Robust Portfolio Optimisation with Filtering Uncertainty



Gonçalo Simões
St Anne's College
University of Oxford

A thesis submitted for the degree of
DPhil Thesis

Michaelmas 2017

To my beloved wife Ana,
for always believing in me,
even when I did not myself.

Acknowledgements

I would like to wholeheartedly thank my supervisor Prof. Raphael Hauser for all his extraordinary support throughout these four years. Without his patience, enthusiasm and guidance this thesis would never have been possible.

I would also like to thank the HSBC FX Research team, who kindly funded my DPhil studentship. Everyone in the team was extraordinary and I will forever be grateful to them for treating me as one of their own. A special thanks, however, to Mark McDonald and Stacy Williams for the uncountable hours spent developing not only my technical knowledge but also my personal skills.

During my time in Oxford I was blessed with too many friends to be able to individually name them. All of them were important at one point and it wouldn't be fair of me to name just a few. I would however like to thank everyone in the Mathematical and Computational Finance Group for all the coffee breaks, everyone at Somerville, St. Anne's and Pembroke College for all the lunches and everyone at the Portuguese Society for all the dinners we spent together.

I would also like to thank my family for always feeling so proud in me. Not wanting to disappoint them has been what has always motivated me in the hardest of times. I am also extraordinarily grateful to my parents for always believing that nothing was beyond my grasp if only I worked hard enough. Finally a big thank you to my wife Ana, whose everyday support overwhelms me.

Abstract

This thesis focuses on how portfolio optimisation can be carried out under different types of uncertainty, which we often measure through the use of filters. Chapter 1 motivates the problem, gives an overview of the thesis and covers some necessary background material. Chapter 2 deals with uncertainty in the covariance matrix and how by identifying different regimes we can solve optimisation problems of interest to practitioners.

Chapter 3 focuses on the uncertainty over tail events and how we can not only extract relevant information by filtering the data but also how we can use that information to construct a portfolio optimisation problem that acts on it. In Chapter 4 we address the lack of tractability for general relative robust portfolio optimisation problems and how one can overcome this so as to make it a viable tool.

Chapter 5 considers the problem of uncertainty in the filter itself and how this uncertainty can be fully incorporated in the portfolio optimisation problem. Finally in Chapter 6 we conclude and propose topics for future research.

Contents

1 Introduction **1**

1.1 Outline of the Thesis 4

1.2 Background 5

1.2.1 Portfolio Optimisation 5

1.2.2 Minimum Volatility Portfolios 8

1.2.3 Filtering 10

2 Robust Portfolio Optimisation with Specified Competitors **14**

2.1 Introduction 14

2.2 Regret Minimisation 15

2.2.1 Modelling Regret 17

2.2.2 Regret as a Constraint 18

2.2.3 Proportional Regret 19

2.3 Empirical Results 21

2.3.1 Minimum Regret in the EMU 21

2.3.2 Maximum Expected Return in the US 25

2.3.3 Model Performance over Multiple Regions 28

2.4 Conclusion 30

3 Portfolio Optimisation for Medium Frequency Trading Strategies and Heavy-Tailed Returns **31**

3.1 Introduction 31

3.2 Tail-Event Constraint via Extreme Value Theory 33

3.2.1 Review of Extreme Value Theory 33

3.2.2 Tail-Event Constraint 35

3.3 A Bimodal Approach to Heavy Tail Modelling 37

3.3.1 Constructing Different Constraints for Different Market States 39

3.4 Use in Robust Portfolio Optimisation 40

3.4.1 Absolute Robust Portfolio Optimisation 40
3.4.2 Relative Robust Portfolio Optimisation 44
3.5 Numerical Experiments 46
3.6 Conclusion 54

4 Approximating Robust Portfolio Optimisation Problems with Multiple Constraints 55
4.1 Introduction 55
4.2 Copositive Sets 57
4.2.1 S-Lemma and Extensions 58
4.2.2 Inner Aproximations 61
4.3 Bound on Robust Portfolio Optimisation Problems 66
4.3.1 Absolute Robust Portfolio Optimisation 67
4.3.2 Relative Robust Portfolio Optimisation 69
4.4 Numerical Examples 72
4.5 Conclusion 76

5 Portfolio Optimisation with DR-Expectation 77
5.1 Introduction 77
5.2 Uncertainty in Filtering 78
5.3 Quadratic Approximation of Penalty κ_t 82
5.4 Portfolio Optimisation with DR-Expectation 84
5.5 Conclusion 90

6 Conclusion 92
6.1 Future Work 93
6.1.1 Relative Robust Portfolio Optimisation with Specified Competitors and Heavy-Tailed Returns 93
6.1.2 Multi-period Relative Robust Portfolio Optimisation 93
6.1.3 Multivariate Portfolio Optimisation with DR-Expectation . . . 96

A List of Future Contracts 98

Bibliography 98

Chapter 1

Introduction

The question of how to preserve one's wealth is probably as old as wealth itself. Store everything away and its relative value will diminish as prices increase over time. Invest all the wealth in a business enterprise and there is a chance of losing it all on a failed enterprise. Use it to buy land and face the risk of not being able turn it back into cash should an emergency arise. The best approach, as the saying goes, is not to put all eggs in one basket.

The concept of diversifying one's wealth is a common theme in financial advice and it is at least 2000 years old – the Talmud famously advises to split one's wealth in equal parts between business, land and cash. It is the cornerstone of risk management – by spreading your wealth among different investments, you can safeguard yourself against the possibility of one of them being a bad one. Unfortunately things are not quite that simple and the prospect of choosing where to invest hard-earned cash can be a daunting one. If one wishes to buy small shares of businesses, the London Stock Exchange trades stock of over two thousand companies'. One might also want to invest in debt, either government's or companies. If one is interested in property, there are thousands for sale in London alone. How about options, commodities, foreign currencies? The possibilities are almost endless.

Such is the complexity of this challenge that a whole industry has emerged to tackle it: financial services. Ranging from pension funds, hedge funds, exchange-traded funds, to many other types of investment funds, there exist plenty of professional firms who specialise in wealth allocation. Each has its own rules – what instruments is it allowed to invest in, how much autonomy it has to make decisions, how long is the investment horizon, how much risk one wishes to take, amongst others. It is then the role of the asset manager to decide how to invest given the fund's rules.

Asset managers are responsible for allocating wealth entrusted to them by investors in a way they see fit. They must evaluate how profitable and risky each

instrument is, how a change in value in one instrument reflects in another and how likely it is that their assessment is correct. They must then turn all this information into decisions, the cumulative result of which will then lead investors to decide whether they want to keep investing or not. It can be a capricious life – it is better to be lucky than to be right.

It is important to point out that although their job is ultimately to produce as good an outcome as possible, asset managers' incentives are not necessarily perfectly aligned with those of their clients. While the rewards of differing from the industry consensus might be huge if proven right, in the form of fame, bonuses and promotions, the cost of that decision can be their career if proven wrong. There is therefore an incentive to keep track of what competitors are doing and, if in doubt, not to deviate significantly from this.

Modern Portfolio Theory is the field of study whose aim is to formalise asset managers' job – given a collection of financial instruments, an objective function and a set of rules, what is the optimal allocation, i.e., which allocation of wealth maximises the objective function while still respecting all the rules? Harry Markowitz effectively created the theory with his paper [45] in 1952. Markowitz considered a one period investment process, where a decision is made at inception and not altered until the end of the period. The objective was taken to be the expected return and a volatility (standard deviation) budget had to be met. This can be theoretically shown to be optimal if one assumes future returns are normally distributed. All that is required, then, is to know the expected value and the covariance matrix of all instruments' returns over the next investment period, which can be statistically estimated. The literature has since flourished – there were improvements in the estimation of various parameters, different notions of risk were introduced, more complex models were considered. But in large part the template remains the same as first introduced by Markowitz.

Asset managers often find that optimal portfolios produced by classical Modern Portfolio Theory are not well diversified and can be sensitive to parameter misspecification. Since statistical estimates of model parameters inherently contain noise, model parameters are not known exactly. This sensitivity leads on occasion to optimal portfolios changing drastically once the model is recalibrated. This not only makes asset managers wary of using such a model but also incurs immediate losses in the form of transaction costs.

There have been several approaches proposed to overcome this problem. One of the most used by practitioners is to impose constraints on the set of portfolios that

are allowed. Common constraints include imposing a maximum of wealth given to each instrument or a bound on the amount of short-selling allowed (or an outright ban). It is also common to impose regularization terms such as lasso (L^1) or ridge (L_2) penalties. Such solutions are often employed due to the fact that they produce desirable results, even if their use might seem ad-hoc. Another approach increasingly popular amongst asset managers is Robust Statistics – by producing statistics that are not affected as much by outliers, the model parameters are less subject to brusque changes, and by extension so is the optimal portfolio.

In contrast Stochastic Programming is an approach favoured mostly by academics. By modelling the parameters as stochastic processes one can first calculate what is the best decision to make in the future and then infer from the parameters' probability distribution what is the best decision to make today. It is however harder in practice to estimate a probability distribution than a single parameter and even if it were easier these problems are usually very computationally demanding.

A separate approach is Robust Portfolio Optimisation, to which this thesis shall be committed. It is motivated by assuming a parameter is not a random variable but instead an unknown constant. While we might not know what this parameter is, we assume it belongs to a so-called uncertainty set. The goal is then to find a portfolio that is in some sense optimal in respect to all parameters in such a set.

Another problem commonly identified by practitioners in Markowitz seminal paper is the assumption that assets' returns are normally distributed. While this might be a decent crude approximation in the best of times, it has become infamous for being not accurate enough in difficult ones. Returns tend to exhibit so called fat or heavy tails, which translates into extreme events happening much more often than would be implied by a normal distribution fitted on historical data. This is particularly aggravated by two factors: first, the tails of the distribution are not even symmetric, as large losses, especially consecutive ones, happen much more often than gains, at least of similar magnitude. Second, the dependence structure between different assets changes over time, with various assets that had displayed independent returns on a given period suddenly displaying high correlations in following periods, usually when associated with large losses mentioned earlier.

This all means that the volatility of a portfolio's future returns is not enough to truly measure the risk of such portfolio. That however does not mean is it fundamentally useless – for periods when positive market conditions remain unchanged it is almost all one needs. It is when these market conditions change that the perils are greatest. This usually happens either because there has been a shock to the market,

or market conditions have significantly changed, or both. The latter can be addressed by taking the view that the market alternates between various scenarios, each with its own underlying structure, and modelling its variance appropriately. The typical asset manager has a short term bias, so there is a bigger emphasis on doing well in the current scenario rather than on the overarching distribution.

Shocks to the market are harder to capture with a measure such as volatility, since it is a problem concerning the tail of the distribution which might not manifest itself on most days and hence not be fully captured by a historical covariance matrix. Hence other risk measures are usually considered such as Value-at-Risk (quantiles of losses) and Conditional Value-at-Risk (expected loss past quantile). In the end it all comes down to measuring the (potentially hidden) risk of the tail of the distribution. As with volatility these measures are not perfect, with each having its pros and cons.

In this thesis we shall tackle both problems. First, by modelling asset's returns as arising from multiple regimes – these might be either multiple volatility profiles, or a more general uncertainty set for the covariance matrix. Second by using Extreme Value Theory to model the tail behaviour of the most common directions of "extreme" market conditions, which are identified through the use of a filter. Finally we look at a very promising new approach of data-driven robust expectation as a potential way of bringing robust optimisation and stochastic programming together.

1.1 Outline of the Thesis

In this section we shall have a very brief outline of this thesis. Besides the Introduction and the Conclusion there are four chapters which albeit connected by the overarching themes of this thesis are intended to be (with some natural adaptations) separate articles published in peer-reviewed journals. All the work therein was done in collaboration with my supervisor Prof. Raphael Hauser, but some chapters are a product of wider collaboration. Chapter 2 has already been submitted and is currently being revised, while the other three will be submitted in the near future.

The six chapters of this thesis read as follows:

1. **Introduction** – Short overview of the problem, outline of the whole thesis and quick review of portfolio optimisation and filtering.

2. **Robust Portfolio Optimisation with Specified Competitors** – Collaboration with Mark McDonald, Stacy Williams and Daniel Fenn from HSBC. We look into the problem of minimising volatility regret in relative robust portfolio

optimisation against a set of well defined benchmarks, instead of the traditional omniscient adversary framework. We also explore a few attractive variants such as proportional regret and using this measure as a constraint, rather than an objective.

3. **Portfolio Optimisation for Medium Frequency Trading Strategies and Heavy-Tailed Returns** – Our goal is to use the information in the tail of a distribution to design a risk metric that can be easily incorporated into portfolio optimisation problems. We start by filtering the data so as to identify an "extreme" dataset, followed by a PCA decomposition to identify the import directions of market correction followed by some extreme value theory motivated modelling of the tails. After some convex relaxation we arrive at a quadratic form that is a proxy for the tail risk of a distribution, which can then be used in a portfolio optimisation problem.

4. **Approximating Robust Portfolio Optimisation Problems with Multiple Constraints** – Unfortunately we are not always able to solve relative portfolio optimisation problems in polynomial time. In fact even upper (or lower) bounds were only known when at most a single quadratic constraint was present. In this chapter we build on previous approximations to build new ones for problems with multiple linear and quadratic constraints.

5. **Portfolio Optimisation with DR-Expectation** – Collaboration with Prof. Samuel Cohen. By using the novel technique of data-driven robust expectation we look into how calibration errors propagate through a Hidden Markov Model and how these errors impact the optimal portfolio of such model.

6. **Conclusion** – Concluding remarks and overview of future research.

1.2 Background

1.2.1 Portfolio Optimisation

Portfolio optimisation is one of the oldest problems in mathematical finance and goes back to the pioneering work of Markowitz [45]. The problem consists of optimally allocating limited budgets of investment capital and risk, subject to conditions that constrain the chosen allocation, and so as to maximize a merit function, typically expected return. For an overview of some of the many models that have been investigated over the years, see [22]. In the one-period setting, where the aim is to find

an optimal allocation of limited investment capital over a fixed investment horizon, most approaches fit the following general scheme,

$$\begin{aligned} &\max_{x\in\mathbb{R}^n} \mathbb{E}\left[u(R^\top x)\right] \\ &\text{s.t.} \quad x \in \mathcal{X}, \end{aligned}$$

where x is a vector of asset positions or of proportions of risk capital allocated to individual assets, R is a vector of random asset returns, u is a utility function and $\mathcal{X}$ a convex set that describes the set of decisions x that satisfy the allocation constraints. For example this might simply be the set of all fully invested long-only equity portfolios, or it might be something different depending on individual circumstances.

The most standard version of the problem when one tries to maximise the expected return within a volatility budget σ,

$$\begin{aligned} &\max_{x\in\mathbb{R}^n} \mu^\top x \\ &\text{s.t.} \quad x^\top Q\, x \leq \sigma^2, \\ &\qquad x \in \mathcal{X}, \end{aligned} \tag{1.1}$$

where $\mu = \mathbb{E}\left[R\right]$ and $Q = \mathbb{E}\left[(R-\mu)(R-\mu)^T\right]$ are known. Other common equivalent formulations are the maximum risk-adjusted return problem

$$\begin{aligned} &\max_{x\in\mathbb{R}^n} \mu^\top x - \lambda\, x^\top Q x \\ &\text{s.t.} \quad x \in \mathcal{X}, \end{aligned} \tag{1.2}$$

where $\lambda > 0$ is the risk-aversion coefficient, and the minimum volatility with guaranteed return problem

$$\begin{aligned} &\max_{x\in\mathbb{R}^n} x^\top Q\, x \\ &\text{s.t.} \quad \mu^\top x \geq c, \\ &\qquad x \in \mathcal{X}, \end{aligned}$$

where c is the minimum return demanded.

This is a simple and intuitive model which has been studied at length, however that same simplicity has prevented it from being widely used due to its limitations, see, for example, [10, 17, 46].

One of the most often mentioned limitations is its sensitivity to input parameters. While μ and Q are assumed to be known, they are in fact unknown constants that need to be estimated from data. Calling them constants is also quite deceitful, since

every market participant knows they actually evolve over time. This sensitivity to inputs is damaging, as they will likely be continuously re-calibrated as new data comes in.

One of the solutions to this problem is robust optimisation (RO). Prior to its application to finance, robust optimisation has been used to tackle various optimisation problems (see for example [6, 7]) and it has been employed on various fields, as the survey [8] shows. It has also been used quite extensively in portfolio optimisation problems – for example [21, 23, 27, 35, 41, 42, 43] among others. The idea behind RO is to try to incorporate parameter uncertainty into the model, but without having to model their entire probability distribution function. This is achieved by having as input an uncertainty set instead of the actual parameters, and dealing in some way with the fact that multiple parameters are available.

The most common type of robust portfolio optimisation is by far the worst-case scenario approach, which we call absolute robust portfolio optimisation. The idea behind it is the following: for each admissible portfolio, there is a set of parameters in the uncertainty region that penalises the portfolio the most. Since any set of parameters in the uncertainty region is possible, we choose the portfolio with the best worst-case scenario.

Let us look back at problem (1.1). Suppose we were confident of our covariance estimate Q, but had doubts about μ. We were however confident that μ belongs to some region $\mathcal{U}$ (for example a confidence region). The absolute robust portfolio optimisation version of (1.1) would be

$$\begin{aligned} \max_{x \in \mathbb{R}^n} \ & \left(\min_{\mu \in \mathcal{U}} \ \mu^\top x \right) \\ \text{s.t.} \ & x^\top Q \, x \leq \sigma^2, \\ & x \in \mathcal{X}. \end{aligned} \tag{1.3}$$

There is growing evidence that these perform better in practice than simple optimization problems with no robustness, even more so when ellipsoidal uncertainty sets are used instead of box-shaped sets [41, 54, 55].

While worst-case scenario might be an appropriate measure in many cases, there are probably others where it is not. Being an absolute measure, it will focus solely on the "catastrophic" scenarios, ignoring the much more abundant and likely "everyday" scenarios. From the point of view of an asset manager, while it is important to be prepared for when the next crisis hit, it is also important to have a respectable

performance in the meantime. A different approach is that of relative robust optimisation, first introduced in [37] and applied to portfolio optimisation in [32]. The idea is to compare portfolios as follows: for each set of parameters in the uncertainty region, there is an optimal portfolio, given to us by solving a non-robust portfolio optimisation problem, such as (1.1). Therefore, for each admissible portfolio and for each set of parameters there exists a regret function, that measures how much better we could have performed if we had known the parameters beforehand (or if we simply had been lucky). Since we don't know the true parameters, for each portfolio we can calculate the maximum regret. By picking the portfolio that minimizes this quantity we know that we've done our best to minimize our regret, i.e., to be as close to the top performer as we can. The relative robust portfolio optimisation equivalent of (1.1) would be

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \max_{\mu\in\mathcal{U}} \max_{y\in\mathcal{X}, y^\top Q\, y\leq\sigma^2} & \left(\mu^\top y - \mu^\top x\right) \\ \text{s.t.} \quad & x^\top Q\, x \leq \sigma^2, \\ & x \in \mathcal{X}. \end{aligned} \tag{1.4}$$

This is much closer to the motivation behind most market practitioners, where success (and lack of it) is often measured in relative, not absolute, terms.

Unfortunately while (1.1) and (1.3) can be solved in polynomial time (at least for linear or quadratically constrained sets $\mathcal{X}$ and $\mathcal{U}$) by readily available software, such as CPLEX, the same is not true of (1.4), where even an approximate solution even exists for the simplest of sets.

1.2.2 Minimum Volatility Portfolios

As mentioned earlier (1.1) is quite sensitive to inputs, but this is particularly true for the expected returns vector μ. Not only that, but μ is also considerably harder to estimate. It is hence tempting to drop it altogether and make an investment decision based solely on Q. Haugen and Baker [30] proposed one should invest in the minimum volatility portfolio, provided we know what the "true" covariance matrix Q is. The problem of finding this optimal portfolio x can be written as

$$\begin{aligned} \min_{x\in\mathbb{R}^n} & \sqrt{x^T Q\, x} \\ \text{s.t.} & \; x \in \mathcal{X}. \end{aligned} \tag{1.5}$$

It is worth pointing out that such strategies are typically considered when deciding between multiple assets which we assume will appreciate over time, such as equities,

as it would be clearly suboptimal to invest in a portfolio where we expect to lose money on average. Also, without some sort of long-only constraint (1.5) will have at least solutions, an optimal strategy and its reverse, which clearly makes no sense.

The idea of investing without taking into account expected returns is not new. Since [30] a lot of research has been published on minimum volatility investing and its performance has been widely studied (for example, [12, 16, 18, 34, 40], amongst others). In fact, some asset managers already offer this solution to their clients in one form or another (see [53]) and it was recently reported in the media that such funds had net inflows of $12.5 billion in the first half of 2016 (see [38]). There has also been an increase in the literature of proposed metrics that solely look at risk, for example [58] and [11].

This model assumes Q is known for the coming investment period, which might be questionable if this period is of sizeable length. While there are many techniques available to provide good estimates, any approach of practical relevance is going to have to deal with the fact that market conditions change, as described in [24] – hence assuming a single covariance matrix as certain is bound to be incomplete. A natural step forward is to assume that the future covariance matrix is an unknown element of a set of finite *scenarios* $\mathcal{U} = \{Q_1, \ldots, Q_k\}$.

The challenge is how to make a decision when given such an uncertainty set. A possible avenue is to take a Bayesian approach by assigning each scenario with a probability p_i. By conditioning on each prospective scenario being realised, we can calculate a minimum volatility portfolio as

$$\begin{aligned} \min_x \ & \left(\sum_{i=1}^{k} p_i x^T Q_i \, x \right)^{1/2} \\ \text{s.t. } & x \in \mathcal{X}. \end{aligned}$$

This however requires us to come up a sensible vector of probabilities p, which will have a significant impact on the optimal portfolio x. Another possibility is, as before, to consider the absolute robust portfolio optimisation version of (1.5).

$$\begin{aligned} \min_x \ & \max_{Q \in \mathcal{U}} \sqrt{x^T Q \, x} \\ \text{s.t. } & x \in \mathcal{X}. \end{aligned} \tag{1.6}$$

The problem is that this model focuses **only** on the worst-case scenario of each portfolio. A different approach is the relative robust portfolio optimisation version of (1.5). Here the notion of regret is slightly different to the one used in (1.4), as the

main purpose is to avoid having to estimate μ. In this instance the goal is to minimise the amount of extra volatility we are forced to accept for not knowing the covariance matrix with certainty. For $Q^* \in \mathcal{U}$, denote by y_{Q^*} as the portfolio we should hold (by solving (1.5)) if we knew Q^* to be the "true" covariance matrix. For any other portfolio x in which we might invest our excess volatility would be

$$\sqrt{x^T Q^* x} - \sqrt{y_{Q^*}^T Q^* y_{Q^*}} = \sqrt{x^T Q^* x} - \min_{y \in \mathcal{X}} \sqrt{y^T Q^* y}.$$

But we do not know which $Q \in \mathcal{U}$ is going to be realised and so we cannot know how much our excess volatility will be. We can however strive to make it as low as possible, by minimising the worst outcome across all possible values of $Q \in \mathcal{U}$, that is, by solving

$$\begin{aligned} \min_{x} \max_{Q \in \mathcal{U}} & \left(\sqrt{x^T Q x} - \min_{y \in \mathcal{X}} \sqrt{y^T Q y} \right) \\ \text{s.t. } & x \in \mathcal{X}. \end{aligned} \tag{1.7}$$

Given that $\mathcal{U}$ is a finite set (1.7) is a simple enough problem to be solved in polynomial time.

1.2.3 Filtering

Given a set of i.i.d. sample returns $r_1, \ldots, r_T$, we are interested in filters $\mathfrak{F}$ that classify these points into either "typical" (encoded as 1) and "extreme" (encoded as 0). We shall list a couple of thumb rules used in practice before delving into a more elaborate example.

Example 1. *Given $q \in (0,1)$ and $v_1, \ldots, v_n$ an orthonormal basis, let u_j be the $(1-q)$-quantile of $\left(|v_j^T r_t|\right)_{1 \le t \le T}$.*

$$\mathfrak{F}(r_t) = \begin{cases} 1 & \text{if } |v_j^T r_t| \le u_j \quad \forall j, \\ 0 & \text{otherwise.} \end{cases}$$

Hence we classify as extreme any data point which has at least one component larger than an acceptable limit.

This is a very versatile filter, as different norms or orthonormal transformations can be used, depending on the particular situation. It is also common to substitute the empirical quantiles for some theoretical ones, such as normal, student-t or chi-square distribution.

Example 2. *Given $q \in (0,1)$,*

$$\mathfrak{F}(r_t) = \begin{cases} 1 & \text{if } |\{\tau : \|r_\tau\|_1 \leq \|r_t\|_1\}| \leq qT \\ 0 & \text{otherwise.} \end{cases}$$

That is, the vector of realised returns r_t is "typical" if its L^1 norm is not in the top $q\%$, and "extreme" otherwise.

Common improvements involve, for example, subtracting the mean from each data series and normalising it by an estimate of the volatility, as well as with picking different norms.

We shall now spend the rest of this section on our filter of choice throughout this thesis: a Hidden Markov Model with normal and t-student distributed observations.

We start by presenting this model. Let $\{X_t\}_{t\geq 0}$ be a 2-state Markov Chain on the unit vectors $\{e_1, e_2\}$ with transition matrix $A^\top$ so that $X_t - AX_{t-1}$ is a martingale. X is not observable, but we do observe the real-valued vector process $\{Y_t\}_{t\geq 0}$. These processes are defined in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where $\mathcal{F}$ denote the filtration generated by both X and Y. Moreover we denote by $\mathcal{Y}$ the observable filtration, which is generated by Y alone.

Given X_{t-1}, the density of Y_t is $f(y; X_{t-1})$, where $f(y; e_1)$ is the normal density with mean μ_1 and covariance matrix Σ_1, while $f(y; e_2)$ is the multivariate t-student distribution with mean μ_2, spread matrix Σ_2 and τ_2 degrees of freedom.

Denote by C_t the diagonal matrix with entries $f(Y_t; e_i)$. We wish to fit this model by using an Expectation–Maximisation (EM)-algorithm, so we need to be able to calculate both the expected log-likelihood and gradient given $\mathcal{Y}$. We can do that by following the work done in [20] – although it was done solely for normal distributions, the same argument works for a general distribution.

Theorem 3. *Let g be a real-valued function and*

$$G_t^r(g) = \sum_{i=1}^{t} \left(e_r^\top X_{i-1}\right)\, g(Y_t). \tag{1.8}$$

Then the expected value of $G_t^r(f)X_t$ given $\mathcal{Y}_t$ can be calculated via the following recursion

$$\begin{aligned} \mathbb{E}\left[G_t^r(g)X_t \mid \mathcal{Y}_t\right] =& \frac{1}{\alpha_t} A^T C_t\, \mathbb{E}\left[G_{t-1}^r(g)X_{t-1} \mid \mathcal{Y}_{t-1}\right] \\ &+ \frac{1}{\alpha_t} f(Y_t; e_r)\left(e_r^\top p_{t-1}\right) g(Y_t) A\, e_r. \end{aligned} \tag{1.9}$$

Moreover, if $N_t^{r,s}$ denotes the number of jumps from state e_r to state e_s, then

$$\begin{aligned}\mathbb{E}\left[N_t^{r,s} X_t \mid \mathcal{Y}_t\right] =& \frac{1}{\alpha_t} A^T C_t \, \mathbb{E}\left[N_{t-1}^{r,s} X_{t-1} \mid \mathcal{Y}_{t-1}\right] \\ &+ \frac{1}{\alpha_t} f(Y_t; e_r) \left(e_r^\top p_{t-1}\right) A_{s,r} \, e_s.\end{aligned}$$

Finally, if $s < t$, then the following recursion holds

$$\mathbb{E}\left[\left(e_r^\top X_s\right) X_t \mid \mathcal{Y}_t\right] = \frac{1}{\alpha_t} A^T C_t \, \mathbb{E}\left[\left(e_r^\top X_s\right) X_{t-1} \mid \mathcal{Y}_{t-1}\right] \tag{1.10}$$

Given data up until t, one can recursively calculate $\mathbb{E}\left[\left(e_r^\top X_s\right) X_t \mid \mathcal{Y}_t\right]$. By summing its components we obtain the probability of X_s being e_r, since $\mathbb{E}$ is a linear operator and X_s takes values in $\{e_1, e_2\}$ (we shall use this in Chapter 3 to classify historical data).

To calculate the expected log-likelihood first we notice that again by linearity, $\mathbb{E}\left[G_t^r(g) \mid \mathcal{Y}_t\right]$ can be obtained by summing the components of $\mathbb{E}\left[G_t^r(g) X_t \mid \mathcal{Y}_t\right]$. The log-likelihood of Y given X can be written using (1.8) as $G_t^1\left(f(\bullet; e_1)\right) + G_t^2\left(f(\bullet; e_2)\right)$, so we recursively calculate the conditional log-likelihood using (1.9). The same works for the first or second (and higher if necessary) derivative of the log-likelihood with respect to parameters μ_1, μ_2, Σ_1, Σ_2 and τ_2 – they can all be decomposed component and state-wise and then computed recursively.

We can then proceed to fit our model via the EM-algorithm, see [19]. Start with an initial parameter guess, A^0, μ_1^0, μ_2^0, Σ_1^0, Σ_0^2 and τ_2^0. Then do the following until convergence

- E-step – Given measure $\mathbb{Q}^n$, generated by A^n, μ_1^n, μ_2^n, Σ_1^n, Σ_n^2 and τ_2^n, calculate the expected log-likelihood of an arbitrary set of parameters given the historical data.

- M-step – Maximise the expected log-likelihood and thus get new estimates A^{n+1}, μ_1^{n+1}, μ_2^{n+1}, Σ_1^{n+1}, Σ_{n+1}^2 and τ_2^{n+1}.

The E-step can be done as described above. The M-step is done in two distinct parts – one first maximises the expected log-likelihood for μ_1, μ_2, Σ_1, Σ_2 and τ_2, and then for A.

To maximise the expected log-likelihood with respect to μ_1, μ_2, Σ_1, Σ_2 and τ_2 for a fixed measure $\mathbb{Q}^n$ we first calculate the expected gradient with respect to those parameters and then implement a quasi-Newton algorithm. We avoid using a Newton algorithm since the parameter space has a significant size, so calculating the expected

Hessian matrix at every step is computationally expensive. For the numerical example shown in this thesis in Chapters 3, 4 and 5 a BFGS algorithm with backtracking-Armijo line search worked well.

The only parameters that are left is the transition matrix A – it was shown in [20] that the following maximises the conditional loglikelihood in the M-step,

$$A_{s,r}^{n+1} = \frac{\mathbb{E}^{\mathbb{Q}^n}\left[N_t^{r,s} \mid \mathcal{Y}_t\right]}{\mathbb{E}^{\mathbb{Q}^n}\left[N_t^{r,1} + N_t^{r,2} \mid \mathcal{Y}_t\right]}.$$

By alternating a E-step with a M-step the model is guaranteed to converge to a local maximum, as it produces a sequence of measures $\mathbb{Q}^n$ with increasing expected log-likelihood. There however no guarantee that a global maximum will be found, so it is a good idea to try different initial parameters.

Chapter 2

Robust Portfolio Optimisation with Specified Competitors

2.1 Introduction

While absolute robust optimisation may be sensible in some situations, we believe it is not well suited for most financial practitioners. Even though it is important to worry about extreme scenarios, typical everyday scenarios are no less important and should not be ignored. Moreover, some professionals such as investment managers are frequently evaluated against the competition and not on absolute terms. For those reasons we believe that relative robust optimisation, introduced in [37] and developed in [32], is more suited to most portfolio managers. In this approach a portfolio's worth is assessed not only on its performance but also on the competitions' performance.

Competition is clearly an ambiguous term which needs to be made precise. In [32] the authors define competition as an "omniscient adversary" that has the same constraints as we do, but knows the "correct" model parameters and hence solves a non-robust portfolio optimisation problem. In this chapter we introduce the idea that competition may be specified independently of our initial problem, i.e., we make it clear that competition may be described according to user's needs, just like the uncertainty set.

Most academic literature on robust portfolio optimisation propose that the objective function be replaced – for example, if an investor is maximising his expected return it is suggested he should aim to maximise their worst-case expected returns instead. However investors have their own objectives and are wary of changing them, which makes the leap from literature to practice much harder to achieve. There is however no reason why an investor cannot maximise his expected return while trying to limit the worst-case volatility, for example. With this in mind we propose

adding an extra constraint instead of a new objective function. This allows investors to keep their framework intact and simply make a minor change, trading an extra constraint for extra robustness. We also promote the use of *proportional regret*, a slightly different measure that is more in line with practitioners common practice.

In line with this philosophy in this chapter we will look at models that deal with volatility, not expected returns. The idea of investing without taking into account expected returns is not new. Since [30] proposed a minimum volatility portfolio a lot of research has been published on it and its performance has been widely studied (see for example [12] and [40]). In fact, some asset managers already offer this solution to their clients in one form or another (see [53]) and it was recently reported in the media that such funds had net inflows of $12.5 billion in the first half of 2016 (see [38]). There has also been an increase in the literature of proposed metrics that solely look at risk, for example [11] and [58].

It is however possible to extend our framework to models that incorporate expected returns or even higher moments. In fact this approach appears particularly promising given how sensitive traditional Portfolio Optimisation is to errors in the means, see [10]. Covariance matrices in comparison are much more stable over time, although they do shift between different regimes, as shown in [24]. We will assume a finite number of regimes, each characterised by a covariance matrix. We will cover a more general setting in a later chapter.

We start introducing our own model, by proposing a different measure of *regret* than the one found in [32]. Initially we consider it an objective function, as is common in the literature, but subsequently we propose using it as a constraint instead, making it more widely applicable. We then pursue the suggestion in [37] and introduce *proportional regret* in the context of portfolio optimisation, a slightly different measure that is more in line with practitioners common practice. At last we run numerical experiments on real data and examine how different model versions we proposed compare against common benchmarks.

2.2 Regret Minimisation

The relative robust portfolio optimisation framework proposed in [32] is interesting, but it might not be clear for practitioners where its value lies. Comparing us against a "more knowledgeable version of ourselves" might be a very interesting theoretical problem, but in practice one gets compared against other players in the market, not

fictional beings. We hence intend to propose a different approach to (1.7) which might be more in line with investors' concerns.

We would like instead to be compared against other investors, as this would be a more realistic portrait of the real world. Unfortunately we do not know how other agents in the market are investing and therefore cannot hope to model precisely the behaviour of all investors.

As an alternative we will compare ourselves against benchmarks, which are naturally unambiguous as both their performance, their constituents and their weights are publicly available. Benchmarks have two properties that make them suitable for this purpose. On the one hand, an asset manager is not only compared against better performing competitors, but also against relevant benchmarks by both clients and stakeholders. This means comparison against benchmarks is a reasonable representation of reality. On the other hand, competitors will themselves be victims of pressure not to fall too short of a particular benchmark. This will in turn pull them towards a similar basket, resulting in a high correlation between the two, hence making benchmarks decent proxies for competitors' unknowable investments.

Notice for example the historical correlation between the HFRI Equity Hedge Fund (Total) Index and the S&P 500 in Figure 2.1. This chart provides support to the claim that benchmarks are a good proxy for agents' returns. We can see that this particular basket of hedge funds have been increasingly correlated with the S&P 500, with recent values surpassing the 90% threshold.

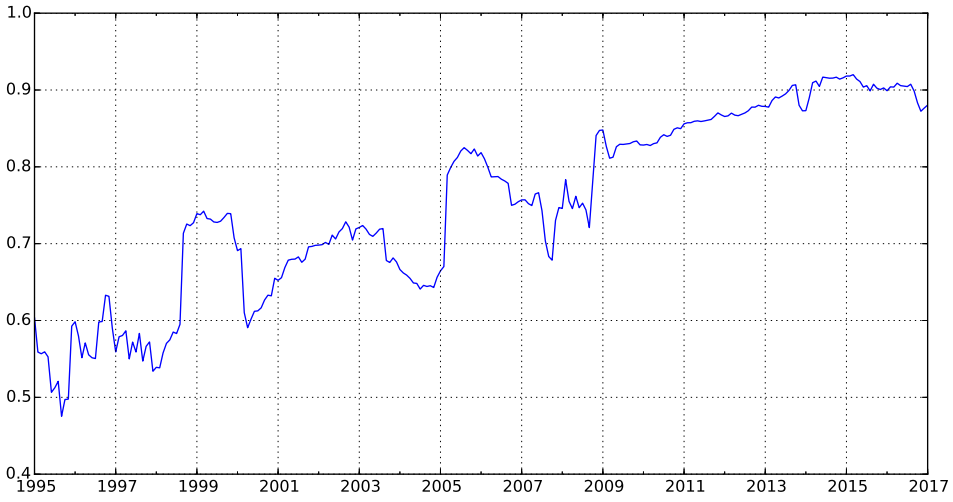


Figure 2.1: Correlation between the HFRI Equity Hedge (Total) Index and the S&P 500.

This is the case despite the fact that most hedge funds do not have explicit benchmarks – their mandate is to deliver positive returns and they are commonly considered a diversified investment.

2.2.1 Modelling Regret

Let $\mathcal{B} = \{b_1, \ldots, b_m\} \subseteq \mathbb{R}^n \setminus \{0\}$ be the set of benchmarks we wish to consider. Following the rationale behind (1.7), we wish to consider our excess volatility to the least volatile benchmark,

$$l_{\mathcal{B}}(x, Q) := \sqrt{x^T Q\, x} - \min_{b \in \mathcal{B}} \sqrt{b^T Q\, b}.$$

Notice that we do not include 0 in our benchmark set, as otherwise excess regret volatility equal actual volatility and the best benchmark would not change irrespectively of Q.

Here we assumed that all the underlyings of all benchmarks are included in the set of n assets considered for investment, and so we can compute their volatility by $\sqrt{b^T Q\, b}$ – in truth this is not strictly necessary, as well will see later on.

Let us now take a finite uncertainty set of covariance scenarios $\mathcal{U} = \{Q_1, \ldots, Q_k\}$. In a similar fashion to [32] we define *regret* to be the maximum of this loss over $\mathcal{U}$, i.e.,

$$\mathrm{Rgrt}_{\mathcal{U}}(x) := \max_{Q \in \mathcal{U}} l_{\mathcal{B}}(x, Q) = \max_{Q \in \mathcal{U}} \left(\sqrt{x^T Q\, x} - \min_{b \in \mathcal{B}} \sqrt{b^T Q\, b} \right). \tag{2.1}$$

Regret can be understood as our distance to the "winner" under the "least desirable" scenario.

Benchmarks choice, just like scenarios, is beyond the scope of this chapter. We can however provide a couple of useful tips that apply to both. First off, they should be picked independently of the portfolio optimisation problem, be it through economic/political considerations, data analysis, or any other technique. Furthermore, they should be as realistic and not too many. Too many different scenarios and benchmarks will lead to overly defensive portfolios, since an "anything can happen" attitude is a sure way to investment paralysis.

If we wish to find the portfolio that has least regret then we have to solve the following *Relative Robust Portfolio Optimisation with Benchmarks* problem.

$$\begin{aligned} \min_{x} \quad & \mathrm{Rgrt}_{\mathcal{U}}(x) \\ \text{s.t.} \quad & x \in \mathcal{X}. \end{aligned} \tag{2.2}$$

Unlike the minimum volatility problem (1.5), minimum regret *might* be negative. It should be clear to the reader that this will be the case if and only if there exists a portfolio that has a lower volatility than all of the considered benchmarks regardless of which scenario is realised.

On a technical level, we actually do not need to have a set of defined benchmarks which we want to be compared against. All we need is a *target volatility* σ_i for each scenario Q_i. In the case where we have a set of benchmarks, our target volatility is

$$\sigma_i = \min_{b \in \mathcal{B}} \sqrt{b^T Q_i \, b} \quad \forall i = 1, \ldots, k.$$

While (2.2) can be expressed as a three-level optimisation problem, it is actually a tractable problem that can be solved as efficiently as the minimum volatility problem (1.5), as long as $\mathcal{X}$ can be cast as a tractable cone (which is true for most admissible sets of practical relevance). This is made possible by the addition of auxiliary variables t and γ that allow (2.2) to be converted into

$$\begin{aligned} \min_{\gamma,t,x} \ & \gamma \\ \text{s.t. } & \gamma - t_i + \sigma_i \geq 0, \quad \forall i = 1, \ldots, k, \\ & \begin{bmatrix} t_i \\ U_i^T x \end{bmatrix} \in L_{n+1}, \quad \forall i = 1, \ldots, k, \\ & x \in \mathcal{X}, \end{aligned} \tag{2.3}$$

where U_i is the Cholesky decomposition of $Q_i = U_i \, U_i^T$ and L_{n+1} is the Lorentz cone defined by

$$L_{n+1} = \left\{ \begin{bmatrix} t \\ x \end{bmatrix} \in \mathbb{R}^{n+1} : \|x\|_2 \leq t \right\}.$$

The optimisation problem found in (2.3) is a second order cone program and can thus be solved using standard solvers (such as CPLEX or MOSEK). This means that in the end we have a much richer approach that can be solved in roughly the same computational time as a standard Markowitz optimisation problem.

2.2.2 Regret as a Constraint

So far we have only seen models that assume volatility as the objective function. Another common approach is to use volatility instead as a constraint. In fact [45] first introduced Modern Portfolio Theory as

$$\begin{aligned} \max_{x} \quad & \mu^T x \\ \text{s.t.} \quad & \sqrt{x^T Q \, x} \leq \kappa, \end{aligned}$$

where μ is the expected future returns and κ is the maximum volatility allowed in our portfolio. κ can thus be as low as the solution of (1.5).

Similarly, we can look at using regret not as an objective in itself, but rather as a constraint to be imposed in our portfolio allocation problem. Say for example we wish to maximise expected return as in the above problem. Then we could solve a similar problem,

$$\begin{aligned} \max_{x} \quad & \mu^T x \\ \text{s.t.} \quad & \text{Rgrt}_{\mathcal{U}}(x) \leq \theta, \end{aligned} \tag{2.4}$$

where θ is the regret we are prepared to accept and can be no lower than the solution to (2.2).

This approach is attractive on the grounds that investors prefer to target more meaningful objectives, such as maximising expected returns or minimising transaction costs. In addition it gives the portfolio manager control of how much risk he is prepared to take in order to increase (or decrease) his objective function. It is also a more flexible option, since adding a constraint is comparatively easier than rewriting from scratch an optimisation problem with a new objective function. In fact there is no reason why we cannot have both constraints – one on volatility and another on regret, so long as they are consistent.

2.2.3 Proportional Regret

As an alternative to regret we could define excess volatility not as the difference in volatilities, but instead the proportion of benchmark volatility we surpass. We can then define *proportional regret* to be the maximum of this loss, i.e.,

$$\text{PRgrt}_{\mathcal{U}}(x) := \max_{Q \in \mathcal{U}} \left(\max_{b \in \mathcal{B}} \left(\frac{\sqrt{x^T Q\, x} - \sqrt{b^T Q\, b}}{\sqrt{b^T Q\, b}} \right) \right),$$

provided $\min_{Q \in \mathcal{U}} \min_{b \in \mathcal{B}} \sqrt{b^T Q\, b} > 0$.

This was first sugested by [37] in the context of general robust optimisation. We believe thinking in proportions is more natural for investors and so this option will be of more practical interest to them.

As before, we only require the volatility of the best benchmark under each scenario to be modelled, not for all of them. Likewise, proportional regret can only be negative if x outperforms all benchmarks over all scenarios.

It should be made clear, however, that this is not just a cosmetic change. While the best benchmark in each scenario remains the same, the trade-off between scenarios does change. To illustrate this let us look at a quick example.

Imagine we have two assets we can invest in, but we can only invest in one of them. We will be compared against a single benchmark – an equally weighted basket of the two assets. Imagine further that over the next year there are two possible scenarios, summarised in the following covariance matrices

$$Q_1 = \begin{bmatrix} 0.02 & 0.006 \\ 0.006 & 0.05 \end{bmatrix}, \qquad Q_2 = \begin{bmatrix} 0.07 & 0 \\ 0 & 0.05 \end{bmatrix}.$$

The first asset has $\sqrt{0.02} \approx 14\text{pp}$[1] volatility in Scenario 1 and 26pp volatility in Scenario 2, while the second asset has 22pp volatility in both Scenario 1 and 2. In addition we have a single benchmark b that has 14pp volatility in Scenario 1 and 17pp volatility in Scenario 2.

One can check that Asset 1 performs worst in Scenario 2, while Asset 2 performs worst in Scenario 1. In Scenario 2 Asset 1 trails b by 9pp, or 53%, while in Scenario 1 Asset 2 trails b by 8pp, or 57%. It should then be clear that if we cared about regret we would invest in Asset 2, while if we cared about proportional regret we would invest in Asset 1. Hence these two approaches to measuring distance to benchmarks yield different decisions.

Proportional regret can be used as an objective simply by substituting $\text{Rgrt}_{\mathcal{U}}$ by $\text{PRgrt}_{\mathcal{U}}$ in (2.2). It can also be used as a constraint, in which case we make the same substitution in s(2.4) instead.

Just like (2.2), all these variations can be efficiently solved using standard tools. For example, a maximum expected return problem with a proportion regret constraint can be shaped into the following second order cone program, where θ is once again a risk budget that has to be chosen large enough for the programme to have feasible points.

$$\begin{aligned} \min\ & \mu^T x \\ \text{s.t. } & (1+\theta)\,\sigma_i - t_i \geq 0, \quad \forall i = 1, \ldots, k, \\ & \begin{bmatrix} t_i \\ U_i^T x \end{bmatrix} \in L_{n+1}, \quad \forall i = 1, \ldots, k, \\ & x \in \mathcal{X}. \end{aligned}$$

[1]To avoid confusion between Regret and Proportional Regret we will refer to volatility units as percentage points (pp), with 1pp = 0.01.

2.3 Empirical Results

In this section we will investigate if the proposed method yields sensible results and, equally importantly, if the results are fundamentally different from other approaches commonly used in the literature. All these models can be solved efficiently using a standard conic optimisation solver, such as MOSEK [1] which we have used.

To do this we will investigate its use in building a fully-invested long-only portfolio of equities across different regions. Since our goal is mostly illustrative we choose to regard as investable asset industry group indices instead of individual companies. For that purpose we follow the recommendations of [9] and use the Global Industry Classification Standard (GICS). We will therefore consider each of the 24 industry groups in GICS as investable assets.

Benchmark choice is a delicate process that will vary substantially from one investor to the next. Given this, we will take, for simplicity's sake, advantage of the hierarchical structure of GICS and use as benchmarks its 10 sectors along with the region wide total return index. We will use daily data from 2001 onwards, with prices quoted in local currency, and for each specific region will only consider industry groups and sectors that had a non-interrupted presence.

Another choice that needs to be made is what are the scenarios we wish to consider. There exists a large number of ways to generate scenarios that take into account all kinds of data and/or expertise. We made the decision to solely use historical data up to 2012 in order to generate our scenarios, which we will then use to compute portfolios under different portfolio optimisation techniques. Data from 2013 to 2016 will then be used to analyse these portfolios' performance. The scenarios will be covariance matrices of subperiods of our training horizon. To determine the change-points we implement the algorithm proposed in [3]. This is an attractive option since it makes no distribution assumption and so fits well in the spirit of these experiments. While we could in theory have "forced" some dates to be change-points (Lehman Brothers collapse or 9/11, just to name a couple of possibilities), we wish to have as little direct influence as possible and so will stick with the dates generated by the algorithm.

2.3.1 Minimum Regret in the EMU

We will start by focusing on the European Economic and Monetary Union equity market. After running the algorithm proposed in [3], we arrive at the following change-points: 31/5/02, 06/06/03, 22/11/04, 27/04/06, 26/11/08, 26/04/10 and 04/01/12. From these we construct 8 sub-periods, which we then summarise into 8 covariance

matrices to be our scenarios. We can visualise them through the correlation heatmaps in Figure 2.2.

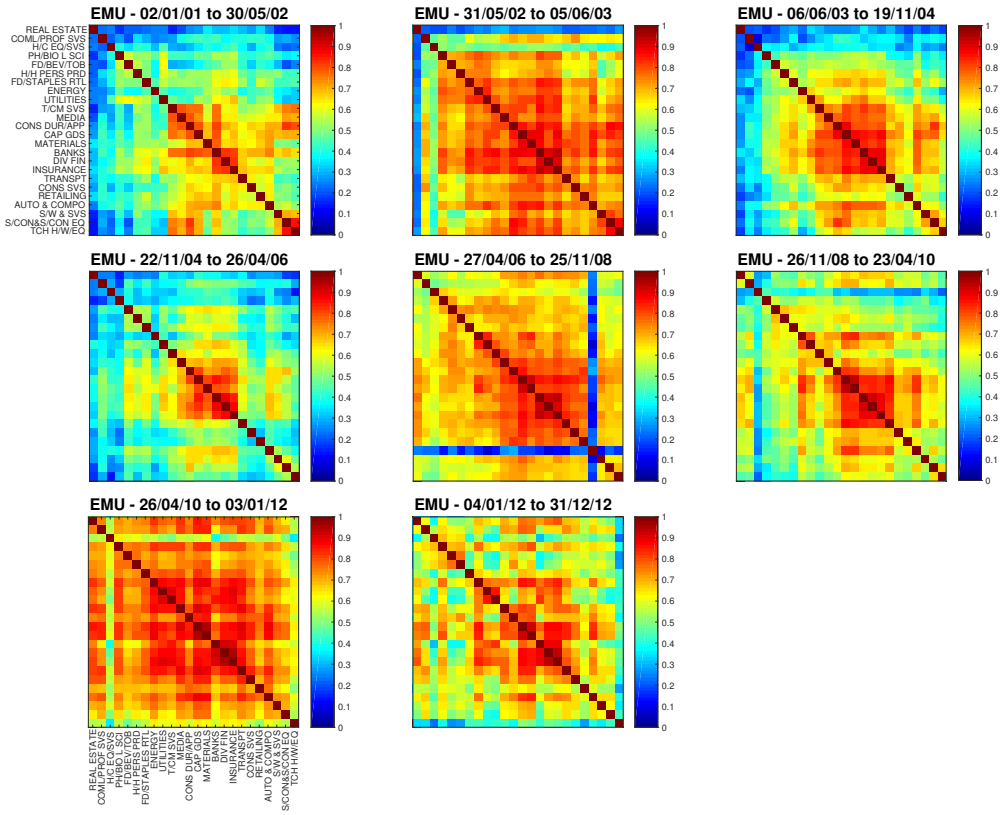


Figure 2.2: Heatmaps of correlation matrix for different scenarios in the EMU.

One can check whether these scenarios are sensible by following the methodology proposed in [24]. In Figure 2.3 we can see the behaviour of the first principal component of a one-year rolling window of returns over different scenarios. The first principal component is the market direction that accounts for the most variance, hence looking at how much of the total variance in the market is explained by a single direction is a powerful tool to visualise how fragmented the market is.

We can see that the changepoint detection algorithm is picking up genuine changes in the first principal component, hence reinforcing our conviction that these constitute true regime changes. The first scenario is purposely absent as the use of a one-year rolling window rules out this analysis over the first year, which is over two-thirds of the total time spent in that scenario.

We are now in a position to run our relative robust portfolio optimisation (RRPO), as in (2.2), and compare it to other methods. Two methods that are reasonable to compare our model against are the absolute robust portfolio optimisation (ARPO), as



Figure 2.3: Proportion of variance explained by the first principal component in EMU over different scenarios.

in (1.6), using the same scenarios as RRPO, and a simple minimum volatility portfolio optimisation (MVPO), as in (1.5), using the whole training period to compute the covariance matrix estimate. These portfolios have their weights detailed in Table 2.1.

Table 2.1: Weights of different portfolios being considered for EMU.

	RRPO	ARPO	MVPO
Food, Beverage & Tobacco	35.84%	32.96%	36.77%
Health Care Equipment & Services	33.28%	50.23%	44.17%
Food & Staples Retailing	10.15%	0%	0%
Media	10.13%	0%	0%
Telecommunication Services	5.74%	0%	2.16%
Real Estate	4.86%	4.09%	12.26%
Transportation	0%	9.65%	4.64%
Automobiles & Components	0%	3.07%	0%

Upon examining this table, one can immediately draw a few conclusions. First, all three portfolios concentrate on a small subset of all investable options (GICS has 24 industry groups, whereas Table 2.1 makes mention of only 8 of them). This is hardly surprising, as all of them are aiming at the same objective: having low volatility. It is then only natural that they tend to concentrate on low volatility assets. Second, RRPO is slightly more diversified than ARPO and MVPO, with a maximum weight

of 35.84% over 6 assets against a maximum of 50.23% and 44.17% over 5 of the other methods. Finally, it is evident that RRPO is yielding a structurally different portfolio to both ARPO and MVPO – the proportion of capital allocated differently is 29.67% and 23.86% respectively, both non-trivial quantities.

Let us now have a look at how these portfolios perform out of sample, over the period starting from 2013 until 2016. In Figure 2.4 we have a plot of the total return over time for each portfolio, while in Table 2.2 we list the realised return and volatility, in yearly terms, of each portfolio.

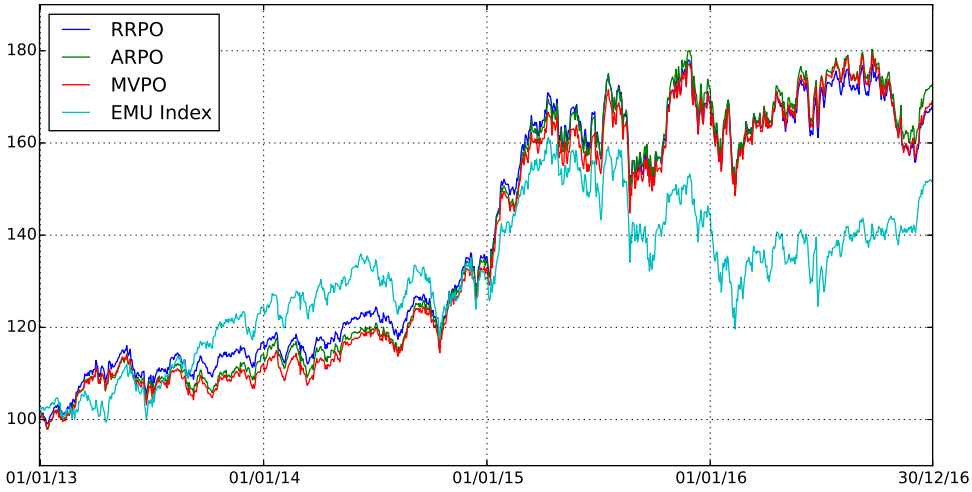


Figure 2.4: Performance of RRPO, ARPO, MVPO and EMU Index.

Table 2.2: Total return series assuming initial capital of 100. Realised return and volatility calculated for each civil year.

	RRPO	ARPO	MVPO	EMU Index
Realised Return 2013	15.04%	12.80%	10.97%	21.09%
Realised Return 2014	14.13%	15.79%	16.34%	4.78%
Realised Return 2015	21.45%	23.66%	23.19%	9.70%
Realised Return 2016	-0.34%	0.65%	0.40%	4.90%
Realised Volatility 2013	11.95pp	12.33pp	12.23pp	14.29pp
Realised Volatility 2014	12.36pp	13.10pp	12.72pp	15.41pp
Realised Volatility 2015	20.94pp	21.55pp	21.45pp	21.21pp
Realised Volatility 2016	17.46pp	18.42pp	17.98pp	20.30pp

It so turns out that for this dataset all three methods behave remarkably similarly,

even if one takes into account the fact that roughly 75% of capital is allocated equally. As we will see later on, this behaviour is purely coincidental – the same methods across different regions have discernible differences in realised return. We can however take note that RRPO was the least volatile portfolio on any of the years considered.

2.3.2 Maximum Expected Return in the US

We now wish to test another of the proposed methods, a proportional regret constraint of 10% in a maximum expected return portfolio optimisation (PRCPO), as in (2.4) but with $\text{PRgrt}_{\mathcal{U}}$ instead of $\text{Rgrt}_{\mathcal{U}}$. This will give us an example of a situation where the objective is already set, and we solely add a constraint to force a more stable behaviour. We will compare this against a Sharpe ratio absolute robust portfolio optimisation (SARPO) (where we aim to have the best worst-case Sharpe ratio) and against a maximum Sharpe ratio portfolio optimisation (MSPO). Moreover we will see how these methods compare when there are other restrictions present. We will not only keep the long-only constraint, but we will also add a 20% cap to force diversification.

We will use US equity market data so we can look in detail at another dataset. Our investable assets will still be the 24 GICS industry groups and our benchmarks the 10 GICS sector groups plus the US Index. The training and testing period also remain unchanged.

As before, we use the change-point detection proposed in [3] to come up with our scenarios. The change-points detected are 31/05/02, 15/06/04, 26/02/07, 03/09/08 and 20/12/11. We can see the heatmaps of each individual scenario in Figure 2.5.

As before, we can see the behaviour of the first principal component over different scenarios, except the first, in Figure 2.6. We can once again see clear differences in behaviour, which justifies our choice of scenarios as true regime changes. It is worth pointing out that while Figure 2.6 might suggest we should either anticipate or delay some of the change-points, their calculation was made using the whole covariance information, not just the first principal component. Making decisions based on Figure 2.6 would throw away all the information contained in the remaining principal components and would hence be suboptimal.

We use the 6 covariance matrices for both PRCPO and SARPO. For MSPO we use the covariance matrix calculated over the whole training period as our estimate. For all three portfolio optimisation problems we will use as the expected return of each individual asset the realised return over the most recent scenario – in this case, the realised return between 20/12/11 and 31/12/12. This is an arbitrary choice that is by

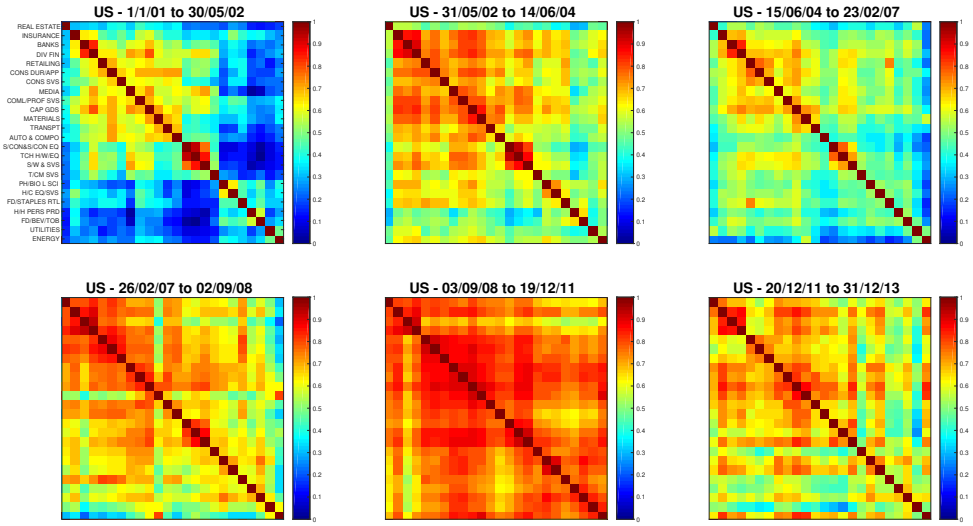


Figure 2.5: Heatmaps of correlation matrix for different scenarios in the US.

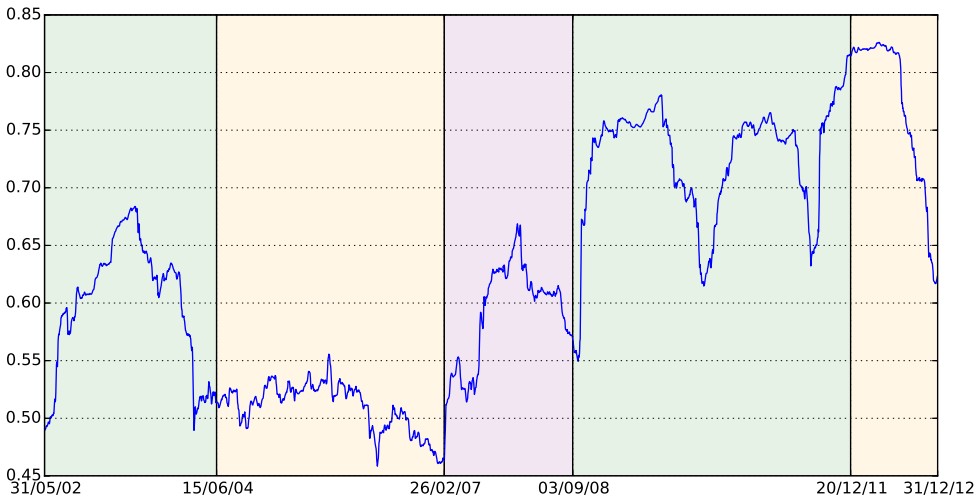


Figure 2.6: Proportion of variance explained by the first principal component in US over different scenarios.

no means a superior predictor to any other prediction, but is simple enough as a rule of thumb for us to use it as an example. Not only is forecasting expected returns beyond the scope of this thesis, but they also have a disproportionate impact on portfolio weights when compared to expected volatility or correlation, as demonstrated in [17] and [46]. Hence different forecasting techniques are likely to considerably impact the results that follow.

After solving these 3 optimisation problems, the resulting portfolios can be found in Table 2.3.

Table 2.3: Weights of different portfolios being considered for US.

	PRCPO	SARPO	MSPO
Food & Staples Retailing	20%	20%	20%
Food, Beverage & Tobacco	20%	0%	0%
Pharmaceuticals, Biotechnology & Life Sciences	20%	20%	20%
Household & Personal Products	19.44%	0%	0%
Telecommunication Services	12.94%	20%	20%
Media	7.62%	20%	20%
Retailing	0%	19.12%	9.40%
Real Estate	0%	0%	5.93%
Diversified Financials	0%	0%	4.66%
Health Care Equipment & Services	0%	0.88%	0%

It is unequivocal that different choices have been made, despite the fact that the objective function (i.e. the expected returns) is the same for all three portfolios. PRCPO focuses more on Food, Beverage & Tobacco, while SARPO and MSPO focus more on Telecommunication Services, Media and Retailing. They are all fairly diversified, but this is merely a symptom of the imposed cap. PRCPO agrees with both SARPO and MSPO on 60.56%, meaning 39.44% of the capital is allocated differently.

We can now analyse how these portfolios perform on the test period in Figure 2.7 and Table 2.4.

We can see that SARPO has a slightly higher annualised Sharpe ratio, but this however comes at the cost of higher volatility than we would consider admissible (this is relevant as we require the portfolio to be fully invested). In 2014, for example, SARPO and MSPO had respectively 17.2% and 13.4% more volatility than the least volatile benchmark, the Consumer Staples sector group, while PRCPO had only 1.93% extra volatility, well below the 10% cap imposed. While both SARPO and MSPO did outperform the market over this period, it could easily have gone the other way. This is a good example of why we think a proportional regret constraint is important – in this case it allowed us to pursue a high Sharpe Ratio while safeguarding against taking too much extra volatility.

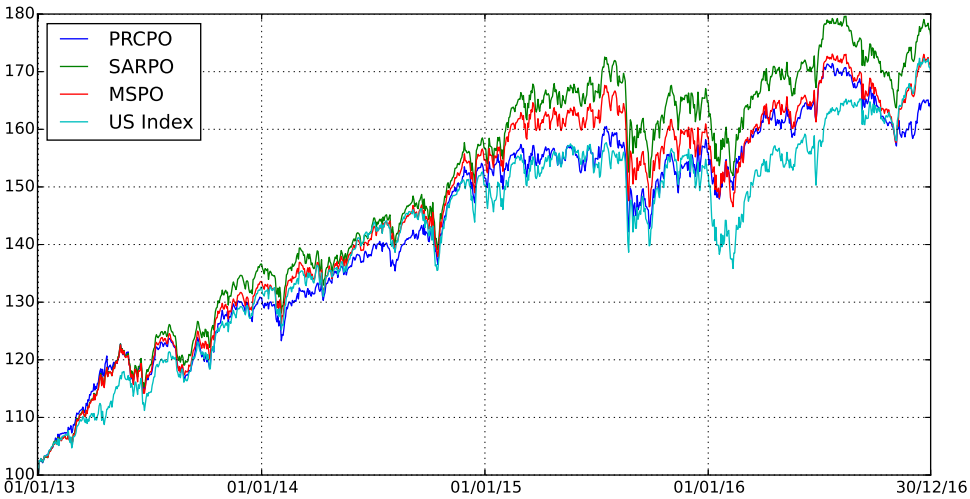


Figure 2.7: Performance of PRCPO, SARPO, MSPO and US Index.

2.3.3 Model Performance over Multiple Regions

As a way of reinforcing the points made earlier we carried out the same experiments on equity data from Australia, Canada, France, Japan, South Korea and United Kingdom. For each we calculated new change-points and recomputed the portfolios above.

Tables 2.5 and 2.6 summarise the results for each region. In the minimum regret problem we outline the percentage of capital that is allocated differently from RRPO to ARPO and MVPO, the biggest weight each allocates to a single industry group, their realised return and volatility for the entirety of the testing period. In the maximum expected value problem we display the percentage of capital allocated differently from PRCPO to SARPO and MSPO, their Sharpe ratio and volatility (the maximum weight is 20% by construction).

A few important points can be taken from tables 2.5 and 2.6. First, it is clear that both RRPO and PRCPO generally offer different solutions not covered by other methods. Secondly, RRPO tends to offer more diversification (by having less weight in a single industry group) than both ARPO and MVPO. In fact, the only instances when it does not is when the amount of capital allocated differently is marginal. This comes naturally from the fact that RRPO takes all scenarios into account, not only the worst-case scenario (as ARPO) and so favours diversification.

Thirdly, the only time one strategy significantly outperforms the others is PRCPO outperforming both SARPO and MSPO in Canada. Finally, it is worth mentioning

Table 2.4: Total return series assuming initial capital of 100. Realised return and volatility calculated for each civil year.

	PRCPO	SARPO	MSPO	US Index
Realised Sharpe Ratio 2013	2.40	2.78	2.56	2.50
Realised Sharpe Ratio 2014	1.48	1.20	1.32	1.08
Realised Sharpe Ratio 2015	0.20	0.33	0.17	0.08
Realised Sharpe Ratio 2016	0.44	0.57	0.61	0.82
Realised Volatility 2013	10.80pp	10.82pp	10.94pp	10.88pp
Realised Volatility 2014	9.61pp	11.05pp	10.69pp	11.22pp
Realised Volatility 2015	13.69pp	14.52pp	14.35pp	15.15pp
Realised Volatility 2016	10.46pp	11.52pp	11.48pp	12.97pp

Summary of performance of RRPO across 8 different regions, from 2013 to 2016.

	AUS	CAN	FRA	JPN	KOR	UK	EMU	US
A1	30.51%	2.33%	33.16%	33.59%	36.32%	5.85%	29.67%	12.62%
A2	20.26%	14.36%	28.85%	14.45%	8.97%	3.07%	23.86%	9.54%
B1	25.15%	38.96%	19.22%	25.12%	35.49%	32.37%	35.84%	53.22%
B2	29.04%	39.74%	26.38%	43.06%	43.60%	33.07%	50.23%	58.37%
B3	28.24%	41.83%	25.84%	30.13%	39.26%	31.78%	44.17%	49.31%
C1	9.64%	12.48%	8.06%	16.15%	4.14%	7.77%	12.57%	11.82%
C2	11.82%	12.44%	6.75%	16.59%	4.13%	8.01%	13.23%	12.22%
C3	10.99%	14.22%	6.83%	15.72%	4.36%	7.46%	12.73%	11.96%
C4	9.11%	8.21%	10.60%	15.07%	1.73%	7.94%	10.12%	12.81%
D1	12.73pp	11.14pp	16.82pp	19.55pp	12.70pp	13.09pp	16.12pp	11.43pp
D2	13.08pp	11.23pp	16.52pp	19.30pp	14.23pp	13.09pp	16.79pp	11.41pp
D3	12.90pp	10.93pp	16.44pp	19.46pp	12.86pp	13.14pp	16.55pp	11.31pp
D4	14.07pp	12.01pp	18.23pp	22.22pp	12.99pp	14.34pp	18.06pp	12.68pp

Table 2.5: A - Percentage of capital RRPO invests differently from ARPO (1) and MVPO (2).
B - Highest weight in RRPO (1), ARPO (2) and MVPO (3).
C - Realised return of RRPO (1), ARPO (2), MVPO (3) and Regional Index (4).
D - Realised volatility of RRPO (1), ARPO (2), MVPO (3) and Regional Index (4).

that over the past few years low volatility assets have outperformed the market. This means that portfolios that invest in these assets, such as the ones examined, have also outperformed the market. There is no reason to believe this to be indefinitely true, and so none of these portfolios is guaranteed to persistently outperform the market.

Summary of performance of PRCPO across 8 different regions, from 2013 to 2016.

	AUS	CAN	FRA	JPN	KOR	UK	EMU	US
E1	23.21%	35.58%	24.41%	22.24%	0.00%	15.32%	18.49%	39.44%
E2	40.00%	43.57%	11.74%	19.00%	8.57%	15.54%	29.51%	39.44%
F1	0.90	0.94	0.58	0.80	0.32	0.79	0.74	1.06
F2	0.99	0.37	0.51	0.72	0.32	0.70	0.76	1.13
F3	0.96	0.53	0.56	0.78	0.36	0.79	0.69	1.08
F4	0.65	0.68	0.58	0.68	0.13	0.55	0.56	1.01
G1	13.24pp	10.96pp	17.15pp	19.75pp	13.89pp	13.73pp	16.70pp	11.26pp
G2	12.65pp	17.38pp	17.37pp	20.63pp	13.89pp	14.36pp	16.72pp	12.08pp
G3	12.38pp	17.35pp	16.61pp	19.70pp	13.13pp	14.73pp	16.49pp	11.97pp
G4	14.07pp	12.01pp	18.23pp	22.22pp	12.99pp	14.34pp	18.06pp	12.68pp

Table 2.6: E - Percentage of capital PRCPO invests differently from SARPO (1) and MSPO (2).
F - Realised Sharpe ratio of PRCPO (1), SARPO (2), MSPO (3) and Regional Index (4).
G - Realised volatility of PRCPO (1), SARPO (2), MSPO (3) and Regional Index (4).

2.4 Conclusion

We demonstrated that regret minimisation does not only produce novel but also sensible results. Indeed, we observed that across different datasets it provides a greater degree of protection than absolute robust optimisation and the more widely used minimum volatility optimisation. A minimum regret portfolio optimisation problem assumes a fresh start and so should be interesting from an academic point of view, although it may also be implemented by a defensive portfolio manager, whose main concern is not to be too heavily outperformed. A proportional regret constraint constructed using a suitable set of benchmarks and scenarios, however, would give investors the confidence to pursue their objectives while at the same time controlling the level of extra volatility one is prepared to take. Consequently it is a tool that deserves to be considered by any portfolio manager.

This is clearly not the El Dorado in asset management, but instead a small cog in a much larger (and hopefully successful) engine. The best results will be achieved by those who possess a profitable trading strategy and combine it with a wide range of techniques including but not limited to robust parameter estimation, inclusion of regularisation terms, transaction costs modelling and tail-risk modelling.

Chapter 3

Portfolio Optimisation for Medium Frequency Trading Strategies and Heavy-Tailed Returns

3.1 Introduction

Consider a portfolio optimisation problem in a one-period setting, where the aim is to find an optimal allocation of investment capital over a fixed investment horizon under the Markowitz's mean-variance approach,

$$\begin{aligned} \max_{x\in\mathbb{R}^n} \ & \mu^\top x \\ \text{s.t.} \ & x^\top Q x \leq \sigma^2, \\ & x \in \mathscr{X}, \end{aligned} \tag{3.1}$$

where x is a vector of proportions of risk capital allocated to individual assets, μ is a vector of expected asset returns, Q is the covariance matrix of asset returns, σ is the volatility budget and $\mathscr{X}$ a convex set that describes the set of decisions x that satisfy the allocation constraints.

If, as in Chapter 2, we are considering investing in indices, a reasonable choice of $\mathscr{X}$ is the set of all long-only portfolios with no leverage (whether fully invested or not). This is a risky market where substantial movements can happen in the medium term (a day, for example), but loss of all capital is virtually impossible. In this chapter we will however be focusing our attention on the futures market – where there is no fundamental difference between a long or a short position and leverage is widely used as contracts are settled daily. In such a market it is reasonable to take $\mathscr{X} = \mathbb{R}^n$, i.e., to impose no constraints.

As mentioned in the Introduction, practitioners often find that optimal portfolios produced by models of the form (3.1) are not well diversified and can be sensitive to parameter misspecification. Since statistical estimates of model parameters inherently contain noise, model parameters are not known exactly. This is particularly problematic for μ, as shown in [17].

Another problem that plagues the mean-variance model is that it typically produces asset allocations with very large positions in a few assets and comparably very small positions in the remaining assets, that is, the optimal portfolios are frequently imbalanced. Non short-selling and non leverage constraints can mitigate this problem, but this effect is particularly pronounced when short-selling and leverage is allowed, such as when considering a portfolio of futures contract.

A portfolio with imbalanced positions would not be fundamentally problematic if the asset returns were approximately multivariate normally distributed with exponentially decaying distribution tails. However, the behaviour of tail events is typically not well described by estimated covariance matrices because these events happen rarely, while covariance matrices tend to be dominated by the bulk of the data. An imbalanced portfolio may thus typically contain higher idiosyncratic tail risk than one might be expecting, that is, it is more strongly exposed to price shocks.

The dependence of asset returns in the tails is often very different from the dependence during "typical" periods, and unless the tail dependence is modelled explicitly, the tail risk carried by a portfolio remains hidden. The approach we take in this chapter is therefore to model the tail dependence of returns separately from the dependence in the region where the bulk of its probability measure lies. Using the modelling of tail events in a novel way, we combine extreme value theory, principal component analysis and convex relaxation to arrive at a *tail-event* constraint that can be cast in second-order programming form, therefore maintaining the same complexity as a standard mean-variance model.

Alternatively we can use this low rank quadratic form to construct a low-dimensional uncertainty set $\mathscr{U}$. Using $\mathscr{U}$ will yield the same complexity as using other ellipsoidal uncertainty sets that have been proposed, for example by [27]. However not only does it incorporate the information on extreme events, which makes the motivation of using an uncertainty set much more clear, but it also produces portfolios that are almost never uninvested, an issue that is largely present with full dimensional uncertainty sets.

3.2 Tail-Event Constraint via Extreme Value Theory

Tail risk has mostly been considered in portfolio optimisation either through Value-at-Risk or, more recently, Conditional Value-at-Risk. CVaR is a very important risk measure in practice, both for risk management and for regulatory purposes, and hence it is regularly present in the minds of asset managers. Unfortunately tail risk does not often play a role in portfolio optimisation – often an optimal portfolio is found and its CVaR is calculated afterwards, rather than it being a factor of consideration when choosing the optimal portfolio.

While it is possible to include CVaR constraints in a portfolio optimisation problem (see [50]), this is often impractical when considering a large number of assets. Not only do most approaches impose a large number of linear and quadratic constraints (which are trivial to solve in small quantities, but lead to ill-conditioned problems in high dimensions due to the exponential growth of auxiliary variables), but they also require a multidimensional distribution of the tails, which is often impossible to provide given the lack of data, by definition, on those tails.

We wish to deal with this issue by relying on Extreme Value Theory, in particular the so called Peak Over Threshold method. The rough idea of this method is to look at uncorrelated univariate principal components, fitting a Pareto distribution on them (as motivated by Picklands-Balkema-de Haan theorem) and finally use convex relaxation to obtain a *tail-event* constraint. While this mainly takes advantage of univariate methods, we motivate its use on a multivariate setting with recent developments in multivariate Peak Over Threshold models.

3.2.1 Review of Extreme Value Theory

Our analysis is only valid for heavy tailed distributions, which intuitively are the ones that decay polynomially, not exponentially, as observed in financial data. If X has such distribution, then

$$\mathbb{P}\left[X > x\right] \approx \frac{c}{x^{1/\xi}}$$

for some constant $c > 0$. This can be made explicit through the use of slowly varying functions. For an introduction to Extreme Value Theory, see [29].

It is worth reviewing the theoretical result that underpins the Peak Over Threshold method.

Theorem 4 (Picklands-Balkema-de Haan [2]). *X is heavy tailed if and only if there exists $\xi > 0$ and $\beta(u)$ positive function such that*

$$\lim_{u\to\infty} \sup_{0\le x<\infty} |F_u(x) - G_{\xi,\beta(u)}(x)| = 0,$$

where

$$F_u(x) = \mathbb{P}\left[X \le u + x \mid X > u\right] = \frac{F(x+u) - F(u)}{1 - F(u)}, \quad x > 0,$$

is the cumulative distribution function conditional on $X > u$ and

$$G_{\xi,\beta}(x) = 1 - \left(1 + \frac{\xi x}{\beta}\right)^{-1/\xi}, \quad x \ge 0,$$

is a generalized Pareto distribution.

Even though the function $\beta(u)$ is not explicitly given in the proof of the above theorem, it is later proven that it can be taken to be $\beta(u) = \xi u$, as seen in [44] – which means the tail of a heavy tail distribution can be approximated by a standard Pareto distribution.

Hence if X is heavy tailed its tail converges to a Pareto and moreover this convergence is uniform. It is hence sensible to model the tail as a Pareto distribution - this is know as the Peak Over Threshold method.

If we were working on a univariate problem, this would be all we needed to know. Given a set of i.i.d. sample points $x_1, \ldots, x_N$ from heavy-tailed random variables $X_i \sim X$, we can estimate the distribution of X as follows.

1. Select $q \in (0, 1)$ and estimate u such that $\mathbb{P}\left[X \le u\right] = q$.

2. Fit a distribution to X empirically on the range $(-\infty, u]$.

3. Estimate β, ξ and model the tail $X > u$ by a Generalised Pareto distribution,

$$\mathbb{P}\left[X \le u + x \mid X > u\right] = G_{\xi,\beta}(x), \quad x > 0.$$

4. Once a distribution has been fitted to the data in this fashion, use this distribution in any calculations of moments (which might not always exist!), quantiles, shortfalls, or any other quantities that might be required in risk or portfolio management[1].

[1]This is however not always theoretical justifiable - for a discussion of shortcomings see [15].

Unfortunately in a multivariate setting a lot more care is required. First of all, we have to formulate what we mean by Peak Over Threshold – do we mean at least one variable is above a certain threshold? Or all of them? Maybe in the L^1 or L^2 norm? Different definitions lead to different limiting distributions, so one needs to consider which is the most appropriate approach.

Since we are interested in financial applications, we will consider a tail event to be one where at least one variable is above a certain threshold. This is the approach taken in [51]. Let Y be a random vector and u a vector of thresholds. We say Y is an *extreme* event if $Y \not\leq u$, i.e., if $Y_j > u_j$ for some j. We are interested in the distribution of Y given it is extreme, i.e., we are interested in $\mathbb{P}\left[Y \leq x + u \mid Y \not\leq u\right]$.

The generalisation of Theorem 4 is beyond the scope of this chapter. We will merely assume that our distribution is in the domain of attraction of a multivariate generalized Pareto distribution with positive ξ. For further details see [52].

There is however one important result worth mentioning. Let Z be a multivariate generalized Pareto distributed random vector. Then

$$\mathbb{P}\left[Z_j \leq x \mid Z_j > 0\right] = G_{\xi_j,\beta_j}(x), \quad x > 0,$$

where β_j and ξ_j are marginal scale and shape parameters. Hence we recover our Pareto distribution in our tail, as long as we are above the threshold.

3.2.2 Tail-Event Constraint

The tail dependence between different assets is hard to estimate due to a lack of significant amounts of data in the tail. Instead of modelling this explicitly and trying to somehow calibrate such a model, we will instead approximate a portfolio CVaR without explicitly modelling that dependence.

As a first step we will consider principal components instead of assets. This is purely an orthogonal transformation that transforms the assets into uncorrelated linear combinations. While these components are not independent in general, particularly in the tail, they are natural objects to work with, as they represent the typical directions of market movement.

Let R be the random vector of returns and Q its $n \times n$ covariance matrix. Then Q is positive definite and has eigenvalue decomposition

$$Q = VSV^\top, \tag{3.2}$$

where $S = \text{Diag}(\sigma_1^2, \ldots, \sigma_n^2)$ with $\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_n^2$ and V an orthogonal matrix whose j-th column v_j is a unit eigenvector associated with eigenvalue σ_i^2. v_j also happens to be the j-th principal component, the directions we wish to consider.

Before talking about extreme events, we have to pick suitable thresholds. In the univariate Peak Over Threshold method, these are often quantiles, either empirical or from the normal distribution. We shall go with the latter – by assuming the bulk of the data is normally distributed, for each direction v_j we can set the threshold u_j at some multiple ι of the standard deviation σ_j. If for example we wanted for a 95% quantile, then we would take $\iota = 2$ and hence $u_j = 2\sigma_j$ for all j.

Let $Y = V^\top R$ be our returns random vector after the principal components transformation. If x is our portfolio, then the return of that portfolio is

$$R^\top x = \sum_{j=1}^{n} (v_j^\top R)\, v_j^\top x = \sum_{j=1}^{n} Y_j\, v_j^\top x.$$

The CVaR of this portfolio would be $\mathbb{E}\Big[\, |R^\top x| \,\Big|\, |R^\top x| > \alpha \Big]$ for some level α. We take absolute values as we wish to allow short-selling – for long-only portfolios, this is not required. As we discussed previously, this is not practical in most situations. Instead we will use an upper bound on $\mathbb{E}\Big[\, |R^\top x| \,\Big|\, |V^\top R| \not\leq u \Big]$ as a proxy for the CVaR term,

$$\begin{aligned}
\mathbb{E}\Big[\, |R^\top x| \,\Big|\, |V^\top R| \not\leq u \Big] &\leq \sum_{j=1}^{n} \mathbb{E}\Big[\, |Y_j| \,\Big|\, |Y| \not\leq u \Big]\, |v_j^\top x| \\
&\leq \sum_{j=1}^{n} \mathbb{E}\Big[\, |Y_j| - u_j \,\Big|\, |Y_j| > u_j \Big]\, |v_j^\top x| + \sum_{j=1}^{n} u_j\, |v_j^\top x|. \qquad (3.3)
\end{aligned}$$

As motivated earlier, the distribution of $\Big(|Y| \,\Big|\, |Y| \not\leq u \Big)$ will be approximated by a multivariate generalized Pareto distribution, which in turn means that the distribution of $\Big(|Y_j| - u_j \,\Big|\, |Y_j| > u_j \Big)$ can be approximated by a univariate generalized Pareto distribution with parameters β_i, ξ_i, which has expected value $\frac{\beta_j}{1-\xi_j}$. Given that $u_j = \iota\, \sigma_j$ we get

$$\mathbb{E}\Big[\, |R^\top x| \,\Big|\, |V^\top R| \not\leq u \Big] \leq \sum_{j=1}^{n} \frac{\beta_j}{1-\xi_j}\, |v_j^\top x| + \iota \sum_{j=1}^{n} \sigma_j\, |v_j^\top x|. \qquad (3.4)$$

We have therefore arrived at a crude linear approximation of CVaR, which can in turn be used as a constraint in a portfolio optimisation problem. We would stop here, if it were not for the fact that box-type constraints are not very well behaved – they lead to corner solutions which have a tendency to be unstable. It is therefore much

better to substitute it with a quadratic constraint. For that purpose, let us introduce the following matrix,

$$\Lambda = V \operatorname{Diag}\left(\frac{\beta_j^2}{(1-\xi_j)^2}\right) V^\top.$$

It follows from the arithmetic-quadratic mean inequality that (3.4) can be further bounded by

$$\mathbb{E}\Big[|R^\top x| \,\Big|\, |V^\top R| \not\leq u\Big] \leq \sqrt{n}\,\sqrt{x^\top \Lambda x + \iota^2\, x^\top Q x}.$$

The $\sqrt{n}$ term emerges because in (3.3) we effectively assumed that we would see an extreme event in all directions simultaneously. This is too pessimistic, so we can tighten the above inequality by dropping the $\sqrt{n}$.

To limit the expected conditional loss for an extreme event to a target ϱ, we could thus impose the following *tail-event* constraint to our portfolio optimisation problem

$$x^\top(\Lambda + \iota^2 Q)x \leq \varrho^2.$$

This is however somewhat redundant, as commonly we already have a volatility constraint of the form $x^\top Q x \leq \sigma^2$. We can hence replace the above constraint by

$$x^\top \Lambda x \leq \eta^2, \tag{3.5}$$

While technically this is a tighter constraint, it gives one flexibility to manage separately the tail risk from the volatility risk.

3.3 A Bimodal Approach to Heavy Tail Modelling

So far we have classified as *extreme* a return vector R if it is above a threshold in at least one of its principal components. We have assumed that we would use the same data to estimate both Q and Λ. In this section we will develop a bimodal approach and hence estimate these two matrices separately. The idea is the following:

- Separate the historical data into two datasets, roughly into "typical" and "extreme" market behaviour;
- From the typical dataset compute covariance $Q^{(t)}$, while from the extreme dataset compute covariance matrix $Q^{(e)}$;
- Perform eigenvector decomposition on $Q^{(e)}$ to find directions v_j;

- Pick a multiple ι so that the threshold in each direction v_j is $u_j = \iota\,\sigma_j$ ($\iota = 2$ means a 95% threshold in each direction);
- For each direction fit a Pareto distribution and find parameters β_j, ξ_j and hence construct matrix Λ;
- Impose the following tail-event constraint to the portfolio optimisation problem

$$x^\top \Lambda x \leq \eta^2.$$

This is similar to what we did in the previous section, where we assumed that $Q^{(e)} = Q^{(t)} = Q$. The underlying idea of this section is that market correlations behave differently during extreme events than it does during normal times, so one is better off by modelling these two separately. This notion of different regimes in the covariance structure has been widely identified both in theory and in practice, see for example [24].

It is worth pointing out at this point that in practice it might not be possible to fit a Generalised Pareto distribution into each individual direction. This might be the case because some directions are more fat-tailed than others, but it is also a consequence of having a limited amount of data, especially in high dimensions. In those circumstances we might either compromise and fit a standard Pareto distribution ($\xi = 1$) or, alternatively, simply consider that particular direction as a non-fat tail direction and not include it in the construction of our matrix Λ. This second alternative will come up later when we the use of such matrix in absolute robust portfolio optimisation.

For the numerical methods that follow we will be using a two state Hidden Markov Model, as seen in Section 1.2.3, with an underlying that is normally distributed in one state and t-distributed in other. This seems like an appropriate choice since a t-distribution is a multivariate heavy tailed distribution, while the normal distribution is appropriate for the "bulk" of the data.

As a short recap, let X_t be a two state Markov Chain that takes values in $\{e_1, e_2\}$, Z_t^1 a sequence of i.i.d. random variables normally distributed for some mean and covariance matrix and Z_t^2 a sequence of i.i.d. random variables t-distributed with some mean, spread matrix and degrees of freedom. Our model for the assets' returns r is

$$r_t = \left(e_1^\top X_{t-1}\right) Z_t^1 + \left(e_2^\top X_{t-1}\right) Z_t^2,$$

i.e., if X_{t-1}, the *market state*, is equal to e_1 then r_t is normally distributed, otherwise it will be t-distributed, a heavy-tailed distribution. If $\mathcal{Y} = \sigma\{r_t, t \leq T\}$ is the observed filtration, then our filter is

$$\mathfrak{F}(r_t) = \begin{cases} 1 & \text{if } \mathbb{E}\left[e_1^\top X_{t-1} \mid \mathcal{Y}\right] \geq 0.5, \\ 0 & \text{otherwise.} \end{cases} \tag{3.6}$$

This is by no means the only or even the most appropriate choice for a filter. We make it for convenience – it can be fitted with an EM-algorithm and does not require any exogenous parameter – and because we shall use this same filter in a later chapter, although in a different context. However this method works with any (reasonable) filter and the best that filter is at separating "typical" market conditions from "extreme" ones, the better the method will perform.

3.3.1 Constructing Different Constraints for Different Market States

Once we pick a filter, we compute two "covariance" matrices, $Q^{(t)}$ and $Q^{(e)}$, from two distinct data series.

$$Q^{(t)} = \operatorname{Cov}\left(\{r_t : \mathfrak{F}(r_t) = 1\}\right),$$
$$Q^{(e)} = \operatorname{Cov}\left(\{r_t : \mathfrak{F}(r_t) = 0\}\right).$$

That is, we compute $Q^{(t)}$ using the "typical" data points and we use the "extreme" ones to calculate $Q^{(e)}$. Any sensible method can be used at this step: taking the sample covariance, using exponential moving averages and using shrinkage methods (see [39]) are only a few options.

Once we have obtained $Q^{(e)}$, we can identify the principal directions of market movement under "extreme" conditions. For each principal direction v_j we consider the threshold $u_j = \iota \sqrt{v_j^\top Q^{(t)} v_j}$ and fit a generalised Pareto distribution on the extreme dataset whose v_j direction is above u_j, ι multiples of standard deviation during "normal" conditions. We can then construct matrix Λ and proceed as in the previous section. Notice that since we took this bimodal approach $Q^{(t)}$ and Λ will not have the same eigenvectors.

In summary:

1. We classify our data as either "extreme" or "typical" using a filter $\mathfrak{F}$;

2. We compute two positive definite covariance matrices $Q := Q^{(t)}$ and $Q^{(e)}$;

3. For each unit eigenvector v_j of $Q^{(e)}$, we project the extreme data onto it, choose the threshold level u_j to be ι multiples of standard deviation of the "typical" dataset in that direction and fit a generalised Pareto distribution on the tail;

4. We construct matrix Λ, which has the eigenvectors as $Q^{(e)}$ and has $\frac{\beta_j^2}{(1-\xi_j)^2}$ as eigenvalues, as justified by (3.5);

5. Finally, impose our tail-event constraint along with a volatility constraint

$$x^\top \Lambda x \leq \eta^2,$$

where η is a target for the mean excess of extreme returns.

6. Given a vector of expected returns μ and a volatility budget σ, the optimal portfolio is the solution of

$$\begin{aligned} \max_x \ & \mu^\top x \\ \text{s.t. } & x^\top \Lambda x \leq \eta^2, \\ & x^\top Q x \leq \sigma^2. \end{aligned} \tag{3.7}$$

3.4 Use in Robust Portfolio Optimisation

The two matrices Λ and Q codify different sources of information from the market. While Q describes the most common market moves, Λ tries to capture the tail behaviour, precisely the events that are not described in Q. This was indeed the motivation for the bimodal approach in the previous section. There is however another way to incorporate the information in Λ beyond using it as a quadratic constraint – using it to generate an uncertainty set for the expected return vector μ. This can be motivated by assuming that an extreme event merely changes the mean of the returns' distribution.

3.4.1 Absolute Robust Portfolio Optimisation

As we saw previously, the idea in Robust Portfolio Optimisation is to assume μ not as fixed, but as to belonging to an uncertainty set $\mathscr{U}$, the set of all "reasonable" values

for μ. The Absolute Robust Portfolio Optimisation solution is then the portfolio that has the best worst-case scenario, i.e.,

$$\begin{aligned} &\max_{x} \left(\min_{\mu \in \mathscr{U}} \ \mu^\top x \right) \\ &\text{s.t. } x^\top Q x \leq \sigma^2. \end{aligned} \tag{3.8}$$

For a review, see [35]. There are two common choices for $\mathscr{U}$. The first one is a hypercube, with lower and upper bounds for the expected value of each asset. These bounds are obtained by assuming the asset returns are normal and then constructing a confidence interval. These type of uncertainty set usually leads corner solutions which tend to be very sensitive to numerical errors, see [41]. The second one is an ellipsoid and is essentially an extension of the first – instead of getting a confidence interval per asset, one assumes the vector of returns is normally distributed with covariance Q and then taking $\mathscr{U}$ to be a confidence region.

The problem is that in practice this is redundant – the behaviour of the portfolio with or without robustness is not significantly different, as the objective function in (3.8) already incorporates Q. A better alternative is to use Λ to generate $\mathscr{U}$ – this way the objective function deals with volatility aversion, while the uncertainty set deals with tail event aversion. This can be motivated through our bimodal approach. Say we have a estimate of μ which we believe is correct for "typical" market behaviour. We know that "extreme" market corrections arise from Λ, so one can assume that these corrections happen because the expected returns vector has been "corrected".

Another common problem with (3.8) is that it is highly sensitive to the choice of confidence level used to define $\mathcal{U}$, for if the origin lies in its interior the optimal portfolio is fully uninvested. This makes intuitive sense – if every possible market direction can be expected, then the best worst-case scenario is achieved by not investing at all. Due to the low-dimensionality of the uncertainty set defined in terms of Λ, this phenomenon has zero probability of occurring in our setting.

Consider the eigenvalue decomposition of $Q^{(e)}$ in (3.2). If $\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_n^2$, then v_j is the j-th biggest direction of market correction. The proportion of variance it is responsible for is $\frac{\sigma_j^2}{\sum_{i=1}^n \sigma_i^2}$. It is therefore much more important to seek protection against moves in the first few principal components, as these are responsible for most of the corrections, rather than in the last ones (in financial data eigenvalues tend to decay exponentially, as exemplified in Figure 3.1). Suppose we only want protection against the directions responsible for $\delta\%$ of the total variance. We can then focus on

the top ρ direction, where ρ is defined as

$$\rho = \min \left\{ j : \sum_{i=1}^{j} \sigma_i^2 \geq \frac{\delta}{100} \sum_{i=1}^{n} \sigma_i^2 \right\}, \tag{3.9}$$

i.e., the first ρ principal directions are responsible for $\delta\%$ of the variance during "extreme" market behaviour. Alternatively one can simply pick ρ directly instead of δ – picking $\rho = 1$, for example, means we are only considering the highest volatile direction during "extreme" periods.

It is worth mentioning that, at least in the case of $Q^{(e)}$ where data will be scarcer, we are not really interested in its eigenvalues, but solely on its eigenvectors. If financial returns exhibit exponentially decaying eigenvalues, as seen for example in [61], then [36] suggests that the leading eigenvectors are much more stable than the trailing ones. This in thus reinforces our original proposal (3.9) to seek protection only against the leading directions of market correction. We can see in Figure 3.1 that seems to hold in data we will analyse in the next section.

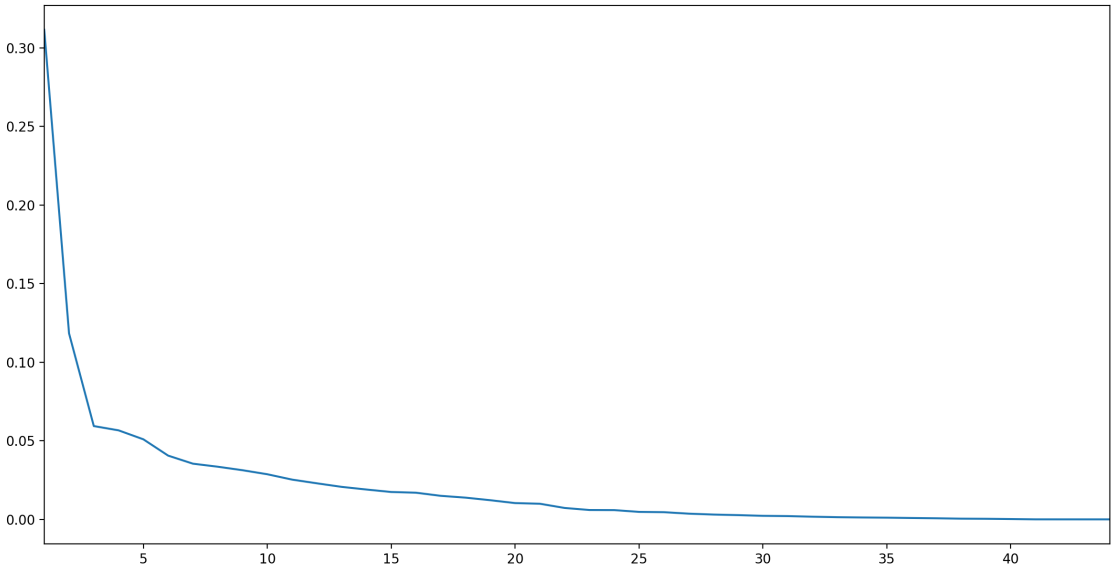


Figure 3.1: Proportion of variance in Q^e at the start of 2015 explained by each eigenvector direction, for the dataset and filter discussed in the next section.

For example, if in the context of the data in Figure (3.1) we were to take $\delta = 80\%$, then $\rho = 12$ out of 44 total eigenvectors would suffice. If we were to pick ρ directly, $\rho = 3$ or $\rho = 6$ seem to be appropriate choices for this particular day and example

After having selected ρ, one can then follow the argument of Section 3.2 assuming movements above the threshold M_j for *only* the first ρ principal directions, leading to matrix $\hat{\Lambda}$

$$\begin{aligned}\hat{\Lambda} &= \begin{bmatrix} | & & | \\ v_1 & \dots & v_\rho \\ | & & | \end{bmatrix} \operatorname{Diag}\left(\frac{\beta_1^2}{(1-\hat{\xi}_1)^2}, \dots, \frac{\beta_\rho^2}{(1-\hat{\xi}_\rho)^2}\right) \begin{bmatrix} | & & | \\ v_1 & \dots & v_\rho \\ | & & | \end{bmatrix}^\top \\ &= \hat{V}\,\Xi\,\hat{V}^\top.\end{aligned} \tag{3.10}$$

Notice that $\hat{V}$ is a $n \times \rho$ matrix that only has the ρ leading eigenvectors, while Ξ is a $\rho \times \rho$ diagonal matrix.

If $\hat{\mu}$ is our estimate for the expected returns vector, then if we assume μ to be normally distributed with mean $\hat{\mu}$ and covariance matrix $\hat{\Lambda}$ (or a multiple of it), then its confidence region $\mathscr{U}$ is

$$\mathscr{U} = \hat{\mu} + \hat{V}\left\{\nu \in \mathbb{R}^\rho : \nu^\top \Xi^{-1}\,\nu \leq \chi^2\right\}, \tag{3.11}$$

where the exact value of χ^2 depends on the level of confidence needed and on the relative size of the covariance matrix (which is inversely proportional to the sample size).

It is worth pointing out that (3.8) can be efficiently solved by analytically finding the worst-case scenario $\mu \in \mathscr{U}$ for each portfolio x,

$$\mu^*(x) = \hat{\mu} - \sqrt{\chi^2}\,\frac{\hat{\Lambda}\,x}{\sqrt{x^\top \hat{\Lambda}\,x}},$$

and then simplifying it into

$$\begin{aligned}\max_x \;& \hat{\mu}^\top x - \sqrt{\chi^2\,x^\top \hat{\Lambda}\,x} \\ \text{s.t.} \;& x^\top Q x \leq \sigma^2.\end{aligned} \tag{3.12}$$

Hence in this setting Absolute Robust Portfolio Optimisation with an ellipsoidal uncertainty set translates into simply penalising by the square root of a quadratic function. This is in turn equivalent to simply adding (3.5) as a constraint for some level η, as one can see by using Lagrange multipliers. Even though we are assuming x is unconstrained, the exact same rationale follows if x is a constrained set.

By using a low-dimensional uncertainty set the probability of it containing the origin is zero, but the following result guarantees that the portfolio obtained for such degenerate input would be well behaved.

Lemma 5. *If $0 \in \mathcal{U}$ then any admissible portfolio x with $\hat{V}^\top x = 0$ is a solution to* (3.12) *and hence an optimal portfolio, including the uninvested portfolio $x = 0$.*

Proof: Let W be the matrix of trailing $(n-\rho)$ eigenvectors, so that the orthonormal matrix $V = \begin{bmatrix} \hat{V} & W \end{bmatrix}$. If $0 \in \mathcal{U}$ there must be $\nu \in \mathbb{R}^\rho$ such that $\hat{V}\nu = \hat{\mu}$ and $\nu^\top \Xi^{-1} \nu \leq \chi^2$. Let $x = \hat{V}y + Wz$ be the decomposition of x onto $\hat{V}$ and W. If we write $w = \Xi^{1/2}\, y$ then we can rewrite (3.12) as

$$\begin{aligned} \max_{w,z} \quad & \left(\Xi^{-1/2}\nu\right)^\top w - \sqrt{\chi^2}\, \|w\|_2 \\ \text{s.t.} \quad & \left(\hat{V}\, \Xi^{-1/2}\, w + W z\right)^\top Q \left(\hat{V}\, \Xi^{-1/2}\, w + W z\right) \leq \sigma^2. \end{aligned}$$

This objective function is a homogeneous function of w and it is maximised when $w = k\, \Xi^{-1/2}\nu$ for some $k \geq 0$. It is however not a function of z. Substituting we get $k\left(\nu^\top \Xi^{-1}\nu - \sqrt{\chi^2\, \nu^\top \Xi^{-1}\nu}\right)$, but this is non-positive since $\nu^\top \Xi^{-1}\nu \leq \chi^2$. Hence the objective function is maximised at $w = 0$. Since the function we aim to maximise does not depend on z, any admissible $x = W\, z$ is an optimal portfolio. □

Notice that if $\mathscr{U}$ were to be generated by Q instead of $\hat{\Lambda}$, problem (3.12) would have simply added a standard deviation penalty to (3.1), which is redundant given we already have one for the variance.

We shall leave the numerical experiments regarding this subsection and the next to Chapter 4.

3.4.2 Relative Robust Portfolio Optimisation

A more interesting version is the Relative Robust Portfolio Optimisation introduced in [32]. This considers to be optimal the portfolio x that minimises our "regret" for not having picked another portfolio y had we known the true value of μ. This can be written as the following optimisation problem

$$\begin{aligned} \min_{x} \max_{\mu \in \mathscr{U}} \max_{\{y : y^\top Q y \leq \sigma^2\}} & \left(\mu^\top y - \mu^\top x\right) \\ \text{s.t.} \quad & x^\top Q\, x \leq \sigma^2. \end{aligned} \tag{3.13}$$

If x has any constraints, then these constraints must also be imposed on y, unless we wish to specify independent benchmarks as in Chapter 2. However even if left unconstrained (3.13) will in general be NP-hard, as shown in [47]. In the next chapter we shall derive an optimisation problem (4.11) which obtains not only a lower bound to (3.13) but also with a portfolio x that attains such lower bound.

As an alternative to maximising expected return with a budget constraint, we might want to consider the unconstrained problem of maximising risk-adjusted return,

as in (1.2). The Relative Robust Portfolio Optimisation version of that problem is

$$\min_{x} \max_{\mu \in \mathscr{U}} \max_{y} \left(\left(\mu^\top y - \lambda\, y^\top Q y \right) - \left(\mu^\top x - \lambda\, x^\top Q x \right) \right). \tag{3.14}$$

This problem can be solved in polynomial time unlike (3.13). As we shall see in the next chapter, this is due to the fact that y is left unconstrained and μ is constrained by only one quadratic function. However, if there are no extra constraints on x and/or y it turns out that (3.14) is equivalent to (1.2), i.e., turning the problem into its relative robust version leaves the optimal portfolio unchanged.

Lemma 6. *If $\mathcal{U}$ is a symmetric set around a point $\hat{\mu}$ and both x and y are left unconstrained in* (3.14)*, then the optimal portfolio x is the same to the unconstrained solution of* (1.2) *for $\mu = \hat{\mu}$.*

Proof: First of all notice that (1.2) with no extra constraints has an explicit solution $x^* = \frac{1}{2\lambda} Q^{-1} \hat{\mu}$. This is exactly the solution to the maximisation in y, so (3.14) can be rewritten as

$$\min_{x} \max_{\mu \in \mathscr{U}} \left(\frac{1}{4\lambda}\, \mu^\top Q^{-1}\, \mu - \left(\mu^\top x - \lambda\, x^\top Q x \right) \right).$$

Let $z = \mu - \hat{\mu}$. Since $\mathcal{U}$ is symmetric around $\hat{\mu}$, if $\hat{\mu} + z \in \mathcal{U}$ then $\hat{\mu} - z \in \mathcal{U}$. We can then express the above as

$$\min_{x} \frac{1}{4\lambda}\, \hat{\mu}^\top Q^{-1}\, \hat{\mu} - \hat{\mu}^\top x + \lambda\, x^\top Q x + \max_{z:\, z+\hat{\mu} \in \mathscr{U}} \left(z^\top \left(\frac{1}{2\lambda}\, Q^{-1}\, \hat{\mu} - x \right) + \frac{1}{4\lambda}\, z^\top\, Q^{-1} z \right).$$

The first term is independent of x and so does not matter for the optimisation process. The next two terms are maximised at x^*, as we are minimising the negative objective function of (1.2). The last term is a maximisation over z, composed of a linear and quadratic term. Denote by $z^*(x)$ the optimal z for each x. Since $\mathcal{U} - \hat{\mu}$ is a symmetric set, both $z^*(x)$ and $-z^*(x)$ are valid choices, therefore the linear term must be positive at $z^*(x)$. Thus, given that the quadratic term does not depend on x, this term is minimised as a function of x at $x = \frac{1}{2\lambda}\, Q^{-1}\, \hat{\mu} = x^*$ – and so the optimal portfolio is x^*. □

As we saw in Chapter 2 it is not true that relative robust portfolio optimisation yields the same results as standard portfolio optimisation. If we were to impose constraints to y this would be no longer the case, but as mentioned above it would then be hard to actually solve the problem. In the next chapter we will derive an approximation to (3.13) that can be efficiently solved for a variety of constraints. This approximation yields not only an upper bound but also an admissible portfolio that attains such a bound.

3.5 Numerical Experiments

In this section we shall use proposed method (3.7) and investigate its performance using real data against (3.1). We use daily data from 44 future contracts quoted in USD kindly made available by the Oxford-Man Institute of Quantitative Finance, all of which had daily closing data for most days from 2000. Pinnacle Data (data source) backwards pastes together multiple contracts by using the Ratio Adjusted method, see [56], therefore constructing a continuous time series by preserving returns but not prices. For a complete list of instruments used see Appendix A.

Future contracts are an asset class that has often evaded the attention of portfolio optimisation. This is due to the fact that most quick fixes to the shortfalls of Markowitz do not work in this asset class. Prohibiting short-selling does not make sense, as there is no fundamental difference between a long and a short position – one agrees to buy at a certain date, another to sell. There is no underlying ownership that is transferred in exchange for payment and the number of outstanding contracts fluctuates through the trading day, as contracts get created whenever two parties agree to do so. Moreover imposing limits on the size of the portfolio is not common practice since this contracts are traded on marginals which are settled daily, not on their notional price. This means that traders effectively hold positions worth many times their capital and can adjust position sizes in accordance with market conditions. Hence we shall look for portfolios that maximise expected return and are unconstrained save for a volatility constraint and, for (3.7), a tail-event constraint.

We start by fitting the filter (3.6) on the return series of the above mentioned 44 future contracts between 2010 and 2014. The probability of each day being in state e_1 given our training data can be seen in Figure 3.2.

We can hence see that the filter broadly classifies the 2010-2011 period into the "extreme" category and the 2012-2014 period into the typical "one", albeit not precisely as clear cut as that. It is not immediately clear why such a classification is a reasonable one, but we might wish to look at one of the more intuitive filters discussed previously. Take for example the filter described in Example 2, with $q = 0.7$. It is fairly easy to understand – if the L^1 norm of a given data point is sufficiently high it gets classified as "extreme", otherwise it does not. This type of filter will not give a split as neat as the one seen in Figure 3.2, as it has no temporal effects. However if we look at the plot of cumulative proportion of extreme days already seen for each filter in Figure 3.3, we see that qualitatively they are not that different. We can see that most of the data points classified as "extreme" by Example 2 occur in 2010 and 2011, with a lot

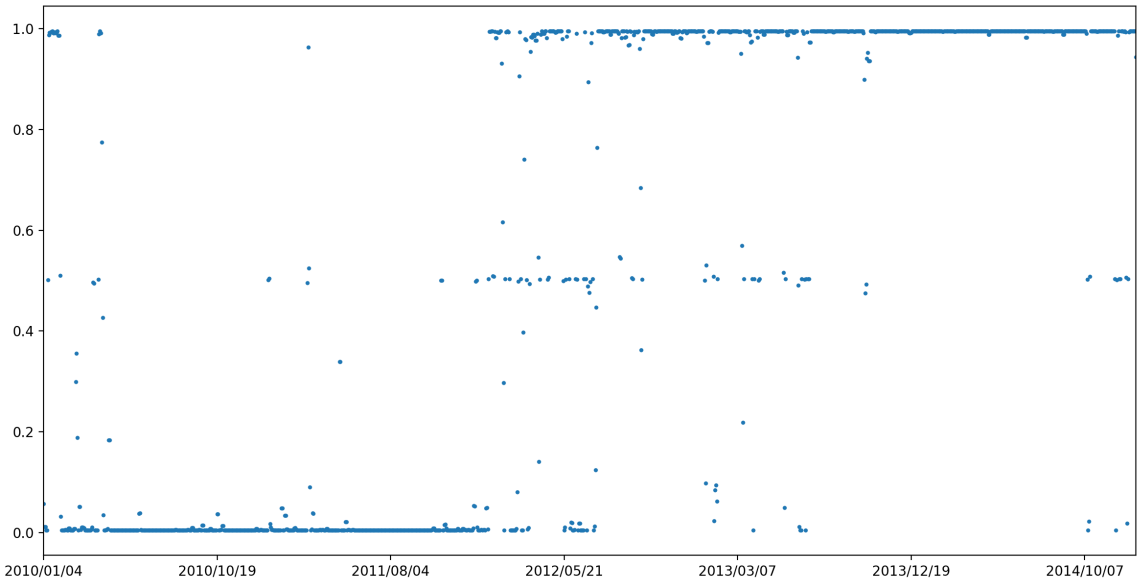


Figure 3.2: Probability of $X_{t-1} = e_1$ given training data.

less being classified in 2013 and 2014. 2012 sits somewhere in the middle, but we can see that happening in Figure 3.2, where there is a lot more uncertainty in classifying data points in 2012 than at any other time.

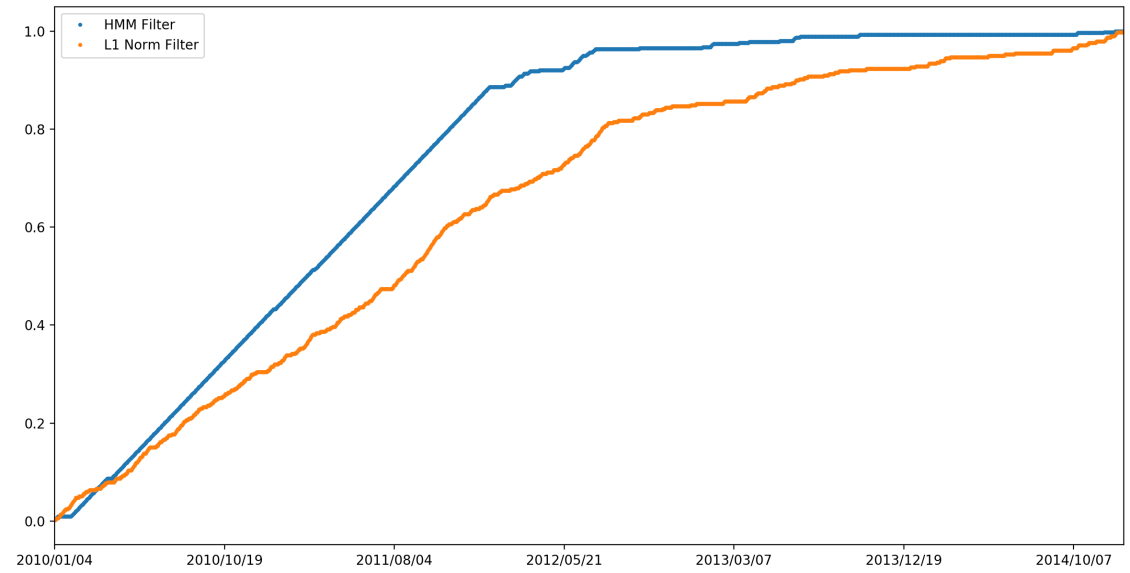


Figure 3.3: Cumulative proportion of days each filter, (3.6) and Example 2 at $q = 0.7$, classifies as "extreme".

Hence our filter (3.6) is a sensible choice and we can move on. Before we solve our optimisation programmes, however, we have to prescribe an expected returns vector

μ. The choice of μ is extremely important – in fact it is the most important decision in the optimal allocation process. If one correctly anticipates where the market is moving, risk-management is of little concern. In fact it is usually detrimental, as it most likely prevented further gains. However if the choice of μ turns out to be wrong, as will inevitably happen to every asset manager at some point, then risk-management prevents a catastrophic outcome. This thesis' focus sits on managing risks and not predicting expected returns, so we shall simply use the momentum signal proposed in [4].

It is worth noticing that while the objective function is the same, the covariance matrix used by the two methods are different – (3.7) uses solely data that comes from the "typical" regime, while Markowitz uses the whole data. In both cases we compute the covariance matrix using the shrinkage method proposed in [39]. All parameters are calculated daily, using a rolling window of the same size as the training data, 5 years. The filter however is not refitted daily – we solely calculate at every new observation which state is it more likely to fall into. We do this so as to preserve the classification of previous observations.

We choose a relatively mild volatility budget at 3% daily volatility. It is worth mentioning that a different volatility budget will not change our results qualitatively – in an unconstrained (3.1) the volatility budget only sets the leverage amount and any different budget value will only scale the same portfolio. Similarly for a different volatility budget (3.7) results in a scaled portfolio for a scaled η.

As mentioned before, the addition of a single quadratic constraint does not drastically change the time it takes to solve an optimisation programme. For a problem with 44 assets, it takes around 0.066 seconds to solve (3.7) and 0.033 seconds to solve (3.1) in a laptop with a 2.8GHz laptop using Mosek in Python 2.7. While it is double the time, it is virtually instantaneous, even on a personal laptop. As we will see in the next chapter, robust optimisation approximations take considerably more time.

We now have to pick η, the mean excess of extreme returns we want to target. In practice this would be set by backtesting against a desirable metric and it would potentially be time varying. Given that we are dealing with futures one would also simply get the relationship between η and σ right and then leverage as necessary. Since there is no right or wrong and each practitioner will have its own viewpoint we shall simply take $\eta = 1\%$ as a start.

We can calculate both optimal portfolios from January 2016 until June 2017, assuming daily recalibration and no transaction costs. The resulting wealth process can be seen in Figure 3.4.

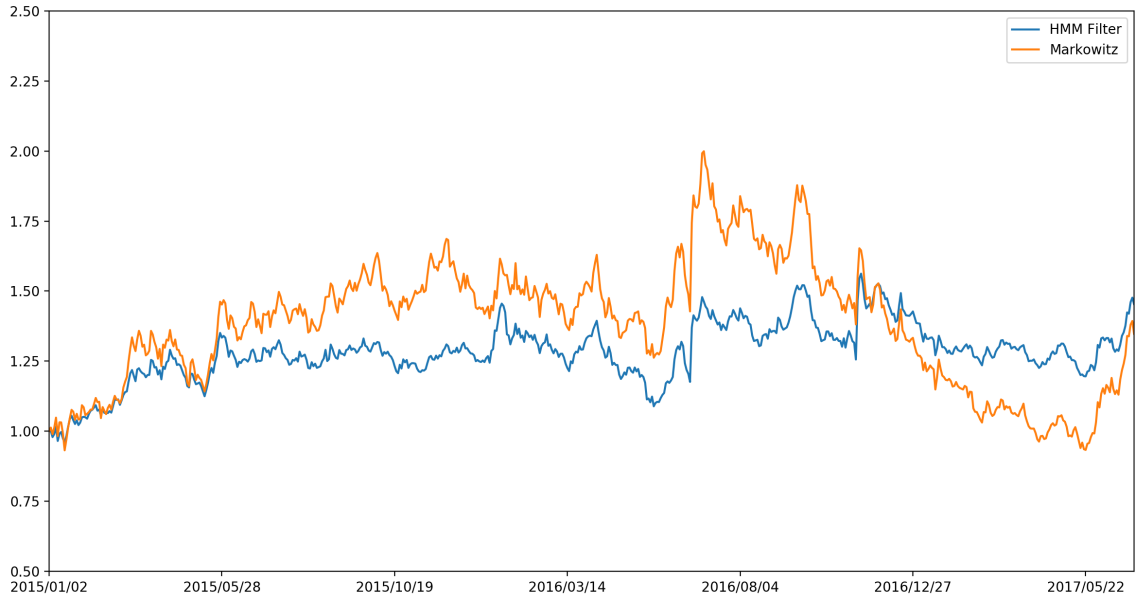


Figure 3.4: Wealth process as a result of investing according to (3.1) with $\sigma = 3\%$ and (3.7) with $\sigma = 3\%$ and $\eta = 1\%$.

In the above mentioned figure we can see that classical Markowitz is a lot more volatile and suffers from more extreme moves than once we add our tail-event constraint. While some of the sharper moves are positive ones, it is important to remember that we are completely independent from the estimation of μ – had our expected return vector been completely opposite we would have held the opposite portfolio and hence suffered a massive loss. μ is in fact, as we have mentioned previous, by far the most important piece of any portfolio optimisation problem. If we could consistently know what tomorrow's return is there would be no need for risk management. Alas, this is not the case, so while μ controls the general direction of investment, it is mostly Q and Λ's job to control the leverage of the portfolio.

In fact we can look at the angle between Markowitz and our portfolio (by taking the arccosine of the ratio between their direct product and the product of their norms) and with the expected return vector μ in Figure 3.5. The angle between both portfolios and the expected return vector is significant, at around one radian. We can also see that by adding our matrix Λ we have have changed Markowitz by roughly a quarter radian. We can however infer that this change is not in the same direction of the one caused by Q, as otherwise any third curve would be the linear combination of the other two, which is clearly not the case. In fact it seems that this correction is almost perpendicular, given how similar the angle between both portfolios and μ are.

Besides running at different leverage, we can look at some different risk measures

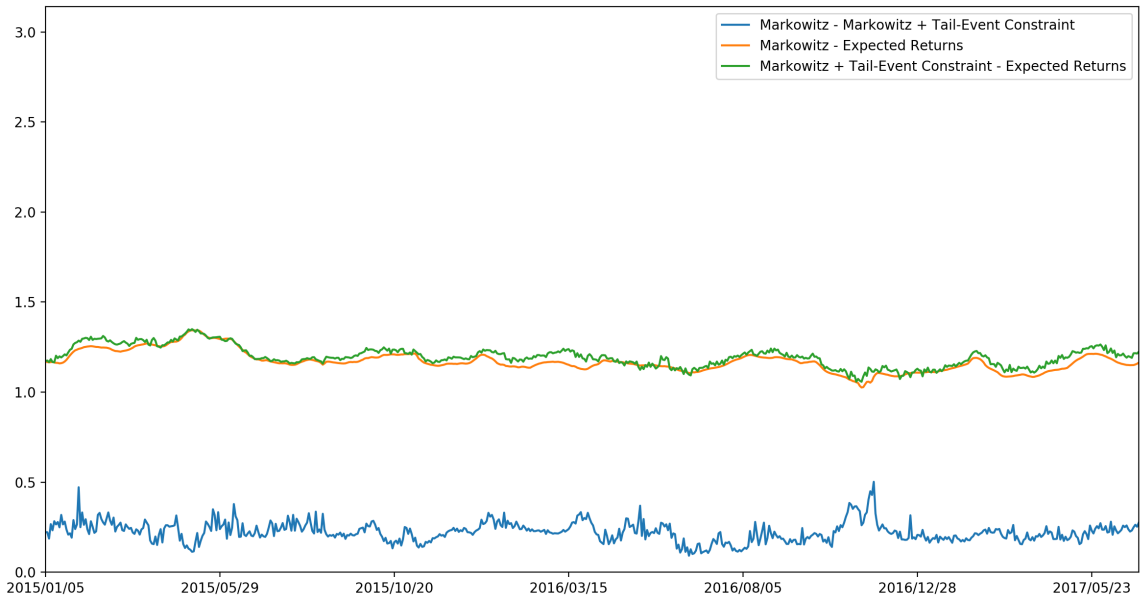


Figure 3.5: Evolution of angle (in radians) of (3.1) with $\sigma = 3\%$, (3.7) with $\sigma = 3\%$ and $\eta = 1\%$ and μ.

to assess the performance of each portfolio. Let us start by the largest single day absolute return. While this might be positive or negative (in this example it is negative) we could have easily have ended up with the opposite μ, so it is important to limit this somewhat. Markowitz has the biggest single day absolute return at 22.1% (!), while ours is not as big but still at 16.8%. These happen in different days and it is worth looking closely into them.

The first one is June 24th 2016, the day after the Brexit referendum. On this day we have a staggering return of 16.8%, while Markowitz actually swings less at 12.4%. This is a good example of the limitations of data only approaches – they cannot prepare against something that has not happened before. It so turns out that we were more vulnerable than Markowitz, perhaps because we were being defensive against tail events we had already seen. These results are positive and hence we actually had bigger gains than Markowitz, however this is in a way completely spurious – since we had no hand in selection μ our only concern is risk management.

The second extreme day is November 9th 2016, the day after the US presidential elections. Again this is a political event for which it is difficult to prepare by solely historical data. Even though we again made considerable profits, this time our return was substantially lower than Markowitz, at 16.4% against 22.1%.

It is not all about different leverages and political events, however. Markowitz managed a Sharpe ratio of 0.5, while we had a realised Sharpe ratio of 0.65. We are

interested in looking at the tail risk metrics, as this was the motivation behind using Λ in the first place. While Markowitz has a 10% quantile of absolute returns at nearly 4%, ours is around 2.5%. In addition the average absolute return over that quantile (i.e. CVaR for absolute returns) is 5.5% for Markowitz and 3.9% for us. The CVaR risk is thus reduced by 30%.

It is interesting however to look at the value of $\sqrt{x^\top (\Lambda + 4\,Q)\,x}$ for the Markowitz portfolio, in order to investigate how much pressure this constraint is exerting. We can see this in Figure 3.6 for both Markowitz and our method.

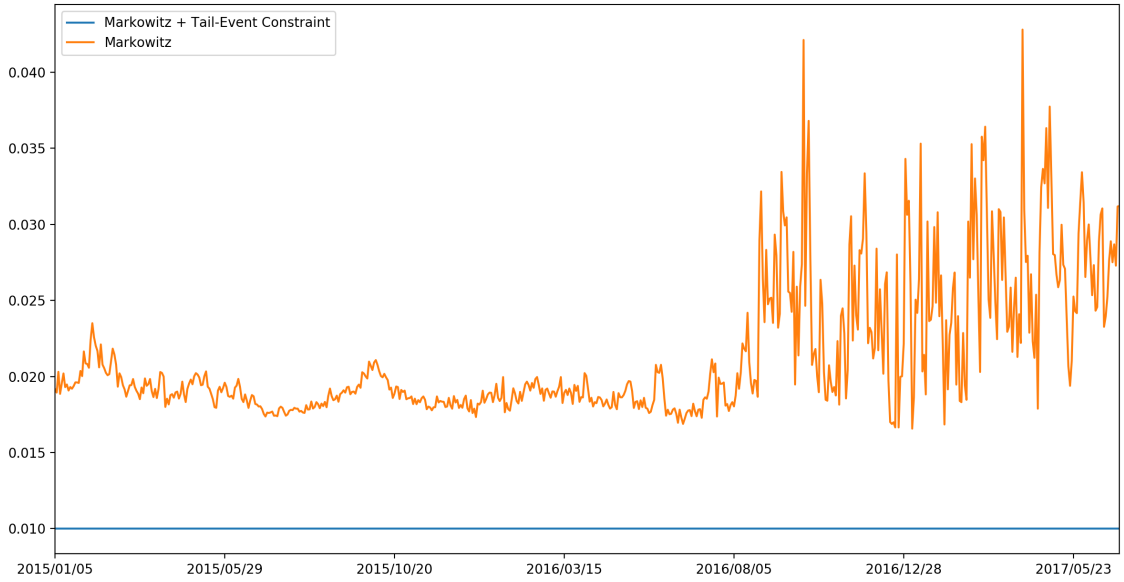


Figure 3.6: Value of $\sqrt{x^\top (\Lambda + 4\,Q)\,x}$ for x the solution to (3.1) with $\sigma = 0.03$, and for x the solution to (3.7) with $\sigma = 3\%$ and $\eta = 0.2$.

As expected our method respects the tail-event bound. In contrast Markowitz completely violates it (as it was not imposed). It is also curious that Markowitz becomes much more erratic around October 2016. We can see in Figure 3.4 that this happens precisely on the same period we observe a sequence of negative returns – had it been constrained by Λ, it could have maintained the considerable lead it had. Given we are dealing with US futures and this starts at the heightening of the US election campaign this shift in performance is likely to have been caused by nervousness in the market over the result of the election, followed by the so-called "Trump trade" (rally in equities, yields and the dollar).

So what would happen if we were to set out η at, say, 2%? In that case we would expect our portfolio to perform roughly like Markowitz, as the constraint will often not be tight and even when it is the excess is small. This is indeed the case

as we can seen in Figure 3.7. We have virtually the same portfolio, with only very occasional deviations in returns. In contrast if we were to drop η to 0.75% we would be constricting too much and we would get a significantly less leveraged portfolio.

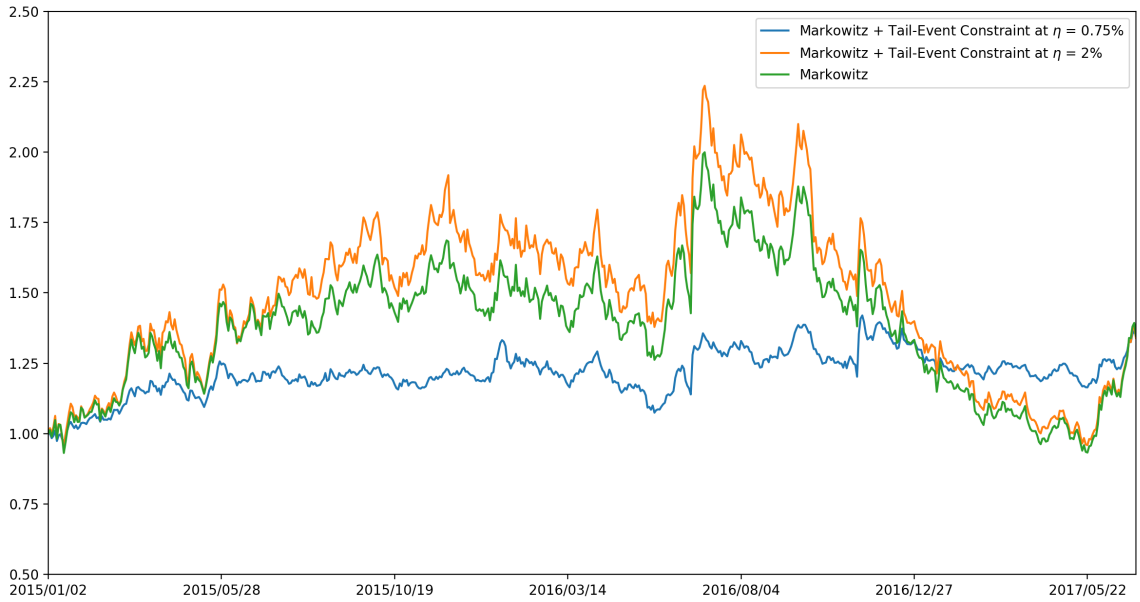


Figure 3.7: Wealth process as a result of investing according to (3.1) with $\sigma = 3\%$ and (3.7) with $\sigma = 3\%$ and $\eta = 2\%$.

There is, naturally, a balance to be struck and as we mentioned previously this balance might even change over time. The goal of our Λ is to limit concentrating our portfolio in dangerous directions and to soften the blow, not to fundamentally reduce the level of risk an investor wants to pursue.

Another question we ask make is how different would the results be had we used a different filter, say Example 1 in the Introduction using the Euclidean basis and taking $q = 0.99$. This means that we classify an observation as extreme if its any of the 44 assets displays an absolute return larger than 99% of its absolute returns in the past 5 years. The wealth process can be seen in Figure 3.8. The results are surprisingly similar in most risk metrics: 10% quantile of absolute returns of 2.9%, average absolute return over that quantile of 4%, maximum absolute return of 11% (fared better in both political events). The big difference is in the Sharpe ratio – 0.2. This is because, similarly to Markowitz, it suffers a sustained downfall from October 2016. This might have happened because we suffer a regime change our filter has captured, but which Example 1 (or Markowitz) is incapable of identifying.

Finally let us conclude by studying the behaviour of our portfolio if instead of Λ we used $\hat{\Lambda}$ as prescribed in (3.10). In this case we would have a low dimensional

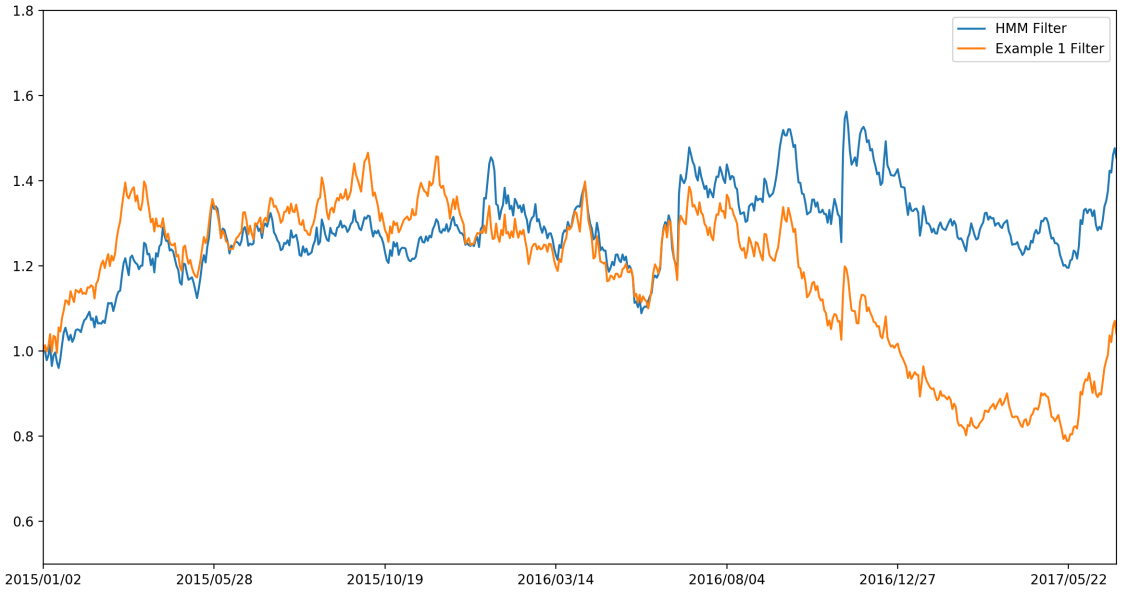


Figure 3.8: Wealth process as a result of investing according to (3.7) using filter (3.6) and Example 2 with Euclidean basis and $q = 0.99$, for $\sigma = 3\%$ and $\eta = 1\%$.

constraint and so we give preference to directions where we hardly seen an extreme event in detriment of those were we regularly see them. We also expect it to be closer to Markowitz, as it will be less constrained. Results for $\delta = 75\%$ are shown in Figure 3.9. As expected, we get something very close to the Markowitz portfolio. The

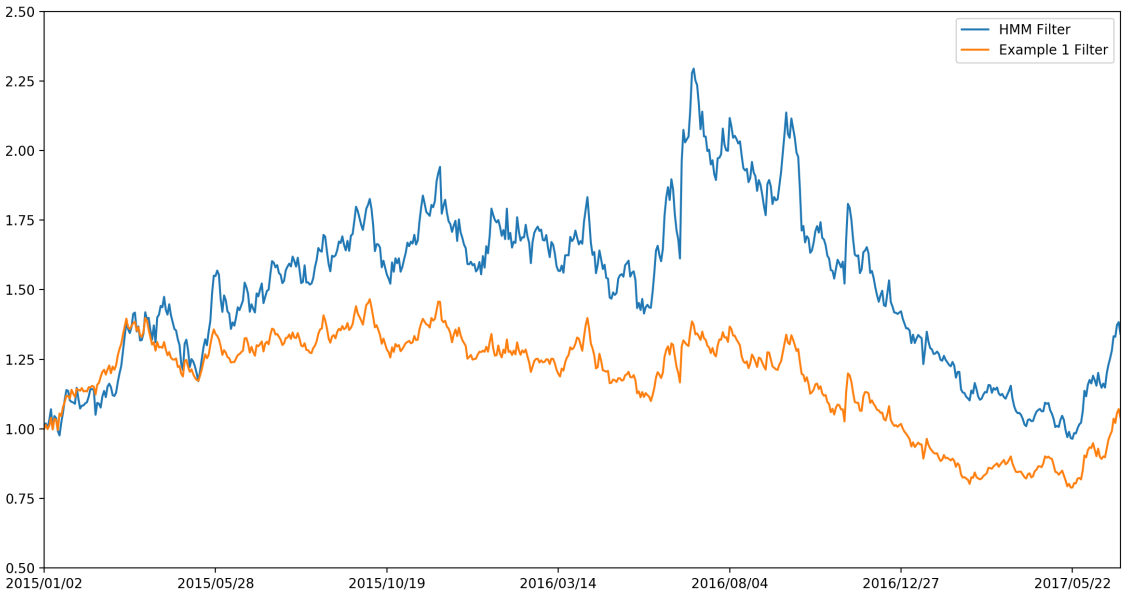


Figure 3.9: Wealth process as a result of investing according to (3.1) with $\sigma = 3\%$ and (3.7) with $\sigma = 3\%$, $\eta = 1\%$ and $\delta = 75\%$.

curious result is that we lose our protection against the October 2016 downfall. This might be explained by the fact that the "Trump trade" only became popular around the time of the election, hence by looking at historical data it was easy to disregard it as not important.

We conclude that using (3.10) is not really relevant in this situation – it is helpful when one really wants a low dimensionality quadratic (such as in absolute robust portfolio optimisation), but when one is not needed there is no reason to throw some directions away and trust nothing eventful will happen in them.

3.6 Conclusion

We have developed a novel approach to incorporate information about the tail of returns distribution into portfolio optimisation in a tractable way. By splitting the data into two we can effectively determine which are the principal components of market corrections during extreme market behaviour. We can then use the Peak Over Threshold method to model the tail of the distribution for each of these directions. Finally, we use convex relaxation to end up with a quadratic form that bounds the expected loss over a certain threshold for each direction, hence effectively a tail-event constraint.

We made an effort to describe this method in as general terms as possible, so as to allow each individual user to adopt whichever tool most suits him. While we made a particular choice of filter and objective function we could have easily made others. Ultimately the success of a portfolio optimisation framework is intimately linked to how good the signal μ is, but our method aims to contain the potential losses of getting this wrong.

Chapter 4

Approximating Robust Portfolio Optimisation Problems with Multiple Constraints

4.1 Introduction

So far we have only considered simple robust portfolio optimisation problems which can be effectively reduced into second-order cone programmes. This is unfortunately not always possible – in practice asset managers will have a large array of constraints, both imposed internally and externally, which might not be possible to effectively solve when converted into an absolute or relative robust portfolio optimisation problem.

For the purpose of clarity let us consider an example before we move on. Suppose we are trading daily in the equity market and want to hold a long-only, fully invested portfolio. Currently our position is x_0, but our expected return for the next day is μ_1 and for the day after is μ_2. We have a volatility budget of σ and a model of linear-quadratic transaction costs, with constants $c_i, d_i \geq 0$ representing respectively the spread and the depth of the limit order book. This problem can be written as

$$\begin{aligned}
\max_{x_1,x_2}\ & \mu_1^\top x_1 + \mu_2^\top x_2 - \sum_{i=1}^{n} \left(c_i|x_{1,i} - x_{0,i}| + d_i(x_{1,i} - x_{0,i})^2 + c_i|x_{2,i} - x_{1,i}| + d_i(x_{2,i} - x_{1,i})^2 \right) \\
\text{s.t. } & x_1^\top Q\, x_1 \leq \sigma^2, \\
& x_2^\top Q\, x_2 \leq \sigma^2, \\
& x_1, x_2 \geq 0.
\end{aligned}$$

Now suppose we are uncertain about our estimates μ_1 and μ_2, and instead we want to protect ourselves by employing relative robust portfolio optimisation with an uncertainty set $\mathcal{U}$ as introduced in the previous chapter. Unfortunately the above

problem does not have a close-form solution, so we cannot take the approach we took in the previous chapter.

To easily formulate the relative robust portfolio optimisation counterpart of the above problem we must first simplify it. Even though we have n assets our decision vector is actually $2n$, since we are trading daily with a two day prediction. Let μ be the $2n$ vector we obtain by gluing μ_1 and μ_2 together, $c = [c_1, \ldots, c_n, c_1, \ldots, c_n]$ the $2n$ vector encoding the linear transaction costs and $D = \text{Diag}(d_1, \ldots, d_n, d_1, \ldots, d_n)$ the diagonal matrix encoding the quadratic transaction costs. If we add auxiliary variable t to represent the absolute value of the $x_1 - x_0$, and $x_2 - x_1$, we can write the above problem as

$$\begin{aligned} \max_{x,t} \ & \mu^\top x - c^\top t - x^\top D\, x \\ \text{s.t. } & (x,t) \in \mathscr{X}, \end{aligned}$$

where our admissible set $\mathscr{X}$ is

$$\mathscr{X} = \Bigg\{ (x,t) : x_1^\top Q\, x_1 \leq \sigma^2, \quad x_2^\top Q\, x_2 \leq \sigma^2, \quad t_1 \geq x_1 - x_0, \\ t_1 \geq x_0 - x_1, \quad t_2 \geq x_2 - x_1, \quad t_2 \geq x_1 - x_2, \quad x \geq 0 \Bigg\}.$$

We can now write the relative robust portfolio optimisation version of the above as

$$\begin{aligned} \min_{x,t,s,\gamma} \ & \gamma \\ \text{s.t. } & \gamma - \mu^\top y + c^\top z + y^\top D\, y + \mu^\top x - c^\top t - s \geq 0, \qquad \forall (\mu, y, z) \in \mathcal{U} \times \mathscr{X}, \\ & x^\top D\, x \leq s, \\ & (x,t) \in \mathscr{X}. \end{aligned}$$

The first constraint is a type of co-positive constraint. If we denote by $q_{x,t,s,\gamma}$ the quadratic

$$q_{x,t,s,\gamma}(\mu, y, z) = \gamma - \mu^\top y + c^\top z + y^\top D\, y + \mu^\top x - c^\top t - s,$$

then our constraint imposes that $q_{x,t,s,\gamma}$ must be non-negative in a subspace of its domain, in this case in $\mathcal{U} \times \mathscr{X}$. Checking whether a quadratic is non-negative in a subspace of its domain is in general a NP-hard problem, see [47], but can also be efficiently solved in some particular cases, as shown in [59] and [57]. In other cases this is not possible, but inner approximations can be obtained, as in [33].

In this Chapter we will first do an overview of the known theory before expanding on the work done in [33]. We then go through how this allows us to get attainable bounds for a wide array of absolute and relative robust portfolio optimisation problems, in particular those with multi-period and transaction costs incorporated. We conclude with a numerical example of such a model.

4.2 Copositive Sets

We denote by $\mathscr{S}^n$ the space of symmetric $n \times n$ matrices. If $A, B \in \mathscr{S}^n$, we denote by $A \bullet B$ its inner product, $A \bullet B = \text{trace}(A\,B)$. If a matrix $A \in \mathscr{S}^n$ is positive definite we write $A \succ 0$, while $A \succeq 0$ means it is positive semidefinite. Moreover we denote by $A^{1/2}$ the symmetric and positive semidefinite root of $A \succeq 0$. The subspace of positive semidefinite $n \times n$ matrices is denoted by $\mathscr{S}^n_+$.

We say a function f is nonnegative over D if $f(x) \geq 0$ for every $x \in D$. Similarly, we say f is copositive with a function g if $f(x) \geq 0$ for every x such that $g(x) \geq 0$, i.e., if f is nonnegative over $\{x : g(x) \geq 0\}$. If q is a quadratic function, i.e., $q(x) = x^\top A\, x + 2\, \mathbf{b}^\top x + c$, then we define $\mathscr{M}(q)$ as its matrix representation,

$$\mathscr{M}(q) = \begin{bmatrix} A & \mathbf{b} \\ \mathbf{b}^\top & c \end{bmatrix}, \qquad \text{since } q(x) = \begin{bmatrix} x \\ 1 \end{bmatrix}^\top \mathscr{M}(q) \begin{bmatrix} x \\ 1 \end{bmatrix}.$$

If D is a set, then $\text{cl}(D)$ is its closure, $\text{conv}(D)$ is its convex hull, while $\text{cone}(D)$ is its conic hull. We denote by $\mathscr{H}(D)$ the homogenization of set D, i.e.,

$$\mathscr{H}(D) := \text{cl}\left\{ \begin{bmatrix} x \\ t \end{bmatrix} \in \mathbb{R}^n \times \mathbb{R}_+ : x/t \in D \right\}.$$

Notice that if D is closed then $\mathscr{H}(D)$ is a closed cone and is convex if and only if D is convex itself. We shall say that a set D is *tractable* if for any x we can decide if $x \in D$ in polynomial time and say it is intractable otherwise.

We will be particularly interested in the set of quadratic functions nonnegative over a set D, the copositive set of D, which shall denote by $\mathscr{FC}_+(D)$,

$$\mathscr{FC}_+(D) := \{\mathscr{M}(q) : q(x) \geq 0,\ \forall x\ \in D\}.$$

One can immediately see that $\mathscr{FC}_+(D)$ is a closed convex cone.

If C is a cone, then C^* represents its dual cone, i.e.,

$$C^* := \{Y : \langle Y, X \rangle \geq 0,\ \forall X \in C\},$$

where $\langle Y, X \rangle$ is the appropriate inner product depending on C being a set of vectors or matrices (note that the notion of dual changes when the inner product is changed). If C is a closed convex cone, then $C^{**} = C$.

We will denote by L_{n+1} the Lorentz cone

$$L_{n+1} = \left\{ \begin{bmatrix} x \\ t \end{bmatrix} \in \mathbb{R}^{n+1} : \|x\|_2 \leq t \right\}.$$

4.2.1 S-Lemma and Extensions

The S-Lemma, also known as Yakubovich's S-Procedure [59], is a basic result of control theory that characterizes the set $\mathscr{FC}_+(D)$ for $D = \{x : p(x) \geq 0\}$, p a quadratic function. Multiple proofs of this result are known, such as [5, 31, 57, 59, 60]. There is also an excellent survey by Pólik and Terlaky [48].

Lemma 7 (S-Lemma)**.** *Let $p, q : \mathbb{R}^n \to \mathbb{R}$ be quadratic functions and suppose there exists $\bar{x}$ such that $q(\bar{x}) > 0$. Then p is copositive with q if and only if there exists $\lambda \geq 0$ such that*

$$p(x) - \lambda\, q(x) \geq 0, \qquad \forall x \in \mathbb{R}^n.$$

Notice that the S-Lemma immediately implies the following set equality

$$\mathscr{FC}_+(\{x : q(x) \geq 0\}) := \mathscr{S}_+^{n+1} + \operatorname{cone}\{\mathscr{M}(q)\}.$$

The S-Lemma is unfortunately not immediately generalised to multiple quadratic functions, however for two quadratic functions some results can be achieved. The following result were proven in [5] and [49] and can be also found in [48].

Lemma 8. *Let $A, B, C \in \mathbb{R}^{n\times n}$, $n \geq 3$ be matrices such that either*

1. *A, B and C commute; or*

2. *there exists a linear combination of A, B and C that is positive definite.*

Furthermore let a, b and c be constants and f, g and $h : \mathbb{R}^n \to \mathbb{R}$ be the following quadratic functions

$$\begin{aligned} f(x) &= x^\top A\, x + a, \\ g(x) &= x^\top B\, x + b, \\ h(x) &= x^\top C\, x + c, \end{aligned}$$

and assume there exists $\bar{x}$ *such that* $f(\bar{x}), g(\bar{x}) > 0$.

Then h *is positive over* $\{x : f(x) \geq 0,\ g(x) \geq 0\}$ *if and only if* $\exists\, \lambda_1, \lambda_2 \geq 0$ *such that*

$$h(x) - \lambda_1\, f(x) - \lambda_2\, g(x) \geq 0, \qquad \forall x \in \mathbb{R}^n.$$

While this is a positive result, in general we do not have equality

$$\mathscr{FC}_+ \left(\{x : f(x), g(x) \geq 0\}\right) \neq \mathscr{S}_+^{n+1} + \text{cone}(\mathscr{M}\,(f)) + \text{cone}(\mathscr{M}\,(g)),$$

as the following example demonstrates.

Example 9.

$$\begin{aligned} f_1(x) &= -x_1^2 - x_2^2 + 1, \\ f_2(x) &= -x_3^2 - x_4^2 + 1, \\ g(x) &= x_1\, x_3 + x_1 + x_3 + 1 = (x_1 + 1)(x_3 + 1). \end{aligned}$$

The matrices that define the quadratic terms have a positive definite linear combination, and $f_1(0) = f_2(0) = 1 > 0$. Moreover, it is easy to see that

$$f_1(x) \geq 0, f_2(x) \geq 0 \Rightarrow g(x) \geq 0.$$

However, there is no $\lambda_1, \lambda_2 \geq 0$ such that

$$g(x) - \lambda_1\, f_1(x) - \lambda_2\, f_2(x) \geq 0, \qquad \forall x \in \mathbb{R}^n.$$

To prove this let

$$\mathscr{M}\,(f_1) = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \qquad \mathscr{M}\,(f_2) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\mathscr{M}\,(g) = \begin{bmatrix} 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 1 \end{bmatrix}, \qquad X = \begin{bmatrix} 4 & 0 & -2 & 0 & -2 \\ 0 & 0 & 0 & 0 & 0 \\ -2 & 0 & 4 & 0 & -2 \\ 0 & 0 & 0 & 0 & 0 \\ -2 & 0 & -2 & 0 & 5 \end{bmatrix}.$$

If such constants existed, this would mean that

$$\mathscr{M}(g) \in \mathscr{S}^5_+ + \text{cone}(\mathscr{M}(f_1)) + \text{cone}(\mathscr{M}(f_2)).$$

This is however not true. On one hand, one can attest that $X \in \mathscr{S}^5_+$ and $X \bullet \mathscr{M}(f_1) = X \bullet \mathscr{M}(f_2) = 1 \geq 0$. Hence

$$X \in (\mathscr{S}^5_+ + \text{cone}(\mathscr{M}(f_1)) + \text{cone}(\mathscr{M}(f_2)))^*.$$

On the other hand, $X \bullet \mathscr{M}(g) = -1 \ngeq 0$, which proves that $\mathscr{M}(g) \notin \mathscr{S}^5_+ + \text{cone}(\mathscr{M}(f_1)) + \text{cone}(\mathscr{M}(f_2))$. Hence no such constants exist.

It is however straightforward from Lemma 8 to prove the following.

Lemma 10. *Let $p, q : \mathbb{R}^n \to \mathbb{R}$ be two quadratic functions, $n \geq 2$. If $\mathscr{M}(p)$ and $\mathscr{M}(q)$ have a positive definite linear combination and if there exists $\bar{x}$ such that $p(\bar{x}), q(\bar{x}) > 0$, then*

$$\mathscr{FC}_+(\{x : p(x), q(x) \geq 0\}) = \mathscr{S}^{n+1}_+ + \mathit{cone}(\mathscr{M}(p)) + \mathit{cone}(\mathscr{M}(q)).$$

Proof. Let $Z \in \mathscr{FC}_+(\{x : p(x), q(x) \geq 0\})$ and $f, g, h : \mathbb{R}^{n+1} \to \mathbb{R}$ the following functions

$$\begin{aligned} f(x) &= x^\top \mathscr{M}(p)\ x, \\ g(x) &= x^\top \mathscr{M}(q)\ x, \\ h(x) &= x^\top Z\ x. \end{aligned}$$

Note that $f((\bar{x}, 1)) = p(\bar{x}) > 0$, $g((\bar{x}, 1)) = q(\bar{x}) > 0$. One can prove that $Z \in \mathscr{FC}_+(\{x : p(x), q(x) \geq 0\})$ is equivalent to $\mathscr{M}(h) \in \mathscr{FC}_+(\{x : f(x), g(x) \geq 0\})$. By Lemma 8, there exists $\lambda_1, \lambda_2 \geq 0$ such that

$$\begin{aligned} h(x) - \lambda_1\, f(x) - \lambda_2\, g(x) &\geq 0, \qquad \forall x \in \mathbb{R}^n, \\ x^\top (Z - \lambda_1\, \mathscr{M}(p) - \lambda_2\, \mathscr{M}(q))\, x &\geq 0, \qquad \forall x \in \mathbb{R}^n, \end{aligned}$$

or, equivalently,

$$Z \in \mathscr{S}^{n+1}_+ + \text{cone}(\mathscr{M}(p)) + \text{cone}(\mathscr{M}(q)).$$

Since the other inclusion is trivial, this concludes the proof.

□

Even though the S-Lemma cannot be extended to two quadratics, it can be extended to one quadratic and one linear function. The following result was proven in [57].

Theorem 11. *Let $q : \mathbb{R}^n \to \mathbb{R}$ be a concave quadratic function, $a \in \mathbb{R}^{n+1}$. Suppose there exists $\bar{x}$ such that $q(\bar{x}) > 0$, $[\bar{x}^\top \; 1]a > 0$. Let*

$$D = \{x : q(x) \geq 0\}, \qquad L = \{x : [x^\top \; 1]a \geq 0\}.$$

Then,

$$\mathscr{FC}_+ (D \cap L) = \mathscr{S}_+^{n+1} + cone(\mathscr{M}(q)) + \{a\,\psi^\top + \psi\,a^\top : \psi \in \mathscr{H}(D)^*\}.$$

4.2.2 Inner Aproximations

Unfortunately $\mathscr{FC}_+ (D)$ is intractable in more general cases – in fact it was proven in [47] that the problem of checking whether a quadratic is in $\mathscr{FC}_+ (D)$ for a simple domain D is NP-hard. It is however possible to obtain tractable inner approximations, which we shall pursue. The inner approximation of $\mathscr{FC}_+ (D)$ for D the intersection of an ellipsoid with multiple half-spaces was derived in [32]. In the rest of this section we shall base ourselves on the following result proven there to extend their inner approximation to the case of multiple ellipsoids. This will be essential when considering multiperiod portfolio optimisation.

Theorem 12.

$$\mathscr{FC}_+ (D) \supseteq \mathscr{S}_+^{n+1} + cl(cone\{v\,w^\top + w\,v^\top : v, w \in \mathscr{H}(D)^*\}).$$

We will in general drop the closure in the above theorem – the result is obviously still valid without it and taking an element in the closure effectively means taking the limit of an infinite sequence, which does not help with our goal of a tractable approximation.

This is a very general theorem, but we are mainly interested in the case when D is an intersection of multiple ellipsoids and half-spaces. By looking at Theorem 11, we can get a tighter inner approximation than the one provided in Theorem 12 for this particular case.

Lemma 13. *Let $q_1, \ldots, q_{k_1}$ be concave quadratic functions and $a_1, \ldots, a_{k_2} \in \mathbb{R}^{n+1}$ vectors. Furthermore let*

$$D = \{x \in \mathbb{R}^n : q_i(x) \geq 0,\ \forall i = 1, \ldots, k_1,\ [x^\top\ 1]\, a_j \geq 0,\ \forall j = 1, \ldots, k_2\}.$$

Then the following inner approximation holds

$$\mathscr{FC}_+(D) \supseteq \mathscr{S}_+^{n+1} + \sum_{i=1}^{k_1} cone(\mathscr{M}(q_i)) + cone\left\{v\, w^\top + w\, v^\top : v, w \in \mathscr{H}(D)^*\right\}.$$

Proof. All we have to do is prove that $\mathscr{M}(q_i) \in \mathscr{FC}_+(D)$, since everything else follows from Theorem 12 and the fact that $\mathscr{FC}_+(D)$ is a closed convex cone. But this is obvious, since

$$D \subseteq \{x : q_i(x) \geq 0\},$$

and so q_i is in fact copositive over D. □

Not all the terms prescribed in Lemma 13 are actually needed. Let us consider the case of an ellipsoid D. Any ellipsoid can be written as $D = \{x : q(x) \geq 0\}$ for q be a concave quadratic function of the form $q(x) = -(x - v)^\top A\,(x - v) + 1$, with A positive definite describing its axes and v its centre. The we know from the S-Lemma that

$$\mathscr{FC}_+(D) = \mathscr{S}_+^{n+1} + \text{cone}\left\{\mathscr{M}(q)\right\}.$$

On the other hand, Lemma 13 tells us that

$$\mathscr{FC}_+(D) \supseteq \mathscr{S}_+^{n+1} + \text{cone}\left\{\mathscr{M}(q)\right\} + \text{cone}\{v\, w^\top + w\, v^\top : v, w \in \mathscr{H}(D)^*\}.$$

This of course implies that

$$\text{cone}\{v\, w^\top + w\, v^\top : v, w \in \mathscr{H}(D)^*\} \subseteq \mathscr{S}_+^{n+1} + \text{cone}\left\{\mathscr{M}(q)\right\},$$

and so this term can be dropped. However this it would be good to know what $\mathscr{H}(D)^*$ looks like – [57] derived this from first principles for ellipsoids centred at $v = 0$, but the same line of though can be followed for a general centre v. We will

explicitly derive this as it will be necessary later. Let $A = U\,U^\top$ be a Cholesky decomposition of A. We start by describing the set $\mathscr{H}(D)$,

$$\begin{aligned}
\mathscr{H}(D) &= \left\{ \begin{bmatrix} x \\ \tau \end{bmatrix} : -(x - \tau\, v)^\top A\,(x - \tau\, v) + \tau^2 \geq 0,\ \tau \geq 0 \right\} \\
&= \left\{ \begin{bmatrix} x \\ \tau \end{bmatrix} : (x - \tau\, v)^\top\, U\,U^\top\,(x - \tau\, v) \leq \tau^2,\ \tau \geq 0 \right\} \\
&= \left\{ \begin{bmatrix} x \\ \tau \end{bmatrix} : \left(U^\top\,(x - \tau\, v)\right)^\top\ \left(U^\top\,(x - \tau\, v)\right) \leq \tau^2,\ \tau \geq 0 \right\} \\
&= \left\{ \begin{bmatrix} x \\ \tau \end{bmatrix} : \begin{bmatrix} U^\top\,(x - \tau\, v) \\ \tau \end{bmatrix} \in L_{n+1} \right\} \\
&= \left\{ \begin{bmatrix} (U^\top)^{-1}\,x + \tau\, v \\ \tau \end{bmatrix} : \begin{bmatrix} x \\ \tau \end{bmatrix} \in L_{n+1} \right\}.
\end{aligned}$$

Notice that the constraint $\tau \geq 0$ is already included in the definition of L_{n+1}. By using the fact that L_{n+1} is a self dual cone, we can now describe $\mathscr{H}(D)^*$ as

$$\begin{aligned}
\mathscr{H}(D)^* &= \left\{ \begin{bmatrix} y \\ \varsigma \end{bmatrix} : \begin{bmatrix} y \\ \varsigma \end{bmatrix}^\top\ \begin{bmatrix} x \\ \tau \end{bmatrix} \geq 0,\ \forall \begin{bmatrix} x \\ \tau \end{bmatrix} \in \mathscr{H}(D) \right\} \\
&= \left\{ \begin{bmatrix} y \\ \varsigma \end{bmatrix} : \begin{bmatrix} y \\ \varsigma \end{bmatrix}^\top\ \begin{bmatrix} (U^\top)^{-1}\,x + \tau\, v \\ \tau \end{bmatrix} \geq 0,\ \forall \begin{bmatrix} x \\ \tau \end{bmatrix} \in L_{n+1} \right\} \\
&= \left\{ \begin{bmatrix} y \\ \varsigma \end{bmatrix} : y^\top\,(U^{-1})^\top\,x + \tau\, y^\top v + \varsigma\,\tau \geq 0,\ \forall \begin{bmatrix} x \\ \tau \end{bmatrix} \in L_{n+1} \right\} \\
&= \left\{ \begin{bmatrix} y \\ \varsigma \end{bmatrix} : \left(\varsigma + v^\top y\right)\tau + \left(U^{-1}\,y\right)^\top\,x \geq 0,\ \forall \begin{bmatrix} x \\ \tau \end{bmatrix} \in L_{n+1} \right\} \\
&= \left\{ \begin{bmatrix} y \\ \varsigma \end{bmatrix} : \begin{bmatrix} U^{-1}\,y \\ \varsigma + v^\top y \end{bmatrix} \in L_{n+1} \right\} \\
&= \left\{ \begin{bmatrix} U\,y \\ \varsigma - v^\top y \end{bmatrix} : \begin{bmatrix} y \\ \varsigma \end{bmatrix} \in L_{n+1} \right\}.
\end{aligned}$$

A similar result can be derived for the case when A is merely positive semidefinite by carefully following the same steps.

Now that we have all the necessary tools, we can finally derive our inner approximation for the cone of quadratics functions non-negative over the intersection of multiple ellipsoids and half-spaces. Let q_1, q_2, … q_{k_1} be concave quadratic functions of the form $q_i(x) = -(x - v_i)^\top A_i\,(x - v_i) + 1$, with A_i positive definite for some centre vector v_i. Moreover let d_1, d_2, …, d_{k_2} be vectors in $\mathbb{R}^n$ and c_1, c_2, …, c_{k_2} constants.

Define the following sets

$$\begin{aligned} D &= \{x : q_i(x) \geq 0,\ \forall i, \quad d_j^\top\, x + c_j \geq 0,\ \forall j\}, \\ D_i^1 &= \{x : q_i(x) \geq 0\}, \\ D_j^2 &= \{x : d_j^\top\, x + c_j \geq 0\}, \end{aligned}$$

so that D is the intersection of all D_i^1, D_j^2. We assume that D has a non-empty interior, i.e., that there exists $\bar{x}$ such that $q_i(\bar{x}) > 0$ for all i and $d_j^\top \bar{x} + c_j > 0$ for all j. Lemma 13 tells us that

$$\mathscr{FC}_+(D) \supseteq \mathscr{S}_+^{n+1} + \sum_{i=1}^{k_1} \operatorname{cone}\{\mathscr{M}(q_i)\} + \operatorname{cone}\{v\, w^\top + w\, v^\top : v, w \in \mathscr{H}(D)^*\}. \quad (4.1)$$

To make this a tractable inner approximation we have to figure out what $\mathscr{H}(D)^*$ is. Let $A_i = U_i\, U_i^\top$ be the Cholesky decomposition of A_i. Following the same argument as in [57] we know that

$$\begin{aligned} \mathscr{H}(D) &= \left\{ \begin{bmatrix} x \\ \tau \end{bmatrix} : -(x - \tau\, v_i)^\top A_i\, (x - \tau\, v_i) + \tau^2 \geq 0,\ \forall i,\ d_j^\top\, x + c_j\, \tau \geq 0,\ \forall j,\ \tau \geq 0 \right\} \\ &= \mathscr{H}\left(D_1^1\right) \cap \cdots \cap \mathscr{H}\left(D_{k_1}^1\right) \cap \mathscr{H}\left(D_1^2\right) \cap \cdots \cap \mathscr{H}\left(D_{k_2}^2\right), \end{aligned}$$

therefore by taking duals we get

$$\mathscr{H}(D)^* = \mathscr{H}\left(D_1^1\right)^* + \cdots + \mathscr{H}\left(D_{k_1}^1\right)^* + \mathscr{H}\left(D_1^2\right)^* + \cdots + \mathscr{H}\left(D_{k_2}^2\right)^*.$$

This is very helpful, because $\mathscr{H}(D_i^1)$ and $\mathscr{H}\left(D_j^1\right)$ are simple enough that we know what their duals are,

$$\begin{aligned} \mathscr{H}\left(D_i^1\right)^* &= \left\{ \begin{bmatrix} U_i\, y \\ \varsigma - v_i^\top y \end{bmatrix} : \begin{bmatrix} y \\ \varsigma \end{bmatrix} \in L_{n+1} \right\}, \\ \mathscr{H}\left(D_j^2\right)^* &= \operatorname{cone}\left\{ \begin{bmatrix} d_j \\ c_j \end{bmatrix} \right\}. \end{aligned}$$

We can now perfectly describe $\mathscr{H}(D)^*$, so we have an inner approximation (4.1). However, as we saw previously, there are a few unnecessary terms in there. In particular we know that

$$\operatorname{cone}\{v\, w^\top + w\, v^\top : v, w \in \mathscr{H}\left(D_i^1\right)^*\} \subseteq \mathscr{S}_+^{n+1} + \operatorname{cone}\{\mathscr{M}(q_i)\},$$

and it is clear that

$$\operatorname{cone}\{v\, w^\top + w\, v^\top : v, w \in \mathscr{H}\left(D_j^2\right)^*\} \subseteq \mathscr{S}_+^{n+1},$$

so these terms can be left out of the inner approximation, simplifying to

$$\begin{aligned}\mathscr{FC}_+(D) \supseteq\ & \mathscr{S}_+^{n+1} + \sum_{i=1}^{k_1} \operatorname{cone}\left\{\mathscr{M}(q_i)\right\} \\ & + \sum_{i=1}^{k_2-1}\sum_{j=i+1}^{k_2} \operatorname{cone}\left\{ \begin{bmatrix} d_i \\ c_i \end{bmatrix} \begin{bmatrix} d_j \\ c_j \end{bmatrix}^\top + \begin{bmatrix} d_j \\ c_j \end{bmatrix} \begin{bmatrix} d_i \\ c_i \end{bmatrix}^\top \right\} \\ & + \sum_{i=1}^{k_1}\sum_{j=1}^{k_2} \operatorname{cone}\left\{ \begin{bmatrix} U_i\, y \\ \varsigma - v_i^\top y \end{bmatrix} \begin{bmatrix} d_j \\ c_j \end{bmatrix}^\top + \begin{bmatrix} d_j \\ c_j \end{bmatrix} \begin{bmatrix} U_i\, y \\ \varsigma - v_i^\top y \end{bmatrix}^\top : \begin{bmatrix} y \\ \varsigma \end{bmatrix} \in L_{n+1} \right\} \\ & + \sum_{i=1}^{k_1-1}\sum_{j=i+1}^{k_1} \operatorname{cone}\left\{ v\, w^\top + w\, v^\top : v \in \mathscr{H}\left(D_i^1\right)^*,\ w \in \mathscr{H}\left(D_j^1\right)^* \right\}. \end{aligned} \tag{4.2}$$

The last cones are, to the best of our knowledge, not tractable, so we can either eliminate them all together or construct their own inner approximation. We can construct a sequence of inner approximations by simply sampling from L_{n+1}. Let $(\tau^l, z^l) \in L_{n+1}$ be random samples for $l = 1, \ldots, N$. Then the cones

$$\operatorname{cone}\left\{ v\, w^\top + w\, v^\top : v \in \mathscr{H}\left(D_i^1\right)^*,\ w \in \mathscr{H}\left(D_j^1\right)^* \right\}$$

have the following inner approximation

$$\sum_{l=1}^{N} \operatorname{cone}\left\{ \begin{bmatrix} U_i\, y \\ \varsigma - v_i^\top y \end{bmatrix} \begin{bmatrix} U_j\, z^l \\ \tau^l - v_j^\top z^l \end{bmatrix}^\top + \begin{bmatrix} U_j\, z^l \\ \tau^l - v_j^\top z^l \end{bmatrix} \begin{bmatrix} U_i\, y \\ \varsigma - v_i^\top y \end{bmatrix}^\top : \begin{bmatrix} y \\ \varsigma \end{bmatrix} \in L_{n+1} \right\}.$$

It is worth pointing out that, unlike some other terms, these cones are actually not redundant, even when the ellipsoids are co-centric. We can see this by going back to Example 9. If we let

$$\begin{aligned} D_1 &= \{x \in \mathbb{R}^4 : f_1(x) \geq 0\}, \\ D_2 &= \{x \in \mathbb{R}^4 : f_2(x) \geq (0)\}, \\ w &= \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \end{bmatrix}^\top, \\ v &= \begin{bmatrix} 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}^\top. \end{aligned}$$

It is easy to see that $w \in \mathscr{H}(D_1)^*$, $v \in \mathscr{H}(D_2)^*$ and $\mathscr{M}(g) = w\, v^\top + v\, w^\top$. Hence one concludes that

$$\{w\, v^\top + v\, w^\top : w \in \mathscr{H}(D_1)^*, v \in \mathscr{H}(D_2)^*\} \nsubseteq \mathscr{S}_+^5 + \operatorname{cone}(\mathscr{M}(f_1)) + \operatorname{cone}(\mathscr{M}(f_2)).$$

Nevertheless this term can be a big drag the speed of an optimisation algorithm, as it can greatly increase the number of variables of the problem. For that reason we

choose to remove it from here on and use instead a slightly worse version of (4.2),

$$\begin{aligned}\mathscr{FC}_+(D) \supseteq\ & \mathscr{S}_+^{n+1} + \sum_{i=1}^{n} \operatorname{cone}\{\mathscr{M}(q_i)\} \\ & + \sum_{i=1}^{k_2-1}\sum_{j=i+1}^{k_2} \operatorname{cone}\left\{ \begin{bmatrix} d_i \\ c_i \end{bmatrix} \begin{bmatrix} d_j \\ c_j \end{bmatrix}^\top + \begin{bmatrix} d_j \\ c_j \end{bmatrix} \begin{bmatrix} d_i \\ c_i \end{bmatrix}^\top \right\} \\ & + \sum_{i=1}^{k_1}\sum_{j=1}^{k_2} \operatorname{cone}\left\{ \begin{bmatrix} U_i\, y \\ \varsigma - v_i^\top y \end{bmatrix} \begin{bmatrix} d_j \\ c_j \end{bmatrix}^\top + \begin{bmatrix} d_j \\ c_j \end{bmatrix} \begin{bmatrix} U_i\, y \\ \varsigma - v_i^\top y \end{bmatrix}^\top : \begin{bmatrix} y \\ \varsigma \end{bmatrix} \in L_{n+1} \right\}.\end{aligned} \tag{4.3}$$

4.3 Bound on Robust Portfolio Optimisation Problems

We are now in a position to be able to solve (or, better said, approximate) problems such as the ones described in the introduction to this chapter. In this section we will first describe the general set of portfolio optimisation problems for which we can not only derive a bound but also an optimal portfolio that guarantees such bound, and then we run a numerical example as a follow-up to the previous chapter.

Let $f(x;\theta)$ be our objective function which we wish to maximise – x is the decision variables over which we have control and θ the parameters we wish to make uncertain. We shall make the following assumption throughout,

- f is a quadratic function of both x and θ, so that there exists matrices M_x, M_θ and $M_{x,\theta}$ and vectors b_x, b_θ of appropriate dimensions such that

  $$f(x;\theta) = -x^\top M_x\, x + \theta^\top M_\theta\, \theta + x^\top M_{x,\theta}\, \theta + b_x^\top x + b_\theta^\top \theta,$$

  where for convenience we have made the decision that $f(0;0) = 0$. Moreover we assume M_x is a positive semidefinite matrix.

- The uncertainty set $\mathcal{U}$ of parameters $\theta \in \mathbb{R}^{n_1}$ we want to consider is an intersection of a finite number of half-spaces and a finite number of ellipsoids. Hence there exists concave quadratic function of the form $q(\theta) = -(\theta - v_i)^\top A_i\,(\theta - v_i) + 1$, for positive definite matrices $A_i = U_i\, U_i^\top$ and vectors v_i for $i = 1, \ldots, k_1$ and vectors d_j and c_j constants for $j = 1, \ldots, k_2$ such that

  $$\mathcal{U} = \left\{ \theta : (\theta - v_i)^\top\, A_i\, (\theta - v_i) \leq 1,\ \forall i,\ d_j^\top \theta + c_j \geq 0,\ \forall j \right\}.$$

- The set of admissible decision variables $\mathscr{X}$ is also an intersection of a finite number of half-spaces and a finite number of ellipsoids.

Notice that we have left out a common geometric component in both the admissible set $\mathscr{X}$ and the uncertainty set $\mathcal{U}$ – hyperplanes. We have done this purposely to avoid increasing the notational complexity even more. If either $\mathscr{X}$ or $\mathcal{U}$ are the intersection of hyperplanes it is possible to re-parametrise the problem in a lower dimensional setting, effectively substituting one or more variables/parameters as a linear combination of their peers.

It might look like the class of problems described above is very restrictive, but most common portfolio optimisation problems can either be directly described this way or are equivalent to such problems. Have a look, for example, at the example displayed at the beginning of this chapter. While the presence of linear transaction costs means the objective function is not quadratic, by lifting the problem to a larger space through the introduction of auxiliary variables t we reach an equivalent problem that satisfies our assumptions.

The classical portfolio optimisation problem for such an objective function and admissible set is

$$\begin{aligned} &\max_x f(x;\theta) \\ &\text{s.t. } x \in \mathscr{X}, \end{aligned} \tag{4.4}$$

where θ is our parameter estimate. Given that f is a concave function of x and $\mathscr{X}$ is a set of linear and convex quadratic constraints this is a second-order cone programme and can be solved for example by interior point methods, see [28]. These are implemented by most standard conic solvers such as Mosek or CPLEX.

4.3.1 Absolute Robust Portfolio Optimisation

Suppose now we want to solve the absolute robust version of (4.4) for an uncertainty set $\mathcal{U}$ as above,

$$\begin{aligned} &\max_x \min_{\theta\in\mathcal{U}} f(x;\theta) \\ &\text{s.t. } x \in \mathscr{X}. \end{aligned} \tag{4.5}$$

It might be the case that we can explicitly find $\theta^*(x) = \text{argmin}_{\theta\in\mathcal{U}} f(x;\theta)$ and so effectively get rid of θ and be left with a problem in x alone. However this is in general not possible, so we must find another way to approach this problem. One

way is by introducing an auxiliary variable γ the above problem can be written as

$$\begin{aligned} \max_{x,\gamma} \ & \gamma \\ \text{s.t.} \ & \gamma \leq f(x;\theta), \quad \forall \theta \in \mathcal{U}, \\ & x \in \mathscr{X}. \end{aligned}$$

It is now tempting to identify $f(x;\theta) - \gamma$ as a quadratic in θ and immediately transform the first constraint above into a constraint of a quadratic being non-negative over $\mathcal{U}$. However this quadratic in θ would have coefficients which would be quadratic in x and this would be problematic later on. Instead we must lift the above problem one more time, by introducing an auxiliary variable s,

$$\begin{aligned} \max_{x,\gamma,s} \ & \gamma \\ \text{s.t.} \ & -s + \theta^\top M_\theta\, \theta + x^\top M_{x,\theta}\, \theta + b_x^\top x + b_\theta^\top \theta - \gamma \geq 0, \quad \forall \theta \in \mathcal{U}, \\ & x^\top M_x\, x \leq s, \\ & x \in \mathscr{X}. \end{aligned}$$

We can now define the quadratic function $q_{x,\gamma,s}$ as

$$q_{x,\gamma,s} = \theta^\top M_\theta\, \theta + x^\top M_{x,\theta}\, \theta + b_x^\top x + b_\theta^\top \theta - \gamma - s,$$

and so we can write problem (4.5) as

$$\begin{aligned} \max_{x,\gamma,s} \ & \gamma \\ \text{s.t.} \ & \mathscr{M}\left(q_{x,\gamma,s}\right) \in \mathscr{FC}_+\left(\mathcal{U}\right), \\ & x^\top M_x\, x \leq s, \\ & x \in \mathscr{X}. \end{aligned} \tag{4.6}$$

The reason why we introduced s to mask $x^\top M_x\, x$ is because it is important that $\mathscr{M}\left(q_{x,\gamma,s}\right)$ be a matrix whose entries are linear in the decision variables x, γ and s.

We do not know how to efficiently solve the problem above, as we have no efficient way of checking membership of $\mathscr{FC}_+\left(\mathcal{U}\right)$. However we can substitute it by the inner approximation in (4.3). By doing such a substitution, we would perhaps be reducing the space of x, γ and s allowed and hence the maximum γ under this new problem might not be the same as under (4.6). Nonetheless, any solution to the new problem would necessarily be admissible to (4.6), as per the definition of an inner approximation. Hence by solving this new problem with not only get a lower bound on the

worst-case objective function, we also get decision variables that (at least) attain that lower bound.

We can hence get a lower bound γ to (4.5) and an admissible variable x that attains such bound by solving the following problem,

$$
\begin{aligned}
\min_{x,\gamma,s,\lambda,\xi,z,\tau} \quad & \gamma \\
\text{s.t. } & \lambda_i \in \mathbb{R}_+, \quad i = 1, \ldots, k_1, \\
& \xi_{i,j} \in \mathbb{R}_+, \quad i = 1, \ldots, k_2 - 1,\ j = i+1, \ldots, k_2, \\
& \begin{bmatrix} z_{i,j} \\ \tau_{i,j} \end{bmatrix} \in L_{n_1+1}, \quad i = 1, \ldots, k_1,\ j = 1, \ldots, k_2, \\
& x^\top M_x \, x \leq s, \\
& x \in \mathscr{X}, \\
& \mathscr{M}\left(q_{x,\gamma,s}\right) - \sum_{i=1}^{k_1} \lambda_i \, \mathscr{M}\left(q_i\right) \\
& \qquad - \sum_{i=1}^{k_2-1} \sum_{j=i+1}^{k_2} \xi_{ij} \left(\begin{bmatrix} d_i \\ c_i \end{bmatrix} \begin{bmatrix} d_j \\ c_j \end{bmatrix}^\top + \begin{bmatrix} d_j \\ c_j \end{bmatrix} \begin{bmatrix} d_i \\ c_i \end{bmatrix}^\top \right) \\
& \qquad - \sum_{i=1}^{k_1} \sum_{j=1}^{k_2} \left(\begin{bmatrix} U_i \, z_{i,j} \\ \tau_{i,j} - v_i^\top z_{i,j} \end{bmatrix} \begin{bmatrix} d_j \\ c_j \end{bmatrix}^\top + \begin{bmatrix} d_j \\ c_j \end{bmatrix} \begin{bmatrix} U_i \, z_{i,j} \\ \tau_{i,j} - v_i^\top z_{i,j} \end{bmatrix}^\top \right) \succeq 0.
\end{aligned}
\tag{4.7}
$$

All the constraints above, except the last, can be cast as part of a second order cone programme. The last constraint may look daunting but it is actually a semidefinite constraint – the matrix we require to be positive semidefinite has entries that are linear in all the decision variables, and so it can be solved by any conic solvers that can handle semidefinite cones, such as Mosek.

4.3.2 Relative Robust Portfolio Optimisation

It is now time to look at the relative robust version of (4.4). For the sake of generality we will follow the ideas introduced in Chapter 2 and consider a set of benchmarks $\mathcal{B} \subseteq \mathbb{R}^{n_2}$ we wish to compare ourselves again, which may or may not be the same set as $\mathscr{X}$. We will make similar assumptions over the shape of $\mathcal{B}$ as we made over the shape of $\mathcal{U}$ – we assume $\mathcal{B}$ is the intersection of a finite number of ellipsoids and a finite number of half-spaces. Hence there exists concave quadratic function of the form $r(y) = -(y - w_i)^\top \Sigma_i \, (y - w_i) + 1$, for positive definite matrices $\Sigma_i = W_i \, W_i^\top$ and vectors w_i for $i = 1, \ldots, l_1$ and vectors g_j and f_j constants for $j = 1, \ldots, l_2$ such that

$$
\mathcal{B} = \left\{ y : (y - w_i)^\top \, \Sigma_i \, (y - w_i) \leq 1, \, \forall i, \; g_j^\top y + f_j \geq 0, \, \forall j \right\}.
$$

The relative robust version of (4.4) is

$$\begin{aligned} \min_{x} \ & \left[\max_{\theta\in\mathcal{U}} \left(\max_{y\in\mathcal{B}} \ f(y;\theta)\right) - f(x;\theta)\right] \\ \text{s.t.} \ & x \in \mathscr{X}. \end{aligned} \tag{4.8}$$

As before we need to lift this problem to a higher space before we can move on. By introducing variables γ and s we can transform (4.8) into

$$\begin{aligned} \min_{x,\gamma,s} \ & \gamma \\ \text{s.t.} \ & \mathscr{M}\left(r_{x,\gamma,s}\right) \in \mathscr{FC}_{+}\left(\mathcal{U}\times\mathcal{B}\right), \\ & x^{\top} M_x \, x \leq s, \\ & x \in \mathscr{X}, \end{aligned} \tag{4.9}$$

where

$$r_{x,\gamma,s}(y,\theta) = \gamma + y^{\top} M_x \, y - y^{\top} M_{x,\theta} \, \theta + x^{\top} M_{x,\theta} \, \theta - b_x^{\top} y + b_x^{\top} x - s$$

is a quadratic function of (y,θ) whose coefficients are linear in x, γ and s.

It is worth pointing out at this point that we could actually make this more general by making the set of benchmarks $\mathcal{B}$ we consider depend on the scenario $\theta \in \mathcal{U}$ we are analysing – we could simply have a set Θ such that if $(\theta, y) \in \Theta$, then we would wish to consider our objective function against that of y under scenario θ. From a technical point of view this would not present a challenge and we would still be able to bound such a problem as long as Θ were an intersection of half-spaces and ellipsoids (and hyperplanes).

Notice that while in the absolute robust version we wish to maximise the worst-case scenario, in the relative robust version the goal is to minimise our distance to the best performing benchmark – hence we have a minimisation problem, not a maximisation one. As before, by substituting $\mathscr{FC}_{+}\left(\mathcal{U}\times\mathcal{B}\right)$ by the inner approximation (4.3) in the first constraint of (4.9) we have a problem with a slightly smaller set of admissible decision variables x, γ and s. Hence the minimal γ of this new problem might not be the same as the minimum of (4.9), but it will definitely be admissible. Hence we can attain not only an upper bound γ but also a decision variable x that attains at most γ by solving the following problem

$$
\begin{aligned}
\min_{x,\gamma,s,\lambda,\xi,z,\tau}\ & \gamma \qquad (4.10)\\
\text{s.t. } & \lambda_i^1, \lambda_j^2 \in \mathbb{R}_+, \quad i = 1,\dots,k_1,\ j = 1,\dots,l_1,\\
& \xi_{i,j}^1 \in \mathbb{R}_+, \quad i = 1,\dots,k_2-1,\ j = i+1,\dots,k_2,\\
& \xi_{i,j}^2 \in \mathbb{R}_+, \quad i = 1,\dots,l_2-1,\ j = i+1,\dots,l_2,\\
& \xi_{i,j}^3 \in \mathbb{R}_+, \quad i = 1,\dots,k_2,\ j = 1,\dots,l_2,\\
& \begin{bmatrix} z_{i,j}^1 \\ \tau_{i,j}^1 \end{bmatrix} \in L_{n_1+1}, \quad i = 1,\dots,k_1,\ j = 1,\dots,k_2+l_2,\\
& \begin{bmatrix} z_{i,j}^2 \\ \tau_{i,j}^2 \end{bmatrix} \in L_{n_2+1}, \quad i = 1,\dots,l_1,\ j = 1,\dots,k_2+l_2,\\
& x^\top M_x\, x \le s, \qquad x \in \mathscr{X},\\
& \mathscr{M}\left(r_{x,\gamma,s}\right) - \sum_{i=1}^{k_1} \lambda_i^1\, \mathscr{M}\left(q_i\right) - \sum_{i=1}^{l_1} \lambda_i^2\, \mathscr{M}\left(r_i\right)\\
& \qquad - \sum_{i=1}^{k_2-1} \sum_{j=i+1}^{k_2} \xi_{ij}^1 \left(\begin{bmatrix} d_i \\ 0 \\ c_i \end{bmatrix} \begin{bmatrix} d_j \\ 0 \\ c_j \end{bmatrix}^\top + \begin{bmatrix} d_j \\ 0 \\ c_j \end{bmatrix} \begin{bmatrix} d_i \\ 0 \\ c_i \end{bmatrix}^\top \right)\\
& \qquad - \sum_{i=1}^{l_2-1} \sum_{j=i+1}^{l_2} \xi_{ij}^2 \left(\begin{bmatrix} 0 \\ g_i \\ f_i \end{bmatrix} \begin{bmatrix} 0 \\ g_j \\ f_j \end{bmatrix}^\top + \begin{bmatrix} 0 \\ g_j \\ f_j \end{bmatrix} \begin{bmatrix} 0 \\ g_i \\ f_i \end{bmatrix}^\top \right)\\
& \qquad - \sum_{i=1}^{k_2} \sum_{j=1}^{l_2} \xi_{ij}^2 \left(\begin{bmatrix} d_i \\ 0 \\ c_i \end{bmatrix} \begin{bmatrix} 0 \\ g_j \\ f_j \end{bmatrix}^\top + \begin{bmatrix} 0 \\ g_j \\ f_j \end{bmatrix} \begin{bmatrix} d_i \\ 0 \\ c_i \end{bmatrix}^\top \right)\\
& \qquad - \sum_{i=1}^{k_1} \sum_{j=1}^{k_2} \left(\begin{bmatrix} U_i\, z_{i,j}^1 \\ 0 \\ \tau_{i,j}^1 - v_i^\top z_{i,j}^1 \end{bmatrix} \begin{bmatrix} d_j \\ 0 \\ c_j \end{bmatrix}^\top + \begin{bmatrix} d_j \\ 0 \\ c_j \end{bmatrix} \begin{bmatrix} U_i\, z_{i,j}^1 \\ 0 \\ \tau_{i,j}^1 - v_i^\top z_{i,j}^1 \end{bmatrix}^\top \right)\\
& \qquad - \sum_{i=1}^{k_1} \sum_{j=1}^{l_2} \left(\begin{bmatrix} U_i\, z_{i,j}^2 \\ 0 \\ \tau_{i,j}^2 - v_i^\top z_{i,j}^2 \end{bmatrix} \begin{bmatrix} 0 \\ g_j \\ f_j \end{bmatrix}^\top + \begin{bmatrix} 0 \\ g_j \\ f_j \end{bmatrix} \begin{bmatrix} U_i\, z_{i,j}^2 \\ 0 \\ \tau_{i,j}^2 - v_i^\top z_{i,j}^2 \end{bmatrix}^\top \right)\\
& \qquad - \sum_{i=1}^{l_1} \sum_{j=1}^{k_2} \left(\begin{bmatrix} 0 \\ W_i\, z_{i,j}^3 \\ \tau_{i,j}^3 - v_i^\top z_{i,j}^3 \end{bmatrix} \begin{bmatrix} d_j \\ 0 \\ c_j \end{bmatrix}^\top + \begin{bmatrix} d_j \\ 0 \\ c_j \end{bmatrix} \begin{bmatrix} 0 \\ W_i\, z_{i,j}^3 \\ \tau_{i,j}^3 - v_i^\top z_{i,j}^3 \end{bmatrix}^\top \right)\\
& \qquad - \sum_{i=1}^{l_1} \sum_{j=1}^{l_2} \left(\begin{bmatrix} 0 \\ W_i\, z_{i,j}^4 \\ \tau_{i,j}^4 - v_i^\top z_{i,j}^4 \end{bmatrix} \begin{bmatrix} 0 \\ g_j \\ f_j \end{bmatrix}^\top + \begin{bmatrix} 0 \\ g_j \\ f_j \end{bmatrix} \begin{bmatrix} 0 \\ W_i\, z_{i,j}^4 \\ \tau_{i,j}^4 - v_i^\top z_{i,j}^4 \end{bmatrix}^\top \right) \succeq 0,
\end{aligned}
$$

where the matrix representation of q_i and r_i must take them to be functions of (θ, y) and not simply θ or y respectively.

Although both the absolute robust and the relative robust approximations derived above can be solved in polynomial time, they are comparatively slower than non-robust alternatives, such as (3.7). This is both due to the large number of auxiliary variables, which grow quadratically with the inclusion of each linear constraint and linearly with each quadratic constraint, but also due to the existence of a semidefinite constraint, since semidefinite programs are much slower than the more common second-order cone programs. Hence both methods will be more appropriate to medium to low frequency trading.

4.4 Numerical Examples

Let us now look back to the setting of Chapter 3, in particular problem (3.13). Whereas before we were not able to make any progress with such problem, we now have the necessary tools to at least attain an upper bound to the minimum regret problem.

We start by first identifying the constraints. There are no linear constraints, either on $\mathcal{U}$ or $\mathcal{B}$, and each has one quadratic constraint, q_1 and r_1, respectively,

$$\begin{aligned} q_1(\nu, y) &= \chi^2 - \nu^\top \Xi^{-1} \nu, \\ r_1(\nu, y) &= \sigma^2 - y^\top Q\, y. \end{aligned}$$

There is also the quadratic $r_{x,\gamma,s}$, which codifies the relative robust portfolio optimisation objective function,

$$r_{x,\gamma,s} = \gamma + \left(\hat{\mu} + \hat{V}\,\nu\right)^\top x - \left(\hat{\mu} + \hat{V}\,\nu\right)^\top y.$$

We can then formulate the approximation (4.10) for the particular case of (3.13),

$$\begin{aligned} \min_{x,\gamma,s,\lambda}\ & \gamma \\ \text{s.t. } & \lambda^1, \lambda^2 \in \mathbb{R}_+, \\ & x^\top Q\, x \leq s, \\ & x \in \mathscr{X}, \\ & \mathscr{M}\left(r_{x,\gamma,s}\right) - \lambda^1 \mathscr{M}\left(q_1\right) - \lambda^2 \mathscr{M}\left(r_1\right) \succeq 0. \end{aligned} \tag{4.11}$$

This does not look nearly as daunting as (4.10). In fact we will be able to solve it for exactly the same dataset as in Chapter 3, hence 44 dimensions. One should take notice, however, of the fact that these methods can be significantly slower than the ones found in the previous chapter. Problem (4.11), for example, runs in over two and a half seconds, which is over thirty times what we saw earlier. If we were to impose no short-selling on how benchmark set $\mathcal{B}$ (which does not make sense in a futures market, but might make sense if we were dealing with equities), this goes up to around 12 seconds, over 180 times slower than any method in Chapter 3. If one wants to run this method on a training dataset for a wide range of parameters this can easily take a day, so it might not be well suited for investment strategies where speed is determinant, such as high frequency trading.

The natural benchmark for (4.11) is (3.8), its absolute robust counterpart.

We shall start by using a low χ^2 value of 1. We also stick with our choice of $\sigma = 0.03$, since as before scaling this leads only to a scaled portfolio.

We can see the evolution of the wealth process in Figure 4.1. Now this is a low

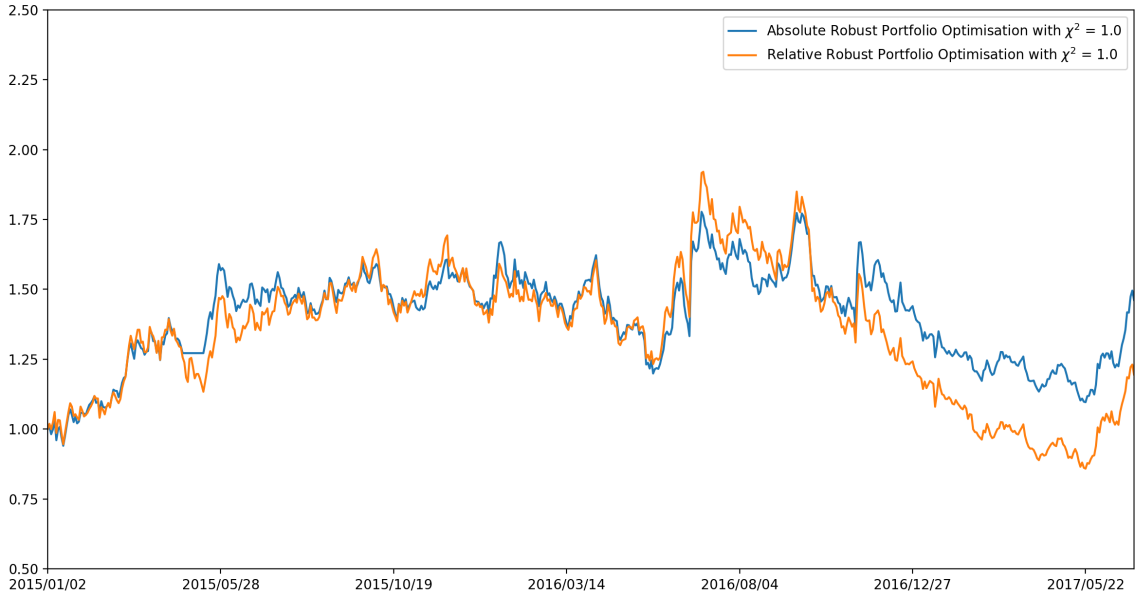


Figure 4.1: Wealth process as a result of investing according to (3.8) and (4.11) for $\sigma = 0.03$ and $\chi^2 = 1$.

value of χ^2, so both portfolios are not very robust and hence close to the non-robust version. Their risk metrics are favourable to absolute robust version as it is the most defensive of the two: 90% quantile of absolute returns at 3.4% against 4%, mean over quantile at 4.8% against 5.4%and maximum absolute return at 20% against 20.1%.

The advantages of using relative robust portfolio optimisation instead of its absolute robust counterpart become clear once we slightly increase χ^2 to 1.5, as can be seen in Figure 4.2. We now have the absolute robust portfolio being uninvested for

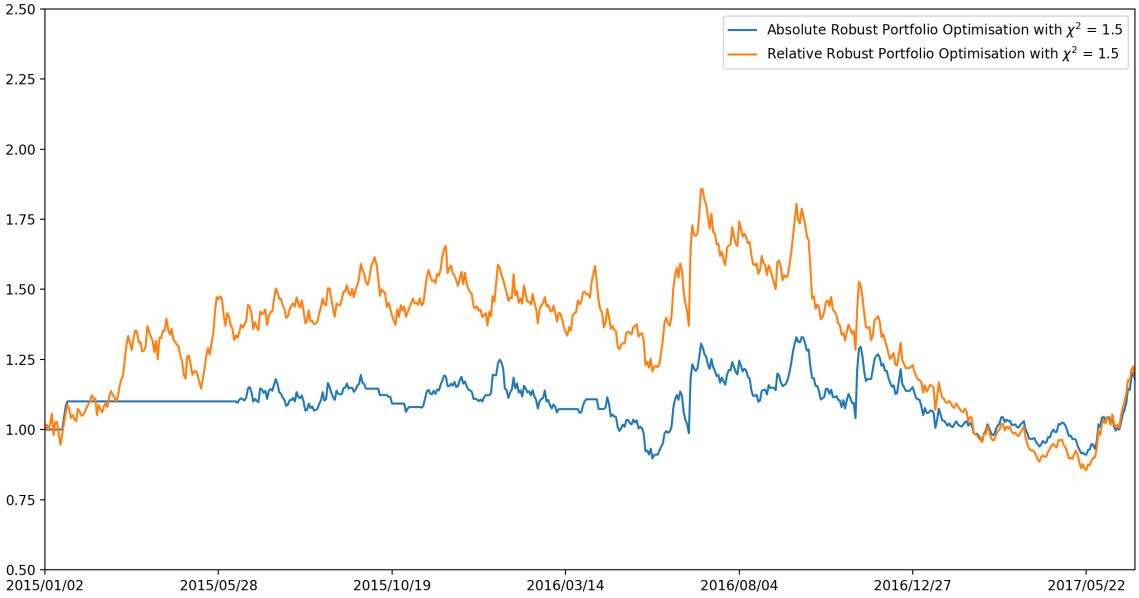


Figure 4.2: Wealth process as a result of investing according to (3.8) and (4.11) for $\sigma = 0.03$ and $\chi^2 = 1.5$.

over 6 months before finally picking up once it gets more confident of its μ – as we saw in the last chapter, we get uninvested portfolios when μ lies inside $\mathscr{U}$. This would never be allowed by an asset manager, whether it is sensible or not, which makes this a hard method to use in practice.

If we now try an ever bigger χ^2, say 2, then the absolute robust portfolio stays uninvested even longer, as seen in Figure 4.3.

Now if we really want to use absolute robust portfolio optimisation, the solution for that was prescribed in the last chapter: using $\hat{\Lambda}$, a low dimensionality version of Λ. That will mean we will need to parametrise $\mathcal{U}$ through ν, as in equation (3.11).

Say we take $\delta = 90\%$ in (3.9). If we now run both the relative robust and the absolute robust portfolio optimisation problems, we obtain the wealth process of Figure 4.4. We now have constantly a fully-invested absolute robust portfolio, as desired. On the other hand the relative robust portfolio has become more risky, as its uncertainty set has dwindled.

It is curious to consider what changes are caused by varying χ^2. When $\delta = 100\%$ (i.e. when we use Λ in full) these changes are visible in full in the absolute robust portfolio – as χ^2 increases it spends more and more time uninvested. For the relative

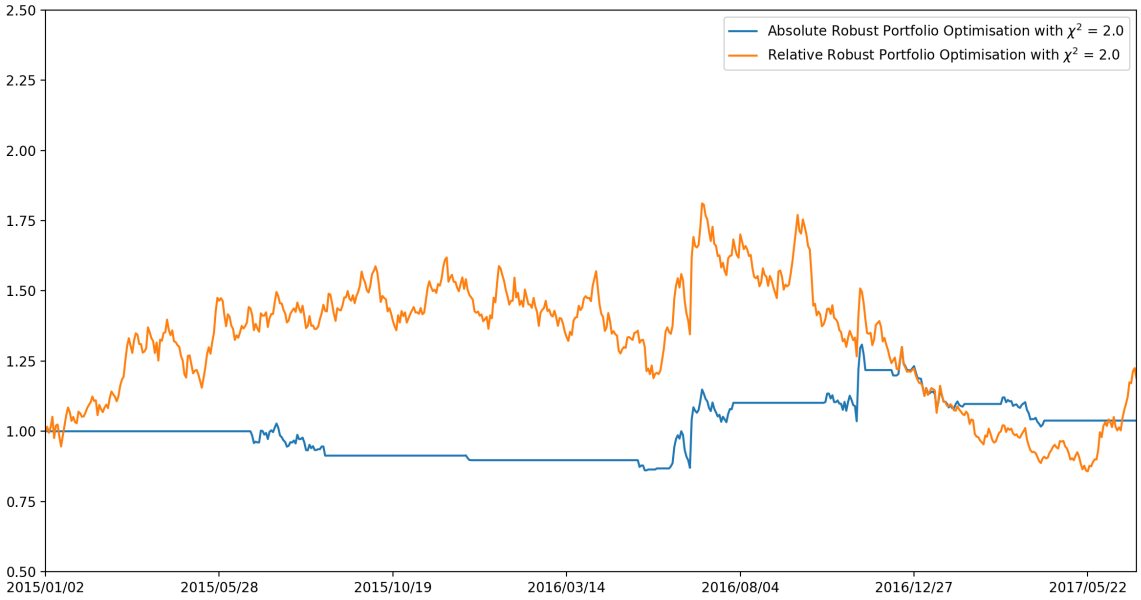


Figure 4.3: Wealth process as a result of investing according to (3.8) and (4.11) for $\sigma = 0.03$ and $\chi^2 = 2$.

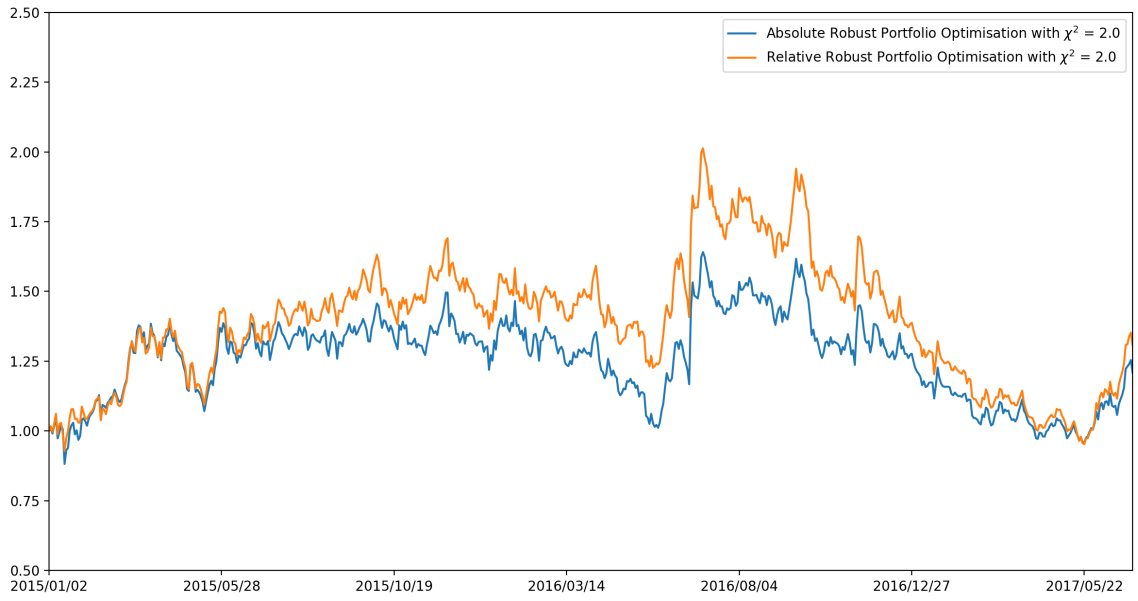


Figure 4.4: Wealth process as a result of investing according to (3.8) and (4.11) for $\sigma = 0.03$, $\chi^2 = 2$ and $\delta = 90\%$.

robust portfolio, however, we simply see a slight change in leverage – as χ^2 increases we can spot its leverage dropping ever so slightly. When $\delta = 90\%$, however, we see no change in the absolute robust portfolio as we move χ^2. This is due to the fact that it is not worth spending volatility budget on an uncertain direction when there

are some directions where no uncertainty is present. For the relative robust portfolio we see the same effect as when we use Λ in full, a slight decrease in leverage as χ^2 increases. It is also natural that as δ goes to zero both our portfolios converge to the non-robust version, as the uncertainty set collapses.

4.5 Conclusion

In this chapter we have built on the work of [33] to reach an inner approximation to the set of quadratic functions non-negative over an intersection of a finite number of ellipsoids and half-spaces. This is relevant because it allows us to approximate absolute robust and relative robust optimisation problems for which we have no way of efficiently solving them and which, in some cases, have been proved to be NP-hard.

This opens the door for a wide range of new uncertainty sets to be considered. One might for example be uncertain about a given estimate, but only wanting to protect itself against risk in one direction. This is particularly promising in equities, where the usual long-only constraints means we might not be so worried over errors that benefit us.

Our results also allow for more freedom in the use of benchmarks, as proposed in Chapter 2. Back then we were limited to a finite set of benchmarks, or else to a single ellipsoid, either in the benchmark set or in the uncertainty set. This has thus allowed us to reach approximate solutions to problems that arise naturally, such as (3.13).

Chapter 5

Portfolio Optimisation with DR-Expectation

5.1 Introduction

We have seen so far a few different models, but they all had in common the fact that in a first step the model parameters were estimated and in a second step the optimal portfolio was calculated using those parameters. It did not matter however for the second step how the first step was conducted, as long as it concluded and we had some parameters. For example, in Chapter 3 we used a Hidden Markov Model to classify historical data into two distinct regimes, one normally distributed and another following a t-student distribution. We did so by first fitting the model to historical data and subsequently by calculating the probability distribution of the hidden state given the whole historical data. From this were merely inferred which hidden state was the most likely.

Once the classification was finished, the filter was of no extra use to us. In fact we could have used a completely different method that arrived at the same classification and this would not alter our method at all. The same could be said for the expected returns vector μ or the covariance matrices Q^e and Q^t – it did not so much matter how they were calculated, as long as they were there. The model fitting is completely separate from the portfolio optimisation step.

In this chapter we shall focus on a portfolio optimisation problem under a Hidden Markov Model, but instead of simply fitting the parameters and getting some uncertainty set, we will use the model fitting process to tell us how likely each set of parameters are. This is particularly interesting for the hidden state probability vector, as we shall see how each new data point affects the next state expected probability vector and hence how the uncertainty propagates through the filter.

For this purpose we shall rely on the theory of data-driven robust (DR) nonlinear expectations introduced in [13]. In the context of portfolio optimisation this theory can be seen as an extension of absolute robust portfolio optimisation. The idea is to consider a family of measures (in a parametric family, such as the ones we shall consider, each element can be thought as a particular choice of parameters) when calculating the optimal portfolio and penalising each measure by how likely it is given historical data. If θ is our parameter vector and α the penalty, then we wish to choose the portfolio that solves

$$\sup_{\pi} \inf_{\theta} \{f(\pi;\theta) - \alpha(\theta)\}$$

for some objective function $f(\pi;\theta)$ in some subset of its domain. If $\mathcal{U}$ is our uncertainty set, then we recover the absolute robust portfolio optimisation problem by taking

$$\alpha(\theta) = \begin{cases} 0, & \text{if } \theta \in \mathcal{U} \\ \infty, & \text{otherwise.} \end{cases}$$

We shall see that in the theory of DR-expectations a penalty as above arises naturally as a limiting case with $\mathcal{U}$ being a confidence region. By taking a less abrupt penalty function we effectively penalise parameters by how likely they are given historical data, with a minimal penalty at the maximum likelihood estimates which increases as we deviate from these.

5.2 Uncertainty in Filtering

First let us start by presenting our model in more detail. Let $\{X_t\}_{t\geq 0}$ be a 2-state Markov Chain on the unit vectors $\{e_1, e_2\}$ with transition matrix $A^\top$ – it follows that $X_t - AX_{t-1}$ is a martingale. X is not observable, but we do observe the real-valued vector process $\{Y_t\}_{t\geq 0}$. These processes are defined in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where $\mathcal{F}$ denotes the filtration generated by both X and Y. We denote by $\mathcal{Y}$ the observable filtration, which is generated by Y alone.

Given X_{t-1}, the density of Y_t is $f(y; X_{t-1})$, where $f(y; e_1)$ is the normal density with mean μ_1 and covariance matrix Σ_1, while $f(y; e_2)$ is the multivariate t-student distribution with mean μ_2, spread matrix Σ_2 and τ_2 degrees of freedom.

Denote by C_t the diagonal matrix with entries $f(Y_t; e_i)$. Bayes' theorem implies that the expected state of the filter, $p_t = \mathbb{E}[X_t \mid \mathcal{Y}_t]$ follows the filtering equation

$$p_t = \frac{1}{\alpha_t} C_t \, A \, p_{t-1}, \tag{5.1}$$

where $\alpha_t = \mathbf{1}^\top C_t A p_{t-1}$. Notice that since X_t takes values in $\{e_1, e_2\}$, p_t completely describes the conditional distribution of $(X_t \mid \mathcal{Y}_t)$.

In this section we will want to impose a prior distribution on the parameter space. Given that parameters have been fitted by maximum likelihood, a natural prior is the maximum likelihood estimate asymptotic sampling distribution, i.e., a normal distribution with the fitted parameters as mean and with the inverse Fisher Information matrix as covariance matrix. This is essentially an "empirical Bayes" approach with DR-expectation.

The Fisher Information matrix for parameters μ_1, μ_2, Σ_1, Σ_2 and τ_2 can be computed in a similar manner to the expected log-likelihood or the expected gradient via (1.9), using measure $\mathbb{Q}^*$ generated by the optimal parameters. As for matrix A, we assume asymptotic independence not only between A and the other parameters but also between A_{s_1,r_1} and A_{s_2,r_2} for $r_1 \neq r_2$, $s_1, s_2, r_1, r_2 \in \{1, 2\}$. Since $A_{2,r} = 1 - A_{1,r}$ their variances are the same and we have the following asymptotic approximation of their variance,

$$\operatorname{Var}(A_{2,r}) = \operatorname{Var}(A_{1,r}) \approx \frac{A_{1,r}\, A_{2,r}}{\mathbb{E}^{\mathbb{Q}^*}\left[N_t^{r,1} + N_t^{r,2} \mid \mathcal{Y}_t\right]}.$$

Instead of fitting a model and using it as the ground truth, [13] suggests that we should consider instead a family of models, but penalise them by how unlikely they are given some historical data. This is precisely what we will do in the following sections, but before we go into details, we shall review the definition of nonlinear expectation along with some useful results. For further details on this theory see [25] and references therein.

Definition 1. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and L^∞ the space of essentially bounded random variables. A nonlinear expectation is a mapping $\mathcal{E} : L^\infty \to \mathbb{R}$ satisfying the following:*

- *For any $Z_1, Z_2 \in L^\infty$, if $Z_1 \geq Z_2$ a.s. then $\mathcal{E}(Z_1) \geq \mathcal{E}(Z_2)$. If in addition $\mathcal{E}(Z_1) = \mathcal{E}(Z_2)$ then $Z_1 = Z_2$ a.s.*
- *For any $k \in \mathbb{R}$, $\mathcal{E}(k) = k$.*
- *For any $k \in \mathbb{R}$ and $Z \in L^\infty$, $\mathcal{E}(Z + k) = \mathcal{E}(Z) + k$.*

Moreover a nonlinear expectation is said to be convex if for any $\lambda \in [0, 1]$, $Z_1, Z_2 \in L^\infty$,

$$\mathcal{E}(\lambda Z_1 + (1 - \lambda Z_2) \leq \lambda \mathcal{E}(Z_1) + (1 - \lambda)\mathcal{E}(Z_2),$$

and lower semicontinuous if for every sequence $\{Z_n\}_{n\geq 1} \in L^\infty$ *with* $Z_n \uparrow Z$ *pointwise for* $Z \in L^\infty$, $\mathcal{E}(Z_n) \uparrow \mathcal{E}(Z)$.

It is worth noting that convex expectations can also be seen as convex risk measures – if $\rho(Z) = \mathcal{E}(-Z)$, then ρ is a convex risk measure. Hence if one wants to choose an optimal control on Z to make decisions based on $\mathcal{E}$ one either minimises $\mathcal{E}(Z)$ or, as we will do in a later section, maximises $-\mathcal{E}(-Z)$.

The next theorem is due to [25] and [26], and it is an essential result, since it lets us talk of nonlinear expectations in terms of their associate penalty functions.

Theorem 14. *Let* $\mathcal{M}_1$ *denote the space of all probability measures absolutely continuous with respect to* $\mathbb{P}$. *Suppose* $\mathcal{E}$ *is a lower semicontinuous convex expectation. Then there exists a penalty function* $\alpha : \mathcal{M}_1 \to [0, \infty]$ *such that for every* $Z \in L^\infty$

$$\mathcal{E}(Z) = \sup_{\mathbb{Q}\in\mathcal{M}_1} \mathbb{E}^{\mathbb{Q}}[Z] - \alpha(\mathbb{Q}). \tag{5.2}$$

Moreover, there is a minimal such function,

$$\alpha_{min}(\mathbb{Q}) = \sup_{Z\in L^\infty} \mathbb{E}^{\mathbb{Q}}[Z] - \mathcal{E}(Z).$$

Notice that since $\mathcal{E}(0) = 0$ we must have $\inf_{\mathbb{Q}\in\mathcal{M}_1} \alpha(\mathbb{Q}) = 0$.

The motivation behind the DR-expectation (Data-driven Robust or Divergence-driven Robust) introduced in [13] is to use as a penalty the negative log-likelihood of some given data under that particular measure. Denote by $\mathcal{Q}$ the family of measures we wish to consider. If $l(\mathbb{Q}; \mathcal{Y}_t)$ is the log-likelihood of $\{Y_s\}_{s\leq t}$ under measure $\mathbb{Q}$, then the DR-expectation with uncertainty aversion parameter $k_1 > 0$ and exponent $k_2 \in [1, \infty]$ is defined as in (5.2) with

$$\alpha_t(\mathbb{Q}) = \begin{cases} \left[\frac{1}{k_1}\left(-l(\mathbb{Q}; \mathcal{Y}_t) + C\right)\right]^{k_2}, & \text{if } \mathbb{Q} \in \mathcal{Q} \\ \infty, & \text{otherwise.} \end{cases} \tag{5.3}$$

where $C = \sup_{\mathbb{Q}\in\mathcal{Q}} l(\mathbb{Q}; \mathcal{Y}_t)$ guarantees that $\mathcal{E}(0) = 0$. Hence the DR-expectation penalises models for being far away from the maximum likelihood model, which is itself not penalised. It also allows us to come up with countless examples of convex expectations.

Imagine for the moment we have a function $\phi(\pi, Y_{t+1})$ such that we want to maximise $-\mathcal{E}(-\phi(\pi, Y_{t+1}))$ over possible choices of π. We can write this as

$$\sup_{\pi} \inf_{\mathbb{Q}\in\mathcal{Q}} \left\{ \mathbb{E}^{\mathbb{Q}}[\phi(\pi, Y_{t+1})] + \left[\frac{1}{k_1}\left(-l(\mathbb{Q}; \mathcal{Y}_t) + \sup_{\mathbb{Q}\in\mathcal{Q}} l(\mathbb{Q}; \mathcal{Y}_t)\right)\right]^{k_2} \right\}. \tag{5.4}$$

If one now leaves k_1 fixed and takes the limit of $k_2 \to \infty$, then we effectively end up with

$$\sup_{\pi} \inf_{\mathbb{Q}\in\hat{\mathcal{Q}}} \mathbb{E}^{\mathbb{Q}} \left[\phi(\pi, Y_{t+1})\right],$$

where $\hat{\mathcal{Q}}$ is the family of measures where the likelihood of $\mathcal{Y}_t$ differs from its maximum by at most k_1. For a parametric family $\mathcal{Q}$ this effectively the same as taking a confidence region of parameters that are most likely. Hence (5.4) can be thought as an extension to the absolute robust optimisation problems whose uncertainty sets are confidence regions.

The family of measures $\mathcal{Q}$ we wish to consider are the ones that arise from the Hidden Markov Model described in the previous section. A recursive formula for the DR-expectation penalty in such models was derived in [14]. If $\mathbb{A}$ is the set of all admissible model parameters,

$$\mathbb{A} = \Big\{ (A, \mu_1, \Sigma_1, \mu_2, \Sigma_2, \tau_2) : \mathbf{1}^\top A = \mathbf{1},\ A_{i,j} \geq 0 \quad \forall i, j = 1, 2, \\ \mu_1, \mu_2 \in \mathbb{R}^n,\ \Sigma_1, \Sigma_2 \in \mathcal{S}_+^n,\ \tau \in \mathbb{R} \Big\}, \tag{5.5}$$

then any measure in $\mathcal{Q}$ can be parametrised by an element $\mathcal{A} \in \mathbb{A}$. In [14] it is shown that the penalty α can be written as

$$\alpha_t(\mathbb{Q}) = \left(\frac{1}{k_1} K_t(p_t, \mathcal{A})\right)^{k_2}, \tag{5.6}$$

where K_t can be computed recursively via

$$K_t(p_t, \mathcal{A}) = \inf_{p_{t-1}: p_t \propto C_t A p_{t-1}} \left\{K_{t-1}(p_{t-1}, \mathcal{A}) - \log\left(\mathbb{E}\left[f(Y_t, X_{t-1}) \mid \mathcal{Y}_t\right]\right)\right\} - m_t,$$

where $K(p_0, \mathcal{A}) = \pi_0(p_0) + \gamma(\mathcal{A})$ is our negative log-prior described in the previous section and m_t is such that $\inf K(\bullet, \bullet) = 0$.

As pointed out in [14], $\mathcal{A}$ grows in dimension quite quickly, so searching the space of all measures through K_t might be prohibitive. The proposed alternative is to allow parameters to change over time – this not only gives the model extra flexibility but also is more in line with market practice.

While it might feel that allowing $\mathcal{A} := \{\mathcal{A}_t\}_{t\geq 0}$ to change over time will increase complexity instead of reducing it, by assuming these are i.i.d. with prior density $\exp(-\gamma(\mathcal{A}))$ (which is a stronger assumption than being Markov), a significant reduction in complexity is attained by minimising at every step of the recursion over all $\mathcal{A}$. This means that the only variable linking different time-steps is p and we can

hence construct a recursive penalty κ that is a function solely of p_t. We can therefore substitute in (5.6) K_t (which depends on both $\mathcal{A}$ and p_t) by such penalty κ_t, which according to [14] satisfies the following recursion

$$\kappa_t(p_t) = \inf_{\mathcal{A}_t \in \mathbb{A},\, p_{t-1}:p_t \propto C_t A p_{t-1}} \{\kappa_{t-1}(p_{t-1}) + \gamma(\mathcal{A}_t) - \log\left(\mathbb{E}\left[f(Y_t, X_{t-1}) \mid \mathcal{Y}_t\right]\right)\} - \tilde{m}_t, \tag{5.7}$$

where $\kappa(p_0) = \pi_0(p_0)$ is our negative log-prior on p_0 and $\tilde{m}_t$ is such that $\inf \kappa = 0$. The dimension of the recursion is now much smaller, so one can actually hope to solve (or approximate) it in order to find κ.

5.3 Quadratic Approximation of Penalty κ_t

Our goal is to not only solve (5.7) but to use κ_t in the portfolio optimisation decision process. This might be difficult to do if we have to solve the full control problem, but it might be easier if we instead approximate κ_t sensibly.

It was shown in [13] that for 'nice' parametric families of measures $\mathcal{Q}$ the divergence (5.3) (with $k_1 = k_2 = 1$) seen as a function of the parameters can be approximated around the maximum likelihood estimate as a quadratic given by the Fisher Information matrix. In the case of a Hidden Markov Model with time independent generators $\{\mathcal{A}_t\}_{t\geq 0}$ the parameter space is prohibitively large, but we can instead approximate $\kappa_t(p_t)$, the infimum of α over all generators $\{\mathcal{A}_t\}_{t\geq 0}$ that lead to p_t, as a quadratic in p_t.

The result is slightly more general but in order for it to hold for a given parametric family it is sufficient for its log density to be L^2 with respect to the parametrisation. In fact under that assumption one can show that substituting (5.7) by its LQ approximation (i.e. linearising (5.1) and taking a quadratic approximation of the cost function in (5.7)) yields an approximation to $\kappa(p_t)$ of the order of $o(|p_t - p_t^*|^2)$, or $O(|p_t - p_t^*|^3)$ if we know the log density is C^3. Showing is easy but a bit gruelling – just Taylor expand (5.1) and (5.7), assume that for every p_t in a neighbourhood of p_t^* the generator that minimises is unique, use the implicit function theorem and voilà.

The results mentioned above are valid for all reasonable parametrisations, so it is worth choosing sensible ones. Approximating functions as quadratics on bounded parameters is quite awkward and leads to poor approximations, especially near said bounds. It is hence worth, for example, to see κ_t not as a function of a vector of probabilities p_t but rather as a function of log-odds $r_t := \lambda(p_t) = \log(\frac{p_{t,1}}{p_{t,2}})$. Similarly

it is worth parametrising $\mathcal{A}$ not as in (5.5) but instead by something more suitable. We shall discuss our own choice when we present numerical results in Section 5.4.

We first start by picking appropriate priors. Let π_0 and γ be

$$\begin{aligned}\pi_0(r_0) &= D_0(r_0 - r_0^*)^2,\\ \gamma(\mathcal{A}) &= (\mathcal{A} - \mathcal{A}^*)^\top \Sigma^{-1} (\mathcal{A} - \mathcal{A}^*),\end{aligned}$$

where $\mathcal{A}^*$ is the MLE estimate we get from fitting the filter (remember that to be able to fit the filter we assume parameters $\{\mathcal{A}_t\}_{t\geq 0}$ are actually the same), Σ is the corresponding Fisher Information matrix, r_0^* is our initial guess for the log-odds and D_0 represents our confidence in that guess. Now suppose by induction that κ_{t-1} is approximately a quadratic around a point r_{t-1}^*,

$$\kappa_{t-1}(r_{t-1}) = D_{t-1}(r_{t-1} - r_{t-1}^*)^2.$$

This is trivially true for $t = 1$ due to our choice of prior. The filter equation (5.1) can be written in terms of log-odds as

$$r_t = \lambda\left(C_t A\, \lambda^{-1}(r_{t-1})\right). \tag{5.8}$$

Let $\bar{r}_t$ be the solution to the filtering equation (5.8) after the model has been fit, i.e., with MLE parameters $\mathcal{A}_t \equiv \mathcal{A}^*$ and initial guess $r_0 = r_0^*$. If we linearise (5.8) around $(\mathcal{A}^*, \bar{r}_{t-1})$ we get

$$\begin{aligned} r_t &\approx \bar{r}_t + \psi_r\,(r_{t-1} - \bar{r}_{t-1}) + \psi_{\mathcal{A}}^\top\,(\mathcal{A} - \mathcal{A}^*),\\ r_{t-1} - \bar{r}_{t-1} &\approx \frac{1}{\psi_r}(r_t - \bar{r}_t - \psi_{\mathcal{A}}^\top\,(\mathcal{A} - \mathcal{A}^*)),\end{aligned}$$

where ψ_r and $\psi_{\mathcal{A}}$ are the appropriate derivatives.

The last step needed is to approximate $\log\left(\mathbb{E}\left[f(Y_t, X_{t-1}) \mid \mathcal{Y}_t\right]\right)$. Notice that this is actually a function of r_{t-1} and $\mathcal{A}$ that can be explicitly calculated since $\mathbb{E}\left[f(Y_t, X_{t-1}) \mid \mathcal{Y}_t\right] = \mathbf{1}^\top C_t\, p_{t-1}$. We can hence linearise it as before around $(\mathcal{A}^*, \bar{r}_{t-1})$

$$\log\left(\mathbb{E}\left[f(Y_t, X_{t-1}) \mid \mathcal{Y}_t\right]\right) \approx l_r\,(r_{t-1} - \bar{r}_{t-1}) + l_{\mathcal{A}}^\top\,(\mathcal{A} - \mathcal{A}^*),$$

where l_r and $l_{\mathcal{A}}$ are the appropriate derivatives.

If we now substitute these approximations into the inner expression in (5.7) we get

$$\begin{aligned}&(\mathcal{A} - \mathcal{A}^*)^\top \Sigma^{-1} (\mathcal{A} - \mathcal{A}^*) - \frac{l_r}{\psi_r}\left(r_t - \bar{r}_t - \psi_{\mathcal{A}}^\top\,(\mathcal{A} - \mathcal{A}^*)\right) - l_{\mathcal{A}}^\top\,(\mathcal{A} - \mathcal{A}^*)\\ &\quad + \frac{D_{t-1}}{\psi_r^2}\left(\left(r_t - \bar{r}_t - \psi_{\mathcal{A}}^\top\,(\mathcal{A} - \mathcal{A}^*)\right) + \psi_r(\bar{r}_{t-1} - r_{t-1}^*)\right)^2\\ &= (\mathcal{A} - \mathcal{A}^*)^\top \Sigma_{\text{full}}^{-1} (\mathcal{A} - \mathcal{A}^*) - 2B_{\text{full}}^\top\,(\mathcal{A} - \mathcal{A}^*) + H_{\text{full}},\end{aligned} \tag{5.9}$$

where

$$\begin{aligned}
\Sigma_{\text{full}}^{-1} &= \Sigma^{-1} + \frac{D_{t-1}}{\psi_r^2}\psi_{\mathcal{A}}\psi_{\mathcal{A}}^{\top}, \\
B_{\text{full}} &= \frac{D_{t-1}}{\psi_r^2}\left((r_t - \bar{r}_t) + \psi_r(\bar{r}_{t-1} - r_{t-1}^*) - \frac{\psi_r}{2}l_r\right)\psi_{\mathcal{A}} + \frac{1}{2}l_{\mathcal{A}}, \\
H_{\text{full}} &= \frac{D_{t-1}}{\psi_r^2}(r_t - \bar{r}_t)^2 + (r_t - \bar{r}_t)\frac{2D_{t-1}(\bar{r}_{t-1} - r_{t-1}^*) - l_r}{\psi_r}.
\end{aligned}$$

If we now minimise (5.9) over $\mathcal{A}$ and write that as a quadratic in r_t, we get an approximate penalty $\tilde{\kappa}_t$,

$$\begin{aligned}
\kappa_t(r_t) \approx \tilde{\kappa}_t(r_t) &= H_{\text{full}} - B_{\text{full}}\Sigma_{\text{full}}B_{\text{full}}^T \\
&= D_t(r_t - \bar{r}_t)^2 + E_t(r_t - \bar{r}_t) + F \\
&= D_t(r_t - r_t^*)^2,
\end{aligned}$$

where

$$\begin{aligned}
D_t &= \frac{D_{t-1}}{\psi_r^2} - \frac{D_{t-1}^2}{\psi_r^4}\psi_{\mathcal{A}}^{\top}\Sigma_{\text{full}}\psi_{\mathcal{A}}, \\
E_t &= -\frac{D_{t-1}}{\psi_r^3}\left[(2D_{t-1}(\bar{r}_{t-1} - r_{t-1}^*) - l_r)\psi_{\mathcal{A}} + \psi_r l_{\mathcal{A}}\right]^T\Sigma_{\text{full}}\psi_{\mathcal{A}} \\
&\quad + \frac{2D_{t-1}(\bar{r}_{t-1} - r_{t-1}^*) - l_r}{\psi_r}, \\
r_t^* &= -\frac{E_t}{2D_t} + \bar{r}_t,
\end{aligned} \tag{5.10}$$

and F is a constant we can assume to be 0, since κ_t is constructed so as to have minimum 0.

5.4 Portfolio Optimisation with DR-Expectation

Having constructed our penalty $\tilde{\kappa}_t$, we are now interested in calculating a one-step optimal portfolio under these families of models. We have throughout this thesis taken as our objective function the expected return of our portfolio, but constrained this to some volatility budget. This was not problematic as we have mainly focused on uncertainty in the expected returns vector μ, but at this time we are exploring joint uncertainty of all parameters. There is therefore slightly more nuance in the idea of a volatility constraint. Is our portfolio to satisfy such a constraint for any admissible covariance matrix? Certainly not, as we allow any semidefinite positive matrix, even though most are heavily penalised and therefore excluded in practice.

Do we only require it to be satisfied for the maximum likelihood estimate of the covariance matrix? This somewhat defeats the point of having joint uncertainty. We shall instead drop the constraint altogether and use as objective function the risk-adjusted return of our portfolio, that is, the expected return of our portfolio minus its variance multiplied by a risk aversion constant λ.

If Y_{t+1} is the future one-step vector of returns, then our risk-adjusted return optimal portfolio π should maximise the one-step DR-expectation (see [14])

$$\sup_{\pi} \inf_{p_t, \mathcal{A}_{t+1}} \left\{ \mathbb{E}^{\mathbb{Q}} \left[Y_{t+1}^{\top} \pi \right] - \lambda \mathrm{VaR}^{\mathbb{Q}} \left[Y_{t+1}^{\top} \pi \right] + \left(\frac{1}{k_1} \left(\tilde{\kappa}_t(p_t) + \gamma \left(\mathcal{A}_{t+1} \right) \right) \right)^{k_2} \right\}. \quad (5.11)$$

Notice that we get a γ penalty in addition to a $\tilde{\kappa}_t$ penalty – this is because by assuming that generators are i.i.d. we have to include our prior at every step. We can compare this to (5.7), with the difference that we have not yet witnessed Y_{t+1}, so we cannot infer anything from it.

Using the tower property and the law of total variance we expand the first two terms above as

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}} \left[Y_{t+1}^{\top} \pi \right] - \lambda \mathrm{VaR}^{\mathbb{Q}} \left[Y_{t+1}^{\top} \pi \right] = \; & p_1 \, \mu_1^{\top} \pi - \lambda \, p_1 \, \pi^{\top} \Sigma_1 \pi + p_2 \, \mu_2^{\top} \pi - \lambda \, p_2 \frac{\nu_2}{\nu_2 - 2} \pi^{\top} \Sigma_2 \pi \\ & - \lambda \, p_1 \, p_2 \, \pi^{\top} (\mu_1 - \mu_2)(\mu_1 - \mu_2)^{\top} \pi \, . \end{aligned}$$

Problem (5.11) is nontrivial to solve, as it is a sup-inf problem where the objective function is not convex-concave (i.e. convex in one variable and concave in the other) due to the fact that $\tilde{\kappa}$ and γ are quadratics in a transformed space, not in the original space. For example, $\tilde{\kappa}$ is a quadratic in $\log\left(\frac{p_{t,1}}{p_{t,2}}\right)$ and as such it is not convex in p_t.

We can however solve it for low dimensional problems. Suppose we have two underlying assets and our portfolios must be long-only and fully invested. Any portfolio is uniquely determined by its first component π_1, which must be between 0 and 1. We can hence solve (5.11) by producing a grid of portfolios with π_1 between 0 and 1 and then solving the inner minimisation for each such portfolio using standard numerical methods. We can choose the optimal portfolio to be the one in the grid who maximises the inner minimisation.

As mentioned in the previous section, in order to approximate penalty κ one has to choose an appropriate parametrisation of all parameters involved. As before we parametrize the probability vector p_t by its log-odds. Generator $\mathcal{A}$ entails one transition matrix, two expected return vectors, two covariance matrices and one degree of freedom. The transition matrix can be parametrised by its off-diagonal elements. Both expected return vectors are unconstrained, so do not need to be transformed.

Each covariance matrix can be described by the standard deviation of each asset and their correlation. Given that the standard deviation has to be positive it is more appropriate to use log standard deviation – similarly for the degree of freedom. Finally correlation has to be between -1 and 1, so a logit transformation as used for p_t is appropriate, after appropriate scaling and translation.

Given we want to impose a prior on these transformed parameters based on the Fisher Information Matrix we get from fitting the static version of the model it helps to use this parametrisation from the start, hence directly fitting the model via maximum likelihood in the transformed space.

We fit this model using daily data obtained from Yahoo Finance of two publicly traded ETFs: iShares Core S&P 500 and iShares Core U.S. Aggregate Bond. The model is fitted on one-year data, January until December 2013. We also use this data to initialise our penalty $\tilde{\kappa}_t$, taking as prior $D_0 = 1$ and $r_t^* = 0$ before updating it through the historical data as prescribed in (5.10).

We then solve (5.11) daily from January 2014 until July 2017, at each day updating the penalty $\tilde{\kappa}_t$. We do not refit our model, so γ is a quadratic penalty constant over time which we take to be the log-prior described at the end of Section 1.2.3.

2013 was a very good year for equities, so given that the two available assets represent equities and bonds and we only consider fully-invested long-only portfolios we need to impose a relatively high risk-aversion λ so as to avoid constantly choosing equity-only portfolios – for that reason we settled with $\lambda = 30$. If we take $k_1 = 12$ (approximately 98% quantile of a χ^2_{12}) and $k_2 = 8$ (as a proxy for $k_2 = \infty$) we are approximately solving the absolute robust portfolio optimisation problem for this model, using as a joint uncertainty set of all parameters the approximate 98% confidence region.

The performance of our portfolio with the above parameters is shown in Figure 5.1, alongside the performance of Markowitz with the same objective function and constraints but no penalty on alternate models, hence simply using the MLE parameters.

We can see that the DR-Expectation portfolio is less volatile and hence does not lose so much during bad periods nor does it gain so much during good ones in comparison to Markowitz. This is of course a characteristic we have seen in previous chapters concerning absolute robust portfolio optimisation – given they are maximising the worst-case scenario they are by nature defensive portfolios. In this case the DR-Expectation portfolio has a realised Sharpe Ratio of 1.05, while the Markowitz portfolio has a realised Sharpe Ratio of 0.68. The DR-Expectation portfolio also

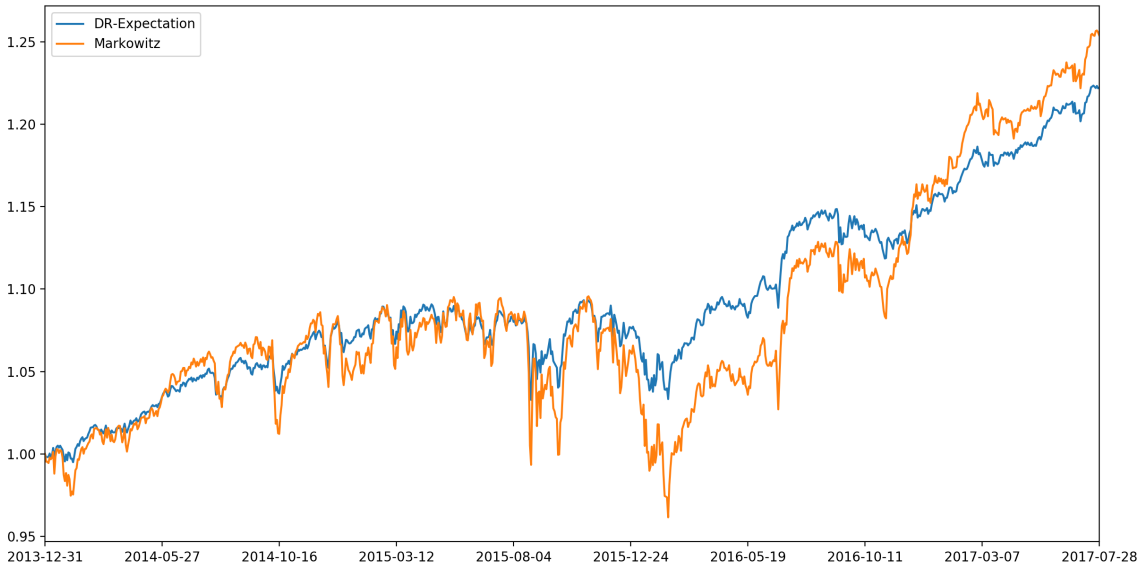


Figure 5.1: Performance of DR-Expectation and Markowitz Portfolios for $\lambda = 30$.

scores highest in terms of realised objective function (realised return minus 30 times realised variance) at -0.03 against -0.23 for the Markowitz portfolio. This is not surprising, given that the model parameters that were calibrated in 2013 did not hold up until 2017.

It is interesting to have a look at how the portfolios composition changes as a function of probability vector $p_t = \mathbb{E}[X_t \mid \mathcal{Y}_t]$, i.e., the probability vector of what distribution we will be getting tomorrow's return. For Markowitz this is the only parameter that changes over time and hence its portfolio is a deterministic function of p_t. For the DR-Expectation we also have to consider D_t, the aversion term that encodes how uncertain we are about the true value of p_t. The proportion of iShares Core U.S. Aggregate Bond each portfolio holds as a function of p_t can be seen in Figure 5.2.

Two things can be seen from Figure 5.2. First, DR-Expectation always holds more bonds than Markowitz. This is in line with our previous argument that due to it being maximising the worst case scenario it tends to be a more defensive portfolio. Second, the portfolio actually varies less as a function of p_t than Markowitz does. This can be explained by the fact that Markowitz takes p_t as the correct probability vector, while the DR-Expectation takes into account different probability vectors close to p_t in order to reach an optimal portfolio. This also reflects one of the motivations behind robust portfolio optimisation – by taking parameters as uncertain rather than ground truths the optimal portfolios are not so sensitive to parameter estimates.

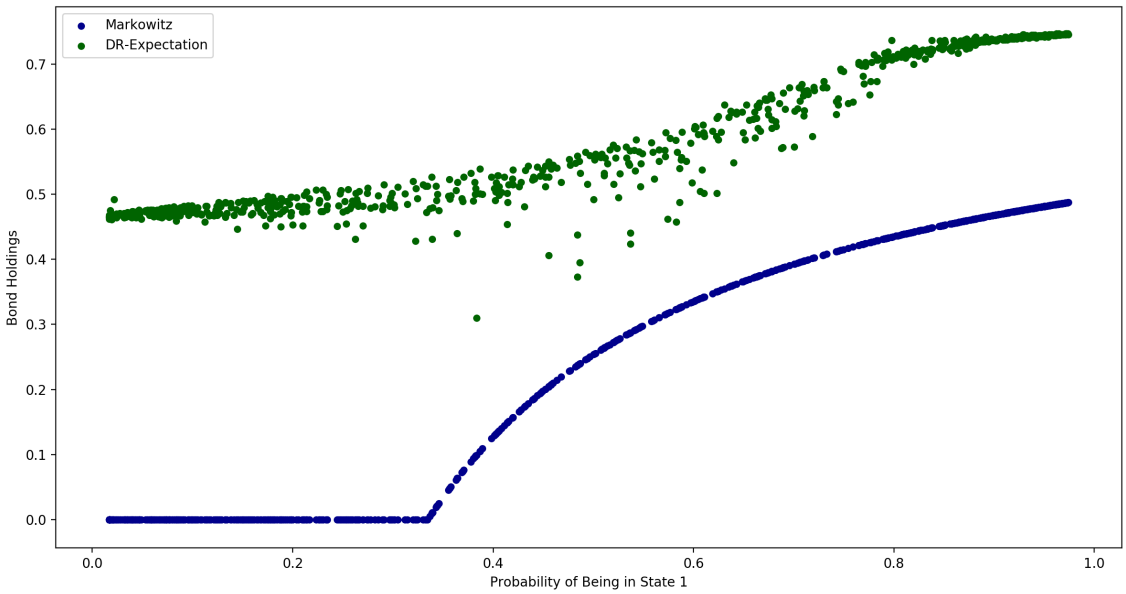


Figure 5.2: Proportion of Bonds held by each portfolio against probability of being in state 1.

It is worth looking at the same plot, of the resulting portfolio as a function of p_t, for $k_2 = 1$ in Figure 5.3. In this case we have not a worst-case scenario approach but rather a simple quadratic penalty in the parameter space, with a minimum at the MLE estimates. While it still holds slightly more bonds than Markowitz, this is no longer significant. This occurs because there is now an effective penalty on moving away from the MLE estimates, even if only slightly. Hence parameter choices that differ non-trivially from the highest likelihood parameters are no longer contemplated.

If instead of $k_1 = 12$ we use $k_1 = 1200$, we still have the same quadratic penalty, although this time considerably smaller. This allows the model to consider a few more options and so gets more defensive than Markowitz, as can be seen by comparing Figure 5.4 and Figure 5.3.

In all the above we made the choice of using the historical data to initialise the state penalty $\tilde{\kappa}_t$. Given that we update this over time, reflecting the fact that not only our state estimate change but also how certain we fell about our estimate, it is interesting to analyse what would have happened had we not initialised it and merely chose some prior D_0 and r_0^*. This will give us some insight into how dependent on the prior our filter uncertainty estimate is.

Suppose we have two penalty functions, $\tilde{\kappa}_t^1$ and $\tilde{\kappa}_t^2$. $\tilde{\kappa}_t^1$ is initialised using data from January to December 2013, while $\tilde{\kappa}_t^2$ is simply started at $D_0 = 1$, $r_0^* = 0$ (notice

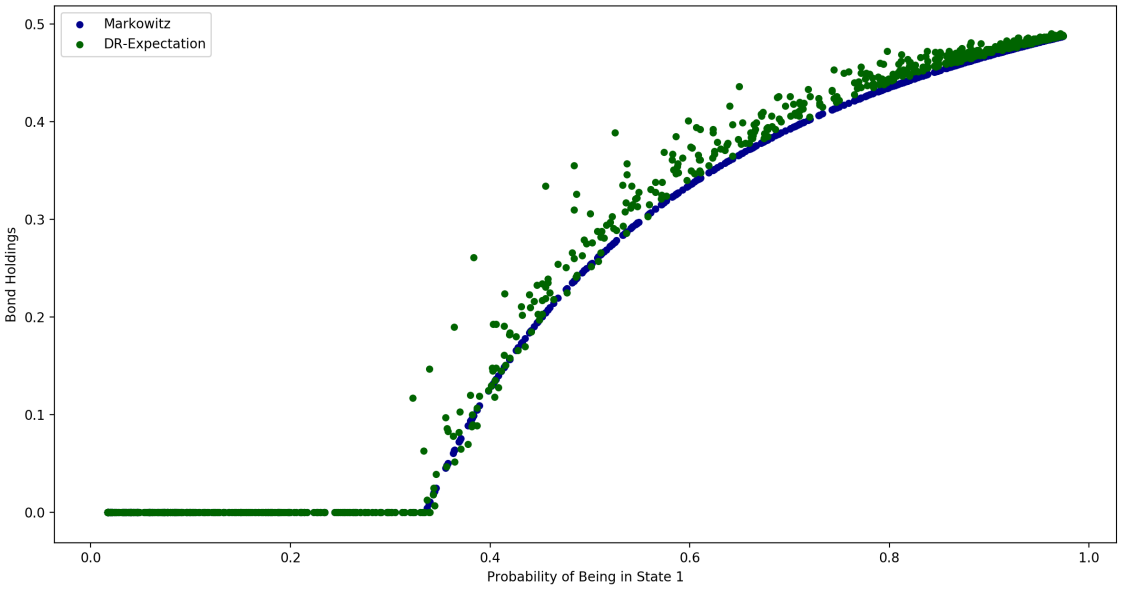


Figure 5.3: Proportion of Bonds held by each portfolio against probability of being in state 1, when $k_1 = 12$ and $k_2 = 1$.

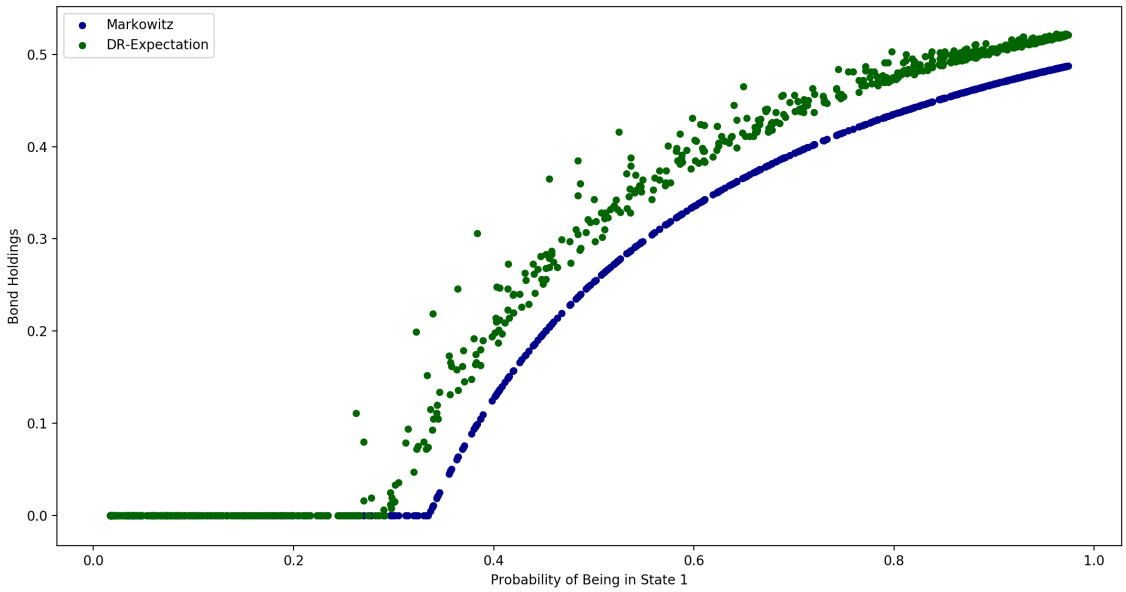


Figure 5.4: Proportion of Bonds held by each portfolio against probability of being in state 1, when $k_1 = 1200$ and $k_2 = 1$.

that log-odds equal to 0 means we start with a prior of 50% probability of being in each state). We then update both penalties for three months, January to March 2014. The difference in D_t and r_t^* between the two filters over time can be seen in Figure 5.5.

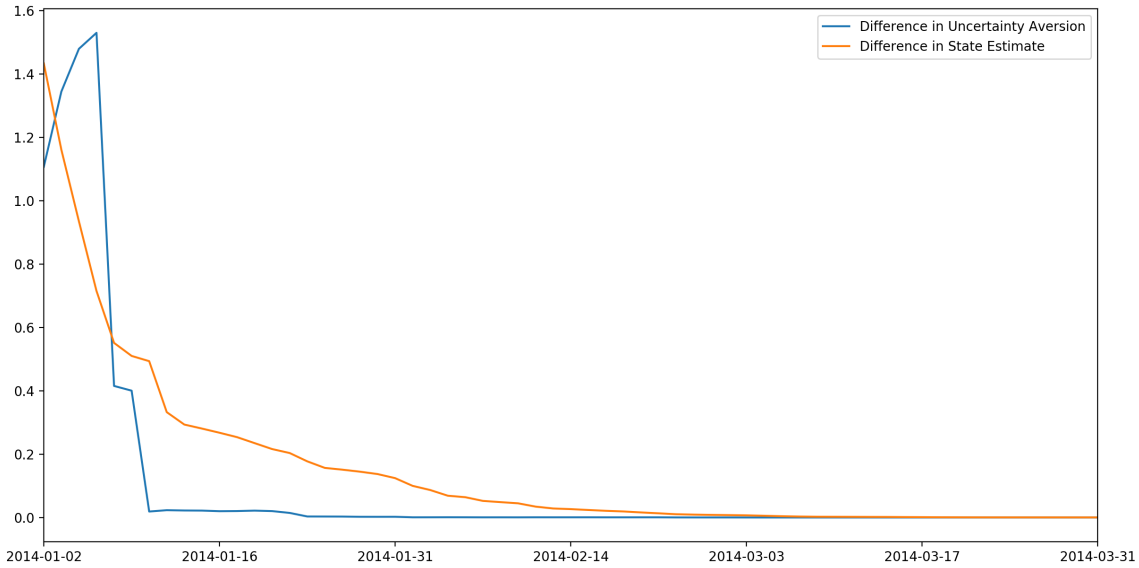


Figure 5.5: Difference in D_t and r_t^* over time between two penalty functions – one initialised using historical data, the other initialised at $D_0 = 1$, $r_0^* = 0$.

After 10 days the uncertainty aversion D_t is the same, while the state estimates r_t^* took roughly two months to converge. This seems to indicate that our penalty has strong ergodic properties and so initial parameters have little impact. In particular this means that if we are using historical data to initialise our penalty, it does not really matter which prior we use – by the time we finished running the initial data through the penalty this choice ceases to matter.

5.5 Conclusion

In this chapter we looked at the problem of portfolio optimisation in a simple Hidden Markov Model and used DR-expectations so as to incorporate the uncertainty arising from our fitted parameters into the decision process directly. While penalising parameters by their log likelihood under historical data leads to a difficult control problem, we approximate this penalty by a quadratic function on a transformed space, leaving us with a simple recursion to solve.

This method acts as a bridge between Markowitz and absolute robust portfolio optimisation – if the penalty is very large we are forced to accept the maximum likelihood parameters and hence we are effectively solving a classical optimisation problem, as in Figure 5.3. However if the penalty is very flat around the MLE and quickly explodes after a certain threshold this is effectively a worst-case scenario

optimisation, as in Figure 5.2. It is when we do a middle ground, as in Figure 5.4, that we get a portfolio which has considered different measures but penalised them by how unlikely they were.

One drawback of this method is that (5.11) is not easily solved, as the inner function does not satisfy the minimax theorem – if it did we could explicitly solve for π and be left with a standard minimisation problem. This means that this method is significantly slower than any other we have seen in this thesis. In the numerical examples we saw each iteration of (5.11) took over 20 seconds. If we were to increase the number of stocks the number of variables would increase quadratically in $\mathcal{A}$ and linearly in π, which would mean an exponential increase in solving time.

Chapter 6

Conclusion

In this thesis we tried to address two of portfolio optimisation most common problems: the uncertain and changing nature of financial markets and the recurrence of bigger than normal tail events.

We started by tackling the former in a simple setting. To deal with the changing behaviour of covariance matrices we considered a finite number of scenarios, any of which could be realised over the next investment period. We then incorporated this uncertainty in a relative robust portfolio optimisation, but taking into account asset managers' worries over divergence from competition. We also proposed a few modifications to the standard robust optimisation framework that put the modelling more in line with the reasoning of market participants.

We then move to tackle the problem of non-normal tails. Most metrics used by practitioners to gauge tail risk are difficult to incorporate on generic portfolio optimisation problems, as the dependency of such functions on individual asset holdings is highly complex. Hence we attempted to overcome this problem by devising an approximate quadratic form that encapsulates such risk. This can then be used not only as a quadratic constraint but also as an ellipsoidal uncertainty set in robust portfolio optimisation problems.

This then led us to the problem of solving (or at least approximating) relative robust portfolio optimisation problems when two quadratic forms are present. These types of problems rely on the set of quadratic functions non-negative over a certain region being tractable, which is not always the case. However we manage to get an inner approximation to such sets, which in turn allows us to calculate bounds for a variety of portfolio optimisation problems.

Finally we look at a different problem that is also usually tackled with robust optimisation: calibration error. The norm in modelling is often to use some historical data to calibrate a given model and henceforth use that calibrated model to make a

decision. When calibrating, however, there might have been a different model that fitted the data almost as well but which gets discarded and never though of again. By using the likelihood inferred by the available data to measure the plausibility of a given model we are able to take a global view of all models before making a decision. This theory is still not developed enough for full use by an asset manager with dozens or hundreds of assets to choose from, but it does hold much promise and the appeal to make decisions purely driven by the data itself.

6.1 Future Work

Following on the work done in this thesis, there are a few of projects that can be carried forward.

6.1.1 Relative Robust Portfolio Optimisation with Specified Competitors and Heavy-Tailed Returns

It would be very interesting to do a collaboration with someone in the asset manager industry on the use of relative robust portfolio optimisation with specified competitors, as in Chapter 2, using the compact uncertainty set developed in Chapter 3. As we saw this problem wouldn't necessarily be tractable, but thanks to the work developed in Chapter 4 we would be able to at least get an approximate solution.

This would be an empirical study in style similar to Chapter 2, as most of the theoretical work can be found in this thesis. We would however require the collaboration of someone familiar with the industry who could provide insight into all the relevant choices, be it in which asset class would this be more appropriate, how to get a signal μ, which benchmarks would one use, what constraints would one impose, etc. One could then, with a real world example, better compare the merits of relative vs absolute robust portfolio optimisation or Q vs Λ as generators of uncertainty sets or the benefits of using a low-dimensionality uncertainty set.

While I was lucky enough to collaborate with HSBC (who kindly funded my studentship) during my thesis, the team that supported me was FX Research, which produces broker research, and hence was not in a position to collaborate in such research.

6.1.2 Multi-period Relative Robust Portfolio Optimisation

Chapter 4 provides the theoretical tools to use Relative Robust Portfolio Optimisation in more complex settings than the ones explored in this thesis. One such setting we

think would be particularly well fitted would be the use of Chapter 3 tools in a multi-period portfolio optimisation problem. This is because a lot of the arguments put forward to model the tail distribution independently actually get exacerbated when we consider more than one periods: not only do large losses happen more regularly than would be expected, they also exhibit higher autocorrelation than otherwise measured in more positive periods. Other reasons to pursue a multi-period model might be the desire to incorporate into their portfolio optimisation problem features like long-term forecasts, transaction costs and other market microstructures.

While in a single period optimisation problem practitioners often use price returns as the underlying variable, in multi-period this can be a bit awkward as returns do not add up over time. The solution is to either use price increments (so called PnL) or log returns. Price increments are attractive especially when dealing with transaction costs or market microstructures as they allow for the easy modelling of real world effects. However it is harder to measure covariance effects due to the difference in scales. Log returns on the other hand are much closer to realised returns and are a natural choice in most of mathematical finance literature. Its main drawback is that a portfolio log return isn't exactly the linear combination of its assets' log returns, but the difference is often so negligible that it is worth using the linear approximation for convenience.

The challenge lies in how to calculate the covariance matrices $Q^{(e)}$ and $Q^{(r)}$, since our methodology works similarly once these matrices are determined. Say we want to include a T-day CVaR-lite constraint into our portfolio optimisation problem. Let x_t be the portfolio we wish to hold t periods from now and R_t the (random) log return vector over that period. We are interested in getting an estimate of

$$\mathbb{E}\left[\left|\sum_{t=1}^{T} R_t^\top x_t\right| \middle| \mathfrak{F}\left(R_1, \ldots, R_T\right) = 0\right],$$

for some multi-period filter $\mathcal{F}$. If we denote $R := (R_1, \ldots, R_T)$ and $x := (x_1, \ldots, x_T)$, then this can be written as $\mathbb{E}\left[\left|R^\top x\right| \mid \mathfrak{F}(R) = 0\right]$. We can then use the same methodology as we did in the single period setting. However this involves designing a multi-period filter, which is much more complex than a single-period one, as autocorrelations and general lack of independence need to be taken into account.

If however we would to keep using the same single period filter $\mathfrak{F}$ instead of developing a new one from scratch, then there are a few adaptations we could do to make that work.

Let us say we want to estimate $Q^{(e)}$, the multi-period covariance matrix of R over extreme market behaviour. This entails not only the covariance between different assets' log returns but also between different time lags. To limit the number of parameters one has to estimate we make the assumption of covariance stationarity,

$$\mathrm{Cov}(R_{t_1,k_1}, R_{t_2,k_2}) = \sigma(t_2 - t_1, k_1, k_2) = \sigma(t_1 - t_2, k_2, k_1),$$

i.e., the covariance between the log return of two assets over two different periods depends only on the lag between these periods (and on the underlying assets themselves). This leads us to the conclusion that $Q^{(e)}$ is a symmetric Toeplitz block matrix, that is,

$$Q^{(e)} = \begin{bmatrix} \Omega_0 & \Omega_1 & \Omega_2 & \dots & \dots & \Omega_{T-1} \\ \Omega_1^\top & \Omega_0 & \Omega_1 & \ddots & & \vdots \\ \Omega_2^\top & \Omega_1^\top & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \Omega_1 & \Omega_2 \\ \vdots & & \ddots & \Omega_1^\top & \Omega_0 & \Omega_1 \\ \Omega_{T-1}^\top & \dots & \dots & \Omega_2^\top & \Omega_1^\top & \Omega_0 \end{bmatrix},$$

where Ω_t are $n \times n$ matrices with $[\Omega_t]_{i,j} = \rho(t, i, j)$. Hence to estimate $Q^{(e)}$ we need only to estimate Ω_t for $t = 0, \dots, T-1$.

Ω_0 is the only one required to be symmetric and is also the most straightforward to estimate. If $T = 1$ then we are back in the single-period setting and we have $Q^{(e)} = \Omega_0$. Therefore we need only to apply $\mathfrak{F}$ to the daily log returns and simply compute the covariance matrix for the set $\{r_s : \mathfrak{F}(r_s) = 0\}$, as previously.

Slightly less straightforward is the task of estimating Ω_t for $t > 0$. Let us take the example $t = 1$. To estimate Ω_1 we have to first consider what makes a vector (R_t, R_{t+1}) be classified as extreme market behaviour. One natural assumption is to consider that (R_t, R_{t+1}) is extreme if and only if $R_t + R_{t+1}$ is extreme and hence if and only if $\mathfrak{F}(R_t + R_{t+1}) = 0$.

In this spirit consider the time series of log returns over t-days (constructed from the log returns over 1-day by taking a rolling cumulative sum of size t). We could now run this time series by $\mathfrak{F}$ to determine which t-days log returns constituted extreme events and which didn't. If we denote by r_s^t the vector of log returns over the time period $[s - t, s]$ we can then calculate an unbiased estimator of

$$\hat{\sigma}(t, k_1, k_2) = \frac{1}{N_t - 1} \sum_{s:\mathfrak{F}(r_s^t)=0} (r_{s-t,k_1} - \bar{r}_{k_1})(r_{s,k_2} - \bar{r}_{k_2}),$$

where $N_t = |\{s : \mathfrak{F}(r_s^t) = 0\}|$. Different methods of covariance estimation may also be employed, such as robust statistics or shrinkage. With this we can construct an estimate of Ω_t and by extension of Q_e. Following the same steps allows us to calculate an estimate of the multi-period covariance matrix under typical market behaviour $Q^{(r)}$ as well.

Given that this is an entry-wise estimation there is unfortunately no guarantee that these matrices will be positive semidefinite in general, even though with enough data this will almost always be the case. If one is employing shrinkage methods [39] or any other technique to ensure stability of the covariance matrix estimate it is best to do this at the very end for $Q^{(e)}$ or $Q^{(r)}$ rather than for (or in addition to) every single Ω_t, as this will improve the stability of the whole matrix. Any other method to force a symmetric matrix to be positive semidefinite could also be employed.

6.1.3 Multivariate Portfolio Optimisation with DR-Expectation

The framework proposed in Chapter 5 combines traditional portfolio optimisation with ideas from stochastic programming while at the same time allowing data to drive the model choice, rather than the other way around. It is therefore a very promising approach we wish to keep developing.

However at the moment it is mostly a proof of concept, as equation (5.11) becomes prohibitively expensive to solve for even only 4 assets. This is due to the fact that the number of parameters in the infimum grows quadratically, hence significantly increasing the time it takes to simply solve the inner problem. On top of that we have an optimisation over π, which itself requires more and more iterations until convergence as the dimension of the problem increases.

There is however hope that by changing slightly the inner function, either by approximating it or by choosing a slightly different shape for the penalty function, it will become feasible to solve for a decent number of dimensions. The underlying idea is that if we could use the minimax theorem, which we cannot at the moment as the function is not convex-concave, then we could swap the supremum and the infimum in (5.11), yielding

$$\inf_{p_t, \mathcal{A}_{t+1}} \sup_{\pi} \left\{ \mathbb{E}^{\mathbb{Q}} \left[Y_{t+1}^{\top} \pi \right] - \lambda \mathrm{VaR}^{\mathbb{Q}} \left[Y_{t+1}^{\top} \pi \right] + \left(\frac{1}{k_1} \left(\tilde{\kappa}_t(p_t) + \gamma \left(\mathcal{A}_{t+1} \right) \right) \right)^{k_2} \right\}.$$

This is now a much easier problem, as the inner optimisation problem can be solved explicitly, leaving a single (although dimensionally sizeable) minimisation problem, which can be solved by standard methods.

There is still a problem when the inner optimisation cannot be solved explicitly, either because of a different objective function or due to additional constraints, but even then we still expect it to be easier computationally to solve than (5.11).

Appendix A

List of Future Contracts

AUSTRALIAN $$, composite	BRITISH POUND, composite
COCOA CANADIAN $$, composite	COTTON # 2 MILK III, Comp.
US DOLLAR INDEX	EURODOLLAR, composite
NASDAQ, MINI	S & P 500, MINI
T-NOTE, 5yr composite	EURO, composite
GOLDMAN SAKS C. I.	JAPANESE YEN, composite
ORANGE JUICE	COFFEE WHEAT, KC
LUMBER	MEXICAN PESO
WHEAT, MINN	NIKKEI INDEX
SUGAR # 11	SWISS FRANC, composite
T-NOTES, 2yr composite	T-NOTE, 10yr composite
T-BONDS, composite	PALLADIUM, electronic
CORN, Electronic	FEEDER CATTLE, Electronic
GOLD, Electronic	HEATING OIL, electronic
SILVER, Electronic	COPPER, electronic
SOYBEAN OIL, Electronic	SOYBEAN MEAL, Electronic
NATURAL GAS, electronic	OATS, Electronic
PLATINUM, electronic	ROUGH RICE, Electronic
SOYBEANS, Electronic	LIVE CATTLE, Electronic
CRUDE OIL, Electronic	WHEAT, Electronic
LEAN HOGS, Electronic	

Bibliography

[1] E. D. Andersen and K.D. Andersen. The mosek interior point optimizer for linear programming: An implementation of the homogeneous algorithm. In *High Performance Optimization*. Springer, 2000.

[2] A. A. Balkema and L. De Haan. Residual life time at great age. *The Annals of Probability*, 2(5):792–804, 1974.

[3] Ian Barnett and Jukka-Pekka Onnela. Change point detection in correlation networks. *Scientific reports*, 6, (2016), 2016.

[4] J. Baz, N. M. Granger, C. R. Harvey, N. Le Roux, and S. Rattray. Dissecting investment strategies in the cross section and time series. *Available at SSRN: https://ssrn.com/abstract=2695101*, 2015.

[5] A. Ben-Tal. *Lectures on modern convex optimization : analysis, algorithms, and engineering applications*. Philadelphia, Pa : Society for Industrial and Applied Mathematics : Mathematical Programming Society, Philadelphia, Pa, 2001.

[6] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805, 1998.

[7] A. Ben-Tal and A. Nemirovski. Robust solutions of uncertain linear programs. *Operations Research Letters*, 25(1):1–13, 1999.

[8] D. Bertsimas, David B. Brown, and Constantine Caramanis. Theory and applications of robust optimization. *SIAM Review*, 53(3):464–501, 2011.

[9] Sanjeev Bhojraj, Charles M. C. Lee, and Derek K. Oler. What's my line? a comparison of industry classification schemes for capital market research. *Journal of Accounting Research*, 41(5):745–774, 2003.

[10] V. K. Chopra and W. T. Ziemba. The effect of errors in means, variances, and covariances on optimal portfolio choice. *The Journal of Portfolio Management*, 19(2):6–11, Winter 1993.

[11] Yves Choueifaty and Yves Coignard. Toward maximum diversification. *The Journal of Portfolio Management*, 35(1):40–51, 2008.

[12] Roger G. Clarke, Harindra de Silva, and Steven Thorley. Minimum- variance portfolios in the U.S. equity market. *The Journal of Portfolio Management*, 33(1):10–24, 2006.

[13] Samuel N. Cohen. Data- driven nonlinear expectations for statistical uncertainty in decisions. *Electronic Journal Of Statistics*, 11(1):1858–1889, 2017.

[14] Samuel N. Cohen. Uncertainty and filtering of hidden markov models in discrete time. 2017. Available at https://arxiv.org/abs/1606.00229.

[15] Matthias Degen and Paul Embrechts. EVT- based estimation of risk capital and convergence of high quantiles. *Advances in Applied Probability*, 40(3):696–715, 2008.

[16] Victor DeMiguel, Lorenzo Garlappi, Francisco Nogales, and Raman Uppal. A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science*, 55(5):798–812, 2009.

[17] Victor DeMiguel, Lorenzo Garlappi, and Raman Uppal. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The Review of Financial Studies*, 22(5):1915–1953, 2009.

[18] Victor DeMiguel and Francisco J. Nogales. Portfolio selection with robust estimation. *Operations research*, 57(3):560–577, 2009.

[19] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B (Methodological)*, 39(1):1–38, 1977.

[20] Robert J. Elliott. Exact adaptive filters for markov chains observed in gaussian noise. *Automatica*, 30(9):1399–1408, 1994.

[21] E. Erdoğan, D. Goldfarb, and G. Iyengar. Robust active portfolio management. *Journal of Computational Finance*, 11:71–71–98, 2008.

[22] Frank J. Fabozzi, Sergio M. Focardi, and Petter N. Kolm. *Financial modeling of the equity market : from CAPM to cointegration.* Wiley, Hoboken, 2006.

[23] Frank J. Fabozzi, Petter N. Kolm, Dessislava A. Pachamanova, and Sergio M. Focardi. Robust portfolio optimization. *The Journal of Portfolio Management*, 33(3):40–48, 2007.

[24] Daniel J. Fenn, Mason A. Porter, Stacy Williams, Mark McDonald, Neil F. Johnson, and Nick S. Jones. Temporal evolution of financial-market correlations. *Physical Review E: Statistical, Nonlinear, Biological, and Soft Matter Physics*, 84(2), 2011.

[25] Hans Föllmer and Alexander Schied. *Stochastic finance : an introduction in discrete time.* De Gruyter, New York, 3rd rev. and extended ed. edition, 2011.

[26] Marco Frittelli and Emanuela Rosazza Gianin. Putting order in risk measures. *Journal of Banking and Finance*, 26(7):1473–1486, 2002.

[27] D. Goldfarb and G. Iyengar. Robust portfolio selection problems. *Mathematics of Operations Research*, 28(1):1–38, 2003.

[28] J. Gondzio. Interior point methods 25 years later. *European Journal of Operational Research*, 218(3):587–601, 2012.

[29] L. Haan and Ana Ferreira. *Extreme value theory : an introduction.* Springer, New York ; London, 2006.

[30] Robert A. Haugen and Nardin L. Baker. The efficient market inefficiency of capitalization-weighted stock portfolios. *The Journal of Portfolio Management*, 17(3):35–40, 1991.

[31] R. Hauser. The S-procedure via dual cone calculus. 2013. Available at https://arxiv.org/abs/1305.2444.

[32] R. Hauser, V. Krishnamurthy, and R. H. Tütüncü. Relative robust portfolio optimization. 2013. Available at https://arxiv.org/abs/1305.0144.

[33] R. Hauser and R. Tütüncü. SDP approximability of relative robust QP. *PAMM*, 7(1):2060043–2060044, 2007.

[34] Ravi Jagannathan and Tongshu Ma. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *NBER Working Paper Series*, page 8922, 2002.

[35] J. H. Kim, W. C. Kim, and F. J. Fabozzi. Recent developments in robust portfolios with a worst-case approach. *Journal of Optimization Theory and Applications*, pages 1–19, 05/16 2013.

[36] Vladimir Koltchinskii and Karim Lounici. Asymptotics and concentration bounds for bilinear forms of spectral projectors of sample covariance. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 52(4):1976–2013, 2016.

[37] Panos Kouvelis and Gang Yu. *Robust Discrete Optimization and Its Applications*. Kluwer Academic Publishers, Norwell, MA, 1997.

[38] Aaron Kuriloff. Stock investors pay up for peace of mind. *The Wall Street Journal*, 2016.

[39] Olivier Ledoit and Michael Wolf. Honey, I shrunk the sample covariance matrix. *The Journal of Portfolio Management*, 30(4):110–119, 2004.

[40] Wai Lee. Risk- based asset allocation: A new answer to an old question? *The Journal of Portfolio Management*, 37(4):11–28, 2011.

[41] Z. Lu. A new cone programming approach for robust portfolio selection. *Technical Report, Deparment of Mathematics, Simon Fraser University, Burnaby, BC*, 2006.

[42] Z. Lu. A computational study on robust portfolio selection based on a joint ellipsoidal uncertainty set. *Mathematical Programming*, 126(1):193–193–201, 2011.

[43] Z. Lu. Robust portfolio selection based on a joint ellipsoidal uncertainty set. *Optimization Methods and Software*, 26(1):89–104, 2011.

[44] Mikhail Makarov. Applications of exact extreme value theorem. *Journal of Operational Risk*, 2(1):115–120, 2007.

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

[46] R. O. Michaud. The markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal*, 45(1):31–42, 1989.

[47] Katta Murty and Santosh Kabadi. Some NP-complete problems in quadratic and nonlinear programming. *Mathematical Programming*, 39(2):117–129, 1987.

[48] I. Pólik and T. Terlaky. A survey of the S-lemma. *SIAM Review*, 49:371, 2007.

[49] B. T. Polyak. Convexity of quadratic transformations and its use in control and optimization. *Journal of Optimization Theory and Applications*, 99(3):553–583, 12/01 1998.

[50] R. Tyrrell Rockafellar and Stanislav Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–41, 2000.

[51] Holger Rootzén, Johan Segers, and Jennifer L. Wadsworth. Multivariate peaks over thresholds models. 2016. Available at http://arxiv.org/abs/1603.06619.

[52] Holger Rootzén, Johan Segers, and Jennifer L. Wadsworth. Multivariate generalized pareto distributions: parametrizations, representations, and properties. 2017. Available at http://arxiv.org/abs/1705.07987.

[53] Bernd Scherer. A note on the returns from minimum variance investing. *Journal of Empirical Finance*, 18(4):652–660, 2011.

[54] K. Schöttle and R. Werner. Robustness properties of mean-variance portfolios. *Optimization*, 58(6):641–663, 2009.

[55] K. Schöttle, R. Werner, and R. Zagst. Comparison and robustification of bayes and black-litterman models. *Mathematical Methods of Operations Research*, 71(3):453–475, 06/01 2010. J2: Math Meth Oper Res.

[56] Thomas Stridsman. Data pros and cons. *Futures*, 27(6):54–56, 1998.

[57] J. F. Sturm and S. Zhang. On cones of nonnegative quadratic functions. *Mathematics of Operations Research*, 28(2):246–267, 2003.

[58] Jé Teiletche, Thierry Roncalli, and Sé Maillard. The properties of equally-weighted risk contributions portfolios. *IDEAS Working Paper Series from RePEc*, 2010.

[59] V. A. Yakubovich. The S-procedure in nonlinear control theory. *Vestnik Leningrad University*, 1:62–77, 1971.

[60] Y. Yuan. On a subproblem of trust region algorithms for constrained optimization. *Mathematical Programming*, 47(1-3):53–63, 1990.

[61] Gilles Zumbach. The empirical properties of large covariance matrices. *IDEAS Working Paper Series from RePEc*, 2009.