Robust Growth-Optimal Portfolios

Napat Rujeerapaiboon

Submitted in partial fulfilment of the requirements for the MPhil degree of Imperial College London

January 2014
The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives license. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the license terms of this work.

©2014 – Napat Rujeerapaiboon
Abstract

A portfolio which has a maximum expected growth rate is often referred to in the literature as a log-optimal portfolio or a growth-optimal portfolio. The origin of the log-optimal portfolio is arguably due to Kelly [27] when he observed that logarithmic wealth is additive in sequential investments and invented a betting strategy for gambling that relies on results from information theory. As a result of the law of large numbers, if investment returns are serially independent and identically distributed, the growth rate of any constant rebalanced portfolio (the log-optimal portfolio included) converges to its expectation. Moreover, under such conditions, one of the strongest advantages of the log-optimal portfolio is that, when implemented repeatedly, the log-optimal portfolio outperforms any other causal portfolio in the long run with probability 1. In other words, if all of these conditions are met, there is no sequence of portfolios that has a higher growth rate than that of the log-optimal portfolio. Stock markets however are different from casinos in the sense that investment returns are not serially independent and identically distributed. Also, since trading incurs transaction costs, investors are discouraged from making frequent trades. Plus, the probability distribution of stock returns is never precisely known, which impedes the calculation of the log-optimal portfolio. In this project, we generalize the results for the log-optimal portfolio. In particular, we establish similar guarantees for finite investment horizons where the distribution of stock returns is ambiguous. By focusing on constant rebalanced portfolios, we exploit temporal symmetries to formulate the emerging distributionally robust optimization problems as tractable conic programs whose sizes are independent of the investment horizon.

Key words. Portfolio Optimization, Value-at-Risk, Robust Optimization, Second-Order Cone Programming, Semidefinite Programming, Distributional Ambiguity.
Acknowledgements

First and foremost, I would like to express my deepest gratitude and highest appreciation towards my thesis supervisor, Dr. Daniel Kuhn, for his constant and immense support throughout the year. I tremendously appreciate his motivation, guidance, and insight. In addition, his creativity, meticulousness, and research experience have contributed greatly to the success of this project. Besides, I would like to thank him for always being there and for offering help to my studies and to my personal life. On the whole, he has given me much more than I ever would have expected from a thesis supervisor. He has become my role model and I want to follow in his footsteps as a scholar and as a person.

I am also immensely thankful to Dr. Wolfram Wiesemann, who has actively participated in the second half of the project. I strongly appreciate his accessibility and his constructive thoughts and comments. To me, both Dr. Daniel Kuhn and Dr. Wolfram Wiesemann have always been an irreplaceable and reliable source of support. I am also indebted to Dr. Victor DeMiguel for agreeing to be my external examiner and for his suggestions on how to improve this thesis. My heartfelt thanks also go to Dr. Berç Rustem and Dr. Panos Parpas for their advice during my presentation to the Computational Optimisation group of Imperial College London. I would like to also thank other former and current members of this research group: Dr. Christos Gavriel, Grani Adiwena Hanasusanto, Chin Pang Ho, Vladimir Roitch, Sei Howe, Vu Ngoc Duy Luong, Micheal Hadjiyiannis, and Iakovos Kakouris for the valuable discussions and all the fun times we had together. My special thanks also go to Kulyancee Jongjairuksa and Xinyuan Zhang for my wonderful time in London and for listening to my hardships patiently.

Finally, a special thank you is wholeheartedly dedicated to my family and my friends, especially Sanprapha Vudhivorn. Were it not for their endless moral and emotional support, this project would not be the success that it is. Thank you.
Declaration of Originality

I hereby certify that this thesis is entirely my own work, except where otherwise indicated in the text. This thesis has not been submitted for the award of any other degree before.
5.5 Out-of-Sample Evaluation .................................................. 65

6 Conclusions ........................................................................71

6.1 Advantages of the Log-Robust Portfolio .............................. 72

6.2 Future Work ..................................................................... 72
List of Figures

1.1 Efficient Frontier ................................................. 18

5.1 VaR comparison between log-robust and log-optimal portfolios (sample estimator) . 63
5.2 VaR comparison between log-robust and log-optimal portfolios (shrinkage estimator) 63
5.3 Sharpe ratio comparison between log-robust and log-optimal portfolios (sample estimator) ................................................. 64
5.4 Sharpe ratio comparison between log-robust and log-optimal portfolios (shrinkage estimator) ................................................. 64
List of Tables

4.1 Tractable forms of uncertainty in mean values ($\mu$) ........................................ 56
4.2 Tractable forms of uncertainty in (co)variances ($\Sigma$) ........................................ 56

5.1 List of datasets ........................................................................................................... 60
5.2 VaR comparison between log-robust and log-optimal portfolios ($T = 120$ months) .... 65
5.3 VaR comparison between log-robust and log-optimal portfolios ($T = 360$ months) .... 66
5.4 VaR comparison between log-robust and log-optimal portfolios ($T = 1,200$ months) . 67
5.5 Out-of-sample evaluation (source: 10Ind) ................................................................. 67
5.6 Out-of-sample evaluation (source: 12Ind) ................................................................. 68
5.7 Out-of-sample evaluation (source: iShares) .............................................................. 68
5.8 Out-of-sample evaluation (source: DJIA) ................................................................. 69
Chapter 1

Introduction

For decades, financial researchers and practitioners have been facing the challenge to construct an optimal asset allocation in stock markets. An early breakthrough in the field by Markowitz [36, 37] in 1950s is highly regarded and won him a prestigious Nobel Memorial Prize in Economic Sciences. It had been long known before that an investor can reduce investment risk by diversifying his or her portfolio, but Markowitz was the first to formalize the idea. Markowitz suggested to measure the investment risk by calculating the variance\(^1\) of the portfolio return. He theorized that an investor can then make a cautious investment by either maximizing the expected return for a target level of risk or minimizing the risk for a target level of expected return. This results in a mean-variance diagram which allows Markowitz supporters, based on their risk preference, to select any portfolio on the efficient frontier. Markowitz’s pioneering work positively impressed financial societies, both academics and practitioners. Moreover, it triggered a lot of research in portfolio theory on the basis of mean and variance analysis. Remarkable examples include the works of Sharpe [47], Lintner [32], and Mossin [41]. All pieces put together give rise to, what is often called, modern portfolio theory, which is undeniably a huge success considering its popularity in the financial industry.

There are many reasons why the modern portfolio theory pioneered by Markowitz is often deemed appealing. For theoretical reasons, an optimal portfolio in the sense of Markowitz is a solution of a simple quadratic program (see Luenberger [34]), and thus selecting portfolio is highly tractable. In addition, only the first two moments of the asset returns’s distribution are required to select the optimal portfolio. Further distributional information is not required, making the

\(^1\)Variance is not the only used risk measure. In fact, there are many alternative risk measures, one of which is the Value-at-Risk (VaR). VaR has been a popular measure of risk among financial regulators (see Jorion [25]). An introductory discussion on VaR can be found in Linsmeier and Pearson [31]. Examples of using VaR in portfolio optimization include Zymler et al. [56] and El Ghaoui et al. [17].
Chapter 1. Introduction

Figure 1.1: Efficient Frontier

approach distributionally robust. In terms of utility theory, mean-variance portfolio optimization is consistent with maximizing a quadratic utility function. This latter point was strengthened by Levy and Markowitz [30], when they pointed out that a quadratic utility function might be an adequate approximation of an investor’s true utility function. However, whether every investor should approximate his or her non-quadratic utility function by a quadratic function is still debatable as some people believe that this should be done only if asset returns are normally distributed (see, for example, Luenberger [34, Chapter 9] and Ingersoll [23, Chapter 4]). In addition, a quadratic utility function is not an increasing function, and it displays an increasing absolute risk aversion. Some investors may thus find it an implausible choice of utility function according (see, for example, Hakansson [21]).

What is more important to note about Markowitz’s model is that it is a single-period framework. Hence, it does not exploit the benefit which could be gained from reinvestments in a multi-period setting. Moreover, the portfolio choice selected from a typically infinite set of efficient portfolios is subjective, and the trade-off between risk and return is not entirely clear. For such reasons, Latané [29] advocated alternative investment criteria which are more objective, and finally suggested that an investor should focus on the terminal wealth rather than the return of one-period investments. The idea gave rise to an investment strategy particularly developed for multi-period investments, namely the log-optimal portfolio or the growth-optimal portfolio. However, the origin of the log-optimal portfolio is sometimes traced back further to Kelly [27] because he had proposed to maximally invest money in the context of gambling three years before Latané did. That being said, long before Kelly published his betting strategy (which incidentally is often referred to as Kelly’s criterion), Williams [52] had already argued investors should really be concerned about the investment’s geometric mean return and not its arithmetic counterpart because of the compounding effect.
More precisely speaking, the log-optimal portfolio maximizes the expected value of the logarithm of the terminal wealth. We explain later in Chapter 3 the specification of the log-optimal portfolio and continue discussing interesting facts about the log-optimal portfolio. In a market where stock returns are serially independent and identically distributed, it has been proved that:

1. The log-optimal portfolio yields a higher growth rate than any other causal investment strategy in the "long" run with probability 1 (see Breiman [8] and Algoet and Cover [1]).

2. The expected time required by the log-optimal portfolio to achieve a "high" target wealth is less than or equal to that required by any other causal investment strategy (see Breiman [8] and Algoet and Cover [1]).

Among many others, these two properties of the log-optimal portfolio seemingly make it a universally best portfolio choice as investors usually aim to get rich or to get money quickly. This conclusion is however far too superficial and resulted in an intense debate between its supporters and its oppositions. Interestingly, one of the supporters is Markowitz [37, 38], who laid the foundation for the modern portfolio theory. Samuelson [45, 46], on the other hand, refused to believe in the log-optimal portfolio. One of the main opposing arguments is that for any finite investment horizon, the log-optimal portfolio is suboptimal for some utility functions no matter how long the investment horizon is. Indeed, the two properties mentioned above are asymptotic results. Such results have to be interpreted carefully. For example, an experiment performed by Rubinstein [44] showed that it can take far longer than a human lifespan for the log-optimal portfolio to beat other strategies with a high probability. Nevertheless, in his original paper, Latané [29] did not argue that the log-optimal portfolio is the only rational investment criterion. He proposed it as an alternative choice for long-term investors.

Since the mean-variance portfolio and the log-optimal portfolio originated from different settings, both are consistent with different utility functions. One might wonder if there is a connection between them. This question was answered by Hakansson [21]. Hakansson found that in some cases the log-optimal portfolio is not even close to the efficient frontier. Estrada [18] made a comparison between the mean-variance portfolio and the log-optimal portfolio. He found out that indeed the log-optimal portfolio attained a higher wealth. However, it also exhibited higher volatility, which is probably the reason for its lack of popularity among practitioners (see Mauboussin [39]).

For enthusiastic readers, more information on the historical developments of the discussed investment criteria can be found in Christensen [11] and Poundstone [43]. An exhaustive list of desirable and undesirable properties of the log-optimal portfolio can be found in MacLean et al. [35].
1.1 Main contributions

The main contribution of this thesis is that we generalize the results for the log-optimal portfolio. The use of the log-optimal portfolio is strongly advocated by the aforementioned asymptotic results, that is, the log-optimal portfolio beats all other investment strategies over infinitely long investment horizons. Particularly, we relax the assumption of infinite investment horizons, and establish similar guarantees for finite horizons. Moreover, we also relax the independence assumption because Greene and Fielitz [20] suggested that it is often violated. These guarantees are valid not just for one probability distribution of asset returns, but for a whole family of distributions that share the same first- and the same second-order moments. The reason that we avoid restricting our analysis to a single distribution, such as, a multivariate normal distribution, is that typically the distribution of stock returns is unknown and many studies, such as Fama [19], provide statistical evidence against the normality assumption.

Unlike mean-variance portfolios, we propose a family of portfolios which could benefit from rebalancing. We extend our results to several practical settings. We show that it is possible to incorporate bounds on the mean vector and the covariance matrix of asset returns to reduce the effect of estimation errors, which is an important aspect of portfolio optimization (see, for instance, Best and Grauer [4] and Broadie [9]). Finally, we provide numerical results for both in-sample and out-of-sample experiments to demonstrate that the proposed investment strategy achieves both low volatility and high terminal wealth.

1.2 Structure of the Thesis

This report is structured as follows. In Chapter 2, we briefly present the historical development of mathematical optimization along with modern concepts in the field. We first discuss the seminal work on the linear programming method. Then, we describe the key ideas of two popular frameworks for modelling optimization problems under uncertainty, namely the stochastic programming approach and the robust optimization approach. We then discuss a relatively new approach which is often referred to as the distributionally robust optimization approach. Some mathematical analysis of this approach is presented because the result is used throughout the thesis.

The main contributions of our work start from Chapter 3. We begin by describing the general setting of the stock market that we study. Afterwards, we give the definition of portfolio growth rate. We then propose a portfolio optimization model. An optimal solution to this model is named a log-robust portfolio. It is a portfolio that achieves a high growth rate even under the worst-
Chapter 1. Introduction

case distribution consistent with the first- and the second-order moment information of the stock returns. Specifically, by exploiting temporal symmetries between the stock returns, we determine an analytical expression for the worst-case value-at-risk of the quadratic approximation of the growth rate of any constant rebalanced investment strategy, and maximizing this expression reduces to solving a tractable second-order cone program. Furthermore, we present some extensions of our portfolio optimization model in Chapter 4. We describe how to determine the log-robust portfolio in practice when the true moments of the random stock returns are not available. Moreover, we give an analysis of how the performance of the log-robust portfolio deteriorates when trading and/or holding stocks is costly. A numerical evaluation of the log-robust portfolio is presented in Chapter 5. We end the report with a short summary of what we have achieved and promising avenues for future research in Chapter 6.

1.3 Notation

Lower-case bold face letters and upper-case bold face letters are used to denote column vectors and matrices, respectively. The space of symmetric matrices of dimension $n$ is denoted by $S^n$, while the space of all positive semidefinite (positive definite) symmetric matrices of the same dimension is denoted by $S^n_+$ ($S^n_{++}$). We also denote by $\mathbf{1}$ and $\mathbf{I}$ the vector of ones and the identity matrix of appropriate dimension, which should be clear from the context, respectively. For any two square matrices $X$ and $Y$ of the same dimension, we let $\langle X, Y \rangle = \text{Tr}(XY)$ be the trace scalar product. The relation $X \succeq Y$ ($X \succ Y$) implies that $X - Y$ is positive semidefinite (positive definite). Random variables are always represented by symbols with tildes, while their realizations are denoted by the same symbols without tildes. The set of all probability distributions over $\mathbb{R}^n$ is denoted by $\mathcal{P}^n_0$. Lastly, we denote the Kronecker delta of two integers $i$ and $j$ by $\delta_{ij}$, i.e.,

$$
\delta_{ij} = \begin{cases} 
1 & \text{if } i = j, \\
0 & \text{otherwise}. 
\end{cases}
$$
Chapter 2

Worst-Case Probabilistic Guarantee

The concept of linear programming was first developed by the Russian mathematician Kantorovich to minimize the cost of military operations during World War II in 1939. Linear programming theory has progressed significantly since then, and a number of applications in different domains, for example, inventory management, machine learning, and finance, benefit from it. Two examples of the early breakthroughs are the simplex algorithm and the duality theory proposed by Dantzig [13] and von Neumann, respectively. Another breakthrough was due to Karmarkar [26], who invented the interior-point method for efficiently solving linear programming problems. The invention of fast linear programming algorithms contributed to the popularity of mathematical optimization in both academia and industrial practice. For the enthusiastic reader keen to learn more about the historical developments of linear programming, we recommend Todd [48] and the references therein.

Despite the impressive advancement of the field, the study of optimization in its early stage usually focused on deterministic optimization problems where the involved parameters are known. More often than not, real-world problems contain some unknown parameters, and some decisions have to be made before the values of those parameters are observed. In order to make such decisions optimally, traditionally there are three frameworks for modelling optimization problems under uncertainty: stochastic programming, robust optimization, and dynamic programming. Here, we discuss only the first two. Stochastic programming assumes exact knowledge of the probability distribution underlying the uncertain parameters, and the goal is to minimize the expectation of some cost function or to maximize the expectation of some profit function. A key challenge in stochastic programming relates to the estimation of such a distribution. One downside of the stochastic programming approach is that if the obtained optimal solution is sensitive to the estimated distribution and the estimation error is high, then the solution may perform poorly in out-of-sample tests. Even worse, the solution might not be implementable.
A different class of stochastic programming problems are chance-constrained programs. In a typical formulation of a chance-constrained programming problem, the objective function is typically deterministic, and therefore there is no need to take the expected value of the objective function. The goal of this framework is to find a solution that optimizes the objective function while remaining feasible with a high probability, for example, 99%. A comprehensive review of stochastic programming can be found in Birge et al. [6].

Even though it may be difficult to obtain an accurate probability distribution of the uncertain parameters, one can still tackle an uncertainty-affected optimization problem by using the robust optimization approach. The robust optimization approach typically does not require any information about the probability distribution of the involved random variables. Instead, it is assumed that the realization of the random parameters falls within a prescribed set, namely the uncertainty set. Put differently, the uncertainty set contains all the future scenarios that can materialize. The aim of the robust optimization approach is to solve a minimax (maximin) problem, which corresponds to the minimum cost (maximum profit) if the worst-case scenario within the uncertainty set is realized. The solution of a robust optimization problem is guaranteed to be feasible for any scenario within the uncertainty set and is therefore always implementable. In this sense, the robust optimization approach can be thought of as a special case of the chance-constrained approach where the produced solution needs to be feasible with probability one. Another advantage of the robust optimization approach is its tractability. The complexity of the robust counterpart and that of the nominal (deterministic) optimization problem are usually comparable due to the duality theory of convex optimization. However, the robust optimization approach has one major disadvantage, which is its tendency to suggest overly conservative decisions as it ignores any information that might be available about the distribution of the random parameters. For a comprehensive review of robust optimization, we refer to Ben-Tal et al. [2].

Recently, there has been a new development in optimization studies. To address the downsides of both stochastic programming and robust optimization, a new framework, namely distributionally robust optimization, combines the ideas of both aforementioned paradigms. This new framework assumes that certain distributional properties of the random parameters are known, such as the mean and the variance, and determines a solution that is optimal under the worst-case probability distribution satisfying those properties. Hence, the distributionally robust optimization approach is based on the worst-case analysis and uses partial distributional information. Therefore, it is often regarded as a mixture between the stochastic programming approach and the robust optimization approach.
In the remainder of this section, we investigate distributionally robust quadratic chance constraints of the form
\[
\inf_{P \in \mathcal{P}} P\left(q^0 + q^\top \tilde{\xi} + \tilde{\xi}^\top Q \tilde{\xi} \leq 0\right) \geq 1 - \epsilon,
\] (2.1)
where \(Q \in S_k, q \in \mathbb{R}^k, q^0 \in \mathbb{R},\) and \(\epsilon \in (0, 1)\) are fixed. Given just the first- and the second-order moment information about the random vector \(\tilde{\xi},\) i.e., for an ambiguity set \(\mathcal{P}\) of the form
\[
\mathcal{P} = \{P \in \mathcal{P}_0 : E_P(\tilde{\xi}) = \mu, E_P(\tilde{\xi}^\top \tilde{\xi}) = \Sigma + \mu \mu^\top\},
\]
the constraint (2.1) is thoroughly investigated by Zymler et al. [55] as they establish the following theorem.

**Theorem 2.1.** The distributionally robust quadratic chance constraint (2.1) holds if and only if
\[
\exists M \in S_{k+1}^+, \exists \beta \in \mathbb{R}, \beta + \frac{1}{\epsilon} \langle \Omega, M \rangle \leq 0,
\]
where \(\Omega\) is the second-order moment matrix of \(\tilde{\xi}\) defined as
\[
\Omega = \begin{bmatrix}
\Sigma + \mu \mu^\top & \mu \\
\mu^\top & 1
\end{bmatrix}.
\]

**Proof.** See Zymler et al. [55].

This result is heavily used in the derivation of the log-robust portfolio. For the sake of completeness, we now prove that (2.1) has a deterministic conic representation which ensures its tractability. Our proof is different from the one in Zymler et al. [55], and therefore we hope that this will be of interest to the reader. We first observe that
\[
P\left(q^0 + q^\top \tilde{\xi} + \tilde{\xi}^\top Q \tilde{\xi} \leq 0\right) \leq P\left(q^0 + q^\top \tilde{\xi} + \tilde{\xi}^\top Q \tilde{\xi} < \gamma\right) \quad \forall \gamma > 0,
\]
for any \(P \in \mathcal{P}.\) Therefore, we have that
\[
\inf_{P \in \mathcal{P}} P\left(q^0 + q^\top \tilde{\xi} + \tilde{\xi}^\top Q \tilde{\xi} \leq 0\right) \leq \inf_{P \in \mathcal{P}} P\left(q^0 + q^\top \tilde{\xi} + \tilde{\xi}^\top Q \tilde{\xi} < \gamma\right) \quad \forall \gamma > 0,
\]
for any ambiguity set \(\mathcal{P}\) which is not an empty set, and
\[
\inf_{P \in \mathcal{P}} P\left(q^0 + q^\top \tilde{\xi} + \tilde{\xi}^\top Q \tilde{\xi} \leq 0\right) \leq \lim_{\gamma \to 0^+} \inf_{P \in \mathcal{P}} P\left(q^0 + q^\top \tilde{\xi} + \tilde{\xi}^\top Q \tilde{\xi} < \gamma\right),
\] (2.2)
We now consider the mapping \( \gamma \mapsto P(q^0 + q^\top \tilde{\xi} + \tilde{\xi}^\top Q \tilde{\xi} \leq \gamma) \). This mapping is non-decreasing, right-continuous, and upper semi-continuous, since it is in fact the cumulative distribution function of a random variable \( q^0 + q^\top \tilde{\xi} + \tilde{\xi}^\top Q \tilde{\xi} \). We conclude that

\[
P(q^0 + q^\top \tilde{\xi} + \tilde{\xi}^\top Q \tilde{\xi} \leq 0) = \lim_{\gamma \to 0^+} P(q^0 + q^\top \tilde{\xi} + \tilde{\xi}^\top Q \tilde{\xi} \leq \gamma).
\]

Moreover, since an infimum of non-decreasing and upper semi-continuous functions is also non-decreasing and upper semi-continuous\(^1\), which implies it being right-continuous, we have that

\[
\inf_{P \in \mathcal{P}} P(q^0 + q^\top \tilde{\xi} + \tilde{\xi}^\top Q \tilde{\xi} \leq 0) = \lim_{\gamma \to 0^+} \inf_{P \in \mathcal{P}} P(q^0 + q^\top \tilde{\xi} + \tilde{\xi}^\top Q \tilde{\xi} \leq \gamma).
\]

Hence, for any non-empty ambiguity set \( \mathcal{P} \), we find

\[
\inf_{P \in \mathcal{P}} P(q^0 + q^\top \tilde{\xi} + \tilde{\xi}^\top Q \tilde{\xi} \leq 0) = \lim_{\gamma \to 0^+} \inf_{P \in \mathcal{P}} P(q^0 + q^\top \tilde{\xi} + \tilde{\xi}^\top Q \tilde{\xi} \leq \gamma)
\geq \lim_{\gamma \to 0^+} \inf_{P \in \mathcal{P}} P(q^0 + q^\top \tilde{\xi} + \tilde{\xi}^\top Q \tilde{\xi} < \gamma),
\]

(2.3)

where the inequality holds because of the non-negativity of probability measures. From (2.2) and (2.3), we conclude that

\[
\inf_{P \in \mathcal{P}} P(q^0 + q^\top \tilde{\xi} + \tilde{\xi}^\top Q \tilde{\xi} \leq 0) = \lim_{\gamma \to 0^+} \inf_{P \in \mathcal{P}} P(q^0 + q^\top \tilde{\xi} + \tilde{\xi}^\top Q \tilde{\xi} \leq \gamma)
\geq \lim_{\gamma \to 0^+} \inf_{P \in \mathcal{P}} P(q^0 + q^\top \tilde{\xi} + \tilde{\xi}^\top Q \tilde{\xi} < \gamma)
\geq \lim_{n \to \infty} \inf_{P \in \mathcal{P}} P(q^0 + q^\top \tilde{\xi} + \tilde{\xi}^\top Q \tilde{\xi} < \gamma_n),
\]

where \( \{\gamma_n\}_{n=1}^\infty \) is a decreasing sequence of strictly positive real numbers converging to 0. The next step of the derivation is due to Calafiore et al. [10]. Define the set \( \mathcal{S}_n \) as

\[
\mathcal{S}_n = \left\{ \xi \in \mathbb{R}^k : q^0 + q^\top \xi + \xi^\top Q \xi < \gamma_n \right\},
\]

and denote by \( I_{\mathcal{S}_n} \) the indicator function of this set, i.e.,

\[
I_{\mathcal{S}_n}(\xi) = \begin{cases} 
1 & \text{if } \xi \in \mathcal{S}_n, \\
0 & \text{otherwise.}
\end{cases}
\]

\(^1\)It is straightforward to prove that an infimum of non-decreasing functions is non-decreasing. To verify upper semi-continuity, we note that a function is upper semi-continuous if and only if its hypograph is closed. In addition, a hypograph of the infimum is identical to the intersection of the hypographs of its member functions. Hence, the hypograph of the infimum of upper semi-continuous functions is closed, which means that the infimum is upper semi-continuous.
Then, \( \inf_{\mathcal{P} \in \mathcal{P}} \mathbb{P} \left( q^0 + q^\top \tilde{\xi} + \tilde{\xi}^\top Q \tilde{\xi} < \gamma_n \right) = \inf_{\mathcal{P} \in \mathcal{P}} \mathbb{E} \left( I_{S_n}(\tilde{\xi}) \right) \), which can be rewritten as the following semi-infinite linear program.

\[
\begin{align*}
\text{minimize} & \quad \int_{\mathbb{R}^k} I_{S_n}(\xi) \mu(d\xi) \\
\text{subject to} & \quad \mu \in \mathcal{M}_+ \\
& \quad \int_{\mathbb{R}^k} \mu(d\xi) = 1 \\
& \quad \int_{\mathbb{R}^k} \xi \mu(d\xi) = \mu \\
& \quad \int_{\mathbb{R}^k} \xi^\top \mu(d\xi) = \Sigma + \mu \mu^\top,
\end{align*}
\]

where \( \mathcal{M}_+ \) represents the cone of nonnegative measures on \( \mathbb{R}^k \) and the decision variable \( \mu \), as a member of this cone, is a nonnegative measure. The first constraint implies that the total probability is equal to 1, while the remaining two constraints are characterized by the first two moments of \( \tilde{\xi} \).

Assigning dual variables \( y^0 \in \mathbb{R}, y \in \mathbb{R}^k, \) and \( Y \in \mathbb{S}^k \) to the constraints and then dualizing the problem, we obtain a maximization problem and strong duality implies that the optimal objective value of this problem is the same whenever the covariance matrix \( \Sigma \) is strictly positive definite (see Isii [24]).

\[
\begin{align*}
\text{maximize} & \quad y^0 + y^\top \mu + \langle Y, \Sigma + \mu \mu^\top \rangle \\
\text{subject to} & \quad y^0 \in \mathbb{R}, y \in \mathbb{R}^k, Y \in \mathbb{S}^k \\
& \quad y^0 + y^\top \xi + \langle Y, \xi \xi^\top \rangle \leq I_{S_n}(\xi), \forall \xi \in \mathbb{R}^k,
\end{align*}
\]

which is equivalent to

\[
\begin{align*}
\text{maximize} & \quad y^0 + y^\top \mu + \langle Y, \Sigma + \mu \mu^\top \rangle \\
\text{subject to} & \quad y^0 \in \mathbb{R}, y \in \mathbb{R}^k, Y \in \mathbb{S}^k \\
& \quad y^0 + y^\top \xi + \langle Y, \xi \xi^\top \rangle \leq 1, \forall \xi \in \mathbb{R}^k \\
& \quad y^0 + y^\top \xi + \langle Y, \xi \xi^\top \rangle \leq 0, \forall \xi \in \mathbb{R}^k : q^0 + q^\top \xi + \xi^\top Q \xi \geq \gamma_n.
\end{align*}
\]
Chapter 2. Worst-Case Probabilistic Guarantee

This maximization problem can thus be rewritten as

\[
\begin{align*}
\text{maximize} & \quad \left\langle \begin{bmatrix} Y & \frac{1}{2} y^T y^0 \\
\frac{1}{2} y^T & y^0 \end{bmatrix}, \left[ \begin{bmatrix} \Sigma + \mu \mu^T & \mu \\
\mu^T & 1 \end{bmatrix} \right] \rightangle \\
\text{subject to} & \quad y^0 \in \mathbb{R}, \ y \in \mathbb{R}^k, \ Y \in \mathbb{S}^k,
\end{align*}
\]

\[
\begin{bmatrix} \xi \end{bmatrix}^T \begin{bmatrix} Y & \frac{1}{2} y^T \\
\frac{1}{2} y^T & y^0 \end{bmatrix} \begin{bmatrix} \xi \\
1 \end{bmatrix} \leq \begin{bmatrix} \xi \\
0 \end{bmatrix}^T \begin{bmatrix} 1 & 0 \\
0 & 1 \end{bmatrix} \begin{bmatrix} \xi \\
1 \end{bmatrix}, \ \forall \xi \in \mathbb{R}^k
\]

\[
\begin{bmatrix} \xi \end{bmatrix}^T \begin{bmatrix} Y & \frac{1}{2} y^T \\
\frac{1}{2} y^T & y^0 \end{bmatrix} \begin{bmatrix} \xi \\
1 \end{bmatrix} \leq 0,
\]

\[
\forall \xi \in \mathbb{R}^k: \begin{bmatrix} \xi \\
1 \end{bmatrix}^T \begin{bmatrix} Q & \frac{1}{2} q^T \\
\frac{1}{2} q^T & q^0 - \gamma_n \end{bmatrix} \begin{bmatrix} \xi \\
1 \end{bmatrix} \geq 0.
\]

For notational convenience, define a combined variable \( N \) and let \( \Omega \) be the second-order moment matrix of \( \tilde{\xi} \), i.e.,

\[
N = \begin{bmatrix} 0 & 0 \\
0^T & 1 \end{bmatrix} - \begin{bmatrix} Y & \frac{1}{2} y^T y^0 \\
\frac{1}{2} y^T y^0 & y^0 \end{bmatrix}, \quad \Omega = \begin{bmatrix} \Sigma + \mu \mu^T & \mu \\
\mu^T & 1 \end{bmatrix}.
\]

We then observe that the first constraint in the maximization problem above can be rewritten as a semidefinite constraint. Also, we can use the \( S \)-lemma to rewrite the second constraint as a semidefinite constraint. By doing that, we obtain a simpler representation of the maximization problem.

\[
\begin{align*}
\text{maximize} & \quad 1 - \langle N, \Omega \rangle \\
\text{subject to} & \quad N \in \mathbb{S}^{k+1}, \ \lambda \in \mathbb{R} \\
& \quad \lambda \geq 0 \\
& \quad N \succeq 0 \\
& \quad N \succeq \begin{bmatrix} 0 & 0 \\
0^T & 1 \end{bmatrix} + \lambda \begin{bmatrix} Q & \frac{1}{2} q^T \\
\frac{1}{2} q^T & q^0 - \gamma_n \end{bmatrix}.
\end{align*}
\]

However, if \( \lambda \) is equal to zero, the optimal objective function value of this maximization problem will be non-positive rendering the distributionally robust chance constraint (2.1) infeasible. Hence we
introduce new decision variables $\beta = -\frac{1}{\lambda}$ and $M = \frac{N}{\lambda}$ and again rewrite the maximization problem.

$$\max 1 + \frac{1}{\beta} \langle M, \Omega \rangle$$

subject to $M \in S^{k+1}$, $\beta \in \mathbb{R}$

$$\beta < 0$$

$$M \succeq 0$$

$$M \succeq \begin{bmatrix} Q & \frac{1}{2}q \\ \frac{1}{2}q^\top & q^0 - \gamma_n - \beta \end{bmatrix}.$$ 

By replacing the original $\beta$ with $\beta - \gamma_n$, the constraint $\inf_{P \in \mathcal{P}} \mathbb{P}(q^0 + q^\top \tilde{\xi} + \tilde{\xi}^\top Q \tilde{\xi} < \gamma_n) \geq 1 - \epsilon$ can be replaced with the following conditions.

$$\exists M \in S^{k+1}, \exists \beta \in \mathbb{R}, 1 + \frac{1}{\beta - \gamma_n} \langle M, \Omega \rangle \geq 1 - \epsilon,$$

$$\beta - \gamma_n < 0,$$

$$M \succeq 0,$$

$$M \succeq \begin{bmatrix} Q & \frac{1}{2}q \\ \frac{1}{2}q^\top & q^0 - \beta \end{bmatrix},$$

which is equivalent to

$$\exists M \in S^{k+1}, \exists \beta \in \mathbb{R}, \beta + \frac{1}{\epsilon} \langle M, \Omega \rangle \leq \gamma_n,$$

$$\beta < \gamma_n,$$

$$M \succeq 0,$$

$$M \succeq \begin{bmatrix} Q & \frac{1}{2}q \\ \frac{1}{2}q^\top & q^0 - \beta \end{bmatrix}.$$  

This existence of $M$ and $\beta$ is equivalent to the non-emptiness of the set $\mathcal{C}_n$ defined as

$$\mathcal{C}_n = \left\{ (\beta_n, M_n) \in \mathbb{R} \times S^{k+1} : \beta_n + \frac{1}{\epsilon} \langle M_n, \Omega \rangle \leq \gamma_n, \right\}$$

$$\beta_n < \gamma_n,$$

$$M_n \succeq 0,$$

$$M_n \succeq \begin{bmatrix} Q & \frac{1}{2}q \\ \frac{1}{2}q^\top & q^0 - \beta_n \end{bmatrix}.$$  

We first observe that $\{\mathcal{C}_n\}_{n=1}^\infty$ is a decreasing sequence of sets, i.e., $\mathcal{C}_{n+1} \subseteq \mathcal{C}_n \ \forall n \in \mathbb{N}$, since $\{\gamma_n\}_{n=1}^\infty$ is a decreasing sequence. Hence the distributionally robust quadratic chance constraint (2.1) holds
Chapter 2. Worst-Case Probabilistic Guarantee

if and only if \( \bigcap_{n=1}^{\infty} C_n \neq \emptyset \). We will now argue that \( \bigcap_{n=1}^{\infty} C_n \neq \emptyset \iff C \neq \emptyset \), where

\[
C = \left\{ (\beta, M) \in \mathbb{R} \times S^{k+1} : \begin{array}{l}
\beta + \frac{1}{\epsilon} (M, \Omega) \leq 0, \\
M \succeq 0, \\
M \succeq \begin{bmatrix} Q & \frac{1}{2} q \\
\frac{1}{2} q^T & q^0 - \beta \end{bmatrix}
\end{array} \right\}.
\]

To prove this, suppose that \( C \) is a non-empty set and pick any pair \((\beta, M)\) from \( C \). We know that \( \beta \leq 0 \) because the trace scalar product between two positive semidefinite matrices \( M \) and \( \Omega \) is nonnegative. Since \( \gamma_n > 0 \), we can then conclude that \( (\beta, M) \in C_n \) for every positive integer \( n \). For the reverse direction, suppose that there exists \((\beta, M) \in \bigcap_{n=1}^{\infty} C_n \). We then have

\[
M \succeq 0, \quad M \succeq \begin{bmatrix} Q & \frac{1}{2} q \\
\frac{1}{2} q^T & q^0 - \beta \end{bmatrix}
\]

and

\[
\beta + \frac{1}{\epsilon} (M, \Omega) \leq \gamma_n, \quad \beta < \gamma_n \quad \forall n \in \mathbb{N}.
\]

Therefore, \((\beta, M) \in C \) because \( \{\gamma_n\}_{n=1}^{\infty} \) converges to 0. Hence the distributionally robust quadratic chance constraint (2.1) holds if and only if

\[
\exists M \in S^{k+1}, \exists \beta \in \mathbb{R}, \quad \beta + \frac{1}{\epsilon} (M, \Omega) \leq 0, \\
M \succeq 0, \\
M \succeq \begin{bmatrix} Q & \frac{1}{2} q \\
\frac{1}{2} q^T & q^0 - \beta \end{bmatrix}.
\]

This is a very important result, and it will be used throughout the thesis. For other variants of the distributionally robust chance constraint, we refer the reader to Zymler et al. [55] and Hu and Hong [22].
Chapter 3

Portfolio Growth

We consider the situation where investors invest their money in a stock market which consists of \( n \) available stocks with the objective to let their wealth grow optimally in some sense. We further assume that the investment horizon is finite and consists of \( T \) rebalancing periods. A safety factor \( \epsilon \in (0, 1) \) reflects the degree of risk aversion of an investor. The detailed interpretation of \( \epsilon \) is discussed later, but we note here that, the lower \( \epsilon \) gets, the more risk averse an investor is. Each element of the vector \( \tilde{r}_t \in \mathbb{R}^n \) represents the rate of return of individual stock during the period \([t - 1, t]\) (0 denotes the present time). Moreover, \( \tilde{\omega}_t \) denotes the portfolio weights at the beginning of the period \([t - 1, t]\). The component \( \tilde{\omega}_{t,i} \) denotes the percentage of total wealth which is invested in asset \( i \). Let \( \mathcal{W} \) denote the set of admissible portfolios. In the simplest case, there are two constraints for portfolio \( \tilde{\omega}_t \in \mathcal{W} \): one of them prevents short selling (\( \tilde{\omega}_t \geq 0 \)), and the other is a budget constraint (\( \mathbf{1}^T \tilde{\omega}_t = 1 \)). However, one could also include other portfolio constraints. The only requirement is that \( \mathcal{W} \) must be representable by an intersection of finitely many second-order cones. In particular, \( \mathcal{W} \) can be any, but not limited to, convex polyhedron. In addition, we are interested in only non-anticipative and self-financing investment strategies. Non-anticipativity implies that portfolio decision is made from only previously observed information. It is a condition that reflects the inability of investors to look into the future. The latter condition (self-financing) requires that investors always reinvest all of their money during each rebalancing period.

Given a sequence of non-anticipative and self-financing portfolios \( \tilde{\omega} = \{\tilde{\omega}_t\}_{t=1}^T \) and a sequence of stock returns \( \tilde{r} = \{\tilde{r}_t\}_{t=1}^T \), we can determine how fast the value of the portfolio grows by determining
the growth rate which is defined as the logarithm of the investment’s geometric mean return

$$\gamma_T (\vec{w}, \vec{r}) = \log \left( \frac{T}{\prod_{t=1}^{T} (1 + w_t^T \vec{r}_t)} \right)$$

$$= \frac{1}{T} \sum_{t=1}^{T} \log (1 + w_t^T \vec{r}_t).$$

The growth rate $\gamma_T (\vec{w}, \vec{r})$ is a random variable because it inherits the uncertainties from the random returns $\vec{r}_t$. The realization of the growth rate is the quantity of our interest as it determines how fast the portfolio grows and the terminal wealth achieved. When the portfolio is constantly rebalanced, i.e., $\vec{w}_t = w$ $\forall t, 1 \leq t \leq T$, it is however possible to determine the value of this random growth rate in a situation where the investment horizon is infinitely long, i.e., $T \to \infty$, and the stock returns are serially independent and identically distributed. In such a situation, by the argument of the strong law of large numbers, the growth rate of any constant rebalanced portfolio $w \in \mathcal{W} \subset \mathbb{R}^n$, which by slight abuse of notation is denoted by $\gamma_T (w, \vec{r})$, converges to its expectation, i.e.,

$$\lim_{T \to \infty} \gamma_T (w, \vec{r}) = E (\log (1 + w^T \vec{r}_1)).$$

(3.1)

with probability 1, where the expectation is taken with respect to probability distribution of single-period rates of returns, say $\vec{r}_1$, without loss of generality. Particularly, the portfolio $w$ which maximizes the expected growth rate in the expression (3.1) is referred to as the log-optimal portfolio or the growth-optimal portfolio.

It is shown that, for a market where $\vec{r}$ is an independent and identically distributed stochastic process, the log-optimal portfolio outperforms any other investment strategy with probability 1 in the long run ($T \to \infty$) in the sense of the following theorem.

**Theorem 3.1** (Asymptotic Optimality of the Log-Optimal Portfolio). Let $\vec{r} = \{\vec{r}_t\}_{t=1}^{T}$ be a sequence of serially independent and identically distributed stock returns and let $w^*$ be the log-optimal investment strategy. Then, for any other causal portfolio strategy $\vec{w} = \{\vec{w}_t\}_{t=1}^{T}$,

$$\limsup_{T \to \infty} \left( \gamma_T (\vec{w}, \vec{r}) - \gamma_T (w^*, \vec{r}) \right) \leq 0 \text{ with probability 1.}$$

**Proof.** We refer to Cover and Thomas [12, Chapter 15] for the proof of this theorem.

Hence, there are three conditions for the log-optimal portfolio to shine which are:

1. The investment horizon is infinitely long.
2. The stock returns are serially independent and identically distributed.
3. The distribution of the stock returns is known.

The last condition is necessary and it enables us to calculate the log-optimal portfolio as well as its expected growth rate, while the other two conditions are the assumptions of Theorem 3.1. In this work, we establish a similar guarantee for the growth rate of the constant rebalanced investment strategy $w$ in a more realistic assumption where: the investment horizon consists of finitely many rebalancing periods and the market is neither serially independent nor identically distributed. The immediate difference from the prior argument is that the growth rate no longer converges. We therefore establish a robust guarantee of the growth rate from a conservative point of view by using Value-at-Risk (VaR) to map the random growth rate to a real value. VaR at level $\epsilon$ is an $\epsilon$-quantile of the input random parameter. Hence, we can maximize this so-called guarantee by solving the optimization problem

$$\sup_{w \in W} \mathbb{P} - \text{VaR}_\epsilon \left( \frac{1}{T} \sum_{t=1}^{T} \log (1 + w^\top \tilde{r}_t) \right),$$

or put differently,

$$\sup_{\gamma} \text{ subject to } w \in W, \gamma \in \mathbb{R}$$

$$\mathbb{P} \left( \frac{1}{T} \sum_{t=1}^{T} \log (1 + w^\top \tilde{r}_t) \geq \gamma \right) \geq 1 - \epsilon,$$

where $\epsilon$ is chosen from an interval $(0, 1)$ and is typically a small value, for example, 5% or 1% to reflect the risk aversion of an investor. Suppose that the optimal solution to this chance constrained program is $(w^*, \gamma^*)$. It is implied that by using a constant rebalanced investment strategy $w^*$ the value of the portfolio after $T$ rebalancing periods grows with a factor of $e^{T\gamma^*}$, at least, with probability $1 - \epsilon$.

**Definition 3.1 (Weak Sense Stationary Process).** Let $\tilde{\xi}_t$ be a stochastic process. $\tilde{\xi}_t$ is a weak sense stationary process, if its mean and its covariance matrix do not vary with respect to time.

The question remains is how to obtain the probability distribution $\mathbb{P}$. In many situations, the probability distribution $\mathbb{P}$ is estimated from historical data and therefore subject to uncertainty resulting from noisy data or estimation errors. It is however much easier to estimate some properties of $\mathbb{P}$ empirically than to estimate the entire distribution function. In addition, we also want to be robust against distributional assumption. Thus, it is reasonable to revise the maximization problem in order to take into account distributional ambiguities by considering a family of distributions
Chapter 3. Portfolio Growth

sharing the same known properties simultaneously.

\[
\sup_{\gamma} \quad \gamma
\]

subject to \( w \in W, \gamma \in \mathbb{R} \)

\[
\mathbb{P} \left( \frac{1}{T} \sum_{t=1}^{T} \log (1 + w^T \tilde{r}_t) \geq \gamma \right) \geq 1 - \epsilon \quad \forall \mathbb{P} \in \mathcal{P},
\]

where \( \mathcal{P} \) is the ambiguity set, i.e., the set of probability distributions which have some properties in common. In this work, we consider all probability distributions of the stock returns that have the same mean \( \mu \) and the same covariance matrix \( \Sigma \) because they can be determined from historical data. Precisely, the ambiguity set \( \mathcal{P} \) is defined as follows.

\[
\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0^T : \begin{array}{l}
\mathbb{E}_{\mathbb{P}} (\tilde{r}_t) = \mu \quad \forall t, 1 \leq t \leq T, \\
\mathbb{E}_{\mathbb{P}} (\tilde{r}_s \tilde{r}_t^T) = \delta_{st} \Sigma + \mu \mu^T \quad \forall s, t, 1 \leq s \leq t \leq T, \\
\mathbb{P} (\tilde{r}_t > -1) = 1 \quad \forall t, 1 \leq t \leq T
\end{array} \right\}, \quad (3.3)
\]

where \( \delta_{st} = 1 \) if \( s = t \) and \( = 0 \) otherwise. In other words, we treat stock returns as a weak sense stationary process because the first- and the second-order moments of \( \tilde{r}_t \) do not vary with time \( t \) and the stock returns are serially uncorrelated, a mild assumption compared to stock returns being an independent and identically distributed process (which is assumed in Theorem 3.1). The condition \( \mathbb{P} (\tilde{r}_t > -1) = 1 \) ensures that the companies will never go bankrupt. In the next section, we transform the maximization problem (3.2) into a deterministic second-order cone program, which is efficiently solvable.

### 3.1 Worst-Case Value-at-Risk of the Growth Rate

In this section, we analyze the Worst-Case Value-at-Risk (WCVaR) of the growth rate of any portfolio \( w \in W \), which is the optimal objective function value of the following optimization problem.

\[
\gamma^*(w) = \sup_{\gamma} \quad \gamma
\]

s.t. \( \mathbb{P} \left( \frac{1}{T} \sum_{t=1}^{T} \log (1 + w^T \tilde{r}_t) \geq \gamma \right) \geq 1 - \epsilon \quad \forall \mathbb{P} \in \mathcal{P} \)

In particular, we aim to find an analytical formula for approximating the WCVaR of the growth rate. Hence, the portfolio optimization model (3.2) can be rewritten as an optimization problem, where \( w \) is a decision, with WCVaR of the portfolio’s growth rate being its objective function. In order to do so, we henceforth assume that the following two assumptions are satisfied.

34
The covariance matrix of asset returns $\Sigma$ is strictly positive definite.

The mean vector $\mu$ and the covariance matrix $\Sigma$ satisfy $1 - w^T \mu > \sqrt{\frac{\epsilon}{(1-\epsilon)T}} \| \Sigma^{1/2} w \|$. These two assumptions are not restrictive and in most cases are satisfied. (A1) is a very realistic assumption. It implies that there is no asset that is truly risk-free or perfectly correlated with the remaining assets. (A2) is a technical assumption required in the derivation of the approximate $\gamma^*(w)$ discussed later (in Theorem 3.2). It is not a demanding assumption either, especially when the rebalancing interval is not too large and the mean values as well as the variances of stock returns are all close to 0. To be more precise, these assumptions are satisfied with the means and the standard deviations of yearly asset returns within their respective typical ranges discussed in Luenberger [34, Chapter 8]. Another example would be a set of portfolios provided in Guide to the Markets reports quarterly published on https://www.jpmorganfunds.com by J.P. Morgan Asset Management.

Under the assumption of the rebalancing interval being small, the quadratic approximation of the growth rate is justified. In addition, we further remove the support constraint from the definition of the ambiguity set in (3.3). The removal of the support results in a portfolio optimization model that is more conservative putting investors in a safer side. Therefore in this section, for the ease of exposition, we will assume that the condition $P(\tilde{r}_t > -1) = 1$ is absent. Hence applying the second-order Taylor approximation to the growth rate in (3.4) and omitting the support constraint give rise to the following optimization problem.

$$
\hat{\gamma}(w) = \sup_{\gamma} \gamma \\
\text{s.t. } P\left(\frac{1}{T} \sum_{t=1}^{T} \left(w^T \tilde{r}_t - \frac{1}{2} (w^T \tilde{r}_t)^2\right) \geq \gamma \right) \geq 1 - \epsilon \quad \forall P \in \mathcal{P},
$$

where, by a slight abuse of notation,

$$
\mathcal{P} = \left\{ P \in \mathcal{P}_0^n : \begin{array}{l}
\mathbb{E}_P(\tilde{r}_t) = \mu \quad \forall t, 1 \leq t \leq T, \\
\mathbb{E}_P(\tilde{r}_s \tilde{r}_t^\top) = \delta_{st} \Sigma + \mu \mu^\top \quad \forall s, t, 1 \leq s \leq t \leq T \end{array} \right\}.
$$

Then, we can then use the result for the distributionally robust quadratic chance constraint with the first- and the second-order moments information developed by Zymler et al. [55] to find a deterministic counterpart of (3.5). For the ease of exposition, we define a truncation operator $P_t \in \mathbb{R}^{n \times n_T}$ through $r_t = P_t r$. Thus $P_t$ maps the return path over all rebalancing periods, i.e., $r$, to the return over rebalancing period $t$. The arising problem is the following semidefinite program.
Chapter 3. Portfolio Growth

\[ \hat{\gamma}(w) = \max \gamma \]

s.t. \( M \in \mathbb{S}^{nT+1}, \beta \in \mathbb{R}, \gamma \in \mathbb{R} \)

\[ \beta + \frac{1}{\epsilon} \langle \Omega, M \rangle \leq 0 \quad (3.7a) \]

\[ M \succeq 0 \quad (3.7b) \]

\[ M - \begin{bmatrix} \frac{1}{2} \sum_{t=1}^{T} P_t^T w w^T P_t & -\frac{1}{2} \sum_{t=1}^{T} P_t^T w \\ -\frac{1}{2} \left( \sum_{t=1}^{T} P_t^T w \right)^T & \gamma T - \beta \end{bmatrix} \succeq 0, \quad (3.7c) \]

where \( \Omega \) is the second-order moment matrix of the random vector \( \tilde{r} \). Observe that (3.7c) contains a quadratic term in \( w \). However, we can linearize (3.7c) by partitioning \( M \) as

\[
\begin{bmatrix} V & v \\
v^T & v_0 \end{bmatrix},
\]

and rewriting (3.7c) as

\[
2M - \begin{bmatrix} 0 & 0 \\ 0 & 2\gamma T - T - 2\beta \end{bmatrix} \succeq \sum_{t=1}^{T} \begin{bmatrix} P_t^T w \\ -1 \end{bmatrix} \begin{bmatrix} P_t^T w \\ -1 \end{bmatrix}^T
\]

\[
= \begin{bmatrix} P_1^T w & P_2^T w & \cdots & P_T^T w \\ -1 & -1 & \cdots & -1 \end{bmatrix} \begin{bmatrix} P_1^T w & P_2^T w & \cdots & P_T^T w \end{bmatrix}^T.
\]

The above inequality constraint can, by a Schur complement argument, be rewritten as

\[
\begin{bmatrix} 2V & 2v \\ 2v^T & 2v_0 - 2\gamma T + T + 2\beta \end{bmatrix} \begin{bmatrix} P_1^T w & \cdots & P_T^T w \\ -1 & \cdots & -1 \end{bmatrix} \succeq 0.
\]

An evident benefit of the above derivation is that finding a portfolio \( w \) which maximizes \( \hat{\gamma}(w) \) becomes too a tractable semidefinite program. However, (3.7) contains a positive semidefinite constraint which involves matrices of dimension \( O(nT) \). This may seem discouraging when an investor want to solve this problem for a long term investment. In the remainder of the section, we show that this problem however can be substantially simplified, and after the simplification we derive an analytical formula for \( \hat{\gamma}(w) \).
Proposition 3.1 (General Projection Property). Let $\xi$ and $\zeta$ be random vectors valued in $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively, and define the ambiguity sets $\mathcal{P}_\xi$ and $\mathcal{P}_\zeta$ as

\begin{align*}
\mathcal{P}_\xi &= \{ P \in \mathcal{P}_p^0 : E_P \left( \left[ \xi^\top 1 \right] \left[ \xi^\top 1 \right] \right) = \Omega_\xi \}, \\
\mathcal{P}_\zeta &= \{ P \in \mathcal{P}_q^0 : E_P \left( \left[ \zeta^\top 1 \right] \left[ \zeta^\top 1 \right] \right) = \Omega_\zeta \},
\end{align*}

where $\Omega_\xi \in S_{+}^{p+1}$ and $\Omega_\zeta \in S_{+}^{q+1}$ are related through

\begin{equation}
\Omega_\zeta = \begin{bmatrix} \Lambda & 0 \\ 0^\top & 1 \end{bmatrix} \Omega_\xi \begin{bmatrix} \Lambda & 0 \\ 0^\top & 1 \end{bmatrix}^\top
\end{equation}

for some matrix $\Lambda \in \mathbb{R}^{q \times p}$. Then, for any Borel measurable function $f : \mathbb{R}^p \to \mathbb{R}$, we have

\begin{equation}
\inf_{P \in \mathcal{P}_\zeta} P \left( f(\zeta) \leq 0 \right) = \inf_{P \in \mathcal{P}_\xi} P \left( f(\Lambda \xi) \leq 0 \right).
\end{equation}

Proof. The projection property of probability distributions was first pioneered by Popescu [42, Proposition 1] and later generalized by Yu et al. [53]. This is an immediate consequence of Yu et al. [53, Theorem 1].

Proposition 3.1 implies that the worst-case probability given the first- and the second-order moments information remains unchanged under a linear transformation (with a linear transformation operator $\Lambda$). This result enables us to simplify the maximization problem (3.4) by identifying $w^\top \tilde{r}_t$ with a new random variable $\tilde{\eta}_t$ with mean $w^\top \mu$ and variance $w^\top \Sigma w$. Thus, we find

\begin{equation}
\hat{\gamma}(w) = \sup_{\gamma} \gamma \quad \text{s.t.} \quad P \left( \frac{1}{T} \sum_{t=1}^{T} \left( \tilde{\eta}_t - \frac{1}{2} \tilde{\eta}_t^2 \right) \geq \gamma \right) \geq 1 - \epsilon \quad \forall P \in \mathcal{P}_{\tilde{\eta}}, \tag{3.8}
\end{equation}

where

\begin{equation}
\mathcal{P}_{\tilde{\eta}} = \left\{ P \in \mathcal{P}_T^0 : \begin{array}{c} E_P(\tilde{\eta}_t) = w^\top \mu \quad \forall t, 1 \leq t \leq T, \\
E_P(\tilde{\eta}_t \tilde{\eta}_t^\top) = \delta_{st} w^\top \Sigma w + (w^\top \mu)^2 \quad \forall s,t, 1 \leq s \leq t \leq T \end{array} \right\}. \tag{3.9}
\end{equation}

Then, we can use the result for the distributionally robust quadratic chance constraint with the first- and the second-order moments information developed by Zymler et al. [55] to find a deterministic counterpart of the maximization problem (3.8). The arising problem is the following semidefinite
program.

\[
\tilde{\gamma}(w) = \max \gamma
\]

s.t. \[ M \in \mathbb{S}^{T+1}, \ \beta \in \mathbb{R}, \ \gamma \in \mathbb{R} \]
\[
\beta + \frac{1}{\epsilon} (\Omega(w), M) \leq 0
\]

\[
M \succeq 0
\]

\[
M - \begin{bmatrix}
\frac{1}{2} \mathbb{I} & -\frac{1}{2} \mathbb{1}^\top \\
-\frac{1}{2} \mathbb{1} & \gamma T - \beta
\end{bmatrix} \succeq 0,
\]

where \( \Omega(w) \in \mathbb{S}^{T+1} \) is the second-order moment matrix of the random vector \( \tilde{\eta} \triangleq [\tilde{\eta}_1, \tilde{\eta}_2, \ldots, \tilde{\eta}_T]^\top \) and is equal to

\[
\Omega(w) = \begin{bmatrix}
(w^\top \Sigma w + (w^\top \mu)^2 & (w^\top \mu)^2 & \ldots & (w^\top \mu)^2 & w^\top \mu \\
(w^\top \mu)^2 & w^\top \Sigma w + (w^\top \mu)^2 & \ldots & (w^\top \mu)^2 & w^\top \mu \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
(w^\top \mu)^2 & (w^\top \mu)^2 & \ldots & w^\top \Sigma w + (w^\top \mu)^2 & w^\top \mu \\
w^\top \mu & w^\top \mu & \ldots & w^\top \mu & 1
\end{bmatrix}.
\]

**Definition 3.2** (Compound Symmetry). A matrix \( M \in \mathbb{S}^{T+1} \) is compound symmetric if it is representable as

\[
M = \begin{bmatrix}
\tau_1 & \tau_2 & \ldots & \tau_2 & \tau_3 \\
\tau_2 & \tau_1 & \ldots & \tau_2 & \tau_3 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\tau_2 & \tau_2 & \ldots & \tau_1 & \tau_3 \\
\tau_3 & \tau_3 & \ldots & \tau_3 & \tau_4
\end{bmatrix}
\]

for some \( \tau_1, \tau_2, \tau_3, \tau_4 \in \mathbb{R} \).

Note that the second-order moment matrix \( \Omega(w) \) is compound symmetric because of the temporal symmetry of the random returns. More generally, the second-order moment matrix of any univariate weak sense stationary process is compound symmetric. Below we argue that there exists an optimal decision matrix \( M \) in (3.10) that is compound symmetric. This has a powerful consequence as the number of decision variables needed to encode \( M \in \mathbb{S}^{T+1} \) reduces from \( \frac{1}{2} T (T + 1) \) to just 4.
Chapter 3. Portfolio Growth

Proposition 3.2. There exists a maximizer \((M, \beta, \gamma)\) of (3.10) where \(M\) is compound symmetric.

Proof. Denote by \(\Pi^{T+1}\) the set of all permutations \(\pi\) of the integers \(\{1, 2, \ldots, T + 1\}\) with \(\pi(T+1) = T + 1\). For any permutation \(\pi \in \Pi^{T+1}\), we also define a permutation matrix as a matrix obtained by permuting rows and columns of the identity matrix \(I \in S^{T+1}\) according to the permutation \(\pi\).

\[
(P_\pi)_{ij} = \begin{cases} 
1 & \text{if } \pi(i) = j, \\
0 & \text{otherwise}
\end{cases}
\]

A permutation matrix can be any matrix that precisely contains a single 1 and 0s elsewhere in each row and each column. \(P_\pi^T\) is therefore a permutation matrix corresponding to an inverse of \(\pi\). A matrix \(K \in S^{T+1}\) is compound symmetric if and only if \(K = P_\pi K P_\pi^T\) for all \(\pi \in \Pi^{T+1}\). Suppose that \((M, \beta, \gamma)\) is a maximizer of (3.10). Since the input matrices in (3.10) are compound symmetric and \(P_\pi\) is non-singular, we have that

\[
\begin{align*}
M - \begin{bmatrix}
\frac{1}{2}I & -\frac{1}{2}
\end{bmatrix} & \geq 0 \iff P_\pi \left( M - \begin{bmatrix}
\frac{1}{2}I & -\frac{1}{2}
\end{bmatrix} \right) P_\pi^T \geq 0 \\
& \iff P_\pi M P_\pi^T - \begin{bmatrix}
\frac{1}{2}I & -\frac{1}{2}
\end{bmatrix} \geq 0.
\end{align*}
\]

Also, since \(\Omega(w)\) is compound symmetric, by using the cycling property of the trace operator we obtain

\[
\langle \Omega(w), M \rangle = \text{Tr}(M\Omega(w)) = \text{Tr}(M P_\pi^T \Omega(w)) = \text{Tr}(P_\pi M P_\pi^T \Omega(w)) = \langle \Omega(w), P_\pi M P_\pi^T \rangle.
\]

Hence, \((P_\pi M P_\pi^T, \beta, \gamma)\), for any permutation \(\pi \in \Pi^{T+1}\), is feasible in (3.10) and yields the same objective value as the optimal solution \((M, \beta, \gamma)\) does. It is therefore also a maximizer of (3.10). Furthermore, we observe that if have several \(M\)'s that are all optimal in (3.10) with the same \(\beta\) and \(\gamma\), then a convex combination between these \(M\)'s constitutes another optimal solution of (3.10).

These observations allow us to produce another maximizer of (3.10) which is

\[
M' = \frac{1}{T!} \sum_{\pi \in \Pi^{T+1}} P_\pi M P_\pi^T.
\]

Moreover, \(M'\) is invariant under any permutation \(\rho \in \Pi^{T+1}\), i.e., \(P_\rho M' P_\rho^T = M'\) because the set \(\{P_\rho P_\pi : \pi \in \Pi^{T+1}\}\) contains \(T!\) different permutation matrices (as \(P_\rho\) is invertible) with the last element fixed and therefore coincides with \(\Pi^{T+1}\). The proof thus completes. \[\square\]
Proposition 3.2 allows us to, without loss of generality, replace $M$ in (3.10) with a compound symmetric matrix defined through (3.12) everywhere. Therefore, all matrices in (3.10) are compound symmetric. Furthermore, each of the semidefinite constraints can be simplified with the help of Proposition 3.3, which enables us to verify the positive semidefiniteness of any compound symmetric matrices through a set of simple linear and quadratic constraints.

**Proposition 3.3.** For any compound symmetric matrix $M \in S^{T+1}$ of the form (3.12), we have that

$$M \succeq 0 \iff \begin{cases} 
\tau_1 \geq \tau_2 \\
\tau_1 + (T-1)\tau_2 + \tau_4 \geq 0 \\
\tau_4 (\tau_1 + (T-1)\tau_2) \geq T\tau_3^2 
\end{cases} \tag{3.13a,b,c}$$

*Proof.* We use the well-known fact that a symmetric matrix is positive semidefinite if and only if all of its eigenvalues are nonnegative. First, it is easy to verify that any vector of the form $v = [v_1, v_2, \ldots, v_T, 0]^T$ with $\sum_{i=1}^{T} v_i = 0$ constitutes an eigenvector of $M$ with eigenvalue $\tau_1 - \tau_2$. Indeed, we have

$$Mv = \begin{bmatrix}
\tau_1 v_1 + \tau_2 (v_2 + v_3 + \cdots + v_T) \\
\tau_1 v_2 + \tau_2 (v_1 + v_3 + \cdots + v_T) \\
\vdots \\
\tau_1 v_T + \tau_2 (v_2 + v_3 + \cdots + v_{T-1}) \\
\tau_3 (v_1 + v_2 + \cdots + v_T)
\end{bmatrix} = \begin{bmatrix}
(\tau_1 - \tau_2) v_1 \\
(\tau_1 - \tau_2) v_2 \\
\vdots \\
(\tau_1 - \tau_2) v_T \\
0
\end{bmatrix} = (\tau_1 - \tau_2)v.$$

There are $T - 1$ linearly independent eigenvectors of the above type. Next, we assume first that $\tau_3 = 0$. In this case, the two remaining eigenvectors can be chosen as $[1, 1, \ldots, 1, 0]^T$ and $[0, 0, \ldots, 0, 1]^T$ with eigenvalues $\tau_1 + (T-1)\tau_2$ and $\tau_4$, respectively. Thus $M \succeq 0$ if and only if

$$\tau_1 - \tau_2 \geq 0, \quad \tau_1 + (T-1)\tau_2 \geq 0 \quad \text{and} \quad \tau_4 \geq 0
\iff \tau_1 - \tau_2 \geq 0, \quad \tau_1 + (T-1)\tau_2 + \tau_4 \geq 0 \quad \text{and} \quad \tau_4 (\tau_1 + (T-1)\tau_2) \geq T\tau_3^2
\tag{3.13b,c}$$

Assume now that $\tau_3 \neq 0$. In this case, the two remaining eigenvectors are representable as $v = [1, 1, \ldots, 1, v]^T$ for some $v \in \mathbb{R}$. Observe that $\lambda$ is an corresponding eigenvalue if and only if $Mv = \lambda v$, which is equivalent to

$$v \lambda = T\tau_3 + v\tau_4 \quad \text{and} \quad \lambda = \tau_1 + (T-1)\tau_2 + v\tau_3.$$ 

The first equation above thus implies that $v(\lambda - \tau_4) = T\tau_3 \neq 0$, and thus $v = \frac{T\tau_3}{\lambda - \tau_4}$. Substituting this expression for $v$ into the second equation above, we obtain

$$\lambda = \tau_1 + (T-1)\tau_2 + \frac{T\tau_3^2}{\lambda - \tau_4}.$$
Solving this equation for $\lambda$ yields the two eigenvalues

$$\lambda = \frac{1}{2} \left( \tau_1 + (T-1)\tau_2 + \tau_4 \pm \sqrt{(\tau_1 + (T-1)\tau_2 + \tau_4)^2 + 4(T\tau_3^2 - \tau_4(\tau_1 + (T-1)\tau_2))} \right).$$

(3.14a)

$$= \frac{1}{2} \left( \tau_1 + (T-1)\tau_2 + \tau_4 \pm \sqrt{(\tau_1 + (T-1)\tau_2 - \tau_4)^2 + 4T\tau_3^2} \right).$$

(3.14b)

Equation (3.14b) shows that the square root term constitutes a strictly positive real number. The two eigenvalues are thus nonnegative if and only if

$$\tau_1 + (T-1)\tau_2 + \tau_4 \geq 0 \quad \text{and} \quad (\tau_1 + (T-1)\tau_2)\tau_4 \geq T\tau_3^2.$$  

(3.15)

The second inequality in (3.15) ensures that the square root term in (3.14a) does not exceed $\tau_1 + (T-1)\tau_2 + \tau_4$. Thus $M \succeq 0$ if and only if $\tau_1 - \tau_2 \geq 0$ and (3.15) is satisfied.

Corollary 3.1. For any compound symmetric matrix $M \in S^{T+1}$ of the form (3.12), a semidefinite constraint $M \succeq 0$ is equivalent to a system of one linear constraint and one second-order cone constraint.

$$M \succeq 0 \iff \begin{cases} \tau_1 \geq \tau_2 \\ \tau_1 + (T-1)\tau_2 + \tau_4 \geq \sqrt{(\tau_1 + (T-1)\tau_2 - \tau_4)^2 + 4T\tau_3^2} \end{cases}$$

(3.16a) (3.16b)

Proof. By Boyd and Vandenberghe [7, Exercise 4.26], (3.13b) and (3.13c) are equivalent to the hyperbolic constraint (3.16b). The claim thus follows.

We are now ready to state the main result of this section. Specifically, Proposition 3.3 enables us to solve (3.10) analytically.

Theorem 3.2 (WCVaR of the Growth Rate). If (A1) and (A2) are satisfied, the Worst-Case Value-at-Risk of the approximate growth rate of portfolio $w$ is given by

$$\hat{\gamma}(w) = \frac{1}{2} \left[ \frac{T-1}{cT} - \left( 1 - w^T\mu + \sqrt{\frac{1-c}{cT} \left\| \Sigma^{1/2}w \right\|^2} \right)^2 - \frac{T-1}{cT} w^T\Sigma w \right].$$

(3.17)

Proof. For notational simplicity, we denote the mean and the standard deviation of the portfolio return by $r_p = w^T\mu$ and $\sigma_p = \sqrt{w^T\Sigma w}$, respectively. Hence, we have that $\sigma_p > 0$ because of (A1), and therefore dividing by $\sigma_p$ is always mathematically possible.

Applying Proposition 3.3 to both semidefinite constraints in (3.10) yields a non-linear program
with only six decision variables.

\[ \hat{\gamma}(w) = \max \gamma \]

s.t. \( \tau_1 \in \mathbb{R}, \tau_2 \in \mathbb{R}, \tau_3 \in \mathbb{R}, \tau_4 \in \mathbb{R}, \beta \in \mathbb{R}, \gamma \in \mathbb{R} \)

\[
\begin{align*}
\beta + \frac{1}{\epsilon} \left[ T \left( \sigma_p^2 + r_p^2 \right) \tau_1 + T(T-1)r_p^2 \tau_2 + 2T r_p \tau_3 + \tau_4 \right] & \leq 0 \\
\tau_1 - \tau_2 & \geq 0 \\
\tau_1 + (T-1) \tau_2 + \tau_4 & \geq 0 \\
\tau_4 (\tau_1 + (T-1) \tau_2) & \geq T \tau_3^2 \\
\left( \tau_1 - \frac{1}{2} \right) - \tau_2 & \geq 0 \\
\left( \tau_1 - \frac{1}{2} \right) + (T-1) \tau_2 + (\tau_4 - \gamma T + \beta) & \geq 0 \\
(\tau_4 - \gamma T + \beta) \left( \left( \tau_1 - \frac{1}{2} \right) + (T-1) \tau_2 \right) & \geq T \left( \tau_3 + \frac{1}{2} \right)^2
\end{align*}
\]

(3.18) is actually a second-order cone program thanks to Corollary 3.1. In the remainder of the proof
we solve (3.18) analytically. We first note that (3.18a) is binding at optimality. Otherwise, we could
always construct a new solution \((\tau_1, \tau_2, \tau_3, \tau_4, \beta, \gamma + \Delta)\) which remains feasible and improves
the objective function value \(\gamma\) for a sufficiently small \(\Delta > 0\). Moreover, (3.18b) is redundant in view
of (3.18e) and can thus be dropped.

Finally, we can prove that there exists an optimal solution for which (3.18e) is binding. Indeed,
for any optimal solution \((\tau_1, \tau_2, \tau_3, \tau_4, \beta, \gamma)\), we can construct another optimal solution \((\frac{\tau_1 + (T-1) \tau_2 - \frac{1}{2}}{T}, \tau_3, \tau_4, \beta, \gamma)\). Without loss of generality, we can thus assume that \(\tau_1 = \tau_2 + \frac{1}{2}\). In
summary, we conclude that

\[ \hat{\gamma}(w) = \max \gamma \]

s.t. \( \tau_1 \in \mathbb{R}, \tau_2 \in \mathbb{R}, \tau_3 \in \mathbb{R}, \tau_4 \in \mathbb{R}, \beta \in \mathbb{R}, \gamma \in \mathbb{R} \)

\[
\begin{align*}
\beta + \frac{1}{\epsilon} \left[ T \left( \sigma_p^2 + r_p^2 \right) \tau_1 + T(T-1)r_p^2 \tau_2 + 2T r_p \tau_3 + \tau_4 \right] & = 0 \\
\tau_1 - \tau_2 = \frac{1}{2} \\
\tau_1 + (T-1) \tau_2 + \tau_4 & \geq 0 \\
\tau_1 + (T-1) \tau_2 + \tau_4 & \geq \frac{1}{2} + \gamma T - \beta \\
\tau_4 (\tau_1 + (T-1) \tau_2) & \geq T \tau_3^2 \\
(\tau_4 - \gamma T + \beta) \left( \left( \tau_1 - \frac{1}{2} \right) + (T-1) \tau_2 \right) & \geq T \left( \tau_3 + \frac{1}{2} \right)^2
\end{align*}
\]

42
We further simplify the problem by using the equality constraints (3.19a) and (3.19b) to eliminate some decision variables. Particularly, we use (3.19b) to remove $\tau_1$ and then we use (3.19a) to remove $\tau_4$ as, at optimality,

$$
\tau_4 = -\beta \epsilon - T(\sigma_p^2 + r_p^2)\tau_1 - T(T - 1)r_p^2\tau_2 - 2Tr_p\tau_3 \\
= -\beta \epsilon - \frac{T}{2}(\sigma_p^2 + r_p^2) - T(\sigma_p^2 + Tr_p^2)\tau_2 - 2Tr_p\tau_3 \\
= -(\theta_0 + \theta_2\tau_2 + \theta_3\tau_3 + \beta \epsilon),
$$

(3.20)

where $\theta_0, \theta_2$ and $\theta_3$ are auxiliary constants defined as follows.

$$
\theta_0 = \frac{T}{2} (\sigma_p^2 + r_p^2), \quad \theta_2 = T (\sigma_p^2 + Tr_p^2), \quad \theta_3 = 2Tr_p
$$

(3.21)

The arising optimization problem thus consists of only four decision variables and can be solved by Karush-Kuhn-Tucker (KKT) optimality conditions.

$$
\hat{\gamma}(w) = \max \gamma \\
\text{s.t.} \quad \tau_2 \in \mathbb{R}, \quad \tau_3 \in \mathbb{R}, \quad \beta \in \mathbb{R}, \quad \gamma \in \mathbb{R} \\
-\frac{1}{2} - T\tau_2 + \theta_0 + \theta_2\tau_2 + \theta_3\tau_3 + \beta \epsilon \leq 0 \\
\gamma T - \beta - T\tau_2 + \theta_0 + \theta_2\tau_2 + \theta_3\tau_3 + \beta \epsilon \leq 0 \\
\tau_2 (\theta_0 + \theta_2\tau_2 + \theta_3\tau_3 + \beta \epsilon + \gamma T - \beta) + \left(\tau_3 + \frac{1}{2}\right)^2 \leq 0 \\
\left(\tau_2 + \frac{1}{2T}\right) (\theta_0 + \theta_2\tau_2 + \theta_3\tau_3 + \beta \epsilon) + \tau_3^2 \leq 0
$$

(3.22)

(3.22) is still a second-order cone program with two hyperbolic constraints, and the Lagrangian function associated with (3.22) is

$$
L(\tau_2, \tau_3, \beta, \gamma, \alpha_1, \alpha_2, \delta, \lambda) = -\gamma + \alpha_1 \left(-\frac{1}{2} - T\tau_2 + \theta_0 + \theta_2\tau_2 + \theta_3\tau_3 + \beta \epsilon\right) + \\
\alpha_2 \left(\gamma T - \beta - T\tau_2 + \theta_0 + \theta_2\tau_2 + \theta_3\tau_3 + \beta \epsilon\right) + \\
\delta \left((\theta_0 + \theta_2\tau_2 + \theta_3\tau_3 + \beta \epsilon + \gamma T - \beta)\tau_2 + \left(\tau_3 + \frac{1}{2}\right)^2\right) + \\
\lambda \left((\theta_0 + \theta_2\tau_2 + \theta_3\tau_3 + \beta \epsilon)\left(\tau_2 + \frac{1}{2T}\right) + \tau_3^2\right),
$$

where $\alpha_1, \alpha_2, \delta$ and $\lambda$ represent the Lagrange multipliers of the constraints in (3.22). Since (3.22) is a convex optimization problem, any point that satisfies its KKT optimality conditions solves (3.22). In the remainder of this proof, we show that there exists $(\tau_2, \tau_3, \beta, \gamma, \alpha_1, \alpha_2, \delta, \lambda)$ that satisfies these conditions where $\alpha_1 = \alpha_2 = 0$. Then, we argue that this solution is one of the optimal solutions (if
there are many). When both $\alpha_1$ and $\alpha_2$ are zero, the KKT conditions, i.e., primal feasibility ((P1) - (P4)), dual feasibility ((D1) - (D2)), complementary slackness ((C1) - (C2)), and stationarity ((S1) - (S4)), are:

\[
- \frac{1}{2} - T\tau_2 + \theta_0 + \theta_2\tau_2 + \theta_3\tau_3 + \beta \epsilon \leq 0 \quad \text{(P1)}
\]
\[
\gamma T - \beta - T\tau_2 + \theta_0 + \theta_2\tau_2 + \theta_3\tau_3 + \beta \epsilon \leq 0 \quad \text{(P2)}
\]
\[
(\theta_0 + \theta_2\tau_2 + \theta_3\tau_3 + \beta \epsilon + \gamma T - \beta)\tau_2 + \left(\tau_3 + \frac{1}{2}\right)^2 \leq 0 \quad \text{(P3)}
\]
\[
(\theta_0 + \theta_2\tau_2 + \theta_3\tau_3 + \beta \epsilon)\left(\tau_2 + \frac{1}{2T}\right) + \tau_2^2 \leq 0 \quad \text{(P4)}
\]
\[
\delta \left(\theta_0 + \theta_2\tau_2 + \theta_3\tau_3 + \beta \epsilon + \gamma T - \beta\right)\tau_2 + \left(\tau_3 + \frac{1}{2}\right)^2 = 0 \quad \text{(C1)}
\]
\[
\lambda \left(\theta_0 + \theta_2\tau_2 + \theta_3\tau_3 + \beta \epsilon\right) \left(\tau_2 + \frac{1}{2T}\right) + \tau_3^2 = 0 \quad \text{(C2)}
\]
\[
\lambda \left(\theta_3 \left(\tau_2 + \frac{1}{2T}\right) + 2\tau_3\right) + \delta \left(\theta_3\tau_2 + 2\tau_3 + 1\right) = 0 \quad \text{(S1)}
\]
\[
\lambda \epsilon \left(\tau_2 + \frac{1}{2T}\right) + \delta \epsilon\tau_2 - \delta \tau_2 = 0 \quad \text{(S2)}
\]
\[
-1 + \delta \tau_2 T = 0 \quad \text{(S3)}
\]
\[
\delta \geq 0 \quad \text{(D1)}
\]
\[
\lambda \geq 0 \quad \text{(D2)}
\]

(S4) implies that $\tau_2 \neq 0$ and $\delta = \frac{1}{\tau_2 T} \neq 0$ and thus, by (S3), $\tau_2 + \frac{1}{2T} \neq 0$ and $\lambda = \frac{1 - \epsilon}{\epsilon \left(\tau_2 + \frac{1}{2T}\right) T} \neq 0$. Hence, (P3) and (P4) are automatically satisfied in view of (C1) and (C2) respectively, and (D1) and (D2) are satisfied as long as $\tau_2 > 0$. Moreover, we can simplify the system of equations (C1), (C2), (S1), and (S2) as

\[
\begin{align*}
(-\tau_4 + \gamma')\tau_2 + \left(\tau_3 + \frac{1}{2}\right)^2 &= 0 \quad \text{(3.23a)} \\
-\tau_4 \left(\tau_2 + \frac{1}{2T}\right) + \tau_3^2 &= 0 \quad \text{(3.23b)} \\
\frac{1 - \epsilon}{\epsilon \left(\tau_2 + \frac{1}{2T}\right)} \left(-\tau_4 + \theta_2\tau_2 + \frac{\theta_2}{2T}\right) + \frac{1}{\tau_2} \left(-\tau_4 + \theta_2\tau_2 + \gamma'\right) &= 0 \quad \text{(3.23c)} \\
\Leftrightarrow \theta_2 + \frac{\epsilon (\gamma' - \tau_4)}{\tau_2} - \frac{(1 - \epsilon)\tau_4}{\tau_2 + \frac{1}{2T}} &= 0 \\
\frac{\theta_3}{\epsilon} + \frac{2(1 - \epsilon)\tau_3}{\epsilon \left(\tau_2 + \frac{1}{2T}\right)} + \frac{2\tau_3 + 1}{\tau_2} &= 0, \quad \text{(3.23d)}
\end{align*}
\]
where $\tau_4$ is given by (3.20) and $\gamma'$ is a newly introduced variable defined as

$$\gamma' = \gamma T - \beta. \quad (3.24)$$

We determine the optimal solution of (3.22) constructively by showing that the following solution solves (3.23a), (3.23b), (3.23c), and (3.23d). The construction is given by

$$\tau_2 = \frac{\epsilon[(1 - \epsilon)(1 - r_p) - c\sigma_p]}{2cT\sigma_p} \quad (3.25a)$$

$$\tau_3 = \frac{(\epsilon r_p - c\sigma_p - \epsilon)(r_p - \epsilon r_p + c\sigma_p)}{2c\sigma_p} \quad (3.25b)$$

$$\tau_4 = \frac{-\tau_3 (r_p - \epsilon r_p + c\sigma_p) T}{1 - \epsilon} \quad (3.25c)$$

$$\gamma' = \frac{(\tau_3 + \frac{1}{\epsilon})(\epsilon r_p - c\sigma_p) T}{\epsilon} + \tau_4 \quad (3.25d)$$

where $c$ is another auxiliary positive constant defined as

$$c = \sqrt{\frac{\epsilon(1 - \epsilon)}{T}}. \quad (3.26)$$

Then, we show that this solution also satisfies the other postulated conditions, which allows us to conclude that this solution is optimal in (3.19). Lastly, we determine the objective value $\gamma$. The detailed steps of the verification are:

1. Ensure that the constructed solution solves the system of equations (3.23).

2. Ensure that $\tau_2$ is strictly positive resulting in (D1) and (D2) as well as (P1) and (P2) being satisfied.

3. Derive the WCVaR of the approximate growth rate of the portfolio $w$, i.e., $\hat{\gamma}(w)$, which is the objective function value of (3.22).
Verification of (3.23a)

\[-\tau_4 + \gamma^\prime \tau_2 + \left( \tau_3 + \frac{1}{2} \right)^2 \]

\[= \left( \tau_3 + \frac{1}{2} \right) \left( \frac{\epsilon r_p - c\sigma_p}{T\tau_2} \right) + \left( \tau_3 + \frac{1}{2} \right)^2 \]

(by (3.25d))

\[= \left( \tau_3 + \frac{1}{2} \right) \left( \frac{(\epsilon r_p - c\sigma_p)((1 - \epsilon)(1 - r_p) - c\sigma_p) + (\epsilon r_p - c\sigma_p - \epsilon)(\epsilon r_p - c\sigma_p + c\sigma_p) + \frac{1}{2}}{2c\sigma_p} \right) \]

(by (3.25a) & (3.25b))

\[= \left( \tau_3 + \frac{1}{2} \right) \left( \frac{\epsilon r_p - c\sigma_p}{2c\sigma_p} + \frac{1}{2} \right) \]

\[= 0 \]

Verification of (3.23a)

\[-\tau_4 \left( \tau_2 + \frac{1}{2T} \right) + \tau_3^2 \]

\[= \tau_3 \left( \frac{r_p - \epsilon r_p + c\sigma_p}{1 - \epsilon} \right) \left( \tau_2 + \frac{1}{2T} \right) + \tau_3^2 \]

(by (3.25c))

\[= \tau_3 \left( \frac{(r_p - \epsilon r_p + c\sigma_p) T}{1 - \epsilon} \left( \tau_2 + \frac{1}{2T} \right) + \tau_3 \right) \]

\[= \tau_3 \left( \frac{(r_p - \epsilon r_p + c\sigma_p) T}{1 - \epsilon} \left( \epsilon(1 - \epsilon)(1 - r_p) + \frac{1}{2T} \epsilon(1 - \epsilon)(1 - r_p) \right) + \frac{(r_p - \epsilon r_p + c\sigma_p - \epsilon)(r_p - \epsilon r_p + c\sigma_p)}{2c\sigma_p} \right) \]

(by (3.25a) & (3.25b))

\[= \tau_3 \left( \frac{(r_p - \epsilon r_p + c\sigma_p) \epsilon(1 - r_p) + (\epsilon r_p - c\sigma_p - \epsilon)(r_p - \epsilon r_p + c\sigma_p)}{2c\sigma_p} + \frac{r_p - \epsilon r_p + c\sigma_p}{2} \right) \]

\[= \tau_3 \left( \frac{-r_p - \epsilon r_p + c\sigma_p}{2} + \frac{r_p - \epsilon r_p + c\sigma_p}{2} \right) \]

\[= 0 \]
Verification of (3.23c)

\[
\frac{\epsilon (\gamma' - \tau_4)}{\tau_2} \frac{(\tau_3 + \frac{1}{2}) (\epsilon r_p - c\sigma_p) T}{\tau_2 + \frac{1}{2T}} + \frac{\tau_3 (r_p - \epsilon r_p + c\sigma_p) T}{\tau_2 + \frac{1}{2T}} = T \left( \frac{(\tau_3 + \frac{1}{2}) (\epsilon r_p - c\sigma_p)}{\tau_2} + \frac{\tau_3 (r_p - \epsilon r_p + c\sigma_p)}{(1-\epsilon)(c-\epsilon r_p + c\sigma_p) 2cT \sigma_p} \right) \tag{by (3.25a)}
\]

\[
= T \left( \frac{(\tau_3 + \frac{1}{2}) (\epsilon r_p - c\sigma_p)}{\tau_2} \right) - \frac{(r_p - \epsilon r_p + c\sigma_p)^2 T}{1-\epsilon} \tag{by (3.25b)}
\]

\[
= T^2 \left( \frac{(\epsilon r_p - c\sigma_p - \epsilon) (r_p - \epsilon r_p + c\sigma_p) + \epsilon r_p - c\sigma_p}{\epsilon(1-\epsilon)(1-r_p) - c\sigma_p} \right) - \frac{(r_p - \epsilon r_p + c\sigma_p)^2}{1-\epsilon}
\]

\[
= T^2 \left( \frac{(1-\epsilon) [(\epsilon r_p - c\sigma_p - \epsilon) (r_p - \epsilon r_p + c\sigma_p) + \epsilon r_p - c\sigma_p]}{\epsilon(1-\epsilon)(1-r_p) - c\sigma_p} \right)
\]

\[
= T^2 \left( \frac{(r_p - \epsilon r_p + c\sigma_p) [c^2 \sigma_p^2 - \epsilon(1-\epsilon) r_p(1-r_p)] + (1-\epsilon) c\sigma_p (\epsilon r_p - c\sigma_p)}{\epsilon(1-\epsilon)(1-r_p) - c\sigma_p} \right)
\]

\[
= T^2 \left( \frac{-\epsilon(1-\epsilon) r_p(1-r_p) (r_p - \epsilon r_p + c\sigma_p) + \epsilon r_p - c\sigma_p}{\epsilon(1-\epsilon)(1-r_p) - c\sigma_p} \right)
\]

\[
= T^2 \left( \frac{-\epsilon(1-\epsilon) r_p(1-r_p) (r_p - \epsilon r_p + c\sigma_p) + \epsilon(1-\epsilon) c\sigma_p r_p}{\epsilon(1-\epsilon)(1-r_p) - c\sigma_p} \right) - \frac{c^2 \sigma_p^2}{\epsilon(1-\epsilon)}
\]

\[
= T^2 \left( \frac{-r_p(1-r_p) (r_p - \epsilon r_p + c\sigma_p) + cr_p \sigma_p}{(1-\epsilon)(1-r_p) - c\sigma_p} \right) - \frac{c^2 \sigma_p^2}{\epsilon(1-\epsilon)}
\]

\[
= T^2 \left( \frac{-r_p^2 - c^2 \sigma_p^2}{\epsilon(1-\epsilon)} \right)
\]

\[
= T^2 \left( \frac{-r_p^2 - \frac{c^2 \sigma_p^2}{T}}{\epsilon(1-\epsilon)} \right)
\]

[47]

(by (3.26))

(by (3.21))
We first observe the following equations.

\[
(2T r_p \tau_2 + r_p + 2 \tau_3 + \epsilon) \tau_2 + \frac{\epsilon (2 \tau_3 + 1)}{2T}
\]

Hence, we conclude the following result.

\[
\theta_3' + \frac{2(1 - \epsilon) \tau_3}{\epsilon (\tau_2 + \frac{1}{2T})} + \frac{2 \tau_3 + 1}{\tau_2}
\]

\[
= \frac{1}{\epsilon \tau_2 (\tau_2 + \frac{1}{2T})} \left( \theta_3 \tau_2 \left( \tau_2 + \frac{1}{2T} \right) + 2(1 - \epsilon) \tau_2 \tau_3 + \epsilon \left( \tau_2 + \frac{1}{2T} \right) (2 \tau_3 + 1) \right)
\]

\[
= \frac{1}{\epsilon \tau_2 (\tau_2 + \frac{1}{2T})} \left( \theta_3 \tau_2^2 + \left( \frac{\theta_3}{2T} + 2 \tau_3 + \epsilon \right) \tau_2 + \frac{\epsilon (2 \tau_3 + 1)}{2T} \right)
\]

\[
= \frac{1}{\epsilon \tau_2 (\tau_2 + \frac{1}{2T})} \left( 2T r_p \tau_2^2 + (r_p + 2 \tau_3 + \epsilon) \tau_2 + \frac{\epsilon (2 \tau_3 + 1)}{2T} \right)
\]

(by (3.21))

\[
= \frac{1}{\epsilon \tau_2 (\tau_2 + \frac{1}{2T})} \left( (2T r_p \tau_2 + r_p + 2 \tau_3 + \epsilon) \tau_2 + \frac{\epsilon (2 \tau_3 + 1)}{2T} \right)
\]

(by (3.25b))

\[
= 0
\]
Verification of the other KKT conditions: (P1), (P2), (D1), and (D2)
The assumption (A2) implies that \((1 - \epsilon)(1 - r_p) > c\sigma_p\). Hence, by (3.25a), we have that \(\tau_2 > 0\). It follows that (D1) and (D2) hold by the discussed arguments. Furthermore, by considering (P3), \(\tau_2\) being positive implies that \(\theta_0 + \theta_2 \tau_2 + \theta_3 \gamma + \beta \epsilon + \gamma T - \beta \leq 0\). Hence, (P2) indeed holds. Similarly from (\(\tau_2 > 0\)) and (P4), we obtain that \(\theta_0 + \theta_2 \tau_2 + \theta_3 \gamma + \beta \epsilon \leq 0\). (P1) thus holds too.

Determination of the WCVaR of the growth rate

\[
\epsilon T \gamma = \epsilon' + \epsilon \beta
\]

(by (3.24))

\[
e\epsilon' - \tau_4 - \theta_0 - \theta_2 \tau_2 - \theta_3 \gamma
\]

= \(\epsilon (\gamma' - \tau_4) - (1 - \epsilon) \tau_4 - \theta_0 - \theta_2 \tau_2 - \theta_3 \gamma\)

= \(\left( \tau_3 + \frac{1}{2} \right)(\epsilon r_p - c\sigma_p) T + \tau_3 (r_p - \epsilon r_p + c\sigma_p) T - \theta_0 - \theta_2 \tau_2 - \theta_3 \gamma\)

(by (3.25c) & (3.25d))

\[
= \frac{1}{2} (\epsilon r_p - c\sigma_p) T + Tr_p \tau_3 - \theta_0 - \theta_2 \tau_2 - \theta_3 \gamma
\]

(by (3.21))

Hence, by dividing \(T\) at both sides of the above equation and substituting the values for \(\theta_0\) and \(\theta_2\) from (3.21), we obtain the following results.

\[
\epsilon \gamma = \frac{1}{2} (\epsilon r_p - c\sigma_p) - \frac{\sigma_p^2 + r_p^2}{2} - \frac{(\sigma_p^2 + Tr_p^2) \tau_2 - r_p \tau_3}{2cT\sigma_p}
\]

(by (3.25a) & (3.25b))

\[
= \frac{1}{2} (\epsilon r_p - c\sigma_p) - \frac{\sigma_p^2 + r_p^2}{2} - \frac{(\sigma_p^2 + Tr_p^2)(\epsilon[(1 - \epsilon)(1 - r_p) - c\sigma_p])}{2cT\sigma_p}
\]

\[
- \frac{r_p (\epsilon r_p - c\sigma_p - \epsilon) (r_p - \epsilon r_p + c\sigma_p)}{2c\sigma_p}
\]

\[
= \frac{1}{2} (\epsilon r_p - c\sigma_p) - \frac{\sigma_p^2 + r_p^2}{2} - \frac{\epsilon (\sigma_p^2 + Tr_p^2)(1 - r_p) + (\epsilon r_p - c\sigma_p - \epsilon) [\epsilon (\sigma_p^2 + Tr_p^2) + Tr_p (r_p - \epsilon r_p + c\sigma_p)]}{2cT\sigma_p}
\]

\[
= \frac{1}{2} (\epsilon r_p - c\sigma_p) - \frac{\sigma_p^2 + r_p^2}{2} - \frac{\epsilon (\sigma_p^2 + Tr_p^2)(1 - r_p) + (\epsilon r_p - c\sigma_p - \epsilon) [\epsilon (\sigma_p^2 + Tr_p^2) + cTr_p \sigma_p]}{2cT\sigma_p}
\]

\[
= \frac{1}{2} (\epsilon r_p - c\sigma_p) - \frac{\sigma_p^2 + r_p^2}{2} - \frac{\epsilon (\sigma_p^2 + Tr_p^2)(1 - r_p) + (\epsilon r_p - c\sigma_p - \epsilon) [\epsilon (\sigma_p^2 + Tr_p^2 + cTr_p \sigma_p)]}{2cT\sigma_p}
\]

\[
= \frac{1}{2} (\epsilon r_p - c\sigma_p) - \frac{\sigma_p^2 + r_p^2}{2} - \frac{-cTr_p^2 \sigma_p + \epsilon (1 - r_p) \sigma_p^2 + (\epsilon r_p - c\sigma_p - \epsilon) [\epsilon \sigma_p^2 + cTr_p \sigma_p]}{2cT\sigma_p}
\]

49
and, by further algebraic manipulations,

\[
\begin{align*}
\epsilon \gamma &= \frac{1}{2} (\epsilon r_p - c \sigma_p) - \frac{\sigma_p^2}{2} - \frac{\epsilon (1 - r_p) \sigma_p + (\epsilon r_p - c \sigma_p - \epsilon) (\epsilon \sigma_p + c T r_p)}{2c T} \\
&= \frac{1}{2} (\epsilon r_p - c \sigma_p) - \frac{\sigma_p^2}{2} - \frac{\epsilon (1 - r_p) (\sigma_p - c \sigma_p - c r_p - \epsilon \sigma_p^2 - c^2 T r_p \sigma_p)}{2c T} \\
&= \frac{1}{2} (\epsilon r_p - c \sigma_p) - \frac{\sigma_p^2}{2} - \frac{\epsilon (1 - r_p) - \epsilon (1 - r_p) \sigma_p}{2c T} + \frac{\sigma_p^2}{2} + \frac{\epsilon r_p \sigma_p}{2} \\
&= \frac{\epsilon}{2} (1 - (1 - r_p)^2 - (1 - r_p) \sigma_p) (\frac{1 - \epsilon}{c T} + \frac{c}{T}) - \frac{1}{2} \sigma_p^2 + \frac{\epsilon \sigma_p^2}{2T} \\
\end{align*}
\]

(by (3.26))

\[
\begin{align*}
\hat{\gamma}(w) &= \frac{1}{2} \left( 1 - \left(1 - w^T \mu + \sqrt{\frac{1 - \epsilon}{c T}} \left\| \Sigma^{1/2} w \right\| \right)^2 - \frac{T - 1}{c T} w^T \Sigma w \right). \\
\end{align*}
\]  

(3.35)

Thus, the constructed solution leads to an optimal objective function value which is equal to

The proof thus completes. ■

**Corollary 3.2.** If the inclusion \( w \in W \) is representable by a finite number of second-order cone constraints and (A1) and (A2) are satisfied, then the problem of finding a portfolio which maximizes the WCVaR of the approximate growth rate

\[
w^* \in \operatorname{argmax}_{w \in W} \hat{\gamma}(w) \\
(3.36)
\]

is a second-order cone program.

**Proof.** Following from (3.35),

\[
w^* \in \operatorname{argmax}_{w \in W} \hat{\gamma}(w) \iff w^* \in \operatorname{argmin}_{w \in W} \left( \left(1 - w^T \mu + \sqrt{\frac{1 - \epsilon}{c T}} \left\| \Sigma^{1/2} w \right\| \right)^2 + \frac{T - 1}{c T} \left\| \Sigma^{1/2} w \right\|^2 \right).
\]

Equivalently put, \( w^* \) minimizes a sum of the squares of two nonnegative terms (see assumption (A2)). By introducing a new variable to denote each square, the objective becomes linear, and the
Chapter 3. Portfolio Growth

claim follows as \( w^* \) also solves the following second-order cone program

\[
\begin{align*}
\text{minimize} & \quad a + \frac{T - 1}{\epsilon T} b \\
\text{subject to} & \quad w \in \mathcal{W}, \ a \in \mathbb{R}, \ b \in \mathbb{R} \\
& \quad a \geq c^2 \\
& \quad b \geq d^2 \\
& \quad c \geq 1 - w^\top \mu + \sqrt{\frac{1 - \epsilon}{\epsilon T}} \mathbf{1}^\top \Sigma^{1/2} w \\
& \quad d \geq \mathbf{1}^\top \Sigma^{1/2} w,
\end{align*}
\]

where the first and the second constraints are hyperbolic constraints and thus they can be reformulated as second-order cone constraints. \( \blacksquare \)

Corollary 3.2 implies that the portfolio optimization model (3.2) approximately reduces to a tractable second-order cone program.

**Remark 3.1** (Relation to the Markowitz model). (3.36) favours a high value of portfolio mean return \( w^\top \mu \) and a low value of portfolio variance \( w^\top \Sigma w \). This is expected since the analysis is based on the quadratic approximation of the portfolio growth rate. This result is similar to the classic Markowitz model where an investor aims to maximize \( \lambda w^\top \mu - \delta w^\top \Sigma w \) for some nonnegative parameters \( \lambda \) and \( \delta \). These values reflect the degree of risk aversion of an investor. However, the choice of \( \lambda \) and \( \delta \) is subjective unlike \( \epsilon \) in our model which can be interpreted as a probabilistic guarantee of the approximate growth rate.

**Remark 3.2** (Asymptotic behaviour of the model). In the limiting case where \( T \) approaches infinity, the expression of \( \hat{\gamma}(w) \) reduces to

\[
\frac{1}{2} - \frac{1}{2} (1 - w^\top \mu)^2 - \frac{1}{2} w^\top \Sigma w.
\]

Observe that the term \( \frac{1}{2} (1 - w^\top \mu)^2 \) is just a quadratic approximation of \( \log(1 + w^\top \mu) \), i.e., the logarithm of the portfolio mean return. We thus conclude that asymptotically the WCVaR of the approximate growth rate at level \( \epsilon \) is equal to the quadratic approximation of the logarithm of the portfolio mean return deducted by \( \frac{1}{2} w^\top \Sigma w \). The lower \( \epsilon \), the larger this risk penalty becomes.

To sum up, in this chapter, by using the quadratic approximation of the growth rate, we can approximately solve (3.2) by solving a highly tractable second-order cone program. The size of the arising second-order cone program (3.36) is independent of the number of rebalancing periods \( T \). Therefore, our model can be efficiently used even by long-term investors.
Chapter 4

Extensions

In this chapter, we develop extensions of the basic model. The aim is to enrich the model by incorporating other information about the probability distribution of the stock returns and information about transaction costs. Two extensions are presented in this chapter. First, we relax the assumption that the mean values and the (co)variances of the random stock returns are known. Rather, we assume that we have some estimates of these moments which are perhaps inaccurate because of outliers and noisy data, and we want to robustly reanalyze $\hat{\gamma}(w)$ in this situation. Second, we show how transaction costs and management fees can affect the WCVaR of the growth rate $\gamma^*(w)$.

4.1 Hedging against Moment Uncertainties

In Chapter 3 we assumed that the first two moments of the stock returns are known. We now consider the case where we only have some estimates of these moments and the true moments are not precisely known. Assume that $\hat{\mu}$ and $\hat{\Sigma}$ are our best estimates of the mean vector and the covariance matrix of the returns, respectively. We assume that the true mean $\mu$ and the true covariance matrix $\Sigma$ of stock returns lie in a convex confidence region parameterized by $\delta_1 \geq 0$ and $\delta_2 \geq 1 \geq \delta_3 > 0$.

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

(4.1)

$\delta_1$ determines how strong we believe in our mean estimate $\hat{\mu}$, whereas $\delta_2$ and $\delta_3$ determine how strong we believe in our variance estimate $\hat{\Sigma}$. This choice of characterizing moment uncertainties is motivated by Delage and Ye [14]. We define a refined WCVaR of the approximate growth rate under the worst-case moments allowed by $\mathcal{U}$ as

$$
\hat{\gamma}_{\mu}(w) = \min_{(\mu, \Sigma) \in \mathcal{U}} \frac{1}{2} \left( 1 - \left( 1 - w^T \mu + \sqrt{\frac{1 - \epsilon}{\epsilon T}} \left\| \Sigma^{1/2} w \right\| \right)^2 - \frac{T - 1}{\epsilon T} w^T \Sigma w \right),
$$

(4.2)
when the assumptions (A1) and (A2) are satisfied for all \((\mu, \Sigma) \in U\). An analytical expression for \(\hat{\gamma}_U(w)\) is derived in Theorem 4.1 with the help from Lemma 4.1.

**Lemma 4.1.** For a given \(\delta \in \mathbb{R}_+, x_0 \in \mathbb{R}^n, c \in \mathbb{R}^n\) and \(A \in S^n_{++}\),

1. \[
\min_{x \in X} c^T x = c^T x_0 - \left\| A^{-1/2} c \right\| \delta^{1/2}
\]
2. \[
\max_{x \in X} c^T x = c^T x_0 + \left\| A^{-1/2} c \right\| \delta^{1/2},
\]

where \(X\) is an ellipsoid centred at \(x_0\) defined as \(\{x : (x - x_0)^T A (x - x_0) \leq \delta\}\).

**Proof.** Optimizing a linear function over an ellipsoid is actually known to be an easy problem that admits an analytical solution. It is an important result that can be used in a wide range of applications. Particularly, it can be used to solve robust linear programs with ellipsoidal uncertainty sets (see Boyd and Vandenberghe [7, Section 4.4]). Here, we prove only the result for the maximization problem as the result for the minimization problem can be proved in a similar fashion. To begin with, we transform the maximization problem into another maximization problem whose feasible set is an ellipsoid centred at origin by simple variable transformation. Then by using that \(A\) is positive definite (and thus invertible), we can reshape the feasible set from an ellipsoid to a ball.

\[
\max_{x \in X} c^T x = \max_x c^T x \\
\text{s.t. } (x - x_0)^T A (x - x_0) \leq \delta \\
= \max_x c^T x + c^T x_0 \\
\text{s.t. } x^T A x \leq \delta \\
= \max_x c^T A^{-1/2} A^{1/2} x + c^T x_0 \\
\text{s.t. } x^T A^{1/2} A^{1/2} x \leq \delta \\
= \max_y c^T A^{-1/2} y + c^T x_0 \\
\text{s.t. } y^T y \leq \delta,
\]

which yields an optimal objective value equal to \(\left\| A^{-1/2} c \right\| \delta^{1/2} + c^T x_0\) due to the Cauchy-Schwarz inequality. The claim thus follows. \(\blacksquare\)

**Theorem 4.1.** If (A1) and (A2) are satisfied with adjusted mean vector \(\hat{\mu} + \sqrt{\delta_1 \Sigma w} \parallel \Sigma^{1/2} w \parallel\) and adjusted covariance matrix \(\delta_2 \hat{\Sigma}\), the Worst-Case Value-at-Risk of the approximate growth rate of portfolio \(w\) (4.2) where the true moments can take any value within \(\mathcal{U}\) of the form (4.1) is given by

\[
\hat{\gamma}_U(w) = \frac{1}{2} \left( 1 - \left( 1 - w^T \hat{\mu} + \left( \sqrt{\delta_1} + \sqrt{\frac{(1 - \epsilon) \delta_2}{\epsilon T}} \right) \left\| \Sigma^{1/2} w \right\|^2 - \frac{\delta_2 (T - 1)}{\epsilon T} w^T \hat{\Sigma} w \right) \right).
\]
Proof. We divide the proof into two parts. The first is a proof of validity of the assumption, while the second is a derivation of \( \hat{\gamma}_U(w) \). We begin the first part by expressing \( U \) as a Cartesian product of \( U_\mu \) and \( U_\Sigma \):

\[
U_\mu = \left\{ \mu : (\mu - \hat{\mu})^T \hat{\Sigma}^{-1} (\mu - \hat{\mu}) \leq \delta_1 \right\},
\]

\[
U_\Sigma = \left\{ \Sigma : \delta_3 \hat{\Sigma} \preceq \Sigma \preceq \delta_2 \hat{\Sigma} \right\}.
\]

Since \( U_\Sigma \) contains only positive definite matrices, the assumption (A1) is satisfied. To verify that (A2) is satisfied for all \((\mu, \Sigma) \in U\), we require

\[
1 - w^T \mu > \sqrt{\frac{\epsilon}{(1 - \epsilon)^T}} \left\| \Sigma^{1/2} w \right\| \quad \forall \mu \in U_\mu, \Sigma \in U_\Sigma.
\]

Put differently, we need

\[
\max_{\mu \in U_\mu} w^T \mu + \sqrt{\frac{\epsilon}{(1 - \epsilon)^T}} \left( \max_{\Sigma \in U_\Sigma} \left\| \Sigma^{1/2} w \right\| \right) < 1,
\]

which in turns is equivalent to

\[
w^T \hat{\mu} + \sqrt{\delta_1} \left\| \hat{\Sigma}^{1/2} w \right\| + \sqrt{\frac{\epsilon \delta_2}{(1 - \epsilon)^T}} \left\| \hat{\Sigma}^{1/2} w \right\| < 1
\]

because the first maximization can be solved analytically due to Lemma 4.1 and the second maximization admits a trivial solution \( \Sigma = \delta_3 \hat{\Sigma} \) as

\[
\left\| \Sigma^{1/2} w \right\|^2 = w^T \Sigma w \leq \delta_2 w^T \hat{\Sigma} w.
\]

Hence, (A1) and (A2) are satisfied for all \((\mu, \Sigma) \in U\) if and only if (A1) and (A2) are satisfied with the adjusted mean vector and the adjusted covariance matrix given above.

We now prove the second part. (4.2) is in favour of a large value of \( \left\| \Sigma^{1/2} w \right\| \) and a small value of \( w^T \mu \). By using similar arguments as in the first part, we can derive an analytical expression of \( \hat{\gamma}_U(w) \). Hence, the claim follows.

Remark 4.1 (Other forms of moment uncertainties). For any constant rebalanced investment strategy \( w \in W \), we derive an analytical expression of the WCVaR of the approximate growth rate of it in Theorem 3.2. Theorem 4.1 further tells us that, by hedging against moment uncertainties of the form (4.1), it is still possible to derive the WCVaR of \( w \)'s approximate growth rate. It is worth noting here that the presented methodology can be applied to other forms of moment uncertainties. For example, we can specify the uncertainties in mean values either way presented in Table 4.1 and specify the uncertainties in (co)variances either way presented in Table 4.2.
Table 4.1: Tractable forms of uncertainty in mean values ($\mu$)

<table>
<thead>
<tr>
<th></th>
<th>Mean uncertainties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Box</td>
<td>$\mu \leq \mu \leq \bar{\mu}$</td>
</tr>
<tr>
<td>Ellipsoid</td>
<td>$(\mu - \hat{\mu})^T \hat{\Sigma}^{-1} (\mu - \hat{\mu}) \leq \delta$ ($\delta \geq 0$)</td>
</tr>
</tbody>
</table>

Table 4.2: Tractable forms of uncertainty in (co)variances ($\Sigma$)

<table>
<thead>
<tr>
<th></th>
<th>(co)Variances uncertainties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Box</td>
<td>$\Sigma \leq \Sigma \leq \bar{\Sigma}$, $\Sigma &gt; 0$ ($\Sigma, \bar{\Sigma} &gt; 0$)</td>
</tr>
<tr>
<td>Matrix cone</td>
<td>$\Sigma \leq \Sigma \leq \bar{\Sigma}$ ($\Sigma, \bar{\Sigma} &gt; 0$)</td>
</tr>
</tbody>
</table>

4.2 Modeling Management Fees

In this section, we briefly discuss the effects of management fees (see for example Bielecki and Pliska [5] and Morton and Pliska [40]). In a market with management fees, an investor is periodically charged proportionally to his or her wealth at a fixed rate $\alpha \in (0, 1)$. It can be seen that the terminal wealth achieved by a constant rebalanced investment strategy $w$ is $(1 - \alpha)^T$ of the wealth achieved by the same portfolio when there are no management fees. Thus, for any $w \in W$, the growth rate of $w$ is reduced by $-\log(1 - \alpha)$ regardless of the sequence of asset returns $\tilde{r}$. This observation implies that the optimal constant rebalanced investment strategy in the view of (3.2) remains invariant in the presence of management fees.

4.3 Modeling Transaction Costs

In contrast to the earlier analysis where it is always assumed that trading is costless, we now use linear programming duality to study the impact of transaction costs on the WCVaR of the growth rate given the first- and the second-order moments of the serially uncorrelated stock returns.

**Theorem 4.2.** For a constant rebalanced investment strategy $w \in W$, the difference between logarithmic portfolio return when there is no transaction cost and that when there is a proportional transaction cost of rate $c$ is no more than $\log \left( \frac{1+c}{1-c} \right)$ for each rebalancing period.

**Proof.** To facilitate the following exposition, we denote by $x_{t-1}$ the value of the portfolio after rebalancing at time $t - 1$; thus, the value of asset $i$ in the portfolio at the beginning of period $t$ is $w_i x_{t-1} (1 + r_{t,i})$. However, in order for the investor to revert back to the original position $w$, he or
she needs to rebalance the portfolio periodically, and after rebalancing the value of asset \( i \) in the portfolio becomes \( w_i x_t \). The rebalancing activities on asset \( i \) at time \( t \) are: to buy it with amount of money \( w_i x_{t-1} \mu_{t,i}^+ (\mu_{t,i}^+ \geq 0) \) and to sell it with amount of money \( w_i x_{t-1} \mu_{t,i}^- (\mu_{t,i}^- \geq 0) \).

The total transaction cost incurred at time \( t \) is the product of \( c \) and the total transaction amount. Therefore, \( x_t = \sum_{i=1}^{n} w_i x_{t-1} (1 + r_{t,i}) - c \sum_{i=1}^{n} w_i x_{t-1} (\mu_{t,i}^+ + \mu_{t,i}^-) \). This equation can also be equivalently rewritten as

\[
\sum_{i=1}^{n} w_i (r_{t,i} - \eta^*_t) = c \sum_{i=1}^{n} w_i (\mu_{t,i}^+ + \mu_{t,i}^-),
\]

where \( \eta^*_t \) denotes the portfolio return during the period \([t - 1, t]\), i.e., \( \frac{w_t - w_{t-1}}{w_{t-1}} \). Also, for any asset \( i \), we have \( w_i x_{t-1} (1 + r_{t,i}) + w_i x_{t-1} \mu_{t,i}^- - w_i x_{t-1} \mu_{t,i}^+ = w_i x_t \), which implies

\[
r_{t,i} - \eta^*_t = \mu_{t,i}^+ - \mu_{t,i}^-.
\]

A sensible way to rebalance the portfolio is to maximize the portfolio return \( \eta^*_t \) while satisfying (4.3) and (4.4). Thus, it can be concluded that \( \eta^*_t \) is an answer to the following linear program.

\[
\begin{align*}
\text{maximize} & \quad \eta^*_t \\
\text{subject to} & \quad \mu_{t,i}^+ \in \mathbb{R}^n, \quad \mu_{t,i}^- \in \mathbb{R}^n, \quad \eta^*_t \in \mathbb{R} \\
& \quad r_t - \eta^*_t \cdot 1 = \mu_{t,i}^+ - \mu_{t,i}^- \\
& \quad w^T (\mu_{t,i}^+ - \mu_{t,i}^-) = c \cdot w^T (\mu_{t,i}^+ + \mu_{t,i}^-) \\
& \quad \mu_{t,i}^+ \geq 0, \quad \mu_{t,i}^- \geq 0,
\end{align*}
\]

where \( \mu_{t,i}^+ = [\mu_{t,i,1}^+, \mu_{t,i,2}^+, \ldots, \mu_{t,i,n}^+]^T \) and \( \mu_{t,i}^- = [\mu_{t,i,1}^-, \mu_{t,i,2}^-, \ldots, \mu_{t,i,n}^-] \). Dualizing this linear program, we achieve a minimization problem that has the same optimal objective value, according to the strong duality of linear programming because the dual feasible set contains at least one point \((\lambda_t, \delta_t) = (w, 1)\) and bounded.

\[
\eta^*_t = \min \lambda_t^T r_t \quad \text{s.t.} \quad \begin{cases} \\
\lambda_t \in \mathbb{R}^n, \quad \delta_t \in \mathbb{R} \\
1^T \lambda_t = 1 \quad (4.5a) \\
(1-c) \delta_t \cdot w \leq \lambda_t \leq (1+c) \delta_t \cdot w \quad (4.5b)
\end{cases}
\]

The interpretation of the dual is that \( \lambda_t \) acts as if it is a transaction cost-adjusted portfolio weight vector, and \( \lambda_t \) approaches \( w \) as \( c \) approaches 0. To determine a conservative approximation of \( \eta^*_t \), we relax the constraints of this linear program. To begin with, we observe that (4.5b) implies that

\[
(1-c) \delta_t = (1-c) \delta_t \cdot 1^T w \leq 1^T \lambda_t \leq (1+c) \delta_t \cdot 1^T w = (1+c) \delta_t.
\]
Chapter 4. Extensions

Hence, by (4.5a), \((1 - c)\delta_t \leq 1 \leq (1 + c)\delta_t\). Thus, we could just omit (4.5a) and replace (4.5b) by \(\frac{1 - c}{1 + c} \cdot \mathbf{w} \leq \mathbf{\lambda}_t \leq \frac{1 + c}{1 - c} \cdot \mathbf{w}\). Therefore, we obtain a lower bound of the logarithmic portfolio return \(\log(1 + \eta_t^c)\) as

\[
\log(1 + \eta_t^c) = \log\left(\mathbf{\lambda}_t^\top (1 + \mathbf{r}_t)\right) \geq \log\left(\frac{1 - c}{1 + c} \cdot \mathbf{w}^\top (1 + \mathbf{r}_t)\right) = \log(1 + \mathbf{w}^\top \mathbf{r}_t) - \log\left(\frac{1 + c}{1 - c}\right).
\]

As a result, for each rebalancing period, the logarithmic portfolio return is smaller than or equal to its nominal value when trading is costless, i.e., \(c = 0\), and the difference between the two is bounded by \(\log\left(\frac{1 + c}{1 - c}\right)\) as claimed. 

This theorem implies that, in the presence of transaction costs, the WCVaR of the growth rate given the first- and the second-order moments of the serially uncorrelated stock returns is bounded from below by \(\gamma^*(\mathbf{w}) - \log\left(\frac{1 + c}{1 - c}\right)\).
Chapter 5

Numerical Experiments

The primary aim of this section is to compare the performance of the constant rebalanced investment strategy suggested by (3.36) with other portfolio strategies from different perspectives. We divide this chapter into two parts. In the first part, we evaluate the proposed portfolio on simulated data, while the second part evaluates it on empirical data. We always assume that each rebalancing period is one month long. All data gathered for moments estimation are also monthly data.

5.1 Implementation Details

We implemented a portfolio optimization software based on Corollary 3.2 in Matlab. In order to use our software, the following prerequisite software must already be installed on the machine.

- **Matlab:**
  Matlab is a scientific programming environment. It can be used in a wide range of applications thanks to its built-in functions and external toolboxes. The version of Matlab used in our experiment is R2011b.

- **SDPT3:**
  SDPT3 is a Matlab software package for solving semidefinite programs and other classes of conic optimization problems. Particularly in this project, we use it to solve second-order cone programs. The version of SDPT3 used in our experiment is 4.0. Further information about SDPT3 can be found in Toh *et al.* [50] and Toh *et al.* [49] SDPT3 can be downloaded from [http://www.math.nus.edu.sg/~mattohkc/sdpt3.html](http://www.math.nus.edu.sg/~mattohkc/sdpt3.html).

- **YALMIP:**
  YALMIP is a Matlab toolbox providing a simple and intuitive way to model mathematical
programs. It is often used in conjunction with a solver package, for example SDPT3. The version of YALMIP used in our experiment is R20120830. For further reference regarding this topic, we recommend Löfberg [33].

YALMIP can be downloaded from http://users.isy.liu.se/johanl/yalmip.

5.2 Datasets

We evaluate the performance of our investment proposal on four empirical datasets. The first two datasets, namely 10 industry portfolios and 12 industry portfolios, are obtained from the Fama French online data library. Each of them consists of portfolios that represent the U.S. stock market. The third dataset we use is a collection of Dow Jones industrial average index components (as of August 2013). The last dataset is comprised of nine iShares exchange-traded funds of international indexes. We obtained the empirical data of the latter two datasets from Yahoo Finance. We refer to these four data sets as 10Ind, 12Ind, DJIA, and iShares respectively.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>#assets</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>10Ind</td>
<td>10</td>
<td>Kenneth R. French</td>
</tr>
<tr>
<td>12Ind</td>
<td>12</td>
<td>Kenneth R. French</td>
</tr>
<tr>
<td>DJIA</td>
<td>30</td>
<td>Yahoo Finance: AA, AXP, BA, BAC, CAT, CSCO, CVX, DD, DIS, GE, HD, HPQ, IBM, INTC, JNJ, JPM, KO, MCD, MMM, MRK, MSFT, PFE, PG, T, TRV, UNH, UTX, VZ, WMT, XOM</td>
</tr>
<tr>
<td>iShares</td>
<td>9</td>
<td>Yahoo Finance: EWG (Germany), EWH (Hong Kong), EWI (Italy), EWK (Belgium), EWL (Switzerland), EWN (Netherlands), EWP (Spain), EQW (France), EWU (United Kingdom)</td>
</tr>
</tbody>
</table>

5.3 Portfolio Candidates

In all experiments, we consider three constant rebalanced investment strategies:

1. *Log-robust portfolio*

We refer to the maximizer of (3.36), which is our proposed portfolio, as the log-robust portfolio.
In all experiments, we set $\epsilon$ to 5%. In the tests with real data, where we have no means to obtain the true moments of the stock returns’ distribution, we use the result from Theorem 4.1 to determine an optimal portfolio, i.e., a vector $\mathbf{w} \in \mathcal{W}$ that maximizes $\hat{\gamma}_U(\mathbf{w})$.

2. Log-optimal portfolio

In Chapter 3, we introduced the log-optimal portfolio as a portfolio that maximizes the expected logarithmic portfolio return. However, as we also argued in the same chapter that the quadratic approximation of the logarithmic portfolio return is justified, we approximate the log-optimal portfolio with the one maximizing $\mathbb{E}(\mathbf{w}^\top \mathbf{r}_t - \frac{1}{2} (\mathbf{w}^\top \mathbf{r}_t)^2) = \mathbf{w}^\top \mathbf{\mu} - \frac{1}{2} \mathbf{w}^\top (\mathbf{\Sigma} + \mathbf{\mu}\mathbf{\mu}^\top) \mathbf{w}$ for the following reasons.

(a) The quadratic approximation of the portfolio’s growth rate is already used in the derivation of the log-robust portfolio, and we want to compare the log-optimal portfolio and the log-robust portfolio on the same ground.

(b) Without the full knowledge of the distribution of stock returns, it is impossible to determine the exact log-optimal portfolio. Moreover, it is generally intractable to find the exact log-optimal portfolio because the calculation involves multivariate integration.

(c) Zhichao [54] showed that this approximation achieved a high growth rate in the long run. Hence it is a good benchmark, and therefore, we find it most suitable to approximate the log-optimal portfolio in this way.

There are other possibilities to determine the approximate log-optimal portfolio. We refer interested readers to Estrada [18] and the references therein.

3. $1/n$ portfolio

A $1/n$ portfolio (or an equally-weighted portfolio) is an investment strategy that is specified by $\mathbf{w} = \frac{1}{n} \mathbf{1}$. This portfolio is well diversified, does not suffer from estimation errors, and delivers satisfactory profits in many cases, see DeMiguel et al. [15].

5.4 Simulated Evaluation

In this part, we analyze the performance of the log-robust portfolio in different investment settings, specifically in comparison with the log-optimal portfolio. First, we report the impact of the length of the investment horizon on the performance of the log-robust portfolio. Second, we investigate the robustness of the log-robust portfolio with respect to the distributional assumptions.
5.4.1 Impact of Investment Horizon

In the first experiment, we assume that future returns are serially independent and identically distributed and that they follow a lognormal distribution, say $\mathcal{F}_{ln}$. We estimate the mean values and the covariance matrix of the assets’ logarithmic returns from the dataset Ind10 during the period 01/2003 – 12/2012 (120 sample points), where the detailed estimation procedures are explained below. This allows us to calculate the mean values and the covariance matrix of the monthly asset returns and simulate the future returns of these assets for any number of months. For a fixed length of the investment horizon, we generate 50,000 scenarios and then determine the VaR of the growth rate at level $\epsilon$, which is set to 5%, for each portfolio strategy.

Figure 5.1 shows the VaR of different portfolio candidates when we use the sample mean and the sample covariance matrix as our moment estimators. Figure 5.2, on the other hand, is obtained by using shrinkage estimators of the moments as proposed by DeMiguel et al. [16]. In their work, the shrinkage estimator of the mean returns is a weighted average of the sample mean and a scaled vector of ones, while the shrinkage estimator of the covariance matrix is a weighed average of the sample covariance matrix and a scaled identity matrix. The shrinkage intensities are obtained from a bootstrap procedure (with 500 bootstrap samples, similar to DeMiguel et al. [16]).

For each generated scenario, we also compare the Sharpe ratio, i.e., the ratio between the mean and the standard deviation of the return, between these two portfolio strategies by calculating the outperformance of the log-robust portfolio defined as

$$\frac{\hat{SR}_{lr} - \hat{SR}_{lo}}{|\hat{SR}_{lr}| + |\hat{SR}_{lo}|},$$

where $\hat{SR}_{lr}$ and $\hat{SR}_{lo}$ are the Sharpe ratios of the log-robust portfolio and the log-optimal portfolio, respectively. We report these outperformances in Figures 5.3 and 5.4 using box-and-whisker plots (the box contains the data between the 25 and the 75 percentiles).

The obtained results suggest that the log-optimal portfolio is risky when compared with the log-robust portfolio in terms of VaR. It can be seen that, with $\epsilon$ probability, the log-optimal portfolio suffers from a more substantial loss (or inferior profit) when the investment horizon is shorter than 29 months or around 42 years. In terms of Sharpe ratio, which is a popular portfolio performance measure, the log-robust portfolio is better by at least 14.24% on average.

5.4.2 Impact of Distributional Assumptions

As in the previous section, we determine the mean values and the covariance matrix of the returns and the logarithmic returns from the same data set (using shrinkage estimators). However, in
contrast to the first experiment, we now fix the number of rebalancing periods $T$. Denote by $w_{lr}$ and $w_{lo}$ the portfolio weights of the log-robust and the log-optimal investment strategies, respectively. By using the result from Vandenberghe et al. [51, Section 2.1], we calculate two probability distributions $P_{lr}$ and $P_{lo}$ which are defined as

1. $P_{lr} \in \text{arginf}_{P \in \mathcal{P}} \mathbb{P} - \text{VaR}_\epsilon \left( \frac{1}{T} \sum_{t=1}^{T} (w_{lr}^T \tilde{r}_t - \frac{1}{2} (w_{lr}^T \tilde{r})^2) > \hat{\gamma}(w_{lr}) \right)$,

2. $P_{lo} \in \text{arginf}_{P \in \mathcal{P}} \mathbb{P} - \text{VaR}_\epsilon \left( \frac{1}{T} \sum_{t=1}^{T} (w_{lo}^T \tilde{r}_t - \frac{1}{2} (w_{lo}^T \tilde{r})^2) > \hat{\gamma}(w_{lo}) \right)$.

For any numbers $\psi_{lr}, \psi_{lo} \in [0, 1]$ such that $\psi_{lr} + \psi_{lo} \leq 1$, we can construct a mixture distribution

$$P(\psi_{lr}, \psi_{lo}) = \psi_{lr} P_{lr} + \psi_{lo} P_{lo} + (1 - \psi_{lr} - \psi_{lo}) P_{ln}.$$ 

For each admissible combination $(\psi_{lr}, \psi_{lo})$, we simulate monthly returns for the next $T$ months and then determine the VaR of the approximate growth rate at level $\epsilon$, which is set to 5%, for
each portfolio strategy. As usual, the outperformance of the log-robust portfolio is defined as the difference between the VaR of the approximate growth rate of the log-robust portfolio and that of the log-optimal portfolio divided by the sum of their magnitudes. Tables 5.2, 5.3, and 5.4 report these outperformances (in percentages) for different investment horizons: 120 months, 360 months and 1,200 months.

When the investment horizon is less than or equal to 30 years, we conclude from Tables 5.2 and 5.3 that the log-robust portfolio attains a higher VaR than the log-optimal portfolio in all of the 66 constructed probability distributions consistent with the first- and the second-order moments of the asset returns. The performance of the log-optimal portfolio gradually increases with the number of rebalancing periods. Nevertheless, when the investment horizon is 100 years (which is already unrealistic), the performances of both investment strategies are comparable for the majority of these
probability distributions, but the log-robust portfolio is still significantly better under some of these distributions.

### 5.5 Out-of-Sample Evaluation

In this part, we use four different datasets listed in Table 5.1 to compare the performance of the log-robust portfolio with the other candidates with regards to different performance criteria and risk measures. Suppose that trading incurs proportional transaction costs at rate $c$. In all experiments, we assume that the transaction cost rate is 50 basis point, i.e., $c = 0.005$. Although the portfolio strategies considered here are all constant rebalanced portfolios, we recalculate the portfolio weights every 12 months to take into account that the moments of the asset returns may change over time. During each re-estimation, the moments of the stock returns are calculated from the estimation window, i.e., the most recent history of length 120 months. Since the true moments are not known, we consider two variants of the log-robust portfolio: with and without moment uncertainty. The log-robust portfolio without moment uncertainty follows directly from (3.36) when we treat the moment estimates as true moments. On the other hand, for the log-robust portfolio with moment uncertainty, we determine $\delta_1$ and $\delta_2$ from an iterative bootstrap procedure. In each iteration, we generate two bootstrap datasets from the estimation window, each of which produces an estimation of the mean vector and the covariance matrix of the asset returns $\left(\hat{\mu}_i, \hat{\Sigma}_i\right), i = 1, 2$. Then we set
Table 5.3: VaR comparison between log-robust and log-optimal portfolios ($T = 360$ months)

<table>
<thead>
<tr>
<th>$\psi_{\text{lo}}$</th>
<th>0.0</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>11.60</td>
<td>11.65</td>
<td>11.53</td>
<td>11.50</td>
<td>11.58</td>
<td>11.80</td>
<td>11.90</td>
<td>12.09</td>
<td>13.27</td>
<td>16.30</td>
<td>81.01</td>
</tr>
<tr>
<td>0.1</td>
<td>11.61</td>
<td>11.54</td>
<td>11.60</td>
<td>11.87</td>
<td>12.00</td>
<td>12.09</td>
<td>12.41</td>
<td>12.21</td>
<td>16.71</td>
<td>68.62</td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>11.42</td>
<td>11.61</td>
<td>11.49</td>
<td>11.46</td>
<td>12.12</td>
<td>12.08</td>
<td>12.05</td>
<td>18.26</td>
<td>65.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>11.42</td>
<td>12.05</td>
<td>12.20</td>
<td>12.56</td>
<td>12.87</td>
<td>13.12</td>
<td>16.20</td>
<td>63.49</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>11.37</td>
<td>12.12</td>
<td>11.76</td>
<td>12.00</td>
<td>13.15</td>
<td>18.69</td>
<td>65.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>11.58</td>
<td>12.06</td>
<td>11.76</td>
<td>12.00</td>
<td>13.15</td>
<td>18.69</td>
<td>65.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>12.07</td>
<td>11.90</td>
<td>12.61</td>
<td>17.24</td>
<td>59.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>12.19</td>
<td>13.62</td>
<td>16.27</td>
<td>59.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>12.78</td>
<td>16.01</td>
<td>58.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>15.40</td>
<td>48.91</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>14.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$\delta_1^b = (\hat{\mu}_2 - \hat{\mu}_1) \Sigma_1^{-1} (\hat{\mu}_2 - \hat{\mu}_1)$, and we solve

$$\begin{align*}
\text{minimize} & \quad \delta_2^b \\
\text{subject to} & \quad \hat{\Sigma}_2 \preceq \delta_2^b \hat{\Sigma}_1
\end{align*}$$

for $\delta_2^b$. We repeat this for 500 iterations and get 500 $\delta_1^b$'s and 500 $\delta_2^b$'s. We finally set $\delta_1$ and $\delta_2$ to the 95 percentiles of $\delta_1^b$'s and $\delta_2^b$'s, respectively. We henceforth refer to the log-robust portfolio with moment uncertainty as log-robust*. We denote by $w_t$ the desired portfolio at time $t$ and denote by $\hat{w}_t$ the portfolio at time $t$ before rebalancing. The performance measures used to evaluate the portfolio candidates when they are used in a $T$-period investment are listed as follows.

1. Mean return:
$$\hat{r}_p = \frac{1}{T} \sum_{t=1}^T \left( (1 + w_t^T r_t) \left( 1 - c \sum_{i=1}^n |w_{t,i} - \hat{w}_{t,i}| \right) - 1 \right).$$

2. Standard deviation:
$$\hat{\sigma}_p = \sqrt{\frac{1}{T-1} \sum_{t=1}^T \left( (1 + w_t^T r_t) \left( 1 - c \sum_{i=1}^n |w_{t,i} - \hat{w}_{t,i}| \right) - 1 - \hat{r}_p \right)^2}.$$

3. Sharpe ratio:
$$\hat{S}R = \frac{\hat{r}_p}{\hat{\sigma}_p}.$$

4. Turnover rate:
$$\hat{T}R = \frac{1}{T} \sum_{t=1}^T \sum_{i=1}^n |w_{t,i} - \hat{w}_{t,i}|.$$
Table 5.4: VaR comparison between log-robust and log-optimal portfolios ($T = 1,200$ months)

<table>
<thead>
<tr>
<th>$\psi_{lo}$</th>
<th>0.0</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>2.49</td>
<td>2.53</td>
<td>2.48</td>
<td>2.50</td>
<td>2.41</td>
<td>2.59</td>
<td>2.58</td>
<td>2.72</td>
<td>2.96</td>
<td>3.03</td>
<td>91.56</td>
</tr>
<tr>
<td>0.1</td>
<td>2.52</td>
<td>2.44</td>
<td>2.46</td>
<td>2.47</td>
<td>2.50</td>
<td>2.46</td>
<td>2.68</td>
<td>3.00</td>
<td>3.87</td>
<td>84.23</td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>2.48</td>
<td>2.48</td>
<td>2.41</td>
<td>2.57</td>
<td>2.54</td>
<td>2.78</td>
<td>2.74</td>
<td>3.34</td>
<td>82.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>2.51</td>
<td>2.54</td>
<td>2.67</td>
<td>2.71</td>
<td>2.66</td>
<td>2.84</td>
<td>3.74</td>
<td>82.98</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>2.56</td>
<td>2.61</td>
<td>2.61</td>
<td>2.64</td>
<td>2.55</td>
<td>3.56</td>
<td>80.90</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\psi_{lr}$</td>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>2.59</td>
<td>2.53</td>
<td>2.64</td>
<td>2.70</td>
<td>2.82</td>
<td>80.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>2.52</td>
<td>2.69</td>
<td>2.87</td>
<td>3.90</td>
<td>78.64</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>2.46</td>
<td>2.61</td>
<td>4.48</td>
<td>79.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>2.93</td>
<td>3.26</td>
<td>75.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>3.48</td>
<td>69.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>23.57</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. Net aggregate return:

$$\hat{NR} = \hat{V}_T, \quad \hat{V}_t = \prod_{i=1}^{t} (1 + w_i^T r_i) \left( 1 - c \sum_{j=1}^{n} |w_{t,j} - \hat{w}_{t,j}| \right).$$

6. Maximum drawdown:

$$\hat{MDD} = \max_{1 \leq \tau < T} \frac{\hat{V}_\tau - \hat{V}_t}{\hat{V}_\tau}.$$

The out-of-sample evaluation results are reported in Tables 5.5 – 5.8. Each of these tables corresponds to one of the datasets listed in Table 5.1.

Table 5.5: Out-of-sample evaluation (source: 10Ind)

<table>
<thead>
<tr>
<th>Candidate portfolios</th>
<th>Dataset: 10Ind, Investment horizon: [2000-01, 2012-12]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{r}_p$</td>
</tr>
<tr>
<td>1/n</td>
<td>0.0050</td>
</tr>
<tr>
<td>Log-robust</td>
<td>0.0062</td>
</tr>
<tr>
<td>Log-robust*</td>
<td>0.0064</td>
</tr>
<tr>
<td>Log-optimal</td>
<td>0.0008</td>
</tr>
</tbody>
</table>

We claim from the obtained results that the log-robust portfolio consistently outperforms other candidates in terms of Sharpe-ratio and the other risk measures. It is noted that the log-robust strategy suffers from a high turnover compared to the $1/n$ portfolio. This is because the $1/n$ portfolio
Table 5.6: Out-of-sample evaluation (source: 12Ind)

<table>
<thead>
<tr>
<th>Candidate portfolios</th>
<th>Dataset: 12Ind, Investment horizon: [2000-01, 2012-12]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{r}_p$</td>
</tr>
<tr>
<td>$1/n$</td>
<td>0.0049</td>
</tr>
<tr>
<td>Log-robust</td>
<td>0.0063</td>
</tr>
<tr>
<td>Log-robust*</td>
<td>0.0065</td>
</tr>
<tr>
<td>Log-optimal</td>
<td>0.0013</td>
</tr>
</tbody>
</table>

Table 5.7: Out-of-sample evaluation (source: iShares)

<table>
<thead>
<tr>
<th>Candidate portfolios</th>
<th>Dataset: iShares, Investment horizon: [2006-04, 2013-07]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{r}_p$</td>
</tr>
<tr>
<td>$1/n$</td>
<td>0.0029</td>
</tr>
<tr>
<td>Log-robust</td>
<td>0.0033</td>
</tr>
<tr>
<td>Log-robust*</td>
<td>0.0033</td>
</tr>
<tr>
<td>Log-optimal</td>
<td>0.0032</td>
</tr>
</tbody>
</table>

is not affected by moment re-estimation. Nonetheless, if not better, the log-robust portfolio, while being less risky, leads to a satisfactory terminal wealth comparable to that achieved by the $1/n$ portfolio, which often delivers high Sharpe-ratios and low turnovers as suggested by DeMiguel et al. [15]. Despite its theoretical appeal, the log-optimal portfolio cannot compete well with the other candidates, which we attribute to its restrictive distributional assumptions.
Table 5.8: Out-of-sample evaluation (source: DJIA)

<table>
<thead>
<tr>
<th>Candidate portfolios</th>
<th>( \hat{r}_p )</th>
<th>( \hat{\sigma}_p )</th>
<th>( \hat{S}R )</th>
<th>( \hat{T}R )</th>
<th>( \hat{N}R )</th>
<th>( \hat{MDD} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/n</td>
<td>0.0066</td>
<td>0.0460</td>
<td>0.1424</td>
<td>0.0527</td>
<td>2.4017</td>
<td>0.4824</td>
</tr>
<tr>
<td>Log-robust</td>
<td>0.0049</td>
<td>0.0381</td>
<td>0.1295</td>
<td>0.0668</td>
<td>1.9553</td>
<td>0.3967</td>
</tr>
<tr>
<td>Log-robust*</td>
<td>0.0057</td>
<td>0.0381</td>
<td>0.1498</td>
<td>0.0651</td>
<td>2.2113</td>
<td>0.4001</td>
</tr>
<tr>
<td>Log-optimal</td>
<td>-0.0026</td>
<td>0.0801</td>
<td>-0.0328</td>
<td>0.1042</td>
<td>0.3757</td>
<td>0.8391</td>
</tr>
</tbody>
</table>

Chapter 6

Conclusions

In this paper, we apply the tractable reformulation for distributionally robust quadratic chance constraints (see Chapter 2 and Zymler et al. [55]) to a multi-period portfolio optimization problem. We begin by proving that the there is a deterministic conic representation of a few distributionally robust quadratic chance constraints. This result in fact coincides with one of the results from Zymler et al. [55]. However, our proof in Chapter 2 is slightly different and may be applicable to other problem domains. The main contribution of our work is the model simplification and the dimensionality reduction of the matrix inequalities (see Chapter 3) where we exploit the temporal distributional symmetries of asset returns. Specifically, we derive an analytical expression for the WCVaR of the approximate growth rate of any constant rebalanced investment strategy in which short sales are not allowed. This expression, which is presented in Theorem 3.2, is elegant and strikingly simple. We then formulate a robust portfolio optimization model that maximizes the WCVaR of the approximate growth rate. The resulting portfolio optimization model (see Corollary 3.2) is a second-order cone program whose size does not depend on the number of rebalancing periods $T$. Hence, our portfolio optimization model can be used even by a long-term investor.

We also provide more insights about the WCVaR of the approximate growth rate. Theorem 4.1 explains how one can decide on the asset allocation without the precise information about the moments of the asset returns. Furthermore, Theorem 4.2 explains how transaction costs reduce the portfolio performance.

We conduct various numerical experiments to position our log-robust portfolio, a solution to (3.36), among a set of portfolio candidates. In a simulated test, we compare the performance of the log-robust portfolio against the log-optimal portfolio, where the latter has great theoretical appeal. Conclusively, compared to the log-robust portfolio, the log-optimal portfolio heavily relies on the underlying distributional assumptions and is much less efficient when the investment horizon is
reasonably long but finite. Even more interestingly, the log-robust portfolio consistently has low risk (according to several risk measures) while achieving a high Sharpe-ratio and terminal wealth compared to other constant rebalanced portfolios.

6.1 Advantages of the Log-Robust Portfolio

We believe that the main advantages of the log-robust portfolio are its strong theoretical guarantees and its tractability. We first establish the guarantee, i.e., the WCVaR of the approximate growth rate of any constant rebalanced investment strategy \( w \), in Theorem 3.2. This guarantee is useful for two reasons. First, it is representable as a quadratic equation in \( w \) and is therefore very simple. Second, the distributional assumption underlying Theorem 3.2 is less restrictive than that underlying Theorem 3.1. We confirm in Section 5.4 that for a realistically large investment horizon, even if the asset returns are serially independent and identically distributed, the log-robust portfolio significantly outperforms the log-optimal portfolio. Finally, the theoretical advantages of the log-robust portfolio are confirmed by our out-of-sample tests in Section 5.5.

6.2 Future Work

In this section, we outline potential extensions of the distributionally robust optimization framework and the log-robust portfolio optimization model.

- One of the most important drawbacks of the moment-based distributionally robust optimization approach is that the worst-case distribution is often a discrete distribution (see for example Vandenberghe *et al.* [51]), which is usually unrealistic in practical applications. This illustrates that moment information alone is typically not sufficient to model an optimization problem under uncertainty realistically. Including more information, such as the number of modalities or the maximum allowable (statistical) distance from a reference distribution, of the partially known distribution could potentially solve the problem.

- One of the ways to inject more information into the well-developed moment-based distributionally robust optimization model discussed above is to encode support information of the unknown probability distribution underlying the random parameters. Although accounting for support information often complicates the derivation of the deterministic counterpart of the distributionally robust optimization problem, there are some approaches proposed to address this issue (see, for example, Zymler *et al.* [55]). On the other hand, ignoring support
information can lead to an overly conservative model that sacrifices too much optimality. Specifically for the log-robust portfolio optimization model, one can specify an ellipsoid that contains the support of the asset returns distribution and derive a new portfolio optimization model which is guaranteed to be at most as conservative as the second-order cone program presented in Corollary 3.2.

- Instead of reinvesting in the same portfolio every month, it may be better for an investor to determine a new portfolio at every rebalancing period in response to constant market changes. Multi-period distributionally robust optimization with decision rules is another possibility to improve the log-robust portfolio optimization model. For multi-period robust optimization problems and multi-period stochastic programming problems, it has been shown that by specifying some functional form, for example, linear decision rules and piecewise linear decision rules, problem modellers can to some extent still simulate recourse possibilities while preserving the tractability of the model (see, for example, Ben-Tal et al. [3] and Kuhn et al. [28]). It is therefore interesting to see to what extent decision rules can improve the performance of the multi-period distributionally robust portfolio optimization problem.
Bibliography


Bibliography


