Robust Methods of Portfolio Optimization Exemplified by the Swiss Market Index

School of Engineering, Zurich University of Applied Sciences

Pascal Aigner Maurice Gerber Basil Rohr

29. November 2020

Abstract

Sed ut perspiciatis unde omnis iste natus error sit voluptatem accusantium doloremque laudantium, totam rem aperiam, eaque ipsa quae ab illo inventore veritatis et quasi architecto beatae vitae dicta sunt explicabo. Nemo enim ipsam voluptatem quia voluptas sit aspernatur aut odit aut fugit, sed quia consequuntur magni dolores eos qui ratione voluptatem sequi nesciunt. Neque porro quisquam est, qui dolorem ipsum quia dolor sit amet, consectetur, adipisci velit, sed quia non numquam eius modi tempora incidunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim ad minima veniam, quis nostrum exercitationem ullam corporis suscipit laboriosam, nisi ut aliquid ex ea commodi consequatur? Quis autem vel eum iure reprehenderit qui in ea voluptate velit esse quam nihil molestiae consequatur, vel illum qui dolorem eum fugiat quo voluptas nulla pariatur?

Keywords: Portfolio optimization, Robust optimization, Markowitz model, Covariance matrix, Shrinkage, Sharpe ratio

1 Introduction

Markowitz' Optimization and other models of Finance require estimates of the expected return, the standard deviation and the correlations of individual assets. These estimates are subject to estimation errors, especially when they are based on short historical time series. This leads to the danger of over- fitting historical data and - in the case of portfolio optimization - of producing portfolios that perform well in backtesting, but perform poorly in the future. To overcome these problems, several approaches have been proposed. This includes Bayesian approaches, including shrinking Estimators for historical returns, volatilites and correlations. It also includes resampling/bootstrapping approaches, where expected returns, volatilites and correlations are modelled as random variables. We will analyze a variety of such approaches, identify those that add most value in the context of investment management, and combine/improve them. The goal is a recommended methodology for computing the strategic asset allocation of pension funds and other institutional investors.

Lange, min. eine halbe Seite

2 Data

For this paper historical data from the Swiss Market Index or (SMI) were used, the data are from the period from 23.05.2001 to 16.10.2020. The SMI is the most important stock index in Switzerland and contains the 20 largest companies traded on the Swiss stock exchange. The SMI covers approximately 80% of the total capitalization of the Swiss stock market. It is also a price index, which means that dividends are not included in the index. The SMI is always reviewed twice a year and, if necessary, reassembled. In our case, we use the composition per end of 2018. The reason for the 2018 composition is, that enough historical data for all those companies is available. The results shown in table 2 are calculated with following two equations, daily logarithmic return (1) and volatility (2). [1] [2]

2.1 Logarithmic return

The reason for using logarithmic returns is that the numbers take smaller values than discrete returns. Due to smaller values there is also a smaller variance, which influences fitting positively. As well historical daily data is used for the calculation of the logarithmic return. For later calculations the risk-free return has to be subtracted from the returns. For the sake of simplicity the risk-free rate is already deducted now. For the risk-free return R_f, shown in table 1, the LIBOR interest rate is commonly used. Due to several negative incidents in the past the Swiss Average Rate Overnight (SARON) is used for the time period 2014-2020. SARON is a rate that is based on daily transactions and is therefore considerably more transparent compared to the LIBOR. [3] [4]

$$R = R_{ln} - R_f = ln\left(\frac{x_t}{x_{t-1}}\right) - R_f \tag{1}$$

R = *return*

R_{ln} = *natural logarithmic return*

R_f = *risk-free rate of return*

x_t = *closing stock price at time t*

Table 1: Average overnight risk-free return per year in %

Year	Rate [%]
2001	3.037
2002	1.083
2003	0.277
2004	0.357
2005	0.722
2006	1.354
2007	2.245
2008	2.035
2009	0.104
2010	0.064
2011	0.049
2012	0.016
2013	-0.005
2014	-0.014
2015	-0.697
2016	-0.736
2017	-0.738
2018	-0.737
2019	-0.729
2020	-0.691

2.2 Volatility (standard deviation)

Volatility describes the risk of a stock or market index and is a statistical measure of the dispersion of the calculated logarithmic returns ((1)). In general, the higher the volatility, the riskier the security.

$$\sigma = \sqrt{\frac{\Sigma(R_i - \overline{R})^2}{n}} \tag{2}$$

σ = *volatility (standard deviation)*

R_i = *each return from the sample*

$\overline{R}$ = *sample mean return*

n = *sample size*

Table 2: SMI with expected return and volatility

Stock	Return [%]	Volatility [%]
ABB	-0.002	2.947
Adecco	-0.016	2.282
Credit Suisse	-0.044	2.465
Geberit	0.051	1.663
Givaudan	0.043	1.336
Julius Baer	-0.013	2.333
LafargeHolcim	-0.018	2.064
Lonza	0.036	1.826
Nestle	0.021	1.186
Novartis	0.003	1.310
Richemont	0.013	2.074
Roche	0.016	1.438
SGS	0.035	1.669
Sika	0.068	1.975
Swatch	-0.001	2.066
Swiss Life	-0.014	2.417
Swiss Re	-0.020	2.258
Swisscom	0.001	1.105
UBS	-0.031	2.325
Zurich Insurance	-0.016	2.202

2.3 Covariance and correlation

The covariance is a numerical measure that describes the linear statistical relationship between two variables. In this case, it represents the strength of the relationship between two return time series. The correlation is the standardized covariance. By definition, the correlation ranges in an interval of [-1, 1]. A value of 1 or -1 means that the returns of two shares move in the same direction respectively in opposite directions. In Markowitz' portfolio theory the dependency (correlation) of two individual stocks is an important matter, which will be discussed later on in chapter [Markovitz Model].

$$cov(x,y) = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{n-1} \tag{3}$$

$cov_{x,y}$ = *covariance between x and y*

x = *variable x*

y = *variable y*

n = *sample size*

$$\rho_{i,j} = \frac{cov_{i,j}}{\sigma_i * \sigma_j} \tag{4}$$

$\rho_{i,j}$ = *correlation coefficient between return i und j*

$cov_{i,j}$ = *covariance between return i und j*

σ_i = *volatility of return i*

σ_j = *volatility of return j*

The matrix below shows the correlations of all 20 stocks included in the SMI. By definition, the diagonal values are always 1. As the SMI only contains shares that are traded on the Swiss stock exchange, there is a rather high correlation between the individual companies. The correlation between companies from the same sector is also higher than that between companies from different sectors or which have less in common. From this it can be concluded that economic similarities are reflected in a higher correlation of stock returns. Another way to demonstrate a connection between two companies is the t-statistic which is described in chapter 2.4.1.

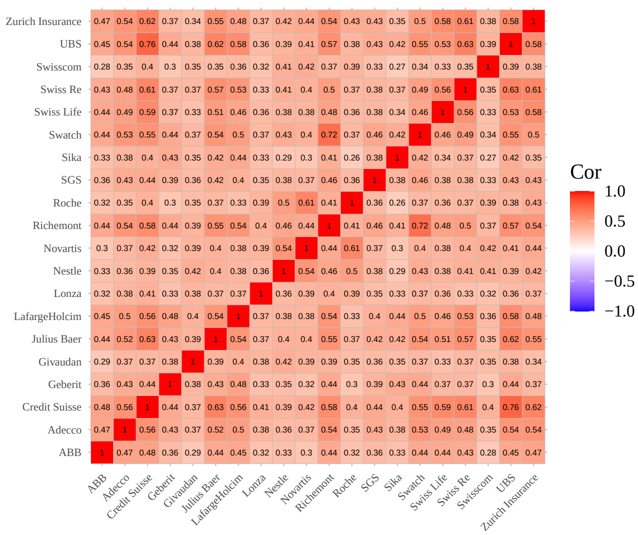

Figure 1: SMI correlation

2.4 Statistical key numbers

2.4.1 T-statistic

$$t = \frac{\rho_{i,j} * \sqrt{n-2}}{\sqrt{1-\rho_{i,j}^2}} \tag{5}$$

t = *t-statistic*

$\rho_{i,j}$ = *correlation coefficient*

n = *sample size*

To show that each correlation coefficient is significantly different from zero, the t-test is performed for the two shares with the correlation closest to zero, which are Sika and Roche.

$$t = \frac{0.264 * \sqrt{4816-2}}{\sqrt{1-0.264^2}} = 18.957 \tag{6}$$

The students t-distribution with 4816 degrees of freedom tells us that the probability of getting a test-statistic in the interval of [-18.957,18.957] equals 1. Therefore, the probability of getting a test-statistic out of the interval equals 0. Since the P-value is smaller than 0.05, we can reject the null hypothesis. There is sufficient statistical evidence at the $\alpha = 0.05$ level to conclude that there is a significant linear relationship between Sika and Roche.

2.4.2 Standard error of expected return

In statistics the standard deviation of the sampling distribution is known as the standard error and it provides a statement about the quality of the estimated parameter. The more individual values there are, the smaller is the standard error, and the more accurately the unknown parameter can be estimated. The standard error for the expected return and the volatility are explained in the following two chapters.

$$\sigma_{\overline{R}} = \frac{\sigma}{\sqrt{n}} \tag{7}$$

$\sigma_{\overline{R}}$ = *standard error*

σ = *standard deviation of sample*

n = *sample size*

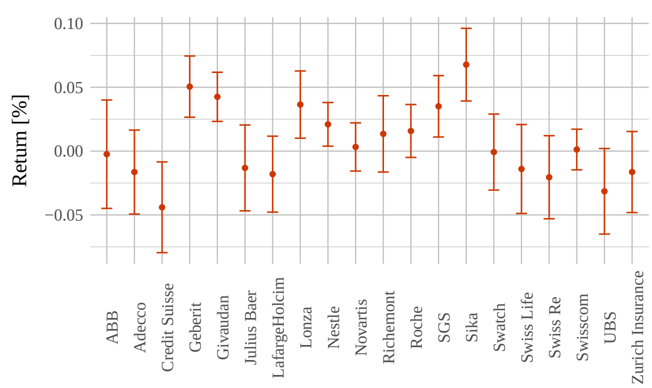

Figure 2: Expected mean return with standard error

As can be seen, the standard error of expected mean return by groups do not show larger values as for example the individual shares *Sika* or *Geberit* do. Due to averaging returns the impact of the outliers is stabilized, therefore the standard error declines when grouping individual stocks.

2.4.3 Standard error of volatility (standard deviation)

This formula is an approximation for the standard error of volatility, which is appropriate for n > 10. [8]

$$\sigma_\sigma = \sigma * \frac{1}{\sqrt{2*(n-1)}} \tag{8}$$

σ_σ = *standard error*

σ = *standard deviation*

n = *sample size*

As shown in table 3 respectively table **??** the values of the standard error of volatility are much lower as the one of the returns. Therefore we will neglect the values for further calculations.

Table 3: Standard error of volatility

Stock	Standard error [%]
ABB	0.03003
Adecco	0.02325
Credit Suisse	0.02512
Geberit	0.01695
Givaudan	0.01362
Julius Baer	0.02378
LafargeHolcim	0.02104
Lonza	0.01861
Nestle	0.01208
Novartis	0.01335
Richemont	0.02113
Roche	0.01465
SGS	0.01701
Sika	0.02012
Swatch	0.02105
Swiss Life	0.02463
Swiss Re	0.02301
Swisscom	0.01126
UBS	0.02370
Zurich Insurance	0.02244

3 Review of classical markovitz optimization

Modern portfolio theory is a theory that deals with the construction of portfolios to maximize the expected return based on a given market risk. Markowitz' portfolio theory shows that efficient risk reduction is only possible if the extent of the correlation of the individual investments is taken into account when putting together the portfolio. Risk reduction through risk diversification is a key finding of the aforementioned portfolio theory. Harry Markowitz pioneered this theory in his article "Portfolio Selection". The main point is that the risk and return characteristics of an investment should not be considered in isolation, but should be evaluated according to how the investment affects the risk and return of the overall portfolio. It is shown that an investor can construct a portfolio of multiple assets that maximizes returns for a given level of risk. Similarly, an investor can construct a portfolio with the lowest possible risk at a desired level of expected return. Based on statistical measures such as volatility and correlation, the performance of an individual investment is less important than how it affects the portfolio as a whole. [5]

The return, volatility and covariance matrix is known. The portfolio volatility is accordingly given as a function of the covariance matrix and the weight vector and it can be minimized as much as desired by sufficient diversification. Also the sum of all weights equals 1. The portfolio weights being searched for are described by the vector $\vec{w} = (w_1, \ldots, w_n)$. The weights that are calculated are those weights that match the portfolio with minimal volatility (variance) to a given expected portfolio return R_p. This is a linear optimization problem, as well a formulation of the fundamental problem of balancing return and risk. Furthermore, negative weightings are defined as short sales. [9]

$$minimize : \frac{1}{2}\vec{w}^T \Sigma \vec{w} \tag{9}$$

With the following two constraints:

I.

$$1 = \vec{w}^T \vec{1} \tag{10}$$

II.

$$R_p = \vec{w}^T \vec{R} \tag{11}$$

According to the method of the Lagrange Multiplier, the Lagrange function is formed with the factors λ and ϵ.

$$L(\vec{w}) = min\frac{1}{2}\vec{w}^T \Sigma \vec{w} - \lambda(\vec{w}^T \vec{R} - R_p) - \epsilon(\vec{w}^T \vec{1} - 1) \tag{12}$$

The disappearance of the gradient is the necessary condition for a minimum. This is together with the two constraints ((10), ((11) an inhomogeneous linear system of equations of the dimension $n+2$ with $n+2$ variables. The solution is a known standard problem from linear algebra.

$$\nabla_w L = \Sigma \vec{w} - \lambda * \vec{R} - \epsilon * \vec{1} = 0 \tag{13}$$

$\vec{w}$ = *weight vector*

Σ = *covariance matrix*

$\vec{1}$ = *all-ones vector*

$\vec{R}$ = *return*

R_P = *total portfolio return*

3.1 Problem

Markowitz' optimization only works under certain assumptions and conditions. Apart from the assumptions concerning particular investors and the capital market, which will not be discussed here, there are also mathematical assumptions that must be made. First assumption is an approximate normal distribution of log-returns. Since it is a stock price over 20 years, it is almost impossible to describe the log-returns as normally distributed. In historical data over a long period of time there are strong stock price movements which can be clearly identified as outliers and therefore reject the assumption of an exact normal distribution. However, a quantile-quantile plot can be used to show that the majority of the data is normally distributed. The y-axis describes the quantiles of the data and the x-axis the quantiles of the normal distribution. The majority of the data is linear near the middle which speaks for a normal distribution. At both ends clear deviations and individual outliers are to be recognized, which is a strong indication of a heavy left tail. This behavior results from stronger stock price volatilities, as they occur for example in times of crisis and uncertain markets.

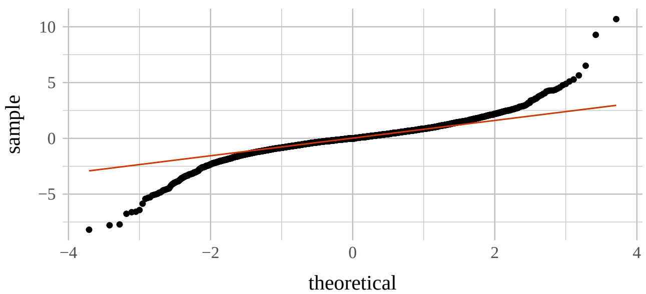

Figure 3: Q-Q plot of daily log-returns for Swisscom

A second condition applies to the linear relationship of the stocks. The log-returns cannot show a perfect correlation between themselves, neither negative nor positive. This is described by a correlation coefficient of -1 respectively 1. If the determinant is 0, the system cannot be solved exactly. This situation occurs if shares have a correlation coefficient of 1 or -1. The system must be uniquely solvable to form the inverse of the covariance matrix Σ^{-1}.

overfitting

3.2 Minimum variance portfolio

The Minimum Variance Portfolio, or MVP for short, describes the portfolio of all possible weightings with the minimum volatility. Since only the volatility is minimized, constraint II. is not included in the equation.

$$\vec{w}_{mvp} = \frac{1}{\vec{1}^T \Sigma^{-1} \vec{1}} * \Sigma^{-1} \vec{1} \tag{14}$$

$\vec{w}_{mvp}$ = *weights*

$\vec{1}$ = *all-ones vector*

Σ = *covariance matrix*

Table 4: Weights of MVP

Stock	Weight
ABB	-0.016
Adecco	-0.013
Credit Suisse	-0.057
Geberit	0.096
Givaudan	0.164
Julius Baer	-0.023
LafargeHolcim	-0.007
Lonza	0.026
Nestle	0.233
Novartis	0.105
Richemont	-0.039
Roche	0.068
SGS	0.072
Sika	0.046
Swatch	-0.001
Swiss Life	-0.004
Swiss Re	-0.001
Swisscom	0.390
UBS	-0.019
Zurich Insurance	-0.018

It is clear to see that returns which have a negative value also have a negative weighting. In practice, this would lead to a short selling.

3.3 Tangency portfolio

The tangency portfolio results from the tangent of the capital market line and the efficient frontier, which will be shown more detailed in chapter Efficient Frontier. The capital market line is an important component of the Capital Asset Pricing Model. The slope of the capital market line indicates how much more return is expected per additional volatility, therefore a steeper slope of the capital market line gives a better Sharpe ratio. This is exactly the situation when the capital market line is tangential to the efficient frontier. Therefore, the best possible diversified portfolio results from the weightings of the tangency portfolio.

In this paper the MVP and TP serve as useful reference points to compare the Sharpe ratio of different optimizations. The tangency portfolio is often referred to in the literature as the market portfolio. In the equation for the capital market line, shown as equation (15), the expected return of the market portfolio respectively the tangency portfolio is written as R_{tp}.

Capital Market Line:

$$R_P(\sigma_P) = R_f + \frac{R_{tp} - R_f}{\sigma_{tp}} * \sigma_P \tag{15}$$

R_P = *total portfolio return as a function of* σ_p

R_f = *return of risk-free asset*

R_{tp} = *return of tangency portfolio*

σ_P = *total portfolio volatility*

σ_{tp} = *tangency portfolio volatility*

Tangency Portfolio:

$$\vec{w}_{tp} = \frac{1}{\vec{1}\Sigma^{-1}(\vec{R} - R_f\vec{1})} * \Sigma^{-1}(\vec{R} - R_f\vec{1}) \tag{16}$$

$\vec{w}_{tp}$ = *weights*

$\vec{1}$ = *all-ones vector*

Σ = *covariance matrix*

$\vec{R}$ = *return*

Table 5: Weights of tangency portfolio

Stock	Weight
ABB	-0.001
Adecco	-0.189
Credit Suisse	-0.309
Geberit	0.533
Givaudan	0.586
Julius Baer	-0.083
LafargeHolcim	-0.341
Lonza	0.230
Nestle	0.338
Novartis	-0.318
Richemont	0.202
Roche	0.165
SGS	0.300
Sika	0.502
Swatch	-0.197
Swiss Life	-0.031
Swiss Re	-0.061
Swisscom	-0.148
UBS	-0.139
Zurich Insurance	-0.039

3.4 Efficient Frontier

The efficient frontier is a set of points that extends in the return-volatility diagram (figure 4) between the minimum variance portfolio at the left edge of the reachable area and the tangency portfolio. All possible weightings of portfolios on this line are efficient because they have the maximum return at a defined level of volatility.

$$\vec{w}_{tp} = \alpha * \vec{w}_{mvp} + (1 - \alpha) * \vec{w}_{tp} \tag{17}$$

$\vec{w}_{tp}$ = *weights*

α = *scale factor*

$\vec{w}_{mvp}$ = *weights of minimum variance portfolio*

$\vec{w}_{tp}$ = *weight of tangency portfolio*

$$R_P = \vec{w}^T * \vec{R} \tag{18}$$

R_P = *total portfolio return*

$\vec{w}$ = *weights*

$\vec{R}$ = *mean return*

$$\sigma_P = \sqrt{\vec{w}^T \Sigma \vec{w}} \tag{19}$$

σ_P = *total portfolio volatility*

$\vec{w}$ = *weights*

Σ = *covariance matrix*

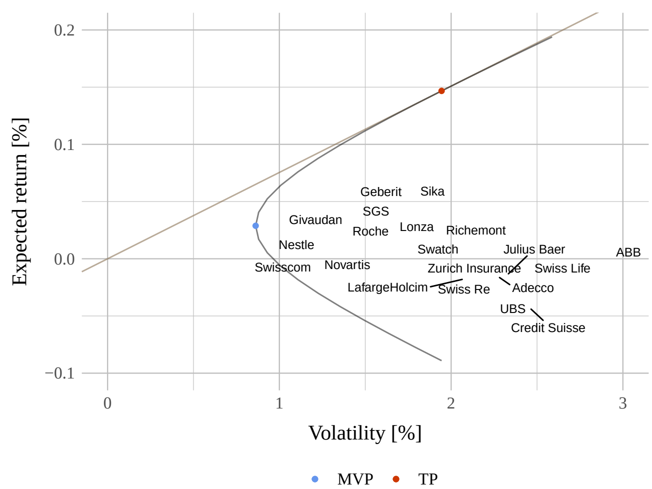

Figure 4: Efficient frontier with stocks

In figure 4 it is clearly visible that the SMI shows a significantly higher return with almost the same volatility and vice versa. This leads to a larger in sample Sharpe ratio of the SMI compared to the groups. The reason for a higher in sample Sharpe ratio of the individual stocks is due to the higher amount of data. There are 20 stocks in the SMI for which the weighting can be optimized, but only four in the grouping. However, it will become apparent in chapter 4.3 that this no longer applies to the out of sample Sharpe ratio, which in our case is much more important than the in sample.

3.5 Sharpe ratio

The Sharpe ratio measures the performance of an investment which means the return of an investment compared to its risk. Generally, the greater the value of the Sharpe ratio, the more attractive the risk-adjusted return. In practice, the value of the Sharpe Ratio is not only positively received. In this paper, however, we will rely on the Sharpe ratio because it is a key figure by which performance can be measured. It is calculated by the average return earned in excess of the risk-free rate per unit of volatility. In this case it is the natural logarithmic return per day R divided by σ. R is already calculated in equation (1). [6] [7]

$$Sharperatio = \frac{R_{ln} - R_f}{\sigma} = \frac{R}{\sigma} \tag{20}$$

R = *natural logarithmic return per day*

R_{ln} = *mean logarithmic return*

R_f = *risk-free rate of return*

σ_i = *volatility of return*

Before any weighting of a portfolio is optimized, the Sharpe ratio of the SMI is calculated. This is done by weighting all stocks equally, 1/20 each, and then applying the weightings to the actual returns of the historical data. This results in a Sharpe ratio of XX. Note that the Sharpe ratio is annualized, which means that it is multiplied by $\sqrt{252}$. As mentioned above, the Sharpe ratio is used as a measure of the optimization and robustness. A robust optimization leads to a lower volatility which results in a higher Sharpe ratio. The higher the value the better the optimization.

4 Methodology

4.1 Grouping

A first way to increase the Sharpe ratio is to create groups. Grouping the stocks also reduces the standard errors, which will be described in more detail later in chapter [Standard error]. Naturally there are different approaches how many groups and how exactly the composition of such groups should be created. In a first approach we formed four groups, which differ in their industries:

Consumer: Adecco, Nestle, Richemont, Swatch, Swisscom

Finance: Credit Suisse, Julius Baer, Swiss Life, Swiss Re, UBS, Zurich Insurance

Industrial: ABB, Geberit, Givaudan, Lafarge Holcim, SGS

Pharma: Lonza, Novartis, Roche, Sika

This method is not mathematical in nature but a simple intuitive decision based on the perception of these companies. Within the grouping, each share is equally weighted. This combination results in the following values for the mean return and volatility of each group as shown in table 6.

Table 6: groups with expected return and volatility

Group	Return [%]	Volatility [%]
Consumer	0.004	1.342
Finance	-0.023	1.902
Industrial	0.022	1.397
Pharma	0.031	1.189

It can already be seen that the values of volatility are lower overall than those of the individual shares, which leads to a higher Sharpe ratio as shown in table **??**. The Sharpe ratio of groups increased significantly compared to the Sharpe Ratio for the SMI. In the matrix below, the correlation of the individual groups is shown. As the returns are averaged within the groups, the values of the individual groups converge, resulting in a larger correlation coefficient.

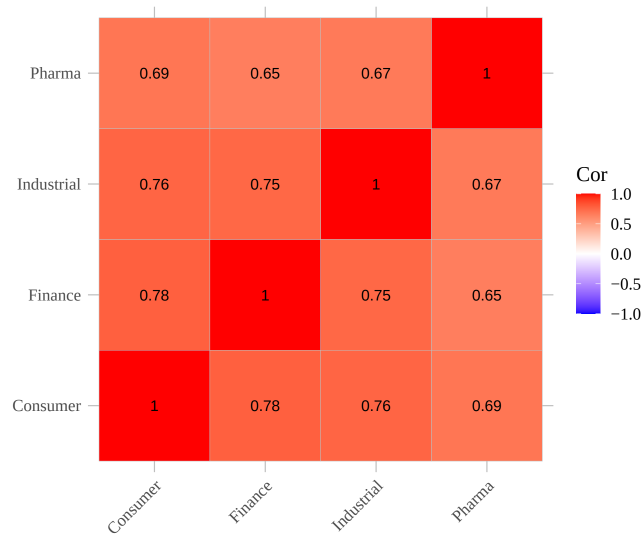

Figure 5: Groups correlation

4.2 Bootstrap

In this part a bootstrap process was made to analyze the robustness of the tangency portfolio weights when resampling all historical returns.

All dates from the historical data were sampled in the same dimension, with replacing included. Which means that certain dates can occur multiple times. From each sampled date, the corresponding returns of each stock are added to the new data, consequently maintaining the daily differences between stocks. If replacing would not be included in the sampling method, the correlation matrix would remain constant to the one of the original data.

One Bootstrap sample contains a different data, with which a new mean return of each stock and group and the corresponding covariance matrix is calculated. Those two variables are needed as input in the optimization function to calculate minimum variance portfolio, tangency portfolio of the 20 stocks and the four groups. With 100 bootstrap samples, 100 MVP's and TP's can be compared and analyzed. Additionally, the standard deviation of the weights over those number of bootstrap samples can be studied. The standard deviation is computed with the 84-Quantil minus the 50-Quantil to achieve a more robust result, because this calculation is less vulnerable to outliers.

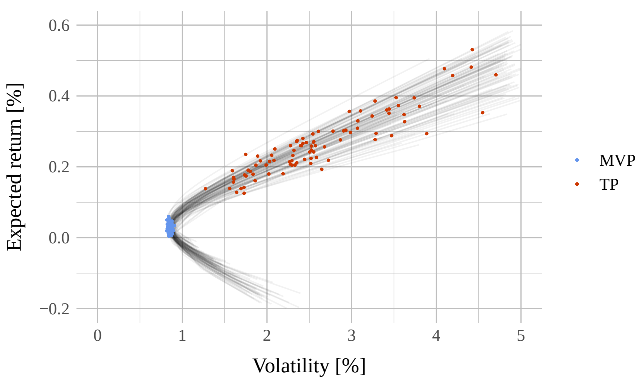

Figure 6: Bootstrap samples efficiency frontier

In figure 6 all bootstrap samples can be seen with their minimum variance portfolio, tangency portfolio and their efficiency frontier. Noticeable is the variance of the tangency portfolio in comparison to the minimum variance portfolio. This is due to the high standard errors of returns. Since the return is included in the calculation of the TP, but not in the MVP, a larger deviation can be seen.

4.3 Cross validation

Cross validation is a model validation technique for assessing how a model will generalize to an independent data set. The data set is split into a training set used to train the model and a test set to evaluate its performance. This procedure is replicated multiple times until all data was once in the test set.

In this case, the data set is split into five sections and consequentially five models are trained as depicted in figure 7. The gray sections are the training sets and the red ones the test sets.

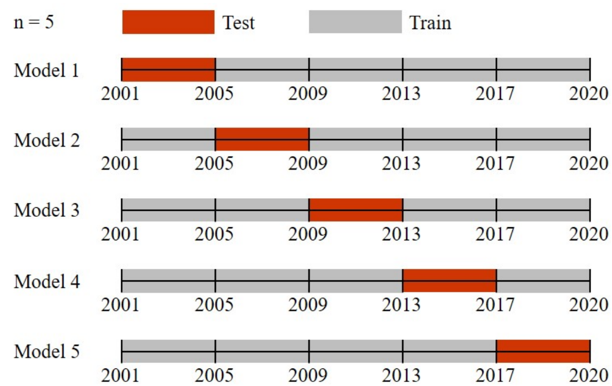

Figure 7: Cross validation training and test sets

In the case of the Markovitz model, the mean returns and covariance matrix of the training set are used to calculate the asset weights. These weights are then applied to the test set asset returns according to (21).

$$R_{i,w} = \vec{w}^{\intercal} * \vec{R_i} \tag{21}$$

$R_{i,w}$ = *weighted return day i*

$\vec{w}$ = *asset weights*

$\vec{R_i}$ = *asset returns day i*

At the end of the cross validation, all five test sets are combined into one time series consisting of the daily weighted returns and equaling the length of the original data set. This time series is completely out of sample and is used to calculate the Sharpe ratio and hence measure the performance of the model.

It is to note that the training and test sets of the five sections vary significantly in some cases. To give one cause for this, the financial crisis from 2007–2008 can be looked at. During this period, most assets yielded negative returns. For example, model 3 in figure 7 includes this time period as part of the test set and model 4 of the training set. This results in fluctuating asset weights across the five models which are illustrated in figure 8.

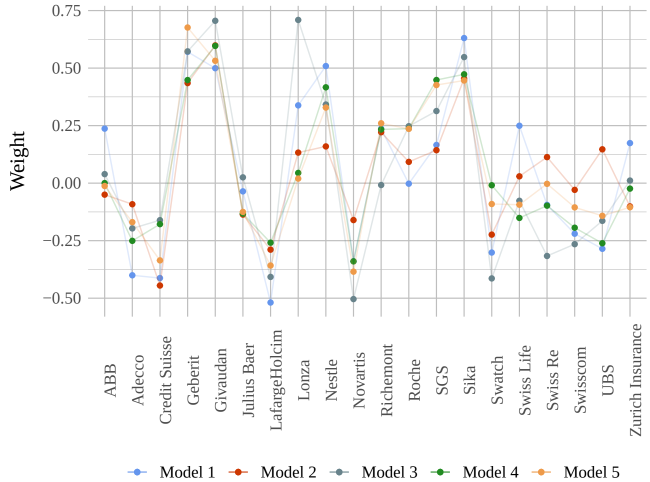

Figure 8: Tangency portfolio weights by model

Stocks like *Julius Baer* and *Sika* have similar weights across all models while *Lonza* for example varies significantly.

4.4 Shrinkage

Shrinkage is an approach with the goal to increase the out of sample Sharpe ratio by altering the inputs of the Markovitz optimization, namely the mean returns and covariance matrix.

4.4.1 Shrinkage of mean returns

The mean returns shrinkage factor is calculated as in (22).

$$\vec{R}(\lambda) = \lambda * \vec{R} + (1 - \lambda) * \vec{\bar{R}} \tag{22}$$

$\vec{R}(\lambda) = \textit{mean returns as a function of } \lambda$

$\lambda = \textit{shrinkage factor}$

$\vec{R} = \textit{mean returns}$

$\vec{\bar{R}} = \textit{mean of mean returns}$

When the shrinkage factor λ is set to one, the mean returns remain unchanged. When it is decreased from one towards zero, the mean returns converge towards their mean value. Once the shrinkage factor λ is zero and therefore all mean returns are equal, the tangency portfolio (16) obtains the same asset weights as the minimum variance portfolio (14).

There is an important constraint in mean returns shrinkage. Take the scaling factor of the tangency portfolio (16) and replace $\vec{R}$ with (22) as in (23).

$$\vec{w}_{tp,sf} = \frac{1}{\vec{1}\Sigma^{-1}(\lambda * \vec{R} + (1 - \lambda) * \vec{\bar{R}})} \tag{23}$$

It is clear that for some value of λ the denominator becomes zero. Computationally, it is already problematic if the denominator is close to zero. When setting the denominator equal to zero, the value of λ at the zero crossing can be calculated as in (24).

$$\lambda = \frac{-\vec{1}\Sigma^{-1}\vec{\bar{R}}}{\vec{1}\Sigma^{-1}\vec{R} - \vec{1}\Sigma^{-1}\vec{\bar{R}})} \tag{24}$$

If the zero crossing is in the shrinkage interval of $[0, 1]$ this results in extreme asset weights. As the denominator approaches zero, the scaling factor becomes very large which carries over to the weights. To illustrate this effect, the weights of the five cross validation models from figure 7 are plotted as a function of λ in figure 9.

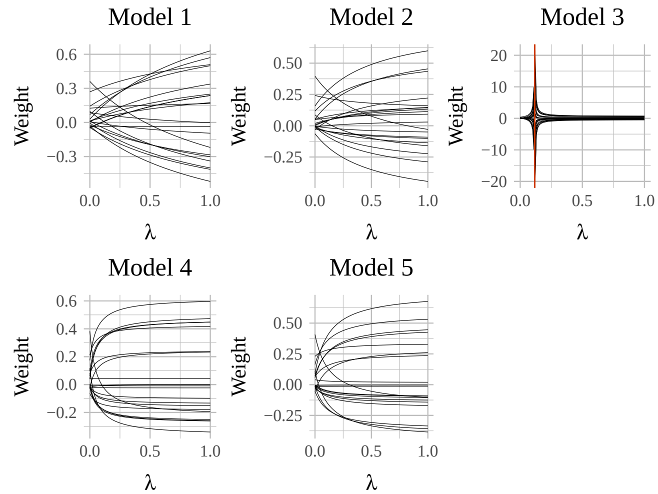

Figure 9: Tangency portfolio weights by model as a function of λ

Model three shows such extreme asset weights. The asymptote is visualized as a vertical red line. These extreme weights influence the out of sample time series and consequentially also the out of sample Sharpe ratio. The other four models do show asset weights in a normal scale.

The scaling factor as a function of λ is shown in 10.

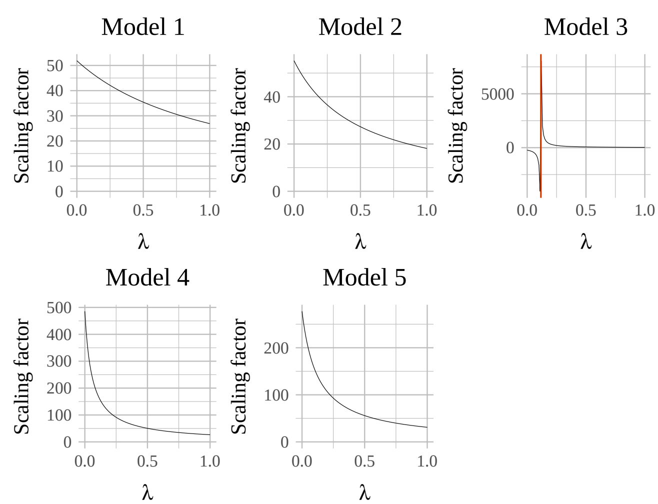

Figure 10: Tangency portfolio weights scaling factor by model as a function of λ

The zero crossing of the scaling factor in model three can be seen clearly. When working with mean return shrinkage one has to be aware of this behavior.

4.4.2 Shrinkage of correlation matrix

The correlation matrix shrinkage factor is calculated as in (25).

$$\rho_{i,j}(\epsilon) = I_{i,j} + \epsilon * \tilde{\rho_{i,j}} \tag{25}$$

$\rho_{i,j}(\epsilon)$ = *correlation coefficient as a function of* ϵ

ϵ = *shrinking factor*

$I_{i,j}$ = *identity matrix*

$\tilde{\rho_{i,j}}$ = *correlation matrix with diagonal zero*

The formula for the correlation shrinking factor is shown in (25). The reason for shrinking the correlation instead of the covariance is that the former is standardized and therefore more suitable. When the shrinking factor is set to one, the correlation is not altered and remains the same. When decreasing the shrinking factor from one to zero, the correlation converges towards zero. This means that all entries of the correlation matrix are equal to zero except the diagonal, which is by definition always equal to one.

5 Results

In this section the different approaches which were desribed in chapter methodology 4 are applied on the historical data. Section I. optimizes the entire SMI with 20 stocks. Section II. includes the optimization of the groups. Section III. applies the shrinkage.

5.1 I.

Table 7: Sharpe ratio

	In sample	Out of sample
SMI	1.198	0.511

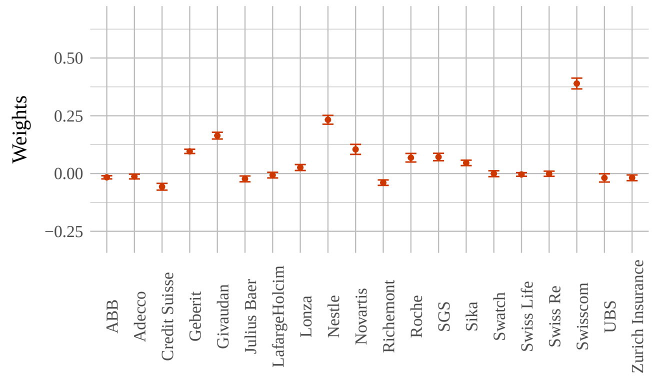

Figure 11: MVP weights with standard error

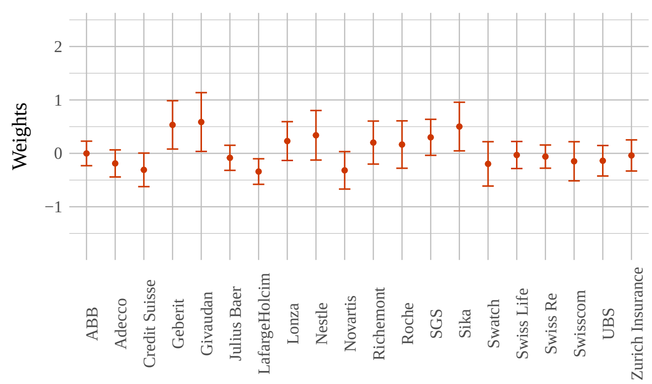

Figure 12: tangency portfolio weights with standard error

5.2 II.

In the following four illustrations the results of the bootstrap can be seen. In all four graphics the mean weights and the standard deviation over the 100 samples are shown. Figure 11 and figure 13 visualize the MVP weights by all 20 stocks and by group. In comparison to figure 12 and figure 14 where the same is depicted with the tangency portfolio.

Table 8: Sharpe ratio

	In sample	Out of sample
Groups	0.815	0.621

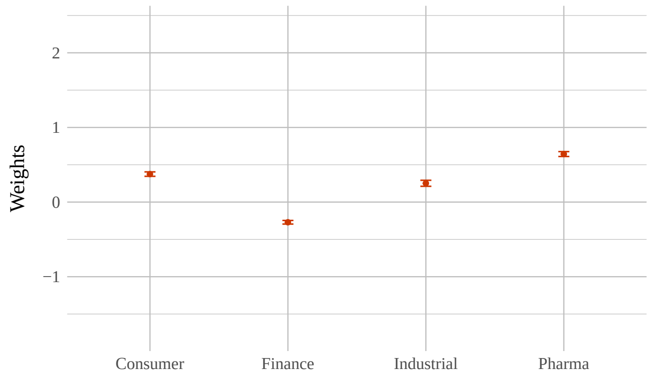

Figure 13: MVP weights with standard error

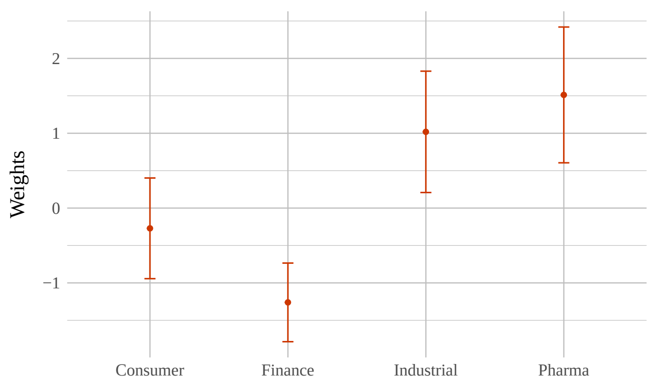

Figure 14: Tangency portfolio weights with standard error by group

As it is obvious to see and also visualized in figure 6, is that TP weights have a higher standard deviation than MVP weights. That means the TP is more sensible to changes of returns than the MVP is. If the equation (14) of MVP and the equation (16) of TP are considered, the MVP does not depend on the returns directly. Consequently different returns samples have not a large impact on the MVP. The small differences arise from the varying covariance matrix. But it can also be noticed that this matrix has not large deviations following the standard deviation of the MVP. Therefore bootsptrap sampling of the returns does not influence the covariance matrix as much as the returns itself.

5.3 III.

5.3.1 Shrinking factors analyzed separately

The Sharpe ratio can be visualized as a function of the shrinking factor. Figure 15 shows such a plot for the return shrinking factor. The global maxima are highlighted with points and their coordinates. It can be seen that the highest Sharpe ratio for the SMI is achieved with a shrinking factor of about two thirds. For the groups on the other hand shrinking does not lead to an improvement of the Sharpe ratio, as the highest Sharpe ratio is achieved at shrinking factor of zero which means that the returns are not altered. For the SMI it is visible that close to the shrinking factor one the line is interrupted. This implies that the corresponding Sharpe ratio is below zero which is not shown in the plot. These outliers are caused by extreme portfolio weights which are obtained at these shrinking factors. Extreme portfolio weights are obtained through near-singularity of the covariance matrix.

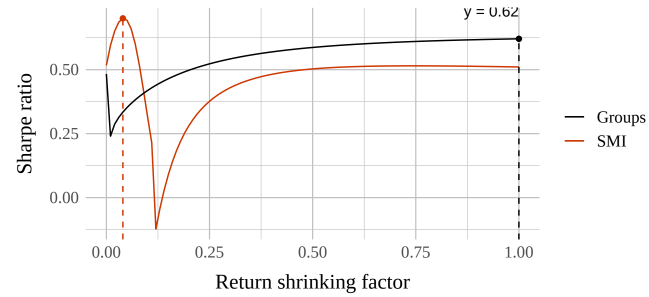

Figure 15: Sharpe ratio as a function of return shrinking factor

Figure 16 shows the same plot for the correlation shrinking factor. The highest Sharpe ratio for the SMI is achieved with a shrinking factor of about one third and for the groups of close to one.

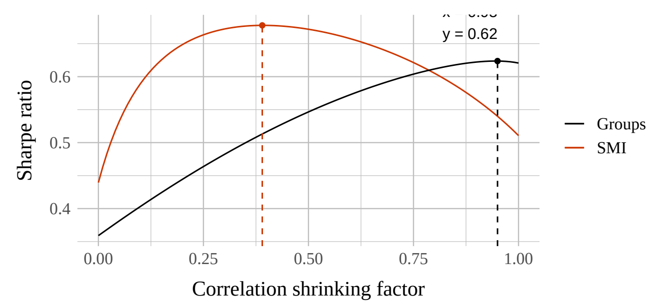

Figure 16: Sharpe ratio as a function of correlation shrinking factor

It can be concluded that for the SMI shrinking does make a noticeable difference in regard of increasing the Sharpe ratio. It is to note that the shrinking factors of the return and correlation are equal in regard of the deviation from the original values. The optimal return shrinking factor is about two thirds (no shrinking at value 0) and the one of the correlation about one third (no shrinking at value 1) which is in both cases a deviation of about two thirds from the original values.

In the case of the groups shrinking does not have proven to make a noticeable impact. Shrinking the returns does worsen the Sharpe ratio and Shrinking the correlation does only make a slight difference. The conclusion of the SMI above also holds true here, as both shrinking factors deviate approximately the same from their original values, which in this case is no or almost no deviation.

5.3.2 Shrinking factors analyzed simultaneously

A further visualization is a three dimensional plot where the Sharpe ratio is plotted as a function of both shrinking coefficients. The points of the maxima and their values are also displayed.

Figure 17 shows such a plot for the SMI. The outliers where the Sharpe ratio drops below zero can also be seen in this plot in the top left corner, although only on the lower third of the correlation shrinking factor. The highest Sharpe ratio is achieved at a correlation shrinking factor of about one third again but a return shrinking factor of zero meaning the original values are used. Varying both shrinking factors together yields the same Sharpe ratio as in 16 where only the correlation shrinking factor is varied. Noticeable is that the major part of the plot area has Sharpe ratios which are very close to each other.

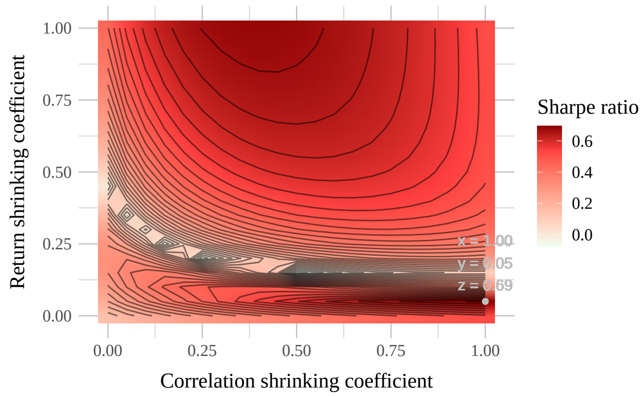

Figure 17: SMI Sharpe ratio as a function of return and correlation shrinking factor

Figure 18 shows the three dimensional plot for the groups. Here, a much smaller fraction of the plot area shows Sharpe ratios which are similarly high. The highest Sharpe ratio is achieved at a correlation shrinking factor of close to one and a return shrinking factor of zero. Varying both shrinking factors toghether yields also the same Sharpe ratio as in 16 were only the correlation shrinking factor is varied.

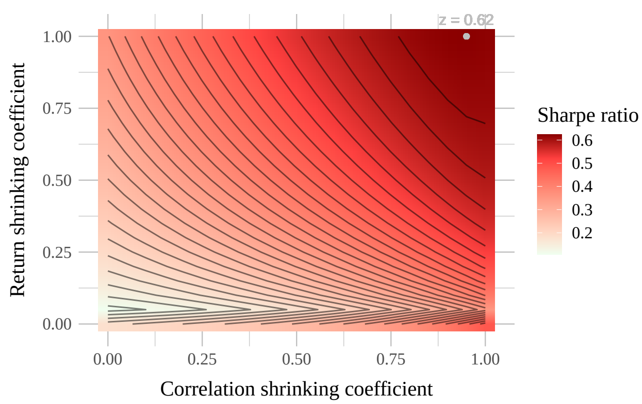

Figure 18: Groups Sharpe ratio as a function of return and correlation shrinking factor

To conclude these results, it can be said that shrinking of the correlation has a greater impact than shrinking of the return and can result in higher Sharpe ratios, although not in all cases. Table 9 provides an overview.

Table 9: Sharpe ratio

	In sample	Out of sample	Out of sample shrinking
SMI	1.198	0.511	0.693
Groups	0.815	0.621	0.624

6 Conclusion

7 References

[1] Yahoo Finance, SMI Equities - EBS in CHF [online]. Available from: https://finance.yahoo.com. [Accessed 28 October 2020]

[2] SIX Group, SMI® – the Blue-Chip Index [online]. Available from: https://www.six-group.com/exchanges/indices/data_centre/shares/smi_en.html [Accessed 15 October 2020]

[3] SIX Group, Swiss Reference Rates (SARON) [online]. Available from: https://www.six-group.com/exchanges/indices/data_centre/swiss_reference_rates/reference_rates_en.html [Accessed 30 October 2020]

[4] global-rates, CHF LIBOR interest rate [online]. Available from: https://www.global-rates.com/en/interest-rates/libor/swiss-franc/swiss-franc.aspx [Accessed 28 October 2020]

[5] H. Makovitz (1952). Portfolio Selection. The Journal of Finance Vol. 7, No. 1, pp. 77-91 (15 pages). Published By: Wiley.

[6] W. F. Sharpe (1966). Mutual Fund Performance. The Journal of Business, Vol. 39, No. 1, Part 2: Supplement on Security Prices, pp. 119-138. Published by: The University of Chicago Press,

[7] A. W. Lo (2003). The Statistics of Sharpe Ratios. Sloan School of Management, Massachusetts Institute of Technology, Cambridge.

[8] Ahn, S. and Fessler, J. (2003). Standard Errors of Mean, Variance and, Standard Deviation Estimators. EECS Department, The University of Michigan, pp.1-2.

[9] D. G. Leuenberger (1998). Investment Science. Published by: Oxford University Press, New York.

List of Figures

List of Tables