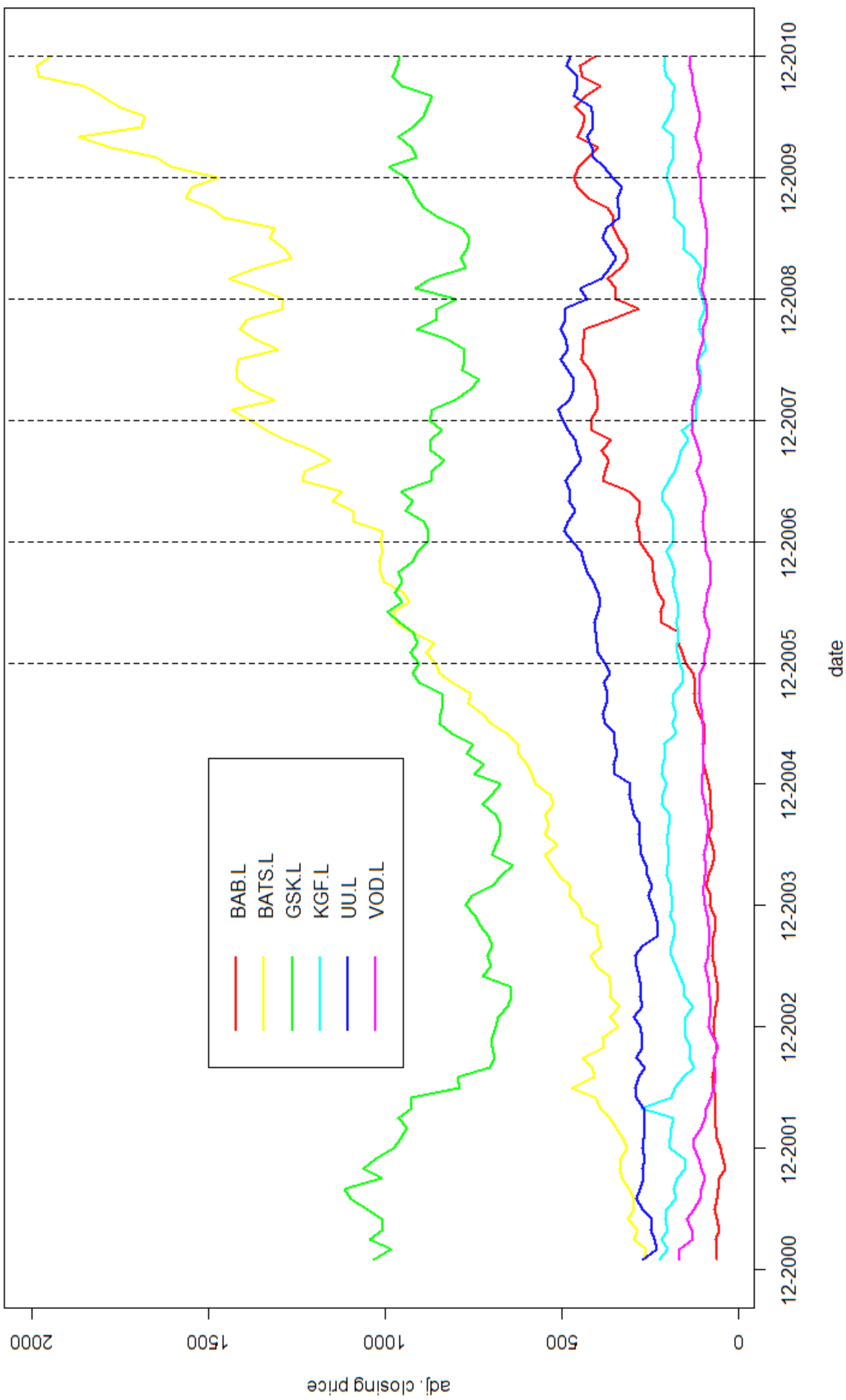

Figure 16: Scenario 2: Adjusted closing prices

If we look at the initial portfolios in Figure 17, we observe a similar pattern as we have seen in Figure 10 for Scenario 1. The robust risk averse portfolios RRAP and RRAP+ are again very “close” to the MVAR portfolio. On the other side, the weight-assignments of the RGOP differs the most from the others, where the RGOP+ again lies somewhere “in between” the robust risk averse and the growth optimal portfolios. This also holds for the RRAP+ (which can not be seen visually but can be verified numerically) and therefore, as in the previous Scenario 1, both robust portfolios with moment ambiguity are closer to the MVAR portfolios than those without moment ambiguity. We also note that this time, the differences between the initial asset-weights of the different strategies are even smaller than they were in Scenario 1.

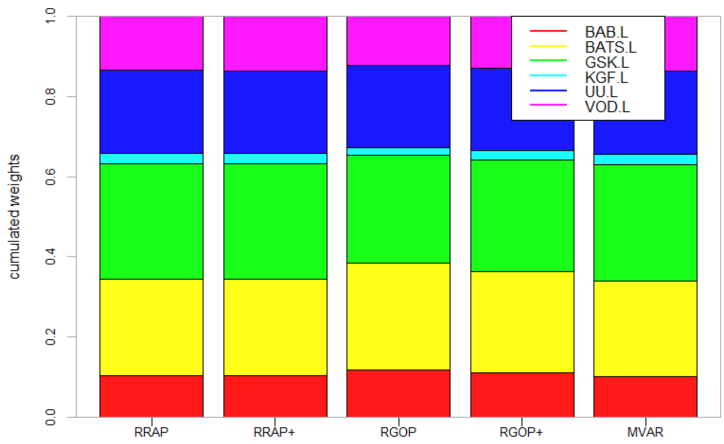

Figure 17: Scenario 2: Initial portfolios

In Figure 18 we see the boxplots of the monthly asset-returns realized in the time period December 2000 to December 2005. The shrinkage estimators for the means, which are used for all portfolio selections, are again marked by bold black points. We see that all strategies neglect to choose KDF.L due to its high volatility (many outliers) compared to the other assets. We also observe that BATS.L is preferred to BAB.L, as both show almost the same estimated mean but the latter has smaller variance. In contrast to the risk averse strategies, the growth optimal portfolios put more weight on BATS.L and less on VOD.L because of the higher expected return.

Based on the weight-assignments in Figure 17, the stocks of BATS.L and GSK.L seem to be the most attractive ones. If we look at the first five years of the price history in Figure 16, the reasons for this preferences are quite obvious. As mentioned above, the prices of the BATS.L performed best in the sense of mean return and also have an attractive, i.e. small, variance. The choice of GSK.L seems counterintuitive at first glance, since the stock lost value over the first five years (2000-2005). Nevertheless, it is chosen by all strategies because it did not behave like the other assets, and therefore helps to minimize the expected portfolio variance.

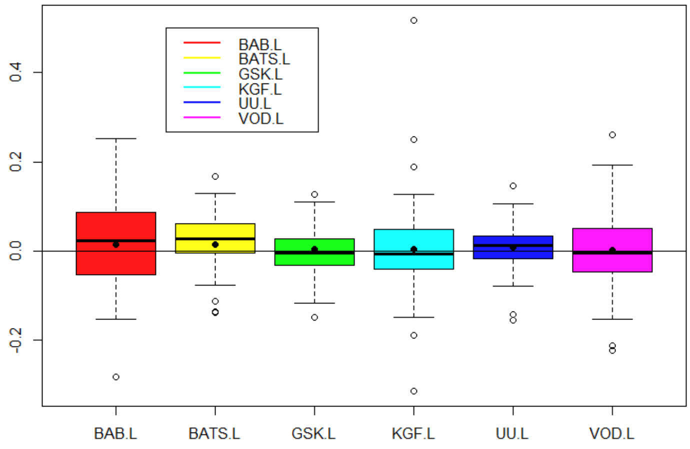

Figure 18: Scenario 2: Boxplots of initial return-history

We want to see how these portfolios develop over time. For this purpose we plot the weight-assignments of the RRAP and RGOP for all rebalancing dates in Figures 19 and 20, starting with December 2005. In these figures one can observe a slightly higher movement in the asset weights of the RGOP compared to the RRAP, which will be confirmed by the higher turnover rate displayed in Table 5. The overall weight distribution at the rebalancing dates December 2006 to December 2009 seem to follow the same pattern as the initial portfolio at December 2005. The most attractive assets are BATS.L and GSK.L, where the latter is chosen to minimize the portfolio variance.

We also want to mention that similar to Scenario 1, our observations for the initial portfolios stay true at all subsequent rebalancing dates. This means that the robust risk averse portfolios are "closer" to the minimum variance portfolio than the robust growth optimal portfolios. The same holds for the robust portfolios with moment ambiguity compared to their "vanilla" counterparts.

Note that although the various strategies are very similar at the beginning, the differences at the last rebalancing date and especially in December 2007 are more distinct, see Figures 19 and 20. Where the robust growth optimal strategy prefers BAB.L due to its appealing performance from December 2005 to December 2008, the robust risk averse strategy relies more on UU.L and KGF.L. From the price history in Figure 16, we see that for this time period the price evolutions of UU.L and KGF.L do not seem to have much in common, which will result in a smaller correlation and hence expected portfolio variance and states the reason why this combination is preferred by the risk averse strategies.

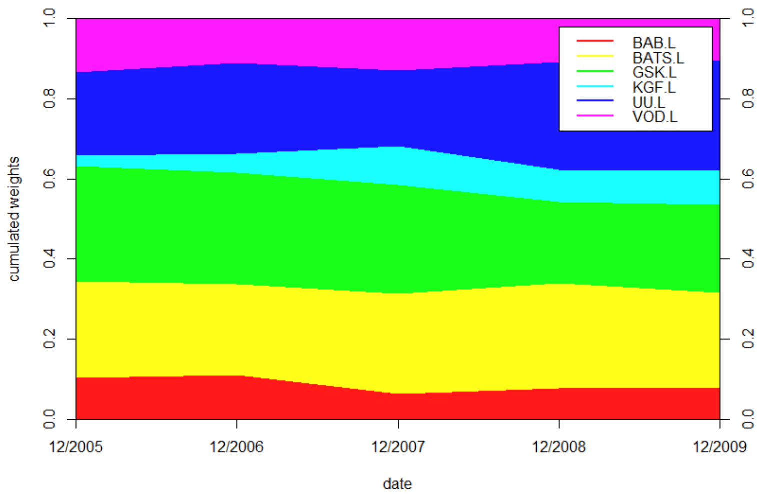

Figure 19: Scenario 2: Asset-weights evolution of the RRAP

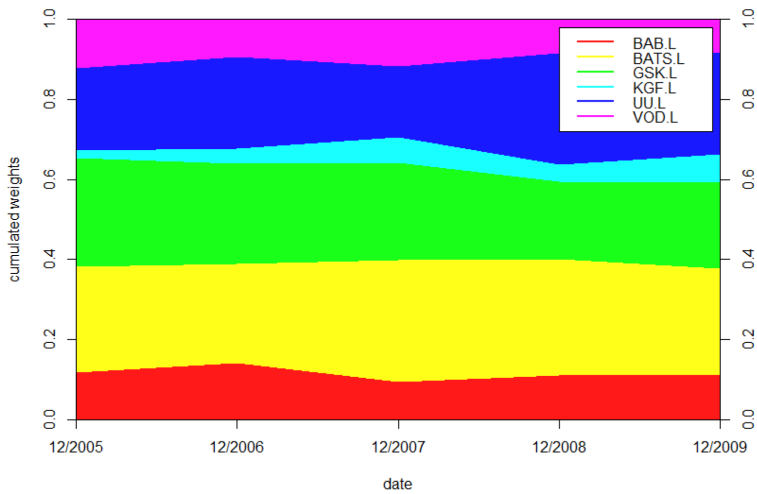

Figure 20: Scenario 2: Asset-weights evolution of the RGOP

In Figures 21 and 22 we again display the influence of the risk aversion parameter γ on the initial robust portfolios with and without moment ambiguity, respectively. As we have already seen in Figure 17 that in this scenario there seems to be less differences between the robust portfolio weight-assignments, this of course holds for all choices of γ. Especially for the portfolios with moment ambiguity (displayed in Figure 22) we can hardly depict any differences visually. Nevertheless, we once again observe the same patterns as in Scenario 1. For increasing γ, the changes in asset weights decrease and hence the biggest differences are observed for γ close to γ^*, where $\gamma^* = 0.02806$ in this example, which corresponds to the RRAP and RRAP+, respectively.

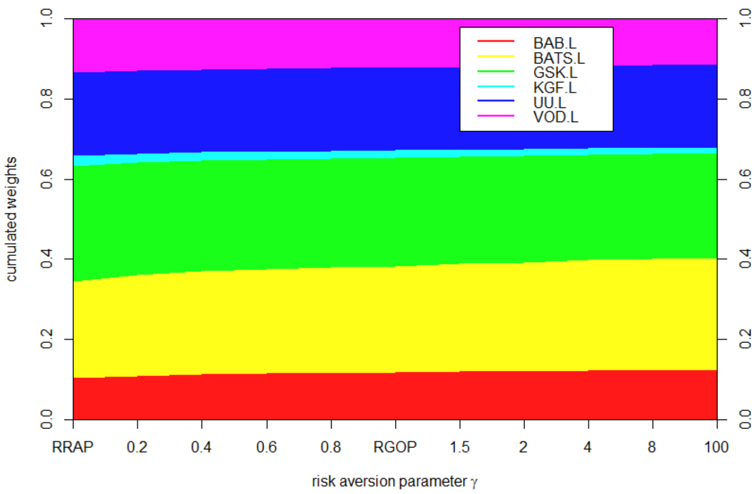

Figure 21: Scenario 2: Initial portfolios without moment ambiguity dependent on γ

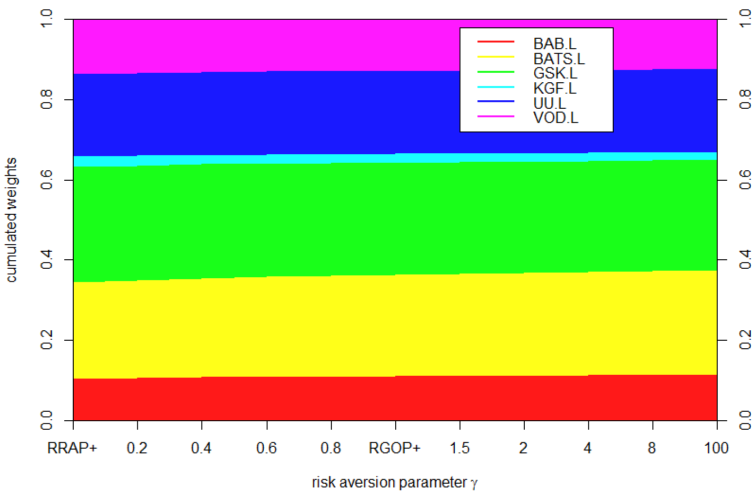

Figure 22: Scenario 2: Initial portfolios with moment ambiguity dependent on γ

We can now compare the performances of the strategies by using the performance measures presented in Section 7.1. The results can be seen in Table 5, where the figures correspond to realized monthly returns.

	$\hat{r}$	$\hat{\sigma}$	$\widehat{SR}$	$\widehat{TR}$	$\widehat{NR}$	$\widehat{MDD}$
RRAP	0.00884	0.03387	0.26098	0.05875	1.63962	0.16820
RRAP+	0.00883	0.03387	0.26075	0.05871	1.63893	0.16776
RGOP	0.00895	0.03407	0.26277	0.06037	1.64999	0.17632
RGOP+	0.00888	0.03390	0.26201	0.05932	1.64354	0.17185
MVAR	0.00882	0.03389	0.26024	0.05864	1.63766	0.16669
1/n	0.00979	0.03780	0.25895	0.05753	1.72048	0.18357

Table 5: Scenario 2: Performance measures of strategies

Once again the equally weighted portfolio strategy performed best in terms of mean and net aggregated return. Where in Scenario 1 this naive diversification dominated the other strategies we see that this time it produced the biggest maximum drawdown and variance out of all considered strategies.
We also observe that the minimum variance portfolio, which dominated the robust strategies in Scenario 1, is now outperformed by the robust risk averse portfolios. If we look at the robust portfolios, the growth optimal outperformed the risk averse portfolios in terms of the return measures $\hat{r}$ and $\widehat{NR}$. On the other hand, the risk averse portfolios realized a smaller variance, turnover rate and maximum drawdown. This perfectly suits their notion of being risk averse.
Similar to Scenario 1, and in accordance with our observations on the initial asset-weights distributions in Figure 17, the performances of the robust portfolios that take moment ambiguity into account deviate from their "vanilla" counterparts towards the minimum variance portfolio.

All in all we observed a similar pattern as in Scenario 1 Section 7.2.1, although this time there was no single strategy dominating the others. The more risk averse the model of a robust portfolio was, the more it resembled the weight assignments of the minimum variance portfolio. This of course directly influenced the observed performances, for which the same assertions hold.

7.2.3 Scenario 3: The DJIA during the Global Financial Crisis

As a last example we apply our investment strategies to the set of assets that form the Dow Jones Industrial Average (DJIA). The DJIA is a stock market index that represents 30 large publicly owned companies based in the United States. We again consider the investment period from December 2005 to December 2010 and use the same method as in the previous scenarios to periodically rebalance the portfolio weight assignments. The evolution of the DJIA over the relevant time period December 2000 to December 2010 can be seen in Figure 23. We immediately notice the big drawdown in 2008, which of course was caused by the global financial crisis.

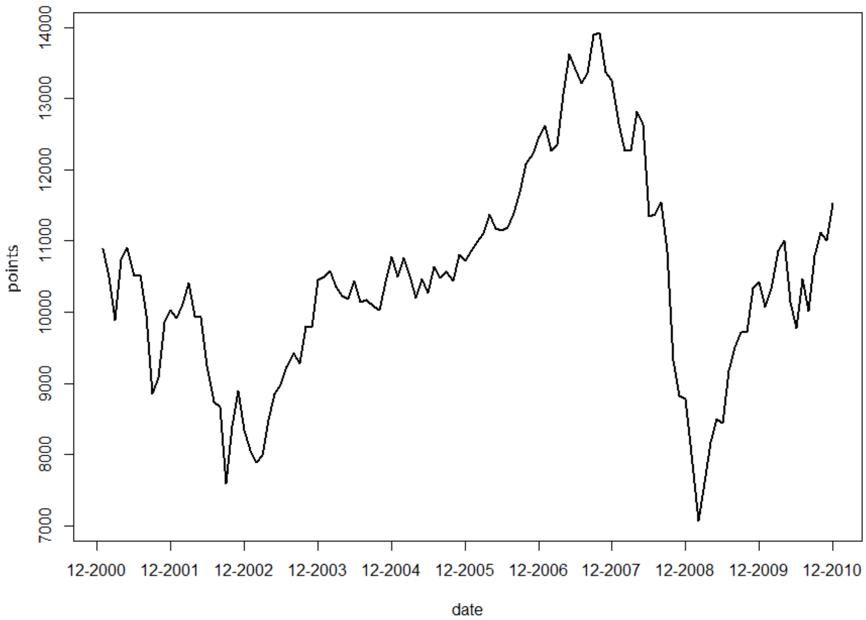

Figure 23: Scenario 3: DJIA from 12/2000 to 12/2010

We want to omit here the interpretation of the portfolio weight assignments since the overall patterns we observed in the previous scenarios stay the same. The robust portfolios RRAP and RGOP are close to the MVAR portfolio and the robust portfolios with moment ambiguity RRAP+ and RGOP+ are even more risk averse in the classical sense.

In addition to the risk measures we introduced in the beginning of this section, we want to compare the equally weighted portfolio to the RRAP and RGOP by the evolution of the net aggregated returns. In Figure 24 we can observe that in periods where there was a steady growth, the equally weighted portfolio outperformed the robust portfolios in terms of aggregated return. On the other hand, from the end of 2007 to the beginning of 2009 and in 2010, i.e. in periods where DJIA dropped down or moved sidewards, the robust portfolios performed significantly better than the equally weighted.

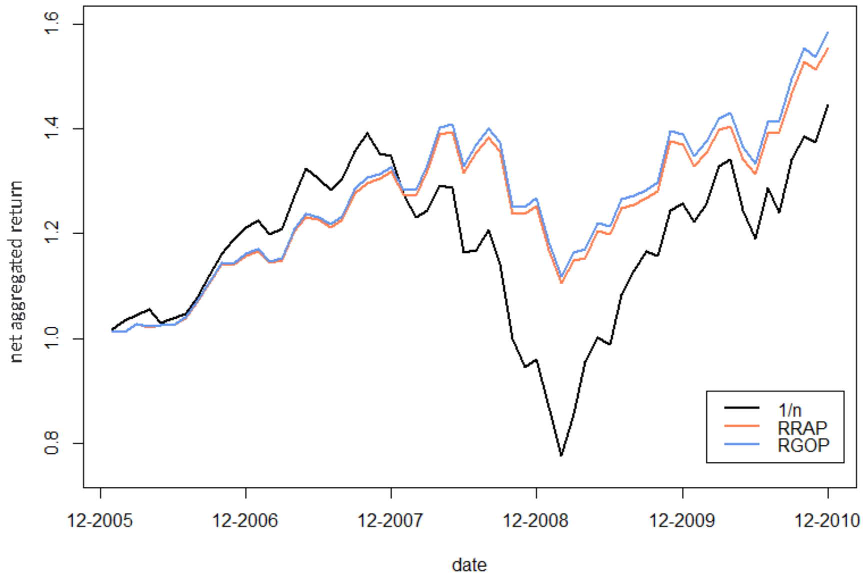

Figure 24: Scenario 3: Net aggregated return of 1/n, RRAP and RGOP

The observation we made on the net aggregated returns in Figure 24 are backed by the performance measures presented in Table 6. The equally weighted portfolio showed the least favourable performance where the RGOP outperformed all other strategies in terms of return and maximum drawdown. We again observe the pattern that the consideration of the moment ambiguity shifted the robust portfolios towards the MVAR portfolio.

	$\hat{r}$	$\hat{\sigma}$	$\widehat{SR}$	$\widehat{TR}$	$\widehat{NR}$	$\widehat{MDD}$
RRAP	0.00785	0.03100	0.25320	0.06930	1.55394	0.20814
RRAP+	0.00782	0.03099	0.25223	0.06929	1.55094	0.20843
RGOP	0.00819	0.03117	0.26265	0.06893	1.58506	0.20516
RGOP+	0.00790	0.03102	0.25460	0.06928	1.55830	0.20769
MVAR	0.00780	0.03098	0.25159	0.06928	1.54893	0.20864
1/n	0.00738	0.04955	0.14901	0.05890	1.44588	0.44274

Table 6: Scenario 3: Performance measures of strategies

8 Conclusion

In this thesis, we derived distributionally robust portfolios which are closely related to the Markowitz mean-variance efficient portfolios. For this purpose we introduced a new method of a posteriori evaluating the performance of a portfolio. We penalized small returns by high values and high returns by small values and defined the average return penalization over all observed portfolio returns as our performance measure. Our choice of penalization function was a quadratic polynomial with positive curvature and a parametrized minimum at $\gamma \in \mathbb{R}$. From this definition we motivated the interpretation of γ as a risk aversion parameter. This a posteriori performance measure depicted an a priori random variable since the portfolio returns are unknown at the beginning of an investment horizon. By the definition of this a priori measure, we preferred portfolios which tend to have a small realization of this random variable. For that reason, we used the (1-ϵ)-quantile of the performance measure's distribution as the objective function for our portfolio optimization problem. By using well known results from robust optimization and without any restrictive assumptions on the asset-return distribution we eventually reformulated this problem as a second-order cone program. In the process of deriving our portfolio optimization problem we obtained a lower bound restriction γ^* on the risk aversion parameter. For all $\gamma > \gamma^*$ the reformulation as a second-order cone program enabled us to solve the optimization problems very efficiently by using one of many already existing solvers. Our approach also allowed us to easily take the moment ambiguity into account, which arises from the necessity of estimating the asset-return's first- and second-order moments. For estimating these moments, we used the technique of the so-called shrinkage estimation, for which again no restrictive assumptions about the asset-return distribution had to be made.

In the presented empirical backtests (and several others) we observed a distinct pattern. The most risk averse robust portfolios tended to behave like the Markowitz minimum-variance portfolio. The consideration of moment ambiguity likewise shifted the asset-weights towards the minimum-variance portfolio. On the other side, the larger the risk aversion parameter γ ($\gamma \geq \gamma^*$), the more the resulting weight distribution differed from the risk averse portfolios. However, this deviation from the risk averse portfolio was bounded since the change in the weight-distribution was decreasing as a function of the risk aversion parameter. We also observed that the robust risk averse portfolios indeed followed the performance of the minimum-variance portfolio but never were as extreme. This means that in scenarios where the minimum-variance strategy showed a poor performance compared to the others, so did the robust risk averse strategies, but not as poorly. On the other hand, in scenarios where the minimum-variance strategy showed a strong performance, the robust risk averse portfolios did as well, but not as strongly. We have also observed the well known result that the equally weighted portfolio is very hard to outperform.

From the empirical backtests and our derivation of the robust portfolios we conclude that our framework contributes a distributionally robust extension to the classical mean-variance efficient portfolios. This interpretation is even more appropriate since all of our robust portfolios are mean-variance efficient in the classical sense. The major advantages of our approach, besides the obvious fact that we do not make any restrictive assumptions on the asset-return distribution, are that it includes the notion of growth optimality and allows for further robustification against moment ambiguity. Due to its reformulation as an second-order cone program, the problem of determining the robust portfolios costs as much computational effort as it does for the classical Markowitz mean-variance efficient portfolios.

We again want to state that our derivation of distributionally robust portfolios is based on the paper "Robust growth-optimal portfolios" which was published by Rujeerapaiboon et al. in 2014. In their work, Rujeerapaiboon et al. approximated the asymptotic growth rate by the second-order Taylor polynomial which, like the portfolio evaluation function introduced in this thesis, is a quadratic polynomial function in the portfolio return. They used the worst-case value-at-risk of the mean (approximated) growth rate as the objective of their portfolio optimization problem, where the worst-case is taken over all asset-return distributions with predefined (estimated) first and second-order moments. In contrast, we simply used a parametrized quadratic polynomial with positive curvature, where the parameter defined the location of the minimum. We have seen that the objective functions of the SOCP-reformulation of the approach of Rujeerapaiboon et al. and our approach are equivalent in terms of optimization if we choose the risk aversion parameter equal to one. This is obvious since for $\gamma = 1$, our polynomial penalization function is nothing but the quadratic Taylor approximation of log-returns multiplied by -2 plus a constant term.

In conclusion, we interpret our approach as a generalization of "Robust growth-optimal portfolios" introduced by Rujeerapaiboon et al.. It enables us to choose a risk aversion parameter which defines the location of the resulting portfolio on the mean-variance efficient frontier and thus the identification of the most risk averse robust portfolio. Like for robust growth-optimal portfolios, our approach allows for easy robustification against moment ambiguity and the computational effort is comparable to that of the classical mean-variance efficiency framework.

References

Alizadeh, F. and Goldfarb, D. (2002). Second-order cone programming. *Mathematical Programming manuscript*, Ser. B 95:3–51.

Bertsimas, D., Gupta, V., and Kallus, N. (2013). Data-driven robust optimization. *INFORMS Nicholson Paper Competition.*

Cover, T. M. and Thomas, J. A. (1991). Elements of information theory. *Wiley Series in Telecommunications. Wiley.*

Delage, E. and Ye, Y. (2010). Distributionally robust optimization under moment uncertainty with application to data driven problems. *Operations Research*, 58(3):595–612.

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

DeMiguel, V., Martin-Utrera, A., and Nogales, F. (2013). Size matters: Optimal calibration of shrinkage estimators for portfolio selection. *Journal of Banking & Finance*, 37(8):3018–3034.

Elton, J. and Gruber, J. (1997). Modern portfolio theory, 1950 to date. *Journal of Banking & Finance*, 21:1743–1759.

Hakansson, N. (1971). Capital growth and the mean-variance approach to portfolio selection. *Journal of Financial and Quantitative Analysis*, 6:517–557.

Kojima, M., Mizuno, S., and Yoshise, A. (1989). A primal-dual interior-point algorithm for linear programming. *Progress in Mathematical Programming.*

Krokhmal, P., Palmquist, J., and Uryasev, S. (2002). Portfolio optimization with conditional value-at-risk objective and constraints. *Journal of Risk*, 4:11–27.

Laurent, M. and Rendl, F. (2005). Semidefinite programming and integer programming. *Handbooks in Operations-Research and Management Science*, 12:393–514.

Lobo, M., Vandenberghe, L., Boyd, S., and Lebret, H. (1998). Applications of second-order cone programming. *Linear Algebra and its Applications*, 284:193–228.

Luenberger, D. (1998). Investment science. *Oxford Univerity Press.*

Luenberger, D. and Ye, Y. (2008). *Linear and Nonlinear Programming*, volume 3.

MacLean, L. C., Thorp, E. O., and Ziemba, W. T. (2010). The kelly capital growth investment criterion: Theory and practice. *World Scientific*, pages 563–574.

Mandelbrot, B. and Hudson, R. L. (2004). The (mis)behaviour of markets: A fractal view of risk, ruin, and reward. *London: Profile Books.*

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

Michaud, R. (1998). Effcient asset management. *Harvard Business School Press, Boston, Massachusetts.*

Monteiro, R. and Adler, I. (1989). Interior path following primal-dual algorithms. part i: Linear programming. *Math. Programming*, 44:27–41.

Rebonato, R. and Jäckel, P. (1999). The most general methodology to create a valid correlation matrix for risk management and option pricing purposes. *Quantitative Research Centre of the NatWest Group.*

Roll, R. (1973). Evidence on the "growth-optimum" model. *Journal of Finance*, 28(3):551–556.

Roy, A. (1952). Safety first and the holding of assets. *Econometrica*, 20(3):431–449.

Rujeerapaiboon, N., Kuhn, D., and Wiesemann, W. (2014). Robust growth-optimal portfolios.

Toh, K., Tütüncü, R., and Todd, M. (2006). On the implementation and usage of sdpt3 – a matlab software package for semidefinite-quadratic-linear programming. (4).

Vandenberghe, L. and Boyd, S. (1996). Semidefinite programming. *SIAM Review*, 38(1):49–95.

Zymler, S., Kuhn, D., and Rustem, B. (2013). Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*, 137(1-2):167–198.