CHALMERS
UNIVERSITY OF TECHNOLOGY

Portfolio optimization with trend following strategies

- A comparative study of methods for portfolio optimization applied on trend following trading strategies in futures markets.

Master's thesis in Engineering Mathematics and Computational Science

Samuel Rubenson

Department of Mathematics
Chalmers University of Technology
Gothenburg, Sweden 2016

Abstract

This thesis investigates how the mean-variance framework for portfolio optimization compares against that of risk-parity and the minimum conditional value-at-risk (CVaR) portfolio. Within the risk measure of portfolio variance, we find that the performance of the mean-variance portfolio is highly dependent on a well-conditioned sample covariance matrix while risk-parity appears to offer increased numerical stability. But with a regularized estimate, no method consistently outperforms the other. We suggest a minor extension to the risk-parity allocation objective with a resulting portfolio that exhibits superior properties in several central aspects. The minimum CVaR portfolio is built around the alternative risk measure conditional value-at-risk and we find that while the original problem formulation is prone to overfitting, a regularized version shows promising results worthy of further investigation.

Acknowledgements

I want to express my gratitude to Lynx Asset Management for making it possible for me to carry out this thesis work and for offering me a friendly work environment. Special thanks go to Tobias Rydén, the main supervisor of this thesis, for his guidance and advise throughout the whole process. I also want to thank Ola Backman for many valuable discussions on the direction of this thesis.

Finally I want to thank Patrik Albin, my supervisor and examiner at Chalmers, for his help administrating and completing this thesis.

Samuel Rubenson, Stockholm, January 2017

Contents

1
Introduction

This thesis is written in cooperation with Lynx Asset Management, a hedge fund that trades on futures markets for equity indices, commodities, currencies and bonds. Such funds are often referred to as Commodity trading advisors, or a CTA funds. Lynx's investment strategies are solely based on quantitative models, and the process of deciding which positions to take in different assets can naturally be divided into two steps. The first step is *prediction*, where based on historical data a prediction model generates signals indicating the direction (usually combined with signal strengths) that prices in different assets are expected to follow. The second step, the *portfolio* step, aims at combining the available assets in a way that yields a favorable risk/return profile based on the signals generated by the prediction model together with some measure of risk. This thesis will focus on the second step, i.e. methods for portfolio optimization, applied to trend following trading strategies.

1.1 Futures contracts

A futures contract is a type of financial derivative, meaning that its current value or future payoff is a function of the price of some underlying asset, for example a stock market index. Specifically it is a standardized forward contract. A forward contract between two parties stipulates that one party must buy and the other one sell, a specific underlying asset for an agreed upon price at a specified future time. Forward contracts are well suited to hedge risk exposure to commodities, currencies etc., in a way that provides physical delivery of the underlying asset at the time of maturity. For example, say that a non-American company receives an order to deliver a product in one year and will be paid \$100 at that time. The value of \$1 in the native currency will most likely be different at the time of payment, hence the company has a risk exposure to the exchange rate between the two currencies. To hedge this risk, the company may enter a forward contract such that it one year from now will receive a specified amount in the native currency in exchange for \$100, thus removing the risk of the US dollar losing value compared to the native currency.

The exchange traded futures contract is better suited for pure speculation in the price of the underlying asset. Similar to a forward contract there is no cost associated with entering a futures contract. But to reduce the risk of default, there is a daily cash settlement corresponding to the price change of the futures contract over that day. This cash settlement is transferred between the two parties' margin accounts associated with the open position. The size of the margin account generally depend on the value of the underlying asset and the volatility of its price.

1.2 Background

Classic portfolio theory mainly deals with the problem of allocating capital. That is, the problem of investing a fixed capital in different available assets, typically by buying stocks and bonds. For a CTA fund the problem becomes a bit different seeing as when trading futures, only a fraction of the position value is required for a margin account and the win/loss of a position is settled daily. Basically no capital is needed to take on positions, thus the interest does not lie in the allocation of capital but rather in finding an interrelation between positions in different assets (the portfolio) that has desired properties based on current information. The absolute position sizes will only scale the realized win/loss and is thus based on how much risk is desired to undertake. Further, when trading on futures markets there is no practical difference between attaining a long or a short position in an asset, making it easier to construct a well hedged portfolio that has low correlation with the ”market portfolio”.

1.3 Thesis outline

The purpose of this thesis is to compare and improve three existing methods for portfolio optimization, with the main aim of investigating if alternatives to the standard Markowitz model could show superior portfolio properties. The second chapter will present the input to these models in the form of data and parameters. Chapter 3 will review three different methods for portfolio selection, their original form along with suggested extensions. Both the theoretical aspects of the optimization problems as well as efficient ways of solving them in practice will be covered in this chapter. Chapters 4 and 5 cover the methods for evaluation together with the results and include an investigation of how the models compare under different circumstances, how sensitive their performance is to changes in the input parameters and finally how the relative performance is dependent on the underlying trading model.

2

Data and parameters

2.1 Price data and returns

The data set considered throughout this thesis contains daily observations of the Open, Close, High and Low futures prices of 74 different underlying assets dating back as far as 1980. The collection of assets are a wide mix of stock market indices, commodities, currencies and bonds. As discussed in section 1.1, a futures contract has a limited time to maturity after which the contract ceases to exist. Thus constructing a continuous time series containing the futures price for an asset involves patching together price series for similar contracts over time. Although there will generally be jumps in a raw combination of such time series, these jumps can be removed in a way that preserves the daily futures price differences (but not the actual price) which is convenient when analysis is based on returns.

Let $\{S_t\}_{t=0}^{T}$ be the price series of an asset over time. To simplify analysis when working with time series, a common practice is to transform the data in such a way that it can be assumed to be a stationary process, implying a time independent mean and variance. In general a financial price series S_t can not be assumed to have a constant mean, but for the return series dS_t defined as

$$dS_t = S_t - S_{t-1},$$

a close to zero mean can generally be assumed. Many empirical studies do however show that financial return series have a time dependent variance, often referred to as volatility clustering [3]. To remove this property and make time series for different assets comparable the return series will be normalized as

$$d\hat{S}_t = \frac{dS_t}{\sigma_{t-1}}, \tag{2.1}$$

where σ_t is the standard deviation of the time series S at time t. Using σ_{t-1} in equation (2.1) keeps $d\hat{S}$ a martingale process. Note that with this normalization returns are additive both over time and over assets where the unit now is standard deviations. To obtain the normalized returns requires an estimate of the standard deviation at each time step t. There are several ways to obtain such an estimate but a popular method making use of both the Open, Close, High and Low price was proposed by D.Yang and Q.Zhang [11].

Let O_t, H_t, L_t, C_t be the time series containing open, high, low and close prices of an asset. The rolling variance estimate $\hat{\sigma}_t^2$ for this asset will be estimated according to

$$\begin{cases} Y_t = (O_t - C_{t-1})^2 + \frac{1}{2}(H_t - L_t)^2 - (2\log 2 - 1)(O_t - C_t)^2 \\ \hat{\sigma}_t^2 = \alpha\hat{\sigma}_{t-1}^2 + (1-\alpha)Y_t, \end{cases} \tag{2.2}$$

where $\alpha \in [0,1]$ is a filter parameter that determines how fast "old" data should be forgotten. Letting $\alpha = 1 - \frac{1}{\tau}$, the parameter τ can loosely be interpreted as the mean of how far back in time the filter weights are placed.

The transformed time series of normalized returns is denoted by

$$dZ_t = \frac{dS_t}{\hat{\sigma}_{t-1}},$$

and is assumed to have constant zero mean and unit variance. This time series dZ_t will be referred to as normalized returns and is considered in all analysis throughout this thesis.

2.2 Dependence structure

In order to construct a portfolio with a desired risk profile some measure of risk needs to be defined. A very common measure is the standard deviation, or volatility, of portfolio returns.

Let our universe consist of M assets for which the normalized returns can be modeled as random variables with zero mean and unit variance. Further, let $w \in \mathbb{R}^M$ be the position vector defining a portfolio P. Each position w_i has the unit of standard deviations and thus $|w_i|$ describes how much risk is taken in asset i; the sign of w_i represents a long (positive) or a short (negative) position. The variance of the portfolio returns R_P can then be defined as

$$\mathrm{Var}(R_P) = \sum_{i=1}^{M}\sum_{i=1}^{M} w_i w_j \sigma_{i,j},$$

where $\sigma_{i,j}$ is the correlation coefficient between assets i and j. In matrix notation this is written as

$$\mathrm{Var}(R_P) = w^T \Sigma w,$$

where $\Sigma \in \mathbb{R}^{M\times M}$ is the correlation matrix and $(.)^T$ denotes the transpose of a vector.

While the considered time series for each asset is assumed to be stationary with a time invariant variance, this is not assumed for correlations between assets. Thus for all $t \in \{1,\dots,T\}$ we seek to estimate

$$\Sigma_t = \mathbb{E}[dZ_t dZ_t^T] \in \mathbb{R}^{M\times M},$$

where we now consider the vector form of returns, i.e. $dZ_t \in \mathbb{R}^M$. Using only the second moment assumes that the squared expected values of returns are if non zero much smaller than the expected squared value, that is

$$\mathbb{E}[dZ_{t,i}dZ_{t,j}] \gg \mathbb{E}[dZ_{t,i}]\mathbb{E}[dZ_{t,j}] \quad \forall i, j.$$

The correlation matrix Σ_t at time t will be taken as a linear combination of daily observations according to

$$\Sigma_t = \sum_{s=0}^{t-1} c_s dZ_{t-s} dZ_{t-s}^T,$$

where c_s are filter coefficients. Again there is a wide variety of options in how to set this coefficient but focus will be kept on variations of an exponential moving average, i.e. a coefficient with exponential decay.

In practice problems can arise from the fact that the estimates for assets at the beginning of their time series can be unstable and to deal with this a burn-in period is introduced, meaning that we enforce that X number of historical data points are required in order to use the correlation estimate. The main reason for this (besides a stable estimate) is to ensure that the estimated matrix will be positive definite which numerically may not hold for daily observations. This is a crucial property of a correlation matrix ensuring a strictly positive variance of portfolio returns.
An exponential moving average of the daily observations can be written as

$$\Sigma_t = \sum_{s=0}^{t-1} (1-\alpha)\alpha^s dZ_{t-s} dZ_{t-s}^T,$$

with $\alpha \in [0,\, 1]$. Now the half-life s of a daily observation can be solved for by

$$(1-\alpha)\alpha^s = \frac{(1-\alpha)}{2} \quad \Rightarrow \quad s = -\frac{\log 2}{\log \alpha}.$$

Letting $\alpha = 1 - \frac{1}{\tau}$, $s \approx \tau \log 2$. The parameter τ is used as input when estimating Σ_t; some multiple of the half-life $\tau \log 2$ could be a reasonable way to set the burn-in period.

2.2.1 Regularized correlation estimate

Estimating a high-dimensional correlation matrix quickly becomes a difficult problem. In a setting with M assets, there are $(M^2 - M)/2$ parameters to estimate. In our case of $M = 74$ that is 2701 parameters and it is safe to assume that there will be large errors in many of these estimates.

A frequently observed property of the sample correlation matrix is a high condition number [4], i.e. that the matrix is ill-conditioned and it is not uncommon to see close to singular estimates. In many applications making use of the correlation matrix, this property may amplify errors. An example is when solving a linear system

of equations where the solution involves the precision matrix (inverse correlation matrix).
As we will see later, this problem applies to some methods for portfolio optimization, and a common approach to increase numerical stability is regularization of the sample correlation matrix.

Recall that for a normal matrix A, the condition number κ_A can be calculated as

$$\kappa_A = \frac{|\lambda_{max}(A)|}{|\lambda_{min}(A)|},$$

where λ refers to eigenvalues. It has been suggested in [7] that the larger eigenvalues of the correlation matrix are generally overestimated whilst the smaller eigenvalues are underestimated. To overcome the problem of an ill-conditioned matrix and reduce the impact of estimation errors, Ledoit and Wolf proposed a shrinkage estimator for the correlation matrix, a linear combination of Σ and the identity matrix I [4]. Consider adding some multiple of the identity matrix to Σ according to

$$\hat{\Sigma} = \Sigma + \lambda I, \tag{2.3}$$

where $\lambda \in \mathbb{R}$. As the correlation matrix is symmetric, it can be decomposed as

$$\Sigma = PDP^T$$

where P is a orthonormal matrix consisting of eigenvectors of Σ, and D is a diagonal matrix with the corresponding eigenvalues. It is easy to see that

$$PDP^T + \lambda I = P(D + \lambda I)P^T,$$

and thus adding some multiple of the identity matrix to Σ will increase all eigenvalues by the constant λ. Note that the eigenvalues' ordering by size stays unchanged. This approach deals with the potential problems of a close to singular estimate, a high condition number, and the confidence placed on directions with very large or small eigenvalues. Note in equation (2.3) that as λ increases towards infinity the properties of the regularized correlation matrix approaches that of the identity matrix. To simplify evaluation, the shrinkage estimator will be calculated as

$$\hat{\Sigma} = (1 - \lambda)\Sigma + \lambda I,$$

where $\lambda \in [0,\, 1]$ will be referred to as the regularization coefficient or the degree of regularization.

2.3 Prediction model

When constructing a portfolio we want to take into account our beliefs about future price movements generated from some prediction model (or trading strategy). CTA funds generally use a variety of trading strategies to meet their investment objectives; the subset that will be considered in this thesis is trend following models. Researching profitable prediction models is however outside the scope of this work, and therefore a well known simple trend following model will be used for all comparisons of portfolio strategies. This model is based on moving averages taken on price series.

Working with the normalized return series dZ_t defined in section 2.1, consider the normalized *price* series Z_t computed as the cumulative sum of returns

$$Z_t = \sum_{\tau=0}^{t} dZ_\tau,$$

this series will have the unit of standard deviations instead of a currency. A moving average of a price series can be interpreted as a trend with respect to some time horizon, a slow moving average can be seen as the long term trend and a fast moving average can be seen as the short term trend. An example is shown in figure 2.1.

Consider a trend following trading strategy where the expected return of an asset is proportional to the difference between a fast exponential moving average (EMA) and a slow EMA. That is, prices are expected to go up if the fast EMA (short term trend) is higher than the slow EMA and vice versa. The absolute difference between the EMA's can be regarded as a trend strength.

As the difference between moving averages on a price series is a sequence of linear operations, and the price series is made up of a cumulative sum of returns (also linear) it is easy to deduce that this procedure can be simplified to one linear filter applied directly to the normalized returns. This trading strategy generating expected returns μ, can thus be described as

$$\mu_t = \sum_{s=1}^{t} c_s dZ_{t-s},$$

where c_s is the filter parameter. An example of how c_s changes with s, the number of days back in time, is shown in figure 2.2 where c_s is scaled to sum to 1. This way the behaviour of c_s (depending on the two parameters for the fast and slow EMA) can be summarized by *one* parameter, its weighted mean τ representing how fast the model forgets historical data. Formally

$$\tau = \sum_{s=1}^{\infty} c_s s,$$

and this parameter will be used as a reference to what time horizon the trading strategy is considering.

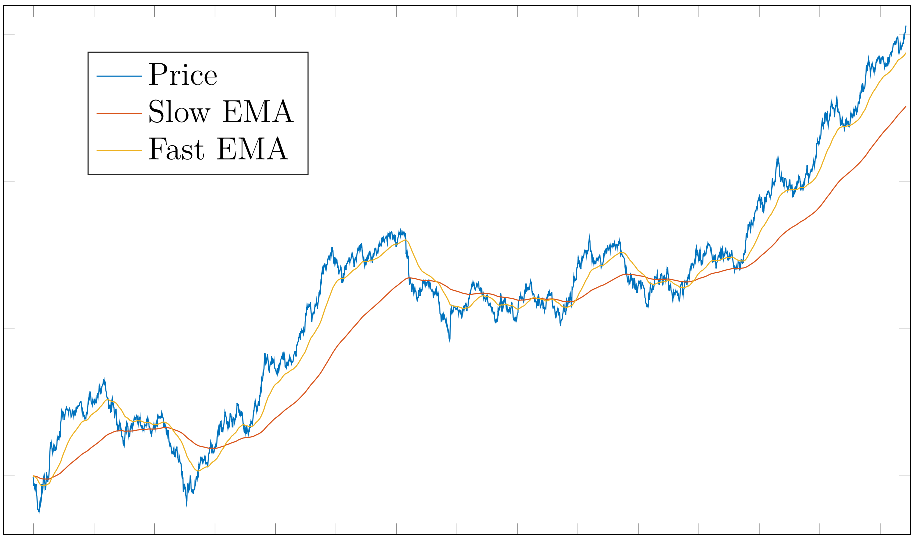

Figure 2.1: An example of a long and a short moving average taken on a price series.

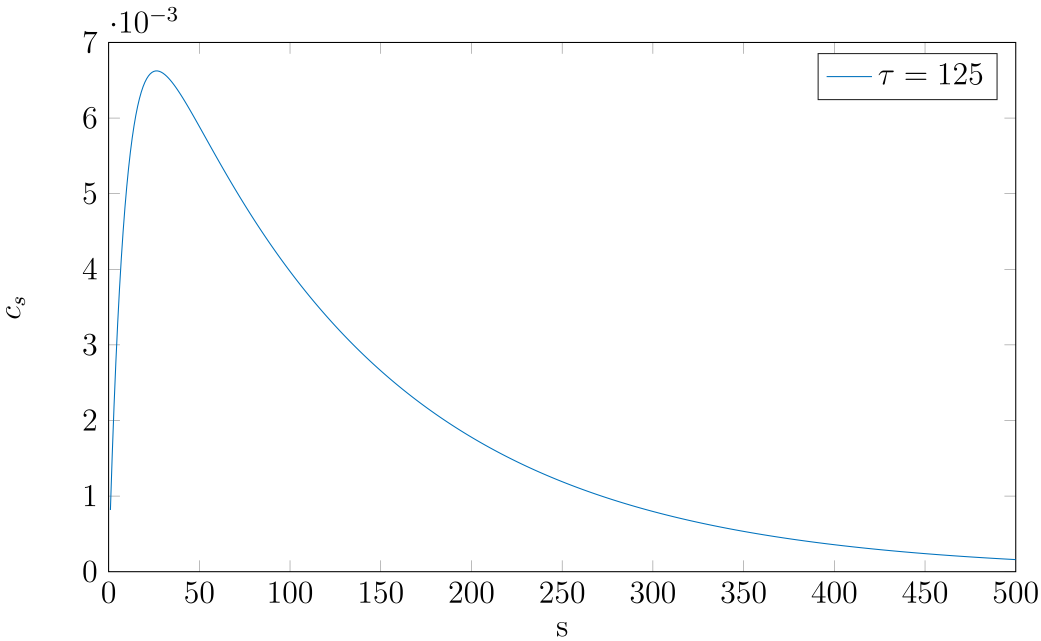

Figure 2.2: The coefficient value c_s placed on returns s time steps back in our considered trend following trading strategy, here with ”mean” $\tau = 125$.

3

Models

3.1 Mean-Variance portfolio

The foundation of modern portfolio theory was laid out by Harry Markowitz in his paper on Portfolio Selection in 1952 [6]. He proposed a model for constructing an optimal portfolio, where *optimal* is under the assumption that investors are rational and seek high returns with low risk.

Let $y \in \mathbb{R}^M$ be a random variable representing the normalized returns (from section 2.1) of the M markets in our universe. Further, let y have an arbitrary distribution but assume that its first and second moments are well defined and finite. That is

$$
\begin{aligned}
-\infty < \mathbb{E}[y_i] &< \infty \quad \forall i, \\
\mathbb{E}[y_i^2] &< \infty \quad \forall i.
\end{aligned}
$$

Consider constructing a portfolio consisting of these M assets, and let $w \in \mathbb{R}^M$ denote the positions vector. Then the random variable $y_p = y^T w$ is the portfolio return and its first two moments are computed as

$$
\begin{aligned}
\mathbb{E}[y_p] &= \sum_{i=1}^{M} w_i \mathbb{E}[y_i] = \mu^T w, \\
\mathbb{E}[y_p^2] &= \sum_{i=1}^{M} \sum_{i=1}^{M} w_i w_j \mathbb{E}[y_i y_j] = w^T \Sigma w,
\end{aligned}
\tag{3.1}
$$

where μ and Σ denote the expected returns and the correlation matrix respectively. In mean-variance (MV) analysis, only the first two moments are considered and the *risk* of a portfolio is quantified by the variance of returns which will be approximated by equation (3.1). The objective is to find an optimal trade-off between risk and return, in the sense that the portfolio weights w maximize the expected return for a given portfolio variance (propensity for risk), i.e. w solves the problem

$$
\begin{aligned}
\max_{w} \quad & \mu^T w \\
s.t. \quad & w^T \Sigma w \le \sigma_{TGT}^2.
\end{aligned}
\tag{3.2}
$$

With a linear objective function maximized over a convex set (Σ is positive definite), this problem is clearly convex. For $\sigma_{TGT}^2 > 0$, the Karush-Kuhn-Tucker (KKT) conditions are both necessary and sufficient to ensure global optimality [1, p. 142-143]. The KKT conditions for (3.2) become

$$\mu + \nu\Sigma w = \mathbf{0}^M, \tag{3.3}$$

$$\nu(w^T\Sigma w - \sigma_{TGT}^2) = 0, \tag{3.4}$$

$$\nu \geq 0,$$

where ν is the Lagrange multiplier. Now, equation (3.3) implies that

$$w^\star = \frac{1}{\nu}\Sigma^{-1}\mu, \tag{3.5}$$

with $\nu > 0$ and it is easy to see that equation (3.4) is satisfied by choosing the scale factor ν such that $w^T\Sigma w = \sigma_{TGT}^2$. The fact of importance is that the input parameter σ_{TGT}^2 only scales the weights equally and does not impact the relative position sizes (the strategy). In other words the solution given by (3.5) (with arbitrary constant ν) gives the interrelation between weights, and the final positions are proportional to this solution, scaled to satisfy the target volatility of the portfolio.

The MV portfolio strategy has been praised in theory but often criticised for poor performance in practice. First of all, there is the obvious problem of estimating the $(M^2-M)/2$ correlation parameters as well as M expected returns. Further, the solution in form of optimal weights is a function of the inverse correlation matrix and as discussed in section 2.2.1 the sample correlation matrix is generally ill-conditioned including close to zero eigenvalues. Because of this, directions corresponding to small eigenvalues will tend to be overrepresented in the optimal portfolio and small changes in the expected returns will cause large changes in the optimal weights resulting in extensive turnovers (and thus large trading costs).

To reduce both the exposure to estimation errors and the condition number of the sample correlation matrix, Σ in equation (3.5) will be replaced by the shrinkage estimator (the regularized matrix) discussed in section 2.2.1. As previously stated the absolute scaling of weights is of no interest in terms of strategy, hence the regularization factor $\lambda \in [0, 1]$ is the only manually set parameter and for the purpose of comparison with other models the weights are scaled to give the portfolio unit variance with respect to Σ.

3.2 Risk-Parity portfolio

Over the last few decades attention has shifted from the Mean-Variance framework balancing risk and expected returns towards models that mainly focus on diversification of risk [5]. The risk-parity (RP) portfolio falls under this category and aims at allocating risk in the sense that each asset should have equal contribution to the overall portfolio risk. Initially we will follow the reasoning and suggestions found in [8][10].

The risk σ_P of a portfolio is defined in the same way as the previous section through the standard deviation of portfolio returns, $\sigma_P = \sqrt{w^T \Sigma w}$. The marginal contribution to risk (MCR) for asset i can thus be obtained by

$$\text{MCR}_i = \frac{\partial \sigma_p}{\partial w_i} = \frac{(\Sigma w)_i}{\sqrt{w^T \Sigma w}}.$$

Since it holds that

$$\sum_{i=1}^{M} w_i \text{MCR}_i = \sum_{i=1}^{M} w_i \frac{(\Sigma w)_i}{\sqrt{w^T \Sigma w}} = \sqrt{w^T \Sigma w} = \sigma_P,$$

the quantity $w_i\text{MCR}_i$ can be interpreted as the contribution from asset i to the overall portfolio risk σ_P. As the objective of risk-parity is to let each asset contribute equally to the portfolio risk, we seek to find a set of weights $w \in \mathbb{R}^M$ such that

$$w_i \text{MCR}_i = \frac{\sigma_P}{M} \quad \forall i,$$

which is equivalent to

$$w_i (\Sigma w)_i = \frac{\sigma_P^2}{M} \quad \forall i. \tag{3.6}$$

Different from the MV objective, this system of equations is non-linear and does not have a closed form solution. Instead it will be shown that a position vector w satisfying the RP objective in equation (3.6) can be found by solving the following optimization problem

$$\begin{aligned} \min_{w} \quad & -\sum_{i=1}^{M} \log(w_i) \\ s.t \quad & w^T \Sigma w \leq \sigma_{TGT}^2, \end{aligned} \tag{3.7}$$

where σ_{TGT}^2 again is an input parameter. For now, assume that all weights are positive so the objective function is well defined. As Σ is a positive-definite matrix the constraint is clearly convex. Further, the objective function is a sum of convex functions and is thus also convex. With $\sigma_{TGT}^2 > 0$ the *Slater* constraint qualification holds and the KKT conditions are both necessary and sufficient for global optimality [1, p. 142-143]. The KKT system for (3.7) becomes

$$-w^{-1} + 2\nu \Sigma w = \mathbf{0}^M \tag{3.8}$$

$$\nu (w^T \Sigma w - \sigma_{TGT}^2) = 0 \tag{3.9}$$

$$\nu \geq 0,$$

where ν is a Lagrange multiplier. It is easy to see that (3.8) is equivalent to

$$w_i(\Sigma w)_i = \frac{1}{2\nu} = c \quad \forall i, \tag{3.10}$$

which is equivalent to the RP objective in equation (3.6) for some value of ν. Further, as $w^{-1} \neq 0$ equation (3.8) implies that $\nu > 0$, thus (3.9) implies that $w^T \Sigma w = \sigma_{TGT}^2$. Using equation (3.10) yields $N/(2\nu) = \sigma_{TGT}^2$, i.e. $\nu^\star = N/(2\sigma_{TGT}^2)$ which agrees with equation (3.6). These results also imply that a scaling of the input parameter σ_{TGT}^2 by a factor β only results in an equal scaling in the weights by the factor $\sqrt{\beta}$.

To summarize we have shown that for positive weights w the problem (3.7) is convex and thus have one and only one global optimum. We have also shown that at this optimum the risk-parity objective in equation (3.6) is satisfied.

3.2.1 Extension to negative weights

The RP optimization problem in equation (3.7) has an implicit constraint that all positions are kept positive. A slight modification to keep the objective function well defined for negative weights is to consider the problem

$$\begin{aligned} \min_{w} \quad & -\sum_{i=1}^{M} \log(|w_i|) \\ s.t \quad & w^T \Sigma w \leq \sigma_{TGT}^2. \end{aligned} \tag{3.11}$$

The objective function is now defined for negative weights, but it is no longer convex. It is however symmetric in each quadrant, i.e. for each individual set of signs on the M weights in our position vector w. As it was previously shown that the problem is convex for the case of positive weights, it follows that problem (3.11) is convex if a sign constraint is enforced on each weight w_i. This further implies that there is one portfolio (defined by w) that satisfies the RP objective in equation (3.6) for *each* possible set of signs, which for M different assets equals 2^M different portfolios that are equivalent in the risk-parity sense. With an increasing number of assets it quickly becomes infeasible to find them all. A suggested approach is to choose the RP portfolio which has the same signs on each asset as the expected returns, i.e. the signs suggested by the underlying trading strategy.

Assume that some expected return μ is provided from a trend following trading model and we want to find the RP portfolio w whit $\text{sign}(w_i) = \text{sign}(\mu_i)$, $\forall i$. Considering that the objective function of problem (3.11) is symmetric in each quadrant the only difference from the case of all positive weights is the variance constraint. In practice a convenient way of setting up this optimization problem is to rotate the correlation matrix Σ with elements $\Sigma_{i,j}$ according to

$$\hat{\Sigma}_{i,j} = \Sigma_{i,j}\, \text{sign}(\mu_i \mu_j),$$

and then solve the original problem (3.7) using positive weights. After obtaining a strictly positive solution $w^\star$ the only modification needed is set $w^\star$ to agree with the signs of μ. The drawback is of course that the weight signs need to specified beforehand.

3.2.2 Modified risk-parity

The risk-parity portfolio is as mentioned a model focusing on risk-allocation and not on expected returns. While this might be a favorable property for example in an index fund, for a CTA fund where research focus lies on building prediction models the portfolio optimization method should preferably make use of these predictions. The previously considered risk-parity model discards information in the form of the "strength" of a signal (prediction), only the directions are used.
We suggest an approach to include the signal strength by modifying the RP-objective in equation (3.6) according to

$$w_i(\Sigma w)_i = c|\mu_i| \quad \forall i, \tag{3.12}$$

thus letting asset i carry a larger contribution to the overall portfolio risk if the signal comes with a high level of confidence.

To find weights w that satisfy equation (3.12) the optimization problem in equation (3.7) needs to be modified. Taking the steps of the KKT-system backwards it is easy to see that the similar problem

$$\begin{aligned} \min_{w} \quad & -\sum_{i=1}^{M} |\mu_i| \log(w_i) \\ s.t \quad & w^T \Sigma w \leq \sigma_{TGT}^2 \end{aligned} \tag{3.13}$$

yields such a solution. This is still a convex problem with the same solution properties as (3.7) in terms of existence and uniqueness. In later evaluations this model will simply be referred to as the modified risk-parity portfolio (RPmod).

3.2.3 Lagrangian relaxation of the RP problem

We will make use of some Lagrangian duality theory in order to get closer to an unconstrained optimization problem. Define the Lagrangian function L for (3.7) as

$$L(w, \nu) = -\sum \log(w_i) + \nu(w^T \Sigma w - \sigma_{TGT}^2).$$

Then (for $\nu \geq 0$) the Lagrangian relaxation of (3.7) becomes

$$\min_{w} \quad L(w, \nu). \tag{3.14}$$

Note that this is in fact a relaxation as $L(w, \nu)$ is less than or equal to the objective function value in (3.7) for feasible w.
As (3.7) is a convex problem, $L(w, \nu)$ is also convex and one can show [1, p. 157-160] that there exists at least one Lagrangian multiplier $\nu^\star \geq 0$ such that

$$w^\star \text{ solves } (3.7) \quad \iff \begin{cases} w^\star \in \operatorname{argmin}_w L(w, \nu^\star) \\ \nu^\star(w^{\star T} \Sigma w^\star - \sigma_{TGT}^2) = 0 \\ w^{\star T} \Sigma w^\star - \sigma_{TGT}^2 \leq 0 \end{cases} \tag{3.15}$$

Further, as $L(w, \nu)$ is differentiable the condition that $w^\star \in \text{argmin}_w L(w, \nu^\star)$ is equivalent to $\nabla_w L(w^\star, \nu^\star) = 0$, and note that the right hand side of (3.15) reduces to the KKT conditions of the original problem (3.7). This can be used by recalling the calculations made below equation (3.10) where we found the explicit value $\nu^\star = N/(2\sigma_{TGT}^2)$. For this $\nu^\star$ it holds that $w^{\star T}\Sigma w^\star = \sigma_{TGT}^2$ where $w^\star = \text{argmin}_w L(w, \nu^\star)$. Hence (3.15) reduces to

$$w^\star \text{ solves (3.7)} \quad \iff \quad w^\star = \text{argmin}_w \, L(w, \nu^\star),$$

and thus the solution to the unconstrained, relaxed (convex) problem (3.14) with $\nu = \nu^\star$ also satisfies the original problem (3.7).
The same applies for the modified RP problem in the previous section. Similar calculations yield $\nu^\star = (\sum_i |\mu_i|)/(2\sigma_{TGT}^2)$. Note that the original RP problem is the special case $\mu_i = 1$ for all i.

3.2.4 Solving the RP optimization problem in practice

We focus on solving the Lagrangian relaxation of the modified RP problem, that is

$$\min_w \quad -\sum_{i=1}^{M} |\mu_i| \log(w_i) + \nu^\star w^T \Sigma w \tag{3.16}$$

and recall that with $\nu^\star = (\sum_i |\mu_i|)/(2\sigma_{TGT}^2)$ the solution to this problem satisfies (3.13) (let $|\mu_i| = 1$ for all i to get (3.7)).

An efficient way to solve this is to use the Alternating Direction Method of Multipliers (ADMM algorithm) that splits the problem into several smaller sub problems that are solved iteratively. The details of this algorithm can be found in [2] and we will follow the steps suggested by the authors. The ADMM algorithm is well suited for solving problems that can be split up on the form

$$\begin{aligned} \min_{x,z} \quad & f(x) + g(z) \\ s.t \quad & Ax + Bz = c, \end{aligned}$$

where typically the variables x and z comes from splitting the original objective variable into two separable parts. Note that we can formulate (3.16) as

$$\begin{aligned} \min_{x,z} \quad & -\sum_{i=1}^{M} |\mu_i| \log(x_i) + \nu^\star z^T \Sigma z \\ s.t \quad & x - z = 0, \end{aligned} \tag{3.17}$$

where the variable w has been split into two parts and at an optimal solution it holds that $x^\star = z^\star = w^\star$.
In the previous section we defined the Lagrangian function for our problem. The *augmented* Lagrangian for (3.17) is defined as

$$L_\rho(x, z, y) = -\sum_{i=1}^{M} |\mu_i| \log(x_i) + \nu^\star z^T \Sigma z + y(x - z) + \frac{\rho}{2} ||x - z||_2^2 \tag{3.18}$$

where y is the Lagrange multiplier and ρ is a penalty coefficient. The last term is the penalty term and is the addition to the ordinary Lagrangian. The penalty term is supposed to give robustness to the procedure of iterating towards a solution [2]. Equation (3.18) can be rewritten in scaled form as

$$L_\rho(x,z,y) = -\sum_{i=1}^{M} |\mu_i| \log(x_i) + \nu^\star z^T \Sigma z + \frac{\rho}{2}||x - z + u||_2^2 + \frac{\rho}{2}||u||_2^2$$

where $u = (1/\rho)y$ is the scaled dual variable. From here, the iterative scheme of the ADMM algorithm can be formulated as

$$x^{k+1} = \arg\min_x \left(-\sum_i |\mu_i| \log(x_i) + \frac{\rho}{2}||x - z^k + u^k||_2^2 \right) \tag{3.19a}$$

$$z^{k+1} = \arg\min_z \left(\nu^\star z^T \Sigma z + \frac{\rho}{2}||x^{k+1} - z + u^k||_2^2 \right) \tag{3.19b}$$

$$u^{k+1} = u^k + (x^{k+1} - z^{k+1})$$

3.2.4.1 The x-update

To solve (3.19a), note that we can separate each x_i and solve M independent one-dimensional problems as it holds that

$$\begin{aligned} &\min_x \quad \sum_i \left(-|\mu_i| \log(x_i) + \frac{\rho}{2}(x_i - z_i^k + u_i^k)^2 \right) \\ &= \sum_i \min_{x_i} \left(-|\mu_i| \log x_i + \frac{\rho}{2}(x_i - z_i^k + u_i^k)^2 \right). \end{aligned}$$

For each x_i this is a problem in one variable and we find the minimum by setting the derivative to zero,

$$\begin{aligned} &\frac{\partial}{\partial x_i} \left(-|\mu_i| \log x_i + \frac{\rho}{2}(x_i - z_i^k + u_i^k)^2 \right) = 0 \\ &\Rightarrow \quad -\frac{|\mu_i|}{x_i} + \rho(x_i - b_i^k) = 0 \quad \Rightarrow \quad x_i^2 - b_i^k x_i - \frac{|\mu_i|}{\rho} = 0, \end{aligned}$$

which has solutions

$$x_i^\star = \frac{1}{2}\left(b_i^k \pm \sqrt{(b_i^k)^2 + \frac{4|\mu_i|}{\rho}} \right),$$

where $b^k = z^k - u^k$. Note that one of these solution is always positive and the other negative (assuming $|\mu_i| > 0$). Since there is an implicit constraint of $w > 0$ in (3.16), an convenient way to ensure this is to always choose positive solution $x_i^{\star +}$.

3.2.4.2 The z-update

The z-update is not as easily separable so we will keep the vector notation. Our objective is to solve (3.19b), which is a quadratic function and the minimum is found by setting the gradient to zero. Letting $b = x^{k+1} + u^k$ yields

$$\begin{aligned} \nabla_z \left(\nu^\star z^T \Sigma z + \frac{\rho}{2} ||z - b||_2^2 \right) = \mathbf{0} \quad &\Rightarrow \quad 2\nu^\star \Sigma z + \rho(z - b) = \mathbf{0} \\ \Rightarrow \quad \left[2\nu^\star \Sigma + \rho I \right] z = \rho b \quad &\Rightarrow \quad z^\star = \rho \left[2\nu^\star \Sigma + \rho I \right]^{-1} b \end{aligned}$$

Thus we need to solve a linear system of equations in each iteration. This procedure can however be simplified by decomposing the matrix Σ *once* into the matrices H and D such that $HDH^T = \Sigma$, i.e. the columns of H contains the orthonormal eigenvectors of Σ and D is a diagonal matrix of the corresponding eigenvalues. Then

$$z^{k+1} = \rho H \left[2\mu^\star D + \rho I \right]^{-1} H^T (x^{k+1} + u^k),$$

where the matrix to be inverted is now a diagonal matrix. Hence the only computations needed in each iteration are a few matrix multiplications.

3.2.5 Exploring different RP portfolios

The risk-parity method was initially presented as long only method (only positive weights) but was extended to deal with negative weights. In doing so, the optimization problem for which the optimal solution is a RP portfolio became non-convex. Recall that in a setup with M different markets, there are 2^M portfolios that satisfy the RP objective in each time step, i.e. one solution for each set of weight signs (each quadrant). The question becomes which of these is the best choice. They are all equal in the risk-parity sense and hence some additional criterion to compare portfolios needs to be introduced.

To include a bit of the mean-variance objective, consider searching for the RP portfolio with the highest expected return. There is however no convex optimization problem that finds this portfolio. Instead, we will use the following greedy iteration scheme

1. Start from the signs suggested by μ,
2. Change the sign in each market one by one and solve for the respective RP portfolio,
3. If no improvement on expected return of the portfolio is found, stop. Otherwise choose the sign change that gave the highest expected portfolio return and go back to 2.

3.3 Minimum conditional value-at-risk portfolio

This model, also called Least Expected Shortfall (LES), is built around a different risk measure; the *value-at-risk*. The value-at-risk, or VaR_β, of a portfolio connected to a confidence level β is informally defined as the smallest amount α such that the probability of a loss larger than α is smaller than or equal to $1-\beta$. The *conditional value-at-risk*, CVaR_β, is then defined as the average loss given that the loss exceeds this α.

In our universe of M assets, assume that the the random variable $y \in \mathbb{R}^M$ represents the returns of all assets in some time step. Given that our portfolio is defined by a set of weights $w \in \mathbb{R}^M$, the portfolio loss corresponding to these weights is $-w^T y$. If we further assume that y has a density $p(y)$, the CVaR_β of this portfolio can be defined as

$$\mathrm{CVaR}_\beta(w) = \frac{1}{1-\beta} \int_{-y^T w \geq \mathrm{VaR}_\beta} (-y^T w) p(y) dy. \tag{3.20}$$

An optimization problem similar to that of mean-variance finding optimal weights based on this risk measure would be on the form

$$\begin{aligned} \max_{w} \quad & \mu^T w \\ s.t. \quad & \mathrm{CVaR}_\beta(w) \leq C, \end{aligned}$$

where μ are the expected returns and C is a constant. But we will follow the steps taken by Rockafellar and Uryasev [9], where focus lies on the dual of problem 3.3 and thus construct a portfolio that minimizes the CVaR for a given β, density and a target return. As equation (3.20) depends on the value-at-risk, this appears to be a complex problem. But it has been shown in [9] that if we define F_β as

$$F_\beta(w, \alpha) = \alpha + \frac{1}{1-\beta} \int_{\mathbb{R}} (y^T w + \alpha)^- p(y) dy, \tag{3.21}$$

where $(x)^- = -\min(0, x)$, it holds that

$$\min_{w} \mathrm{CVaR}_\beta(w) = \min_{w,\alpha} F_\beta(w, \alpha),$$

where the right hand side is a convex problem. Thus the problem to solve for optimal weights is

$$\begin{aligned} \min_{w,\alpha} \quad & F_\beta(w, \alpha) \\ s.t. \quad & \mu^T w \geq \mu_{TGT}. \end{aligned} \tag{3.22}$$

The density $p(y)$ included in equation (3.21) is however unknown. To deal with this, it has been suggested in [9] to approximate F_β using q samples y_k from $p(y)$ taken as historical data. Define

$$\tilde{F}_\beta(w, \alpha) = \alpha + \frac{1}{q(1-\beta)} \sum_{k=1}^{q} (y_k^T w + \alpha)^-,$$

where $y_k \in \mathbb{R}^M$ are historical returns. The approximation of problem (3.22) using $\tilde{F}_\beta(w, \alpha)$ can now be formulated as a linear program on the form

$$\begin{aligned} \min_{u,\alpha,w} \quad & \alpha + \frac{1}{q(1-\beta)} \sum_{k=1}^{q} u_k \\ \text{subject to} \quad & y_k^T w + \alpha + u_k \geq 0 \quad \text{for } k = 1, \dots, q \\ & u_k \geq 0 \quad \text{for } k = 1, \dots, q \\ & \mu^T w \geq \mu_{TGT}, \end{aligned} \tag{3.23}$$

where u_k are auxiliary variables, μ_{TGT} is the target return for the portfolio and μ contains the expected returns for each asset. For both mean-variance and the risk-parity portfolios we found that the input parameter of a target variance (σ^2_{TGT}) only scaled all weights equally. This also holds for μ_{TGT} in problem (3.23). Note that it is possible to scale the problem (3.23) with a factor C as

$$\begin{aligned} \min_{u,\alpha,w} \quad & C\left(\alpha + \frac{1}{q(1-\beta)} \sum_{k=1}^{q} u_k\right) \\ \text{subject to} \quad & C\left(y_k^T w + \alpha + u_k\right) \geq 0 \quad \text{for } k = 1, \dots, q \\ & Cu_k \geq 0 \quad \text{for } k = 1, \dots, q \\ & C\mu^T w \geq C\mu_{TGT}, \end{aligned}$$

is an equivalent problem with the same solution as (3.23). Thus it is easy to see that scaling μ_{TGT} with a factor C will only result in all variables (w, α, u) being scaled with the same factor.

3.3.1 Extensions

The original model in equation (3.23) solving an approximation of the problem to minimize the conditional value at risk has shown poor results in practice. This section will walk through suggested steps to tackle issues including sensitivity to the time horizon q, large turnovers and huge spread positions.

3.3.1.1 Using a fixed α

The first step will be to drop the direct connection to the definition of conditional value at risk. Instead of minimizing the mean of historical portfolio returns below the value at risk for some given confidence level, consider setting a fixed value for α and thus removing one variable from the optimization problem. This problem would be to find a portfolio that solves

$$\begin{aligned} \min_{w} \quad & \sum_{k=1}^{q} c_k (y_k^T w + \alpha)^- \\ \text{s.t} \quad & \mu^T w \geq \mu_{TGT}, \end{aligned} \tag{3.24}$$

where c_k can be taken as a simple mean ($c_k = 1/q$) or be decreasing in time, i.e. old data would carry less weight. Note that this formulation is similar to the dual of the mean-variance problem but with a different risk measure (the sum in equation (3.24)).

Choosing a positive α in problem (3.24), the objective becomes to minimize the average of portfolio losses larger than α. If instead α would be negative, the aim moves towards pushing all historical portfolio returns towards some target return defined by α.

There are a few things to consider in order for (3.24) to be a well formulated problem and useful in practice,

- Without constraints on variance the norm of the solution is not really controlled, meaning that the linear constraints might be easy to manipulate and the resulting portfolio may be very far from the positions suggested by the trading model (μ).
- Only considering extreme events (as intended with CVaR) will lead to a small number of observations having an impact on the objective function and the result will be sparse portfolios since $w = 0$ is an attractor (optimum of the unconstrained problem).
- The terms $(y_k^T w + \alpha)$ and $(\mu^T w - C)$ are assumed to carry the same meaning, i.e. be comparable in some way. Thus μ and y_k should have at least similar norms. Further, the values on α and C should preferably be connected.

3.3.1.2 Regularization

The potential issues laid out in the previous section need to be addressed. Considering that the linear program basically tries to fit a portfolio to historical data, this problem might be prone to overfitting causing poor performance in practice. To regularize the problem in equation (3.24), consider the following reformulation for $\lambda > 0$,

$$\begin{aligned} \min_{w} \quad & \lambda w^T w + \sum_{k=1}^{M} c_k (y_k^T w + \alpha)^- \\ s.t \quad & \mu^T w \geq ||\mu||. \end{aligned} \tag{3.25}$$

A penalty on the norm of the position vector is introduced to avoid manipulation of the linear constraints. Further, the target return has been set to the norm of the expected returns μ. This will have the effect that when $\lambda \to \infty$, the solution $w^\star$ approaches μ and when $\lambda \to 0$, $w^\star$ approaches the solution to the previously considered problem in equation (3.24). Hence λ is a way to control how far the solution can go in directions agreeing well with historical returns, and also control the angle between the vectors μ and $w^\star$. This is intended to reduce overfitting to the input data.

To be a well formulated problem, α should be set in relation to $||\mu||$. Moreover, the norms of μ and y_k should preferably be of the same order of magnitude. For simplicity consider letting $||\mu|| = ||y_k|| = 1$. Then, both $\mu^T w$ and $y_k^T w \in [-||w||, ||w||]$.

As $||w|| \geq 1$, reasonable values for α could be $\alpha \in [-1, 1]$ or larger depending on λ. The parameter λ will cause the optimal solution to over time have some average norm $\overline{||w||}$ where deviations from $||\mu||$ suggest how large role the risk term (the sum in equation (3.25)) plays in determining the optimum. The impact of this terms will vary over time depending on how well μ agrees with the historical returns y_k.

The problem of fixating a value on α remains. While the new penalty term on the norm in equation (3.25) might help to avoid sparse portfolios, the value on α should not be set to high. To guarantee that enough data points is used during the optimization, α could be assigned a negative value. Then in order to keep the model within the class of "tail risk optimizers", while letting α be negative, a suggested extension is to put a power on the terms in the sum of equation (3.25) and thus place a larger penalty on losses (that increases non-linearly with the size of the loss).

3.3.1.3 Piecewise linear penalty

To approximate the penalty on the terms in the sum of equation (3.25) with a power larger than one, consider a sum of penalties for different values of α,

$$
\begin{aligned}
\min_{w} \quad & \lambda w^T w + \sum_{i=1}^{n} d_i \sum_{k=1}^{q} c_k (\hat{y}_k^T w + \alpha_i)^- \\
s.t \quad & \hat{\mu}^T w \geq 1,
\end{aligned}
\tag{3.26}
$$

where $\{\alpha_i\}_{i=1}^n$ is an increasing sequence and $\alpha_1 < 0$ (here $||\hat{\mu}|| = ||\hat{y}_k|| = 1$). The coefficients d_i will be used to control which power is approximated and with this approach the need to specify the parameter α directly is eliminated. Instead, specify which power to put on the deviations of $y_k^T w$ from something close to its maximum value (which will depend on λ). The coefficients d_i can be obtained though regression after choosing a suitable power with a given sequence of α's, in later presented results the power $(.)^{\frac{3}{2}}$ has been used.

3.3.1.4 Efficient solver

The optimization problem in equation (3.26) can be formulated as a quadratic program and there is a variety of standard software that deals with problems of this kind. However, (3.26) becomes very computationally expensive to solve seeing as with around one year of historical data together with a few different values on α the quadratic program formulation will include over a thousand variables.

In this setting many of the solvers in standard software become inefficient to use in practice, thus we suggest an ADMM formulation with faster convergence.

The unconstrained formulation of problem (3.26) is

$$
\min_{w} \quad \lambda w^T w + \sum_{i=1}^{n} d_i \sum_{k=1}^{q} c_k (y_k^T w + \alpha_i)^- + I_\infty(\mu^T w \geq 1) \tag{3.27}
$$

where $I_\infty(.)$ is an indicator function taking the value zero if the condition is true and taking the value of infinity otherwise. To simplify notation below, let $Y' \in \mathbb{R}^{q\times M}$ denote the historical data of returns (q observations of M assets) and let $Y \in \mathbb{R}^{nq\times M}$ contain n copies of Y'. Further, let

$$b = \begin{bmatrix} \left.\begin{matrix}\alpha_1 \\ \alpha_1 \\ \vdots \\ \alpha_1\end{matrix}\right\} q \\ \alpha_2 \\ \vdots \\ \vdots \\ \alpha_n \end{bmatrix} \quad \text{and} \quad f = \begin{bmatrix} c_1 d_1 \\ c_2 d_1 \\ \vdots \\ c_q d_1 \\ c_1 d_2 \\ \vdots \\ \vdots \\ c_q d_n \end{bmatrix},$$

hence both b and $f \in \mathbb{R}^{nq}$. Now problem (3.27) can be rewritten on ADMM form as

$$\begin{aligned} \min_{x,z} \quad & \lambda x^T x + \sum_{k=1}^{nq} f_k(z_k + b_k)^- + I_\infty(\mu^T w \geq 1) \\ s.t \quad & Yx - z = 0, \end{aligned} \tag{3.28}$$

where $z \in \mathbb{R}^{nq}$. Taking the steps of the ADMM formulation as laid out in [2], the augmented Lagrangian of problem (3.28) on scaled form is defined as

$$L_\rho = \lambda x^T x + \sum_{k=1}^{nq} f_k(z_k + b_k)^- + I_\infty(\mu^T w \geq 1) + \frac{\rho}{2}||Yx - z + u||_2^2,$$

where u is the scaled dual variable. Now the iterative scheme of the ADMM algorithm can be formulated as

$$x^{i+1} = \arg\min_x \left(\lambda x^T x + I_\infty(\mu^T w \geq 1) + \frac{\rho}{2}||Yx - z^i + u^i||_2^2\right) \tag{3.29a}$$

$$z^{i+1} = \arg\min_z \left(\sum_{k=1}^{nq} f_k(z_k + b_k)^- + \frac{\rho}{2}||Yx^i - z + u^i||_2^2\right)$$

$$u^{i+1} = u^i + (Yx^{i+1} - z^{i+1}).$$

The x-update

The objective is to solve problem (3.29a), which has the constraint $\mu^T w \geq 1$. With a Lagrangian relaxation of this problem, the x-update is taken as the minimizer of

$$L_x = \lambda x^T x + \nu(1 - \mu^T w) + \frac{\rho}{2}||Yx - z + u||_2^2,$$

where $\nu \geq 0$ is the Lagrange multiplier. Note that this is a quadratic problem in x, hence the minimum is found by setting the gradient to zero.

$$\begin{aligned} \nabla_x L_x &= 2\lambda x + \rho Y^T(Yx - z + u) - \nu\mu = 0 \\ &\Rightarrow (2\lambda I + \rho Y^T Y)x = \rho Y^T(z-u) + \nu\mu \\ &\Rightarrow x^\star = g_0 + \nu g_1, \end{aligned} \tag{3.30}$$

where

$$\begin{aligned} g_0 &= \rho(2\lambda I + \rho Y^T Y)^{-1}(Y^T(z-u)) \\ g_1 &= \nu(2\lambda I + \rho Y^T Y)^{-1}\mu. \end{aligned} \tag{3.31}$$

To satisfy the explicit constraint of $\mu^T w \geq 1$, note that with the obtained solution $x^\star$,

$$\mu^T x^\star = \mu^T g_0 + \nu \mu^T g_1,$$

and if we can show that $\mu^T g_1$ always is positive, then this constraint can be satisfied by increasing ν (if necessary). To see that this holds, note that

$$\mu^T g_1 = \mu^T(2\lambda I + \rho Y^T Y)^{-1}\mu \; > 0,$$

assuming that either Y has full column rank (the matrix $Y^T Y$ is positive definite) *or* that $\lambda > 0$. To summarize, the x-update $x^\star$ is given by equations (3.30) and (3.31) where $\nu \geq 0$ is taken as the smallest value such that $\mu^T x^\star \geq 1$. Note that the matrix inversion in each iteration can be avoided by decomposing the matrix $Y^T Y$ once, thus reducing the time complexity using a similar approach as discussed in section 3.2.4.2.

The z-update

The z-update is taken as the minimizer of

$$L_z = \sum_{k=1}^{nq} f_k(z_k + b_k)^- + \frac{\rho}{2}||Yx - z + u||_2^2.$$

Letting $r = Yx + u$, this can be written as

$$L_z = \sum_{k=1}^{nq}\left[f_k(z_k + b_k)^- + \frac{\rho}{2}(z_k - r_k)^2\right],$$

which shows a convenient way of solving nq independent one-dimensional problems

$$\min_{z_k} \quad f_k(z_k + b_k)^- + \frac{\rho}{2}(z_k - r_k)^2. \tag{3.32}$$

Recall that $(t)^- = -\min(0, t)$; thus three possibilities need to be considered.

Case $\boldsymbol{z_k^\star < -b_k}$
The solution is found from

$$\begin{aligned} &\frac{d}{dz_k}\left[-f_k(z_k + b_k) + \frac{\rho}{2}(z_k - r_k)^2\right] = 0 \\ &\Rightarrow z_k^\star = \frac{f_k}{\rho} + r_k, \end{aligned}$$

which is only a valid candidate if $z_k^\star < -b_k$.

Case $\boldsymbol{z_k^\star > -b_k}$
Here,

$$\begin{aligned} &\frac{d}{dz_k}\left[\frac{\rho}{2}(z_k - r_k)^2\right] = 0 \\ &\Rightarrow z_k^\star = r_k, \end{aligned}$$

which is only a valid candidate if $z_k^\star > -b_k$. The third case if of course that $z_k^\star = b_k$. The one of the valid candidates that minimizes equation (3.32) is taken as the z_k-update.

4

Methods

4.1 Choice of filter parameters

4.1.1 Variance estimate

Section 2.1 discussed normalizing the time series of asset returns using estimates $\hat{\sigma}_t^2$ for each asset. As described in equation (2.2), this estimate is based on an exponential moving average (EMA) with a parameter τ that roughly describes how many days back in time the average weight is placed. Figure 4.1 shows the realized average variance of the normalized return series for different values of τ; recalling the objective of unit variance reasonable values of τ appears to be between 15 and 60.

4.1.2 Correlation estimate

Recalling section 2.2, the correlation matrix Σ will be estimated using EMAs taken on daily observations $dZ_t dZ_t^T \in \mathbb{R}^{M \times M}$, i.e.

$$\Sigma_t = \sum_{s=1}^{t-1} c_s dZ_{t-s} dZ_{t-s}^T.$$

The coefficients in a single EMA may however decrease to fast (Σ contains a lot of parameters) and it is possible to explore variations such as a double EMA (data filtered recursively), or a combination of these. An example of the filter coefficients behavior for different options is shown in figure 4.2 and the choice between these will depend on how fast the correlation estimate should include new information and forget old data.

All results presented in the next chapter is based on a correlation matrix constructed as the average of a single and a double EMA, several values of the filter parameter τ are tested.

4.2 Model evaluation

All methods for portfolio selection have been evaluated on the set of historical data from section 2.1, using the same underlying trading strategies. All strategies are trend following models generating expected returns $\mu_t \in \mathbb{R}^M$ for each time step t and recall that daily time steps is considered. For practical reasons, it is assumed

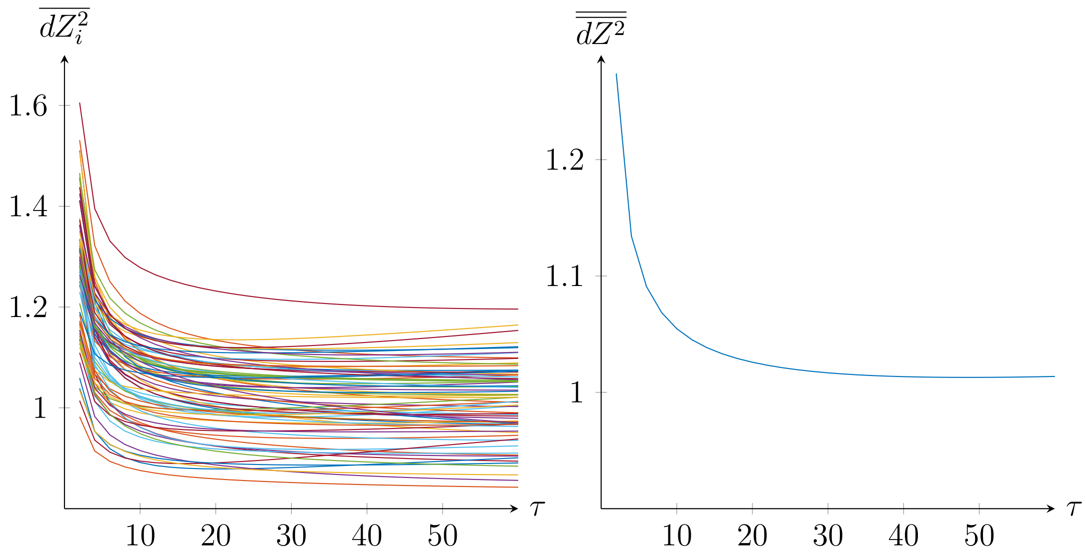

Figure 4.1: Guidance in how to choose the filter parameter τ for the variance estimate. Left: Average variance per market. Right: Average variance over all markets.

that information received at time t cannot be used until time $t+1$. That is, based on information at time t the portfolio model calculates optimal positions w, and the next day (time $t+1$) we are able to obtain these positions by trading on the opening price. The unit of standard deviations is kept when calculating revenues making it possible to sum results over different assets that are traded in different currencies.

Sharpe ratio

The main key value that will be used to compare portfolio strategies is the Sharpe ratio of portfolio returns. Let R_t, $t = 1, \ldots, T$, denote the resulting revenue at time t after implementing a portfolio strategy. The Sharpe ratio measures the return (revenue) per unit of risk (standard deviation of returns), and is simply defined as the ratio between the mean and the standard deviation of the time series R_t. A common measure of reference is the annualized Sharpe ratio obtained by multiplying with the factor $\sqrt{252}$ as there is roughly 252 trading days per year. That is,

$$\text{Annualized Sharpe ratio} = \frac{\text{Mean}(R_t)}{\text{Std}(R_t)} \frac{252}{\sqrt{252}}.$$

Drawdown

A complement to the Sharpe ratio will be to look at the drawdown properties of a portfolio strategy, where a drawdown refers to a time period where the equity curve is below the ”all time high”. For a managed fund a good selling point to present to potential investors would be small and short drawdowns. The equity curve $E(t)$ is the cumulative sum of portfolio returns R_t and to simplify comparisons it will be scaled to have an annualized standard deviation equal to 1,

$$E(t) = \frac{1}{\sqrt{252} \cdot \text{Std}(R_t)} \sum_{\tau=1}^{t} R_\tau.$$

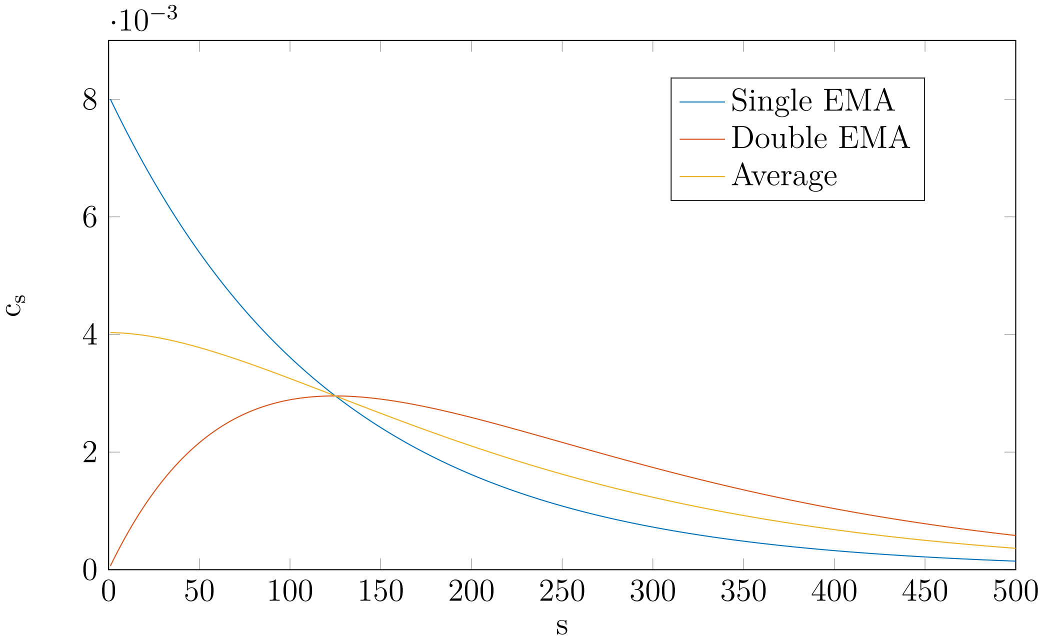

Figure 4.2: A visualization of the filter coefficient as a function of s, the number of time steps back. This example is for a filter "mean" of $\tau = 125$.

The drawdown series $D(t)$ is then defined as

$$D(t) = E(t) - \max_{\tau \leq t} E(\tau),$$

note that $D(t) \leq 0$ and has the unit of annualized standard deviations. The quantity that will be compared is the average of $D(t)$.

Holding time

The final key value in evaluations is the holding time. This is meant to give a sense of how fast a portfolio strategy changes its positions (weights w). A large holding time indicates that positions are held over a longer period of time and will thus result in lower trading costs. Let

$$\text{Holding time} = 2\frac{\sum_{i=1}^{M}\sum_{t=1}^{T}|w_{t,i}|}{\sum_{i=1}^{M}\sum_{t=1}^{T}|w_{t,i} - w_{t-1,i}|}. \tag{4.1}$$

5

Results

5.1 Comparing MV and RP

This section will show a comparison between mean-variance and the risk-parity portfolios, evaluating the key values described in section 4.2 and their sensitivity to the input parameters. This is a natural setup since these methods are more alike, they use the same measure of risk and takes the same input. Results will be based on *one* underlying trading strategy, namely the trend following model described in section 2.3 with $\tau = 200$, which is considered to be following the long trend. TF will be the abbreviation of the trend following model without portfolio optimization and recall that RPmod is the modified risk-parity method from section 3.2.2.

To give an indication of how sensitive the portfolio models performance is to the input, the evaluation will be made for several randomly drawn pairs of input parameters. Recall from section 4.1.1, that for the filter parameter τ estimating the variance of price series using the Yang-Zhang method, reasonable values was found to be between 15 and 60. Hence this parameter will be drawn uniformly in this range, $yz_\tau \in [15, 60]$. The estimate of the correlation matrix has a similar filter parameter τ. Since this estimate includes considerably more parameters a slower filter will be used. This parameter will be uniformly drawn as $\Sigma_\tau \in [80, 180]$. For each randomly drawn pair of parameters, the portfolio methods will be evaluated for several degrees of regularization (denoted $\lambda \in [0, 1]$) of the correlation matrix Σ, see section 2.2.1.

Figure 5.1 shows the comparison of the Sharpe ratio. For each portfolio method, the shaded area contains the mean $\pm 2\sigma$ of results from 1000 randomly drawn pairs of parameters. For the trend following model (TF) without portfolio optimization only the mean is shown. But as the parameter yz_τ also effects the result of TF, this variation is included in results for the portfolio strategies. A better measure for comparison purposes may therefore be the *marginal* Sharpe, i.e. the difference in Sharpe from applying the portfolio strategy. This is shown in figure 5.2 and gives a clearer picture of the portfolio models contributions. It is clear that the modified risk-parity model outperforms both MV and RP which show similar results. RP-mod shows a smaller band width, less sensitivity to an ill-conditioned correlation matrix and a significantly higher best case (for the λ resulting in the highest Sharpe).

Figure 5.3 shows the corresponding results for the average drawdown. Here both

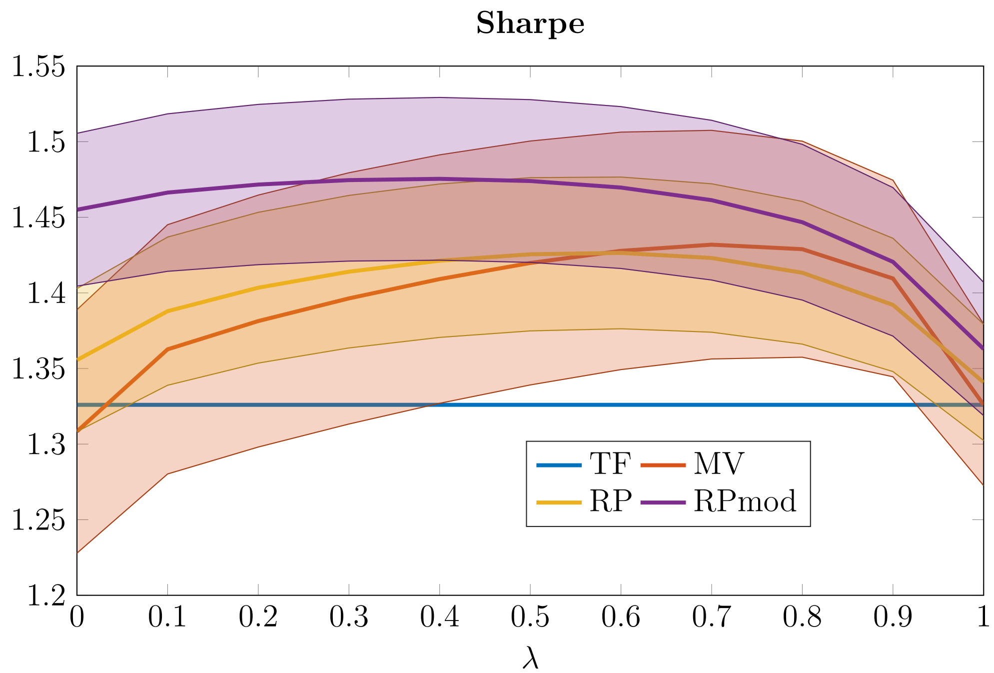

Figure 5.1: Trend following filter using $\tau = 200$. The shaded area contains the mean $\pm 2\sigma$ based on random parameters $yz_\tau \in [15, 60]$ and $\Sigma_\tau \in [80, 180]$. The results are shown as a function of the regularization factor λ.

RP models show superior performance to the MV portfolio, and appear to be much less sensitive to an ill-conditioned correlation matrix. The modified risk-parity again shows the lowest sensitivity to input parameters in the form of a tight band width.

The resulting holding times as defined in equation (4.1) is shown in figure 5.4. All models exhibit similar dependence on the regularization factor λ, but that a lower condition number on Σ leads to smaller variations in portfolio positions should come as no surprise. The original RP naturally has a lower holding period since it switches positions in a binary way when μ changes sign. Note that when comparing MV and RPmod it must be taken into account that RPmod showed better performance for lower values on λ relative to MV, where it might take on larger trading costs. The impact of trading costs will be further discussed in coming sections.

5.1.1 Factors that have an impact on results

Results over the full 35 year period do however not tell the whole story. The question becomes when and why does one portfolio model perform better? To get an idea of how the models differ in portfolio positions, consider the objective differences for the very simplified case of full regularization of the correlation matrix,

$$\begin{array}{lllll} \text{MV:} & (\Sigma w)_i = c\mu_i & & & w = \mu \\ \text{RP:} & w_i(\Sigma w)_i = c & \Rightarrow \{\Sigma = I\} \Rightarrow & & w = \text{sign}(\mu) \\ \text{RPmod:} & w_i(\Sigma w)_i = c|\mu_i| & & & w = \text{sign}(\mu)\sqrt{|\mu|}. \end{array} \tag{5.1}$$

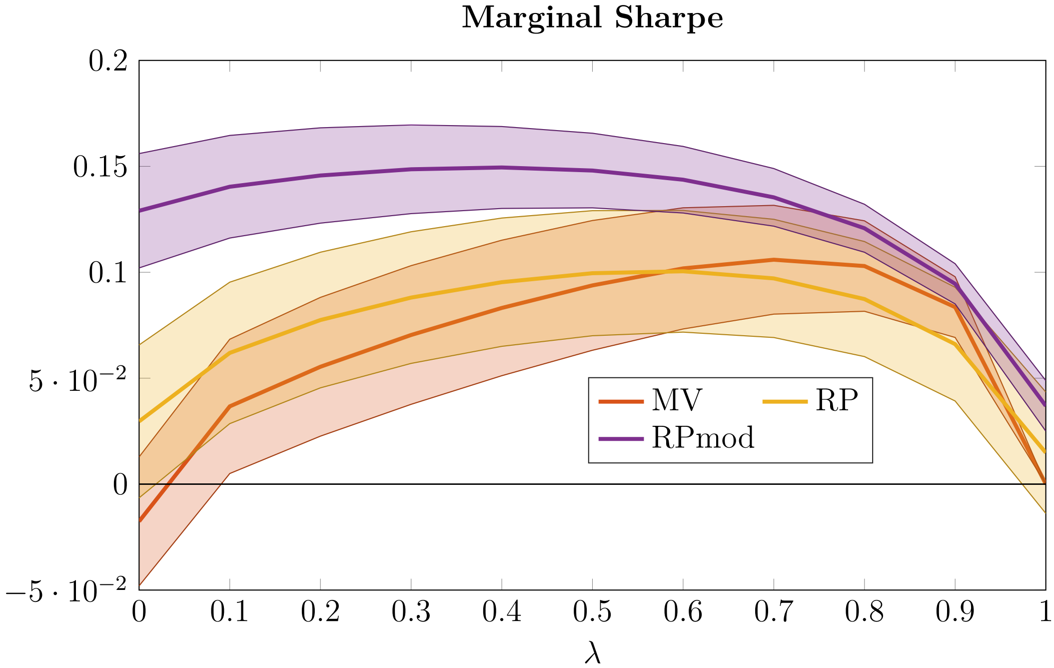

Figure 5.2: Trend following filter using $\tau = 200$. The shaded area contains the mean $\pm 2\sigma$ based on random parameters $yz_\tau \in [15, 60]$ and $\Sigma_\tau \in [80, 180]$. The results are shown as a function of the regularization factor λ.

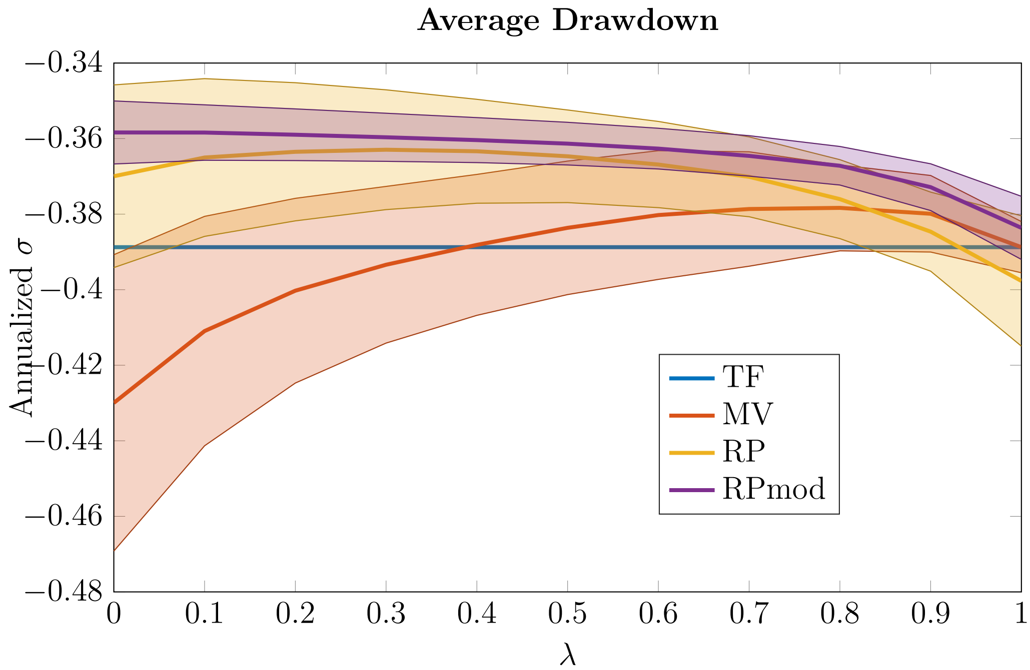

Figure 5.3: Trend following filter using $\tau = 200$. The shaded area contains the mean $\pm 2\sigma$ based on random parameters $yz_\tau \in [15, 60]$ and $\Sigma_\tau \in [80, 180]$. The results are shown as a function of the regularization factor λ.

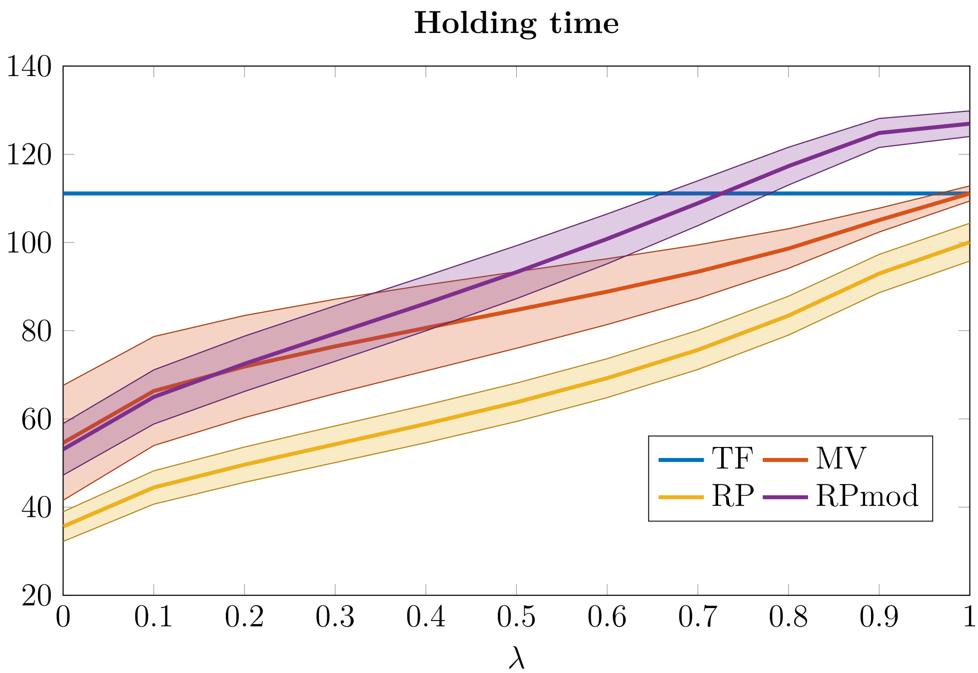

Figure 5.4: Trend following filter using $\tau = 200$. The shaded area contains the mean $\pm 2\sigma$ based on random parameters $yz_\tau \in [15, 60]$ and $\Sigma_\tau \in [80, 180]$. The results are shown as a function of the regularization factor λ.

Equation (5.1) shows that if ignoring correlations, the positions of the MV portfolio is taken as the expected returns μ. The RP model becomes a binary version of the underlying trend follower while the modified RP is something in between MV and RP. RPmod "shrinks" the expected returns while keeping the signs and could be considered to place less confidence in the signal strength $|\mu|$. A reasonable hypothesis of when the RP methods outperforms MV in terms of Sharpe could therefore be when the signal strength is a poor estimate of future price movements. This could be quantified as when a binary version of the prediction model outperforms the original model (TF) in terms of higher Sharpe. Comparing RPmod and MV, figure 5.5 shows the difference in Sharpe between RPmod and MV plotted against the Sharpe difference between a binary version of TF and the original TF (trading sign (μ) instead of μ). The two quantities show a high degree of correlation as indicated by the correlation coefficient $\rho = 0.875$. Further investigation however reveals a parameter dependence for how the Sharpe of RPmod compare to that of MV. The colormap of figure 5.5 shows the filter parameter yz_τ in each evaluated pair of parameters which clearly has an impact on these quantities. Figure 5.6 shows how the marginal Sharpe of the respective model depend on this parameter and suggests that MV has a linear dependence while RPmod appears less sensitive.

Keeping the hypothesis of when RPmod outperforms MV but reducing the impact of input parameters, we will look at the rolling Sharpe over time. Figure 5.7 shows the difference in rolling Sharpe one year back in time for the same quantities considered in figure 5.5. These results are for the same underlying TF model, with parameters $yz_\tau = 30$, $\Sigma_\tau = 100$ and (for MV and RPmod) the regularization factor λ that gave

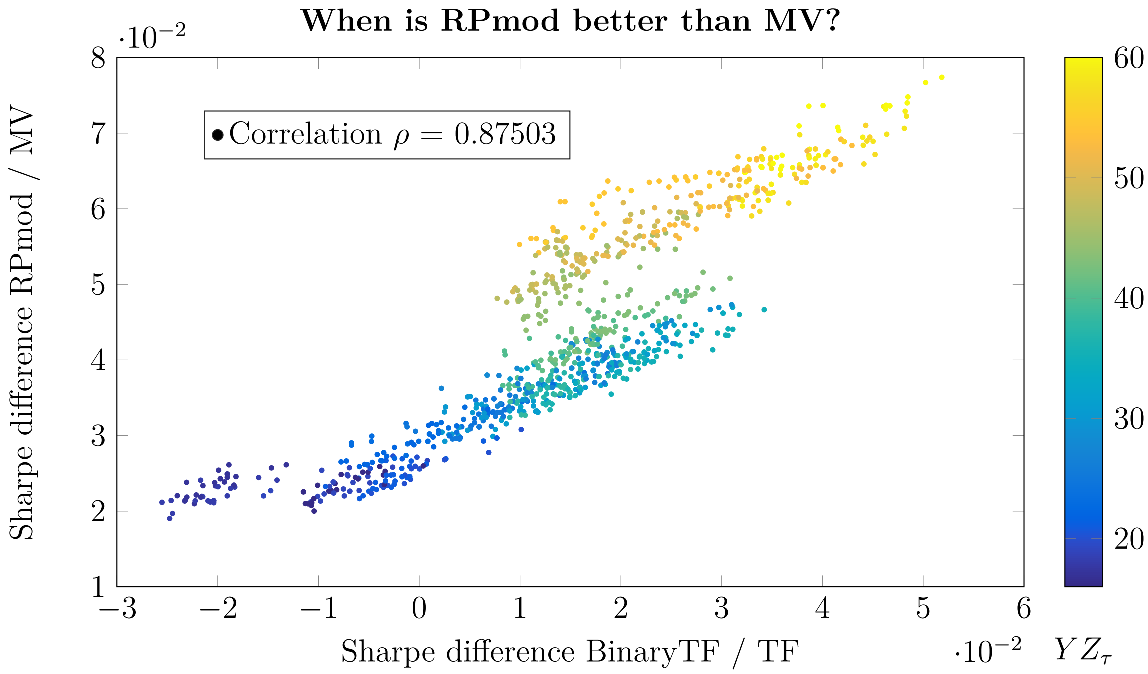

Figure 5.5: How the difference in Sharpe between MV and RPmod depend on the marginal contribution of $|\mu|$. The observations are the 1000 random pairs of parameters, using the degree of regularization resulting in the highest Sharpe for the respective model.

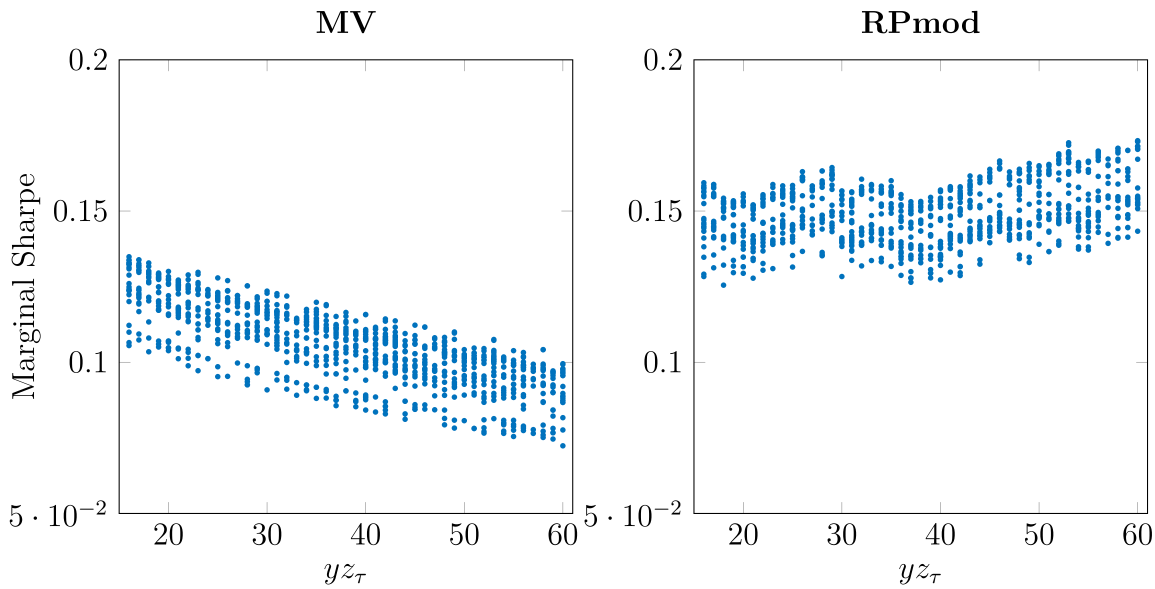

Figure 5.6: How the marginal Sharpe depend on the input parameter yz_τ used in the variance estimate. Showing results for 1000 random pairs of input parameters.

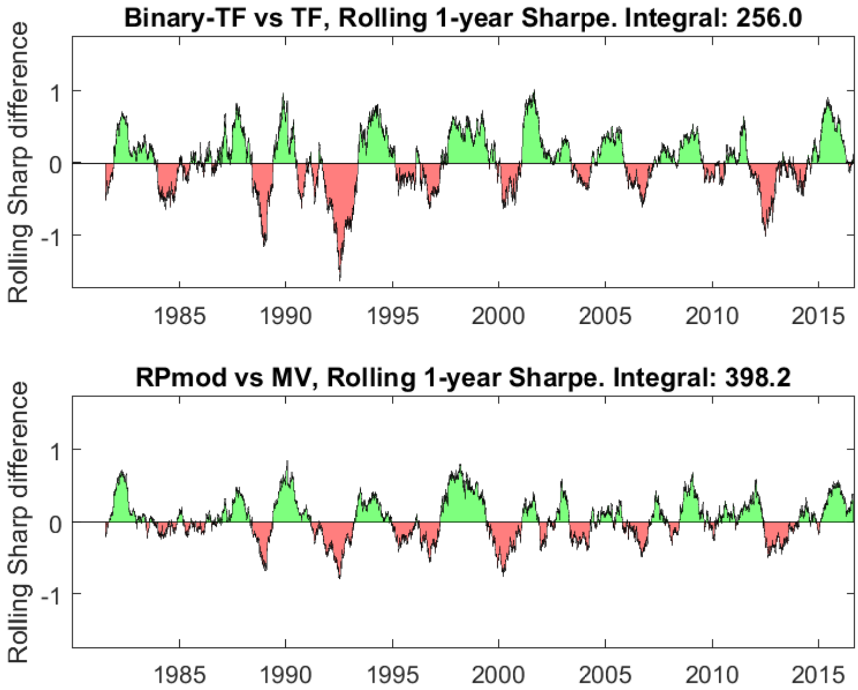

Figure 5.7: Showing similarities between the Sharpe difference between RPmod and MV, and the marginal contribution of the signal strength $|\mu|$.

the highest Sharpe ratio for the respective model. It is obvious that how the Sharpe ratio compares between MV and RPmod during a time period greatly coincides with the quality of the estimated signal strength $|\mu|$. But it is unclear if this property is the reason for the superior Sharpe properties of the modified RP portfolio. Figure 5.8 compares RPmod and MV and shows the corresponding results for the two parts of the Sharpe ratio, average return and standard deviations of returns. Comparing to figure 5.7 it appears as if the discussed factor, the quality of $|\mu|$, only has an impact on the mean return of the portfolio, for which the difference roughly seems to even out over time. Instead this figure implies that the excess in Sharpe comes from a lower realized standard deviation of returns (recall that both models were scaled to have the same theoretical standard deviation of returns).

It is possible to speculate in why the modified risk-parity model shows lower realized standard deviation of returns compared to mean-variance. As the name implies, RP focuses on risk allocation while MV "streches" out in directions with high expected portfolio return. For the MV portfolio this usually results in spread positions (taking a long and a short position in two correlated assets) making it possible for the position vector to have a larger norm under the same variance. Since the RP models currently have a sign restriction, meaning that $\text{sign}\, w_i = \text{sign}\, \mu_i \; \forall i$, the RP portfolio is relatively free from spread positions. A contributing factor to this property is that since μ is based on a trend following trading strategy, correlated assets will tend to have the same sign.

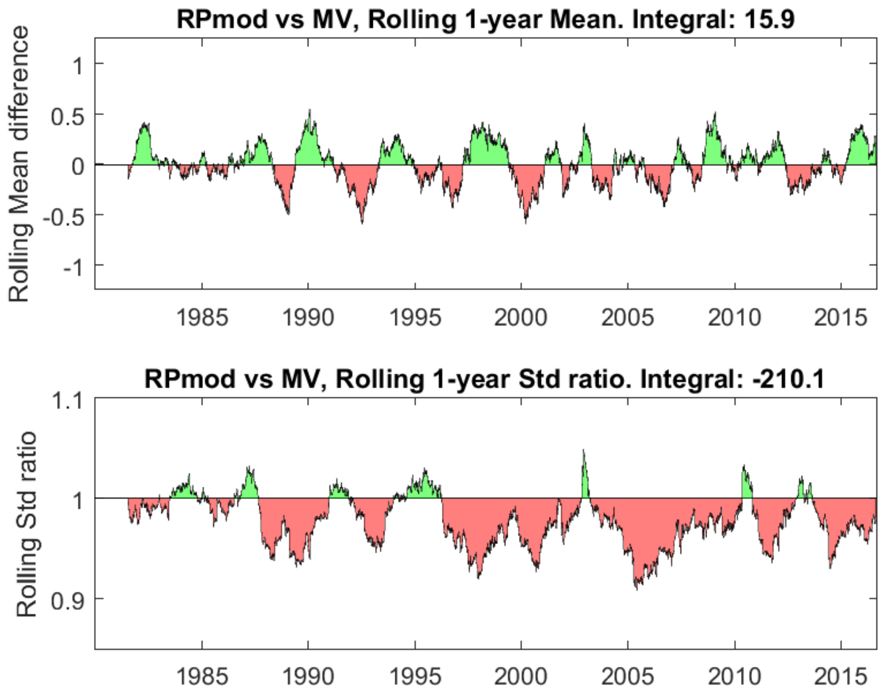

Figure 5.8: Parts of the rolling Sharpe ratio, comparing RPmod and MV. The excess in Sharpe for the RPmod model appears to come from a lower realized variance of returns.

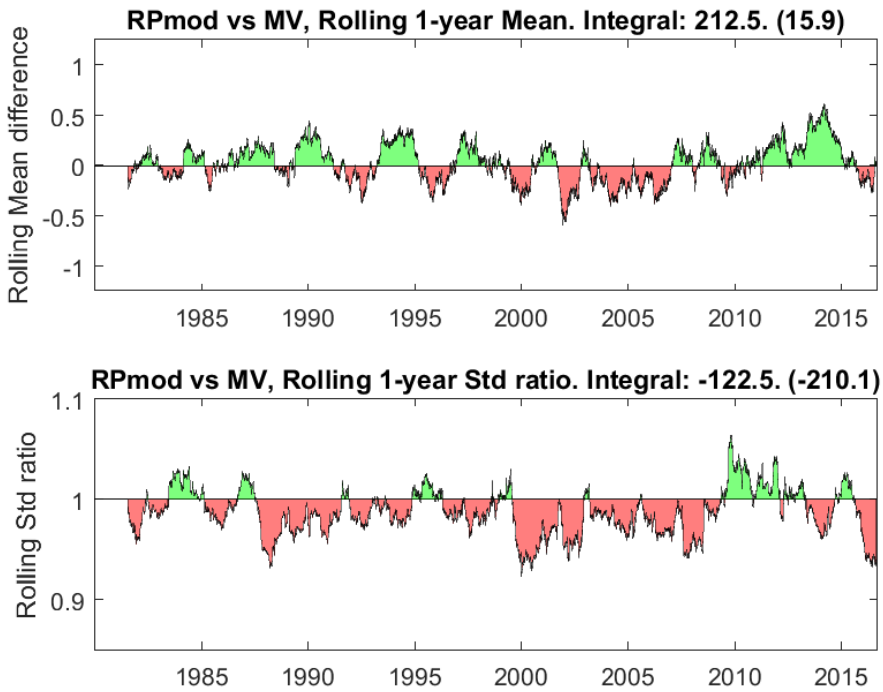

Figure 5.9: Parts of the rolling Sharpe ratio, comparing RPmod and MV when searching for the RPmod portfolio with the highest expected return.

It may however be beneficial for the RP models to include the option of taking spread positions. This theory has been tested through the method described in section 3.2.5, searching for a RPmod portfolio with a high expected return. The results from this approach corresponding to figure 5.8 (same parameters) is shown in figure 5.9, and shows a trade-off between a higher average return but also a higher realized standard deviation of returns. Experimental simulation does however suggest that this approach has a positive impact on the Sharpe ratio over longer periods of time, but also that the current iteration scheme results in much higher turnover (trading costs). Taking the RP model down this path looks promising but requires further development of the algorithm to maximize expected returns over the non-convex set of RP portfolios.

5.2 LES

The portfolio model described in section 3.3 that approximately minimizes the conditional value-at-risk takes different input parameters compared to mean-variance and risk-parity, and will be evaluated separately. As previously discussed, this original model suggested in [9] has not shown any promising performance. Shortcomings include a negative marginal Sharpe, high sensitivity to the input data, and large turnover. Figures 5.10 and 5.11 show the Sharpe ratio and the holding time of this portfolio as a function of its input parameters. Figure 5.10 suggests that no positive marginal Sharpe can be expected, while figure 5.11 shows that this model has a very low holding time compared to the prediction model (and also comparing to the portfolio models in the previous section). As discussed in section 3.3, these issues appear to come from overfitting to historical data. This section will instead focus on results from the final version presented in equation (3.26) which will still be referred to as Least Expected Shortfall (LES).

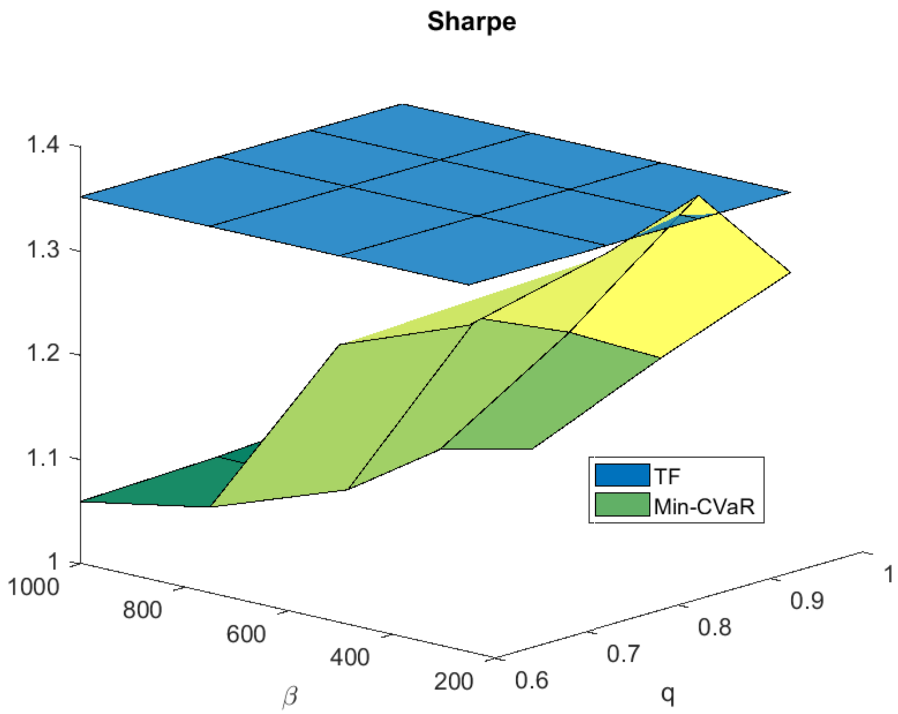

Figure 5.10: The Sharpe ratio of the minimum CVaR portfolio as a function of its parameters q and β, compared to the prediction model TF without portfolio optimization.

Results are based on the same underlying trend follower used in the previous section. As discussed in section 3.3.1.3, an increasing sequence of $\alpha \in [-1, 1]$ is used and the coefficients d_i are fitted to approximate the power $(.)^{\frac{3}{2}}$. Figure 5.12 shows the Sharpe ratio as a function of the time horizon q and the penalty parameter λ. For comparison, the best results from previously considered models are included (best in the sense of the highest Sharpe using this input). The LES model shows a significant

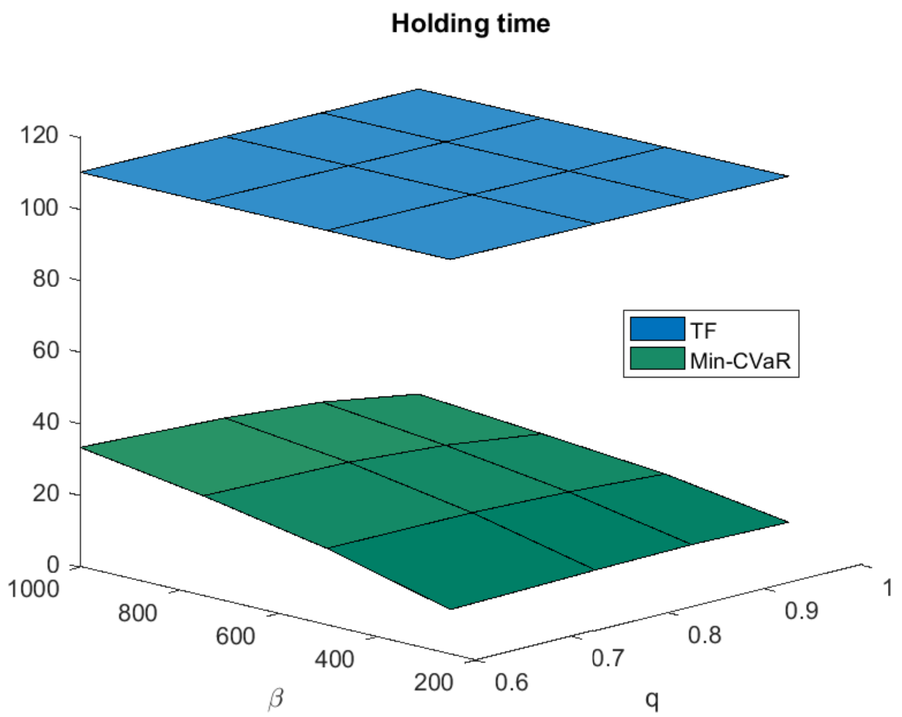

Figure 5.11: The holding time of the minimum CVaR portfolio as a function of its parameters q and β, compared to the prediction model TF without portfolio optimization

boost in Sharpe with reasonably stable results for a range of values on q and λ. Recall that as $\lambda \to 1$, the LES model returns the input μ, and note that this figure only shows λ-values up to 0.9. Similar results are exhibited for the average drawdown, shown in figure 5.13. This quantity does however appear to be more sensitive to the number of historical observations, q, for low values on λ. Figure 5.14 shows the resulting holding times and it is clear that LES is a ”faster” model, i.e. has a higher turnover. This will of course reduce the marginal contribution to Sharpe, but this is not the main problem affiliated with implementing this portfolio method. The next section will show that some reformulation of the optimization problem (3.26) is required in order to be well suited for diversified portfolio selection.

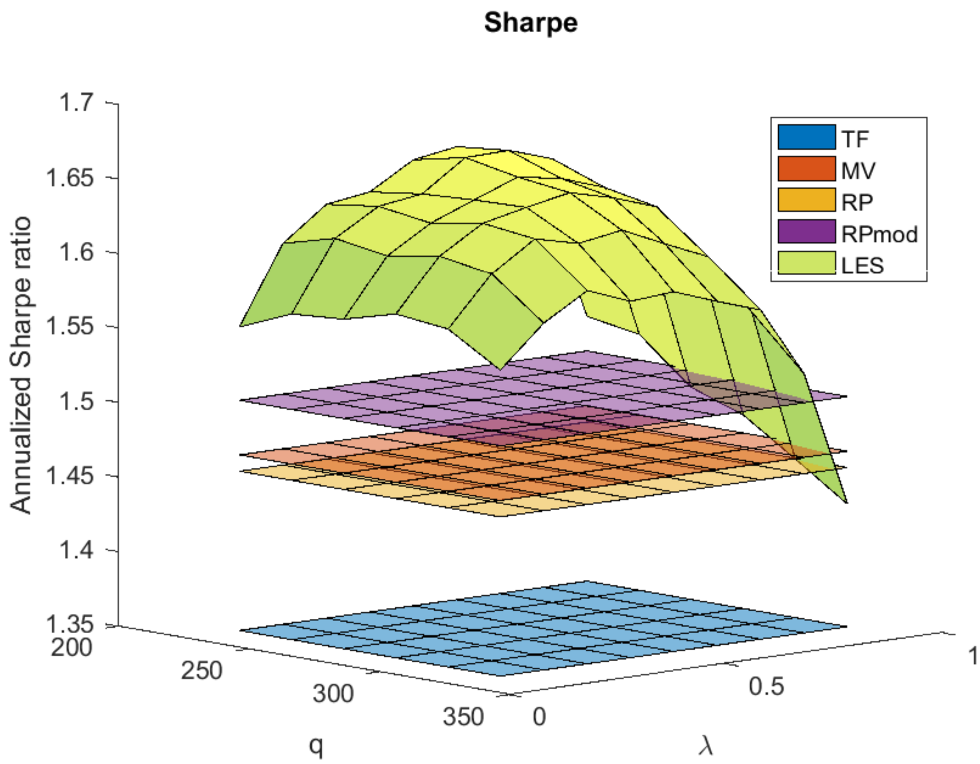

Figure 5.12: The Sharpe ratio of the LES model as a function of its parameters q and λ. This is compared to the highest value obtain from other models for the same prediction model (TF).

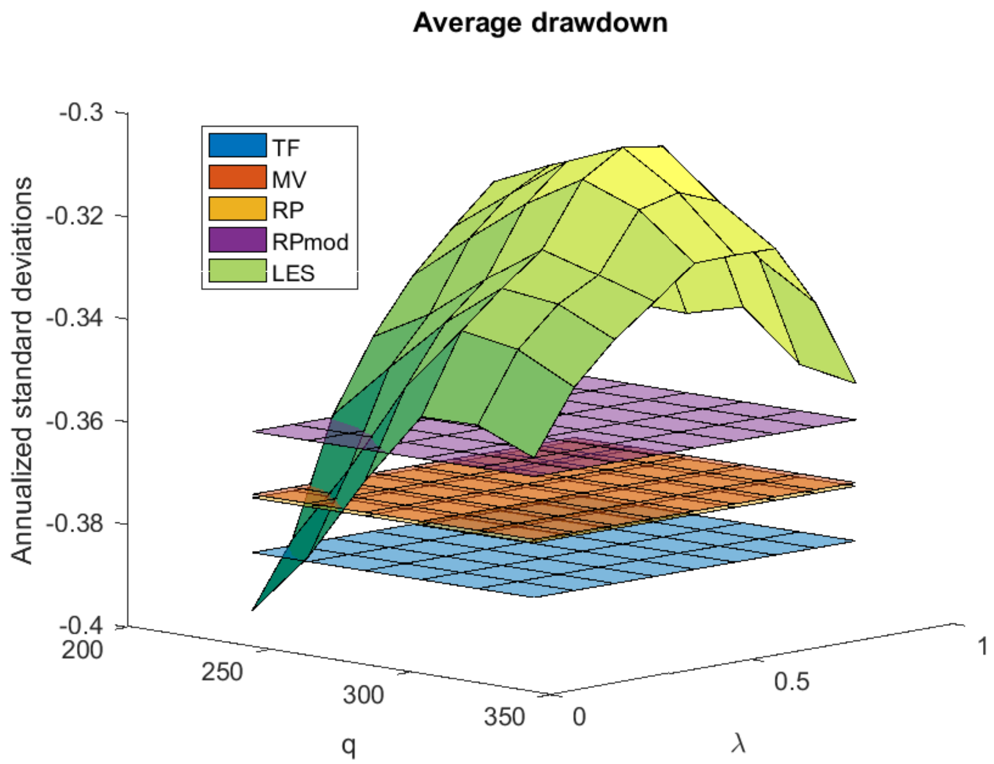

Figure 5.13: The average drawdown of the LES model as a function of its parameters q and λ. This is compared to the value obtain from other portfolio models for the same TF (μ) and the degree of regularization yielding the highest Sharpe in the respective model.

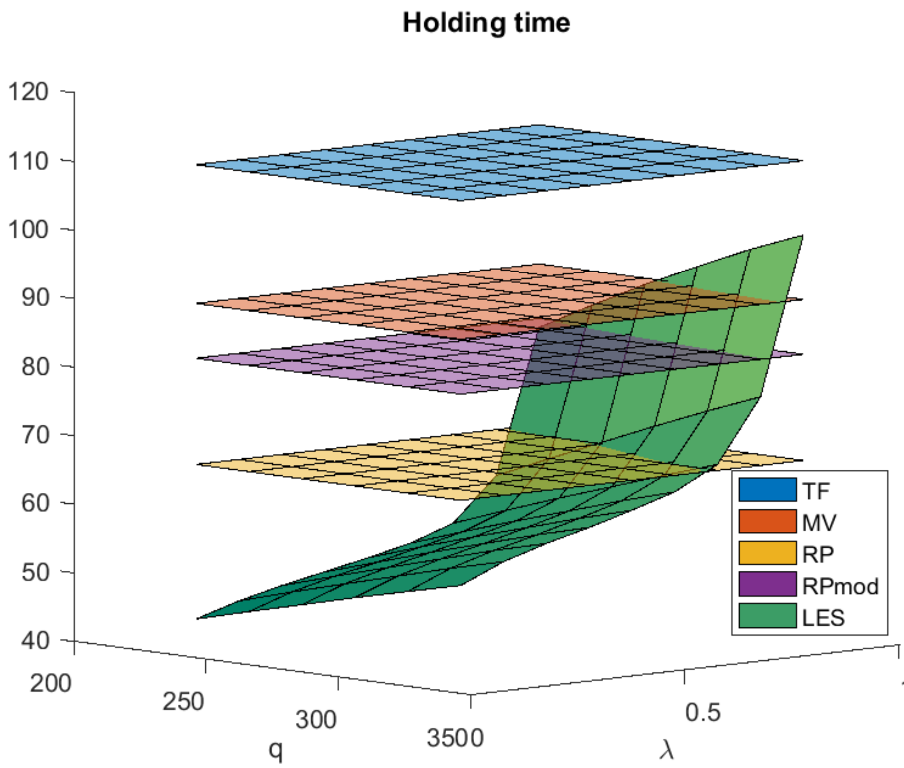

Figure 5.14: The holding time of the LES model as a function of its parameters q and λ. This is compared to the value obtain from other portfolio models for the same TF (μ) and the degree of regularization yielding the highest Sharpe in the respective model.

5.3 Results for different prediction models

All previous results has been for portfolio methods applied on *one* and the same trend following prediction model. This section will test if the results follow when varying this input μ. Recall from section 2.3 that the TF model (μ) was characterized by its filter parameter τ describing the considered time horizon. The other filter parameters will be kept fixed.

In the previous section the LES model showed stable results for $q \in [230, 330]$. Fixing this parameter to the middle point of this interval allows for an easier comparison. All models should now be evaluated for different TF models (TF_τ) and their regularization factor $\lambda \in [0, 1]$. Figure 5.15 shows the maximum Sharpe ratio for any value on λ as a function of the TF parameter τ. The observation that the modified RP portfolio outperforms MV and RP in terms of Sharpe appears to hold over the different trading models while MV and RP shows similar performance. The results for the LES model stands out as it appears to be almost independent of the input μ. Unfortunately this is the result of poor diversification, as the LES model finds basically the same positions for similar inputs μ which is a terrible property for a portfolio selection method. This feature is shown in figure 5.16 that shows the average correlation of portfolio returns between models, before (TF) and after applying the different methods of portfolio optimization. Obviously the LES model destroys all possible diversification from trading different strategies (different TF models), unless for λ close to unity for which it returns the input μ. Another observation from figure 5.16 is that both RP models appear to increase diversification contrary to the MV model. Excluding the LES model from the remaining results on the grounds of unacceptable diversification properties, figure 5.17 shows the average drawdown for the degree of regularization λ yielding the highest Sharpe in the respective portfolio model (in other words the value on λ used in figure 5.15). This figure suggests that the MV portfolio will exhibit better drawdown properties than the RP models for "fast" prediction models, i.e. following the shorter trend. Recall that section 5.1 used a trend following prediction model with $TF_\tau = 200$.

For a final reality check, a trading cost will be included in evaluations. Keeping the unit of returns, the trading cost will be in terms of daily standard deviations of returns. The theory behind this is that since the information at time t is used to trade at time $t+1$, prices can be expected to have moved some amount proportional to the standard deviation during this time leap (over night). We will examine if the marginal contribution to Sharpe is still positive when adding trading costs and also if these costs affect the portfolio models differently. Figure 5.18 shows the highest marginal Sharpe (for any λ) of the different portfolio models as a function of the trading cost and the underlying trading model. These results are consistant with previous analysis as the trading costs affect MV and RPmod in a similar way and does not change how these models compare in terms of Sharpe. But as previously discussed the original RP model naturally has a lower holding time and is thus more affected by trading costs.

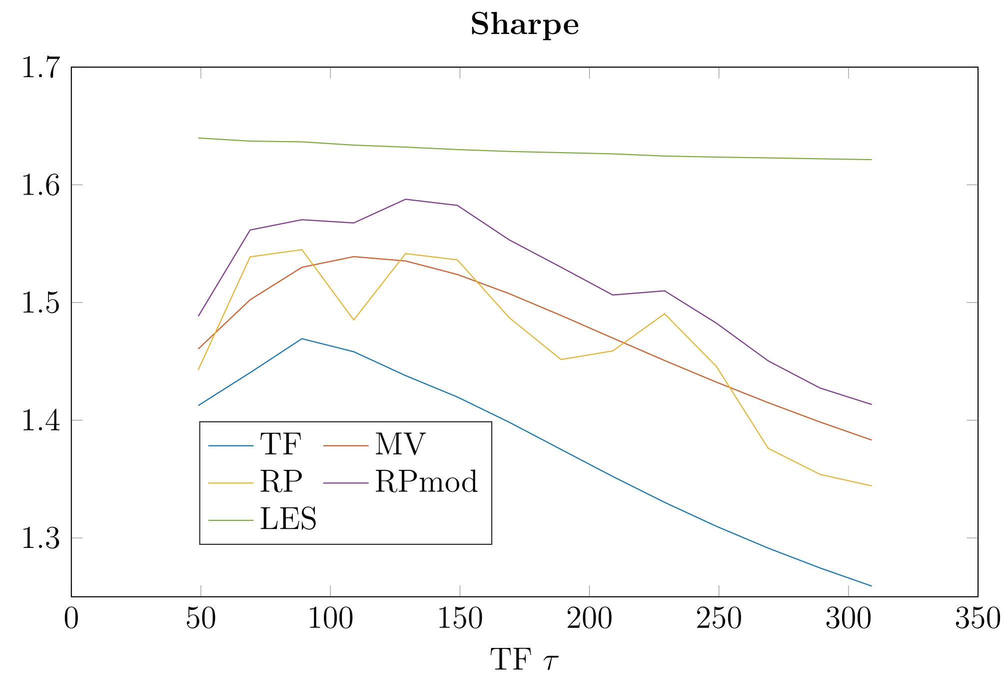

Figure 5.15: The maximum Sharpe ratio, for any value on λ, as a function of the TF parameter τ.

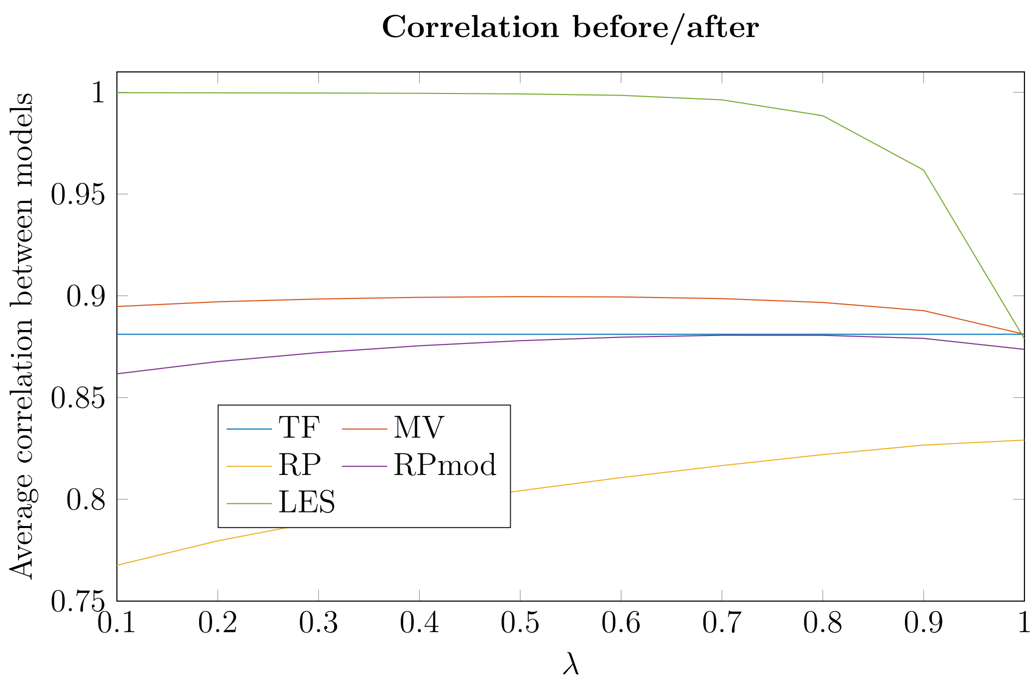

Figure 5.16: The average correlation of returns between trading models, before (TF) and after applying the different methods of portfolio optimization, as a function of λ.

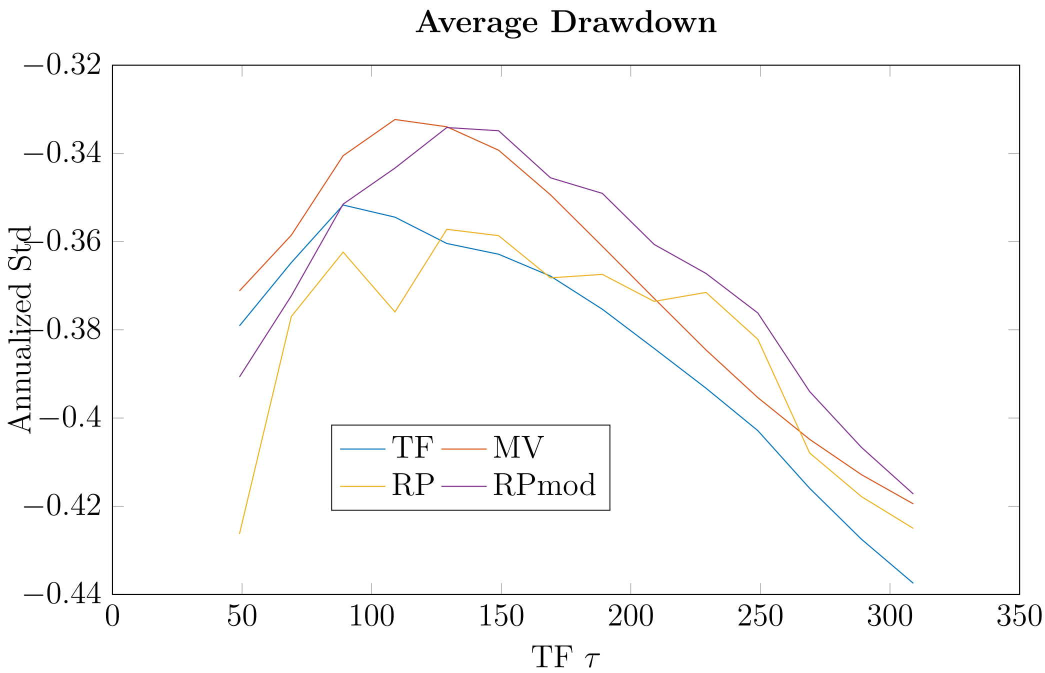

Figure 5.17: The average drawdown, for the value on λ yielding the highest Sharpe, as a function of the TF parameter τ.

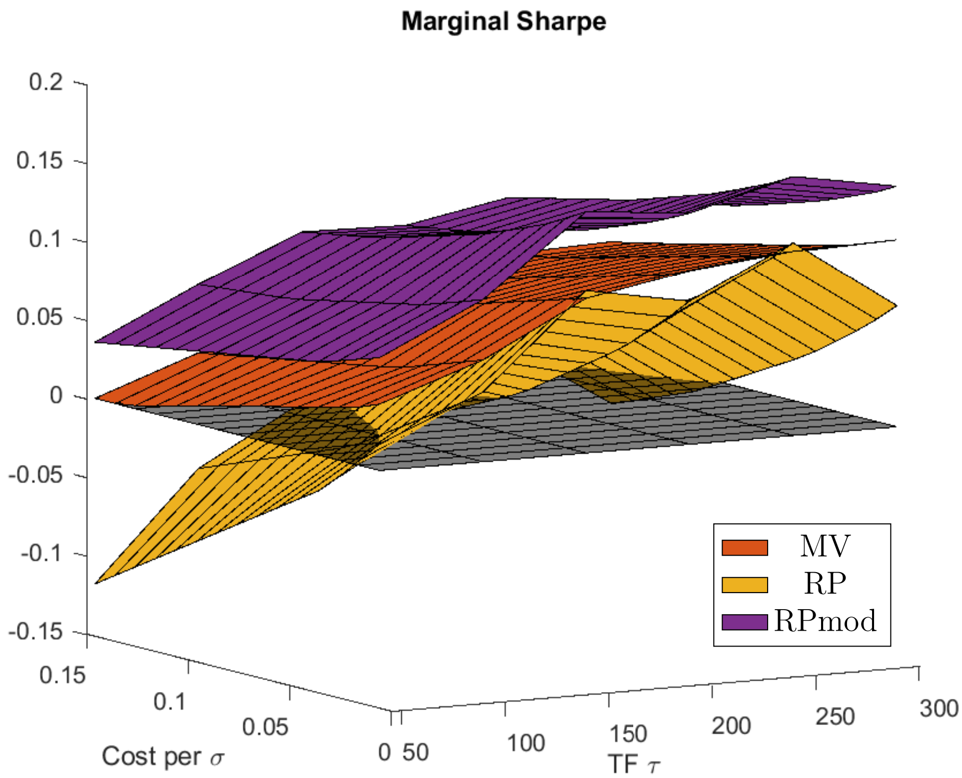

Figure 5.18: The highest marginal Sharpe (for any λ) of the different portfolio models as a function of the trading cost and the TF parameter τ.

6

Discussion

The aim of this thesis has been to compare and improve three existing methods for portfolio optimization, with the main purpose of investigating if alternatives to the standard mean-variance model could show superior portfolio properties. The Sharpe ratio of portfolio returns based on historical data has been the primary measure of performance followed by the average drawdown of the equity curve defined in section 4.2. While the mean-variance portfolio has been criticised for poor performance in practice, the results in sections 5.1 and 5.3 suggests that this is mainly because of sensitivity to an ill-conditioned covariance matrix and with a regularized estimate this model exhibits a positive contribution to both the Sharpe ratio and the average drawdown over time. It can thus be considered a very viable option for portfolio section.

Perhaps the most significant result in this thesis is that the modified RP portfolio suggested in section 3.2.2 has shown superior performance compared to both MV and the original RP in several central aspects. The results in section 5.3 suggests that RPmod exhibits a higher marginal Sharpe for all considered time horizons of the underlying prediction model. The drawdown results were however not unambiguous and suggest that the MV portfolio will have better drawdown properties for ”faster” prediction models. But how likely are these results to hold in practice? The comparison in section 5.1 considered one underlying prediction model and focused on investigating the sensitivity of the portfolio performance when varying the other input parameters (filter parameters for the variance and correlation estimates). As the correlation estimates is a grave approximation of reality, and small changes in the variance estimate causes small variations in the return series, models for which the performance is very sensitive to this input will be likely to perform worse out of sample. The results point to the modified RP portfolio as the most stable to changes in these inputs considering both the Sharpe ratio and especially the average drawdown, while MV and RP show similar behaviour. This suggests that the observed performance of RPmod is more likely to hold in practice.

A closer look at how the rolling Sharpe ratio over time compare between the modified RP model and MV was investigated in section 5.1.1 and revealed some interesting results. Risk-parity focuses on allocating risk, and our results show that RPmod exhibits almost a consistently lower realized variance of returns *while* over time keeping an average return very close to that of MV. During shorter periods of time, how the average return compares between MV and RPmod will depend on the quality on the estimated signal strength in which RPmod places less confidence. Evidence has been found suggesting that there is potential gain to be made from searching for a RPmod portfolio with a high expected return (recall that for M assets there is 2^M portfolios satisfying the RP objective in each time step). Finding the RP portfolio with the highest expected return is however a non-convex optimization problem. The greedy iteration scheme presented in section 3.2.5 requires modifications to be useful in practice, the main issue being increased trading costs. Other possibilities includes stochastic optimization algorithms with a well formulated fitness function.

The model from section 3.3.1.3, built upon the framework of minimizing the conditional value-at-risk of a portfolio was evaluated in section 5.2 and showed some very promising results suggesting that this alternative risk measure can contribute greatly to both Sharpe and drawdown properties. However, it turned out that the final form of the optimization problem was poorly suited for portfolio selection. The issue being that the risk term (the sum) of equation (3.26) has a dynamic impact on the solution, having the consequence that with two similar inputs μ the model returns the same positions. In practice this will lead to a lack of diversification when trading several different (but similar) prediction models. It may be possible to counteract this behaviour by enforcing a more strict, explicit constraint on the norm of the position vector. In other words enforcing that the angle between μ and the position vector be smaller than some suitable constant. Further work could focus on reformulating the ADMM algorithm in section 3.3.1.4 to fit this purpose.

Bibliography

[1] Andreasson, N., Evgrafov, A., Patriksson, M., Gustavsson, E., and Onnheim, M. *Introduction to Continuous Optimization.* Edition 2 : 1. Studentlitteratur AB, 2013.

[2] Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning 3*, 1 (Jan. 2011), 1–122.

[3] Cont, R. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance 1* (2001), 223–236.

[4] Ledoit, O., and Wolf, M. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis 88*, 2 (2004), 365 – 411.

[5] Lee, W. Risk-Based Asset Allocation: A New Answer to an Old Question? *The Journal of Portfolio Management* (June 2011), 11–28.

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

[7] Mestre, X. Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates. *IEEE Transactions on Information Theory 54*, 11 (Nov 2008), 5113–5129.

[8] Qian, E. Risk parity portfolios: Efficient portfolios through true diversification. *Panagora Research Paper* (September 2005).

[9] Rockafellar, R. T., and Uryasev, S. Optimization of conditional value-at-risk. *Journal of Risk 2* (2000), 21–41.

[10] Sébastien Maillard, T. R., and Teïletche, J. The properties of equally weighted risk contribution portfolios. *The Journal of Portfolio Management 36*, 4 (2010), 60–70.

[11] Yang, D., and Zhang, Q. Drift-independent volatility estimation based on high, low, open, and close prices. *Journal of Business 70*, 3 (2000), 477–491.