Robust Methods of Portfolio Optimization Exemplified by the Swiss Market Index Constituents

School of Engineering, Zurich University of Applied Sciences

Pascal Aigner   Maurice Gerber   Basil Rohr

17. December 2020

Abstract

Markowitz optimization requires estimates of mean return, standard deviation, and correlation of individual stocks, all of which are subject to estimation error. This leads to the risk of overfitting historical data and, in the case of portfolio optimization, creating portfolios that perform well in backtesting but poorly in the future. To overcome these problems, several robust approaches are investigated, such as different asset grouping methods and shrinkage factors for mean return and correlation. The goal is to provide a recommended approach for optimizing a portfolio that includes all assets from the Swiss Market Index. In summary, three key observations are made. First, asset grouping leads to significant noise reduction and thus better portfolio performance compared to individual assets. Likewise, shrinkage leads to increased performance due to major noise reduction. Using both methods simultaneously achieves a further, albeit smaller, increase in performance, as noise can only be reduced to a certain extent. All code is published on https://github.com/pascalaigner/markovitz and the Shiny app can be accessed at https://pascalaigner.shinyapps.io/markovitz.

Keywords: Portfolio optimization, Markovitz model, Robust methods, Shrinkage

1   Introduction

In order to make the best possible use of returns and volatilities when selecting assets in a portfolio, the US economist Harry M. Markowitz developed a mathematical method in 1952 to determine efficient portfolios. This method has since been known as modern portfolio theory. Since then, it is formalized how to create the most efficient portfolio from different stocks with given variables such as returns and volatilities. The disadvantage of this optimization is that the calculations are based on historical data. This data, namely the returns, are strongly affected by noise and therefore will not perform well in future estimations.

The fundamental question is how to create an optimization from historical data that delivers robust results on independent data. In this paper, the Sharpe ratio is used as a performance measurement of the optimization. The data used in this paper are the assets included in the SMI, afterwards a short review of Markovitz theory is presented. To achieve more robust results, several approaches are investigated. The first method described is grouping where a qualitative and a quantitative approach are applied and compared. Followed by a description of the bootstrap which is used to determine the standard errors of the asset weights. To measure the performance on to an independent data set, cross validation is applied to obtain an out of sample time series. As a final approach, shrinkage of the mean returns and correlations is discussed.

All these methods are applied to 20 years of historical data and their performance in terms of Sharpe ratio is compared. At the end of this paper, the aim is to provide a general approach to create a robust portfolio optimization.

2 Data

In this paper, daily historical data from the Swiss Market Index, short SMI, and its constituents is used. The data stems from the period 23.05.2001 to 15.10.2020. The SMI is the most important stock index in Switzerland and contains the 20 largest companies traded on the Swiss stock exchange. It covers approximately 80% of the total market capitalization of the Swiss stock market. The SMI is also a price index which means that dividends are not included. It is reviewed twice a year and, if necessary, reassembled. In this case, the composition per end of 2018 is used. The reason for the 2018 composition is that enough historical data for all companies is available. The key numbers are shown in table 2. [1], [2]

2.1 Logarithmic return

The reason for calculating logarithmic returns is that stock returns are generally assumed to be log-normally and not normally distributed. In later calculations, the risk-free rate of return has to be subtracted from the return and for the sake of simplicity, this is already done from the start as illustrated in eq. (1). For the risk-free rate R_f shown in table 1, the LIBOR is commonly used. Due to several negative incidents in the past, the Swiss Average Rate Overnight (SARON) is used for the time period 2014 to 2020. The SARON is based on daily transactions and is therefore considerably more transparent compared to the LIBOR. [3], [4]

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

R = return

R_{ln} = natural logarithmic return

R_f = risk-free rate of return

x_t = closing stock price at time t

Table 1: Annualized risk-free rate of return

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)

The volatility in eq. (2) describes the risk of an asset and is a statistical measure of the dispersion of its returns. In general, the higher the volatility, the riskier the asset.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(R_i - \overline{R})^2}{n-1}} \qquad (2)$$

σ = volatility (standard deviation)

R = return

$\overline{R}$ = mean return

n = sample size

Table 2: Mean return and volatility of SMI constituents

Stock	Return [%]	Volatility [%]
ABB	-0.003	2.947
Adecco	-0.017	2.282
Credit Suisse	-0.044	2.465
Geberit	0.050	1.664
Givaudan	0.042	1.336
Julius Baer	-0.014	2.333
LafargeHolcim	-0.019	2.064
Lonza	0.035	1.825
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	-0.021	2.258
Swisscom	0.001	1.105
UBS	-0.032	2.325
Zurich Insurance	-0.017	2.202

2.3 Covariance and correlation

The covariance in eq. (3) is a numerical measure that describes the linear relationship between two variables. In this case, it represents the strength of the relationship between two return time series. The correlation in eq. (4) is the standardized covariance and ranges in an interval of $[-1, 1]$ with the diagonal elements of a correlation matrix always being one. A value of one or minus one means that the returns of two assets move in the same direction respectively in opposite directions. In Markowitz' portfolio theory the correlation of two assets is an important matter which will be discussed later in chapter 3.

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

$cov_{x,y} = covariance\ between\ returns\ of\ assets\ x\ and\ y$

$x = return\ of\ asset\ x$

$y = return\ of\ asset\ y$

$n = sample\ size$

$$\rho_{i,j} = \frac{cov_{i,j}}{\sigma_i * \sigma_j} \qquad (4)$$

$\rho_{i,j} = correlation\ between\ returns\ of\ assets\ i\ and\ j$

$cov_{i,j} = covariance\ between\ returns\ of\ assets\ i\ and\ j$

$\sigma_i = volatility\ of\ return\ of\ asset\ i$

$\sigma_j = volatility\ of\ return\ of\ asset\ j$

The correlation matrix of all 20 stocks is shown in figure 1. As the SMI only contains stocks that are traded on the Swiss stock exchange, there is generally a rather high correlation between them.
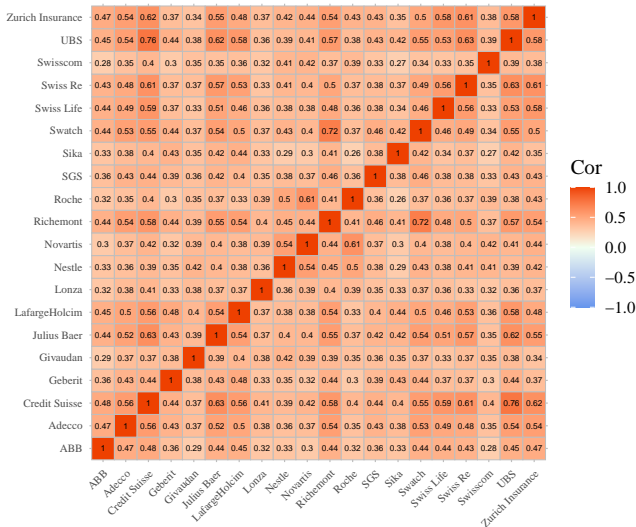
	ABB	Adecco	Credit Suisse	Geberit	Givaudan	Julius Baer	LafargeHolcim	Lonza	Nestle	Novartis	Richemont	Roche	SGS	Sika	Swatch	Swiss Life	Swiss Re	Swisscom	UBS	Zurich Insurance
Zurich Insurance	0.47	0.54	0.62	0.37	0.34	0.55	0.48	0.37	0.42	0.44	0.54	0.43	0.43	0.35	0.5	0.58	0.61	0.38	0.58	1
UBS	0.45	0.54	0.76	0.44	0.38	0.62	0.58	0.36	0.39	0.41	0.57	0.38	0.43	0.42	0.55	0.53	0.63	0.39	1	0.58
Swisscom	0.28	0.35	0.4	0.3	0.35	0.36	0.32	0.41	0.42	0.37	0.39	0.33	0.27	0.34	0.33	0.35	0.35	1	0.39	0.38
Swiss Re	0.43	0.48	0.61	0.37	0.37	0.57	0.53	0.33	0.41	0.4	0.5	0.37	0.38	0.37	0.49	0.56	1	0.35	0.63	0.61
Swiss Life	0.44	0.49	0.59	0.37	0.33	0.51	0.46	0.36	0.38	0.38	0.48	0.36	0.38	0.34	0.46	1	0.56	0.33	0.53	0.58
Swatch	0.44	0.53	0.55	0.44	0.37	0.54	0.5	0.37	0.43	0.4	0.72	0.37	0.46	0.42	1	0.46	0.49	0.34	0.55	0.5
Sika	0.33	0.38	0.4	0.43	0.35	0.42	0.44	0.33	0.29	0.3	0.41	0.26	0.38	1	0.42	0.34	0.37	0.27	0.42	0.35
SGS	0.36	0.43	0.44	0.39	0.36	0.42	0.4	0.35	0.38	0.37	0.46	0.36	1	0.38	0.46	0.38	0.38	0.33	0.43	0.43
Roche	0.32	0.35	0.4	0.3	0.35	0.37	0.33	0.39	0.5	0.61	0.41	1	0.36	0.26	0.37	0.36	0.37	0.39	0.38	0.43
Richemont	0.44	0.54	0.58	0.44	0.39	0.55	0.54	0.4	0.45	0.44	1	0.41	0.46	0.41	0.72	0.48	0.5	0.37	0.57	0.54
Novartis	0.3	0.37	0.42	0.32	0.39	0.4	0.38	0.39	0.54	1	0.44	0.61	0.37	0.3	0.4	0.38	0.4	0.42	0.41	0.44
Nestle	0.33	0.36	0.39	0.35	0.42	0.4	0.38	0.36	1	0.54	0.45	0.5	0.38	0.29	0.43	0.38	0.41	0.41	0.39	0.42
Lonza	0.32	0.38	0.41	0.33	0.38	0.37	0.37	1	0.36	0.39	0.4	0.39	0.35	0.33	0.37	0.36	0.33	0.32	0.36	0.37
LafargeHolcim	0.45	0.5	0.56	0.48	0.4	0.54	1	0.37	0.38	0.38	0.54	0.33	0.4	0.44	0.5	0.46	0.53	0.36	0.58	0.48
Julius Baer	0.44	0.52	0.63	0.43	0.39	1	0.54	0.37	0.4	0.4	0.55	0.37	0.42	0.42	0.54	0.51	0.57	0.35	0.62	0.55
Givaudan	0.29	0.37	0.37	0.38	1	0.39	0.4	0.38	0.42	0.39	0.39	0.35	0.36	0.35	0.37	0.33	0.37	0.35	0.38	0.34
Geberit	0.36	0.43	0.44	1	0.38	0.43	0.48	0.33	0.35	0.32	0.44	0.3	0.39	0.43	0.44	0.37	0.37	0.3	0.44	0.37
Credit Suisse	0.48	0.56	1	0.44	0.37	0.63	0.56	0.41	0.39	0.42	0.58	0.4	0.44	0.4	0.55	0.59	0.61	0.4	0.76	0.62
Adecco	0.47	1	0.56	0.43	0.37	0.52	0.5	0.38	0.36	0.37	0.54	0.35	0.43	0.38	0.53	0.49	0.48	0.35	0.54	0.54
ABB	1	0.47	0.48	0.36	0.29	0.44	0.45	0.32	0.33	0.3	0.44	0.32	0.36	0.33	0.44	0.44	0.43	0.28	0.45	0.47

Cor

1.0
0.5
0.0
−0.5
−1.0

Figure 1: Correlation between SMI constituents

2.4 Statistical key numbers

2.4.1 t-Test

A t-test for the correlation as in eq. (5) is conducted to investigate if the true correlation is significantly different from zero.

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

$t = test\ statistic$

$\rho_{i,j} = correlation\ between\ returns\ of\ assets\ i\ and\ j$

$n = sample\ size$

The null hypothesis says that the correlation is not significantly different from zero and the alternative hypothesis that it is significantly different from zero.

$H_0 : \rho = 0$

$H_1 : \rho \neq 0$

The test is performed for the two stocks with the smallest correlation which are *Sika* and *Roche*. When inserting the numbers in eq. (5) as shown in eq. (6), one obtains 18.951.

$$t = \frac{0.264 * \sqrt{4815 - 2}}{\sqrt{1 - 0.264^2}} = 18.951 \qquad (6)$$

With the help of the t-distribution in figure 2 it is seen clearly that a value of approximately 2 is already below the threshold of $\alpha = 0.05$. Therefore, with the obtained test statistic value of 18.951 which is far to the right in figure 2, the p-value is almost zero. Consequentially, the null hypothesis can be rejected, meaning that the true correlation is significantly different from zero.
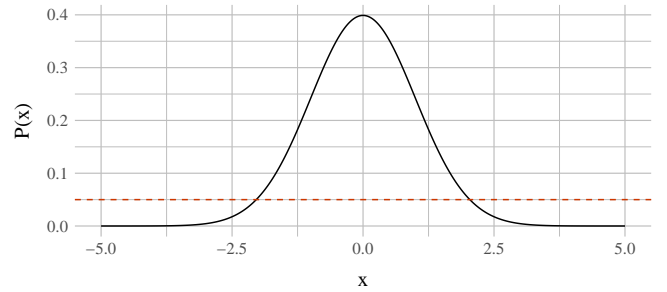
Figure 2: t-distribution with $4815 - 2$ degrees of freedom

2.4.2 Standard error of mean return

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 observations there are, the smaller the standard error is and the more accurately the unknown parameter can be estimated. The standard error of mean return is shown in eq. (7).

$$\sigma_{\overline{R}} = \frac{\sigma}{\sqrt{n}} \qquad (7)$$

$\sigma_{\overline{R}} = standard\ error\ of\ mean\ return$

$\sigma = standard\ deviation$

$n = sample\ size$

Figure 3 depicts the standard error of mean return for each stock. It is seen clearly that stocks as *Nestle* and *Swisscom* with a low volatility in table 2 also have a lower standard error. Likewise, *ABB* and *Credit Suisse* with a high volatility show a higher standard error.
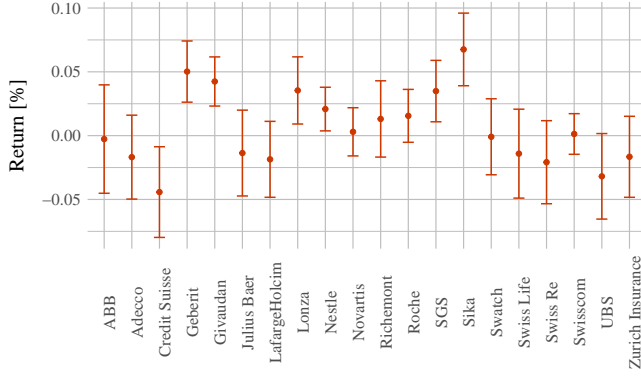

Figure 3: Standard error of mean return

2.4.3 Standard error of volatility (standard deviation)

Eq. (8) is an approximation for the standard error of volatility which is appropriate for n > 10. [5]

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

σ_σ = *standard error of volatility (standard deviation)*

σ = *volatility (standard deviation)*

n = *sample size*

As shown in table 3, the standard errors of volatility are much lower than the ones of the mean returns in figure 3. This property will be clearly visible in chapter 4.2. Furthermore, the standard errors of volatility are neglected for further investigations as they can be estimated almost exactly.

Table 3: Standard error of volatility

Stock	Standard error [%]
ABB	0.03004
Adecco	0.02325
Credit Suisse	0.02513
Geberit	0.01695
Givaudan	0.01362
Julius Baer	0.02378
LafargeHolcim	0.02104
Lonza	0.01860
Nestle	0.01208
Novartis	0.01335
Richemont	0.02113
Roche	0.01466
SGS	0.01701
Sika	0.02012
Swatch	0.02106
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 deals with the construction of portfolios that maximize return for a given level of risk. Harry Markowitz pioneered this theory in his article *Portfolio Selection*. His portfolio theory shows that efficient risk reduction is only possible if the degree of correlation between the individual assets is considered in the construction of the portfolio. Risk reduction through risk diversification is a central insight of the aforementioned theory. The main point is that risk and return characteristics of an asset should not be considered in isolation but be judged by how the asset affects risk and return of the whole portfolio. It is shown that an investor can construct a portfolio of multiple assets that maximizes return for a given level of risk. Likewise, an investor can construct a portfolio with the lowest possible risk for a desired level of return. [6]

It is assumed that return and covariance are known. The portfolio volatility is given as a function of covariance matrix and weights vector and can be minimized by sufficient diversification as shown in eq. (9). The sum of all weights equals one as given in eq. (10). The weights calculated are those which match the portfolio with a given R_p as in eq. (11). This is a linear optimization problem as well as a formulation of the fundamental problem of balancing return and risk. Negative weightings are defined as short sales.

$$min \frac{1}{2} \vec{w}^\intercal \Sigma \vec{w} \tag{9}$$

With constraints:

I.

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

II.

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

With the Lagrange multiplier, the function in eq. (12) is formed.

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

The gradient is set equal to zero as in eq. (13). Together with the two constraints (10) and (11), a linear system of equations with $n + 2$ equations and variables is formed.

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

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

Σ = *covariance matrix*

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

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

R_p = *portfolio return*

3.1 Problems

In order to perform the optimization, some assumptions have to be made. As already mentioned before, return, volatility and correlation must be known. The volatility is described with the standard deviation and therefore no assumption about a normal distribution has to be made. Another condition applies to the linear relationship of the assets. Their correlation is not allowed to be either one or minus one. The determinant of such a correlation or covariance matrix would be zero and therefore it would not be invertible. To calculate the minimum variance and tangency portfolio, the covariance matrix Σ must be invertible.

3.2 Minimum variance portfolio

The minimum variance portfolio, short MVP, describes the portfolio with the lowest volatility. Since only the volatility is minimized, the condition II (11) is not required. Negative returns do not necessarily lead to negative weights. The calculation of the MVP does not depend directly on the returns, but they are reflected in the covariance matrix. The MVP is calculated as illustrated in eq. (14) and the corresponding weights are shown in table 4.

$$\vec{w}_{mvp} = \frac{1}{\vec{1}^\mathsf{T}\Sigma^{-1}\vec{1}} * \Sigma^{-1}\vec{1} \qquad (14)$$

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

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

Σ = covariance matrix

Table 4: Minimum variance portfolio weights

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	0.389
UBS	-0.019
Zurich Insurance	-0.019

3.3 Tangency portfolio

The tangency portfolio, short TP, results from the capital market line and efficient frontier, which will be discussed in more detail in chapter 3.4. The capital market line is an important component of the capital asset pricing model (CAPM). The slope of the capital market line indicates how much more return is expected per additional risk. The optimal ratio is found where the capital market line is tangential to the efficient frontier, hence the tangency portfolio. It is often referred to in the literature as market portfolio. [7]

The capital market line is shown in eq. (15) and the tangency portfolio in eq. (16) with its corresponding weights in table 5. Note that in this paper the risk-free rate R_f is already deducted from the returns directly as shown in eq. (1). Therefore, it can be set to zero in all further calculations.

Capital market line:

$$R_p(\sigma_p) = R_f + \frac{R_{tp} - R_f}{\sigma_{tp}} * \sigma_p = \frac{R_{tp}}{\sigma_{tp}} * \sigma_p \qquad (15)$$

R_p = portfolio return as a function of σ_p

σ_p = portfolio volatility

R_{tp} = TP return

σ_{tp} = TP volatility

R_f = risk-free rate of return

Tangency portfolio:

$$\vec{w}_{tp} = \frac{1}{\vec{1}^\mathsf{T}\Sigma^{-1}\vec{R}} * \Sigma^{-1}\vec{R} \qquad (16)$$

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

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

Σ = covariance matrix

$\vec{R}$ = mean returns

Table 5: Tangency portfolio weights

Stock	Weight
ABB	-0.001
Adecco	-0.190
Credit Suisse	-0.307
Geberit	0.533
Givaudan	0.592
Julius Baer	-0.085
LafargeHolcim	-0.344
Lonza	0.222
Nestle	0.339
Novartis	-0.321
Richemont	0.200
Roche	0.165
SGS	0.301
Sika	0.505
Swatch	-0.194
Swiss Life	-0.029
Swiss Re	-0.062
Swisscom	-0.143
UBS	-0.141
Zurich Insurance	-0.038

3.4 Efficient Frontier

The efficient frontier is a set of points that extends in the return-volatility diagram illustrated in figure 4 between the MVP at the left edge of the reachable area and the TP. All portfolio above the MVP on this line are efficient because they have the highest possible return at a defined level of volatility.

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

$\vec{w}_{ef} = $ *efficient frontier weights*

$\alpha = $ *factor*

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

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

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

$R_p = $ *portfolio return*

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

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

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

$\sigma_p = $ *portfolio volatility*

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

$\Sigma = $ *covariance matrix*
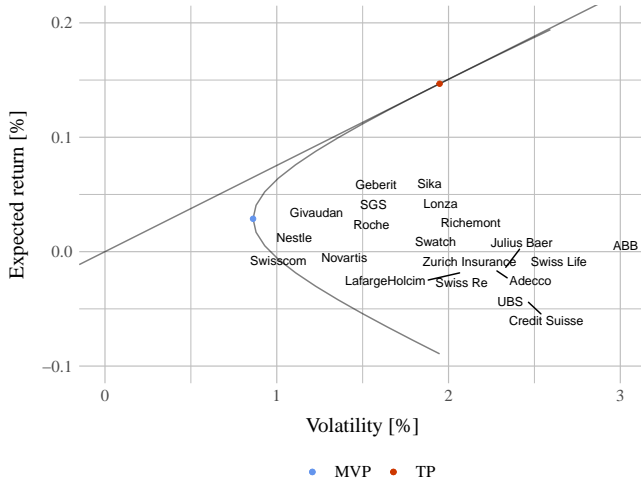
Figure 4: Return-volatility diagram with efficient frontier of SMI constituents

In figure 4 is visible that the efficient frontier covers the range in which all possible portfolios are located. The TP achieves a significantly higher return than the stocks individually. The MVP is at the very left of the achievable area of the efficient frontier which means that there is no portfolio with a lower volatility.

3.5 Sharpe ratio

The Sharpe ratio measures the performance of an asset, meaning its return compared to its risk. Generally, the greater the value of the Sharpe ratio, the more attractive the asset. In practice, the Sharpe ratio is not only positively received. In this paper however, the Sharpe ratio is relied on because it is a simple performance indicator. It is calculated by the return in excess of the risk-free rate per unit of volatility. As daily historical data is used in this paper, the Sharpe ratio is annualized by multiplication with $\sqrt{252}$. [7]

$$Sharpe\,ratio = \frac{R - R_f}{\sigma} = \frac{R}{\sigma} \tag{20}$$

$R = $ *return*

$R_f = $ *risk-free rate of return*

$\sigma_i = $ *volatility of return*

4 Methodology

4.1 Grouping

Grouping of assets is an approach to reduce the noise given by individual assets and therefore increase their performance. A qualitative and a mathematical procedure are applied.

4.1.1 Grouping by industry

An intuitive approach is grouping stocks by industry. In the case of the SMI, stocks can be grouped as follows:

Consumer: *Adecco, Nestle, Richemont, Swatch, Swisscom*

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

Industrial: *ABB, Geberit, Givaudan, LafargeHolcim, SGS*

Pharma: *Lonza, Novartis, Roche, Sika*

To obtain the group returns, the mean value of the daily returns of all stocks within the group is calculated. Therefore, the group returns are a time series equaling the length of the historical data and all stocks within the group are weighted equally. From this time series, the mean returns, volatility and correlation can be calculated as shown in table 6.

Table 6: Mean returns and volatility of groups by industry

Group	Return [%]	Volatility [%]
Consumer	0.003	1.342
Finance	-0.024	1.902
Industrial	0.021	1.397
Pharma	0.030	1.189

4.1.2 Grouping with cluster analysis

In a second approach grouping is based on the similarity measure which is calculated with Euclidean distances between two correlation coefficients as shown in eq. (21).

$$d_{i,j} = \sqrt{\vec{1}^{\intercal}(\vec{x}_i - \vec{x}_j)^2} \tag{21}$$

$d_{i,j}$ = Euclidean distance between asset i and asset j

$\vec{x}_i$ = correlation coefficients of asset i (column in cor. matrix)

$\vec{x}_j$ = correlation coefficients of asset j (column in cor. matrix)

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

When eq. (21) is applied to the correlation matrix shown in figure 1, it results in a distance matrix with the same dimensions. The algorithm sorts the values of the distance matrix with the complete-linkage method.

Step 1 Group the two assets with the smallest distance value.

Step 2 Compare the created group with individual assets or other created groups. The distance between a group i, j and an individual asset k is the largest possible distance between an asset within the group and the individual asset.
$d\langle\{i, j\}, k\rangle = max(d\langle i, k\rangle, d\langle j, k\rangle)$.
For distances between groups, the largest possible distance is taken from the distances between each asset of a group with an asset of another group. For distance between group i, j and k, l with two assets applies:
$d\langle\{i, j\}, \{k, l\}\rangle = max(d\langle i, k\rangle, d\langle i, l\rangle, d\langle j, k\rangle, d\langle j, l\rangle)$.

Step 3 Repeat the previous steps with the newly created distance matrix until only two groups remain. The outcome is a dendrogram which is also part of the illustration in figure 5. The first step of grouping is the connection which is drawn lowest. The color scaling indicates the different distances. [8]
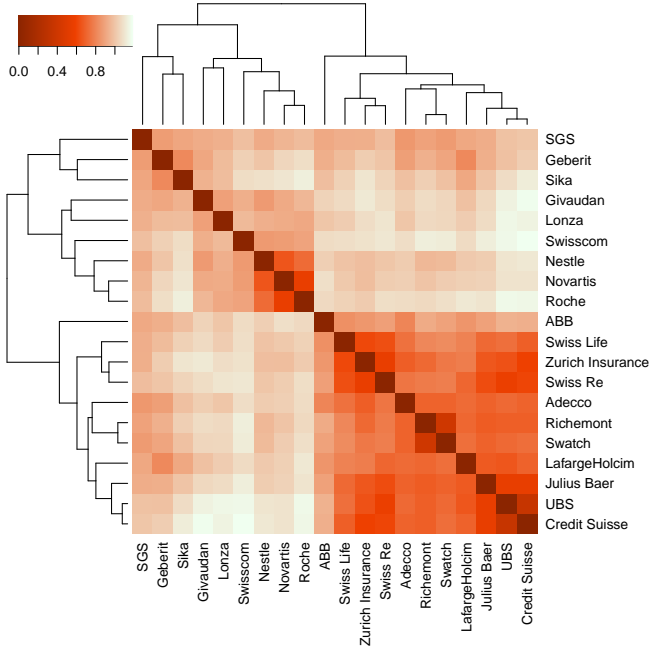
Figure 5: Stocks grouped by Euclidean distance of correlation

Group 1: *ABB, Adecco, Credit Suisse, Julius Baer, LafargeHolcim, Richemont, Swatch, Swiss Life, Swiss Re, UBS, Zurich Insurance*

Group 2: *Givaudan, Lonza, Nestle, Novartis, Roche, Swisscom*

Group 3: *Geberit, SGS, Sika*

Using the algorithm described, the 20 stocks can now be divided into three groups. The dendrogram in figure 5 shows the allocation. This division makes mathematical sense but is hardly ever used in practice. First, the groups are different in terms of size. *Group 1* consists of eleven stocks but *Group 3* of only three. This means that with equal weights within the group, the stocks in *Group 3* are weighted higher overall than those in *Group 1*. Second, one would have to explain to an investor why stocks from different industries are grouped together. For example, *ABB* is grouped with *UBS* but not with *Geberit* which is intuitively contradictory.

Figure 6 gives some visual intuition. The groups contain stocks with similar volatility and return, so it makes sense that *ABB* is grouped with *UBS* because and not with *Geberit*.
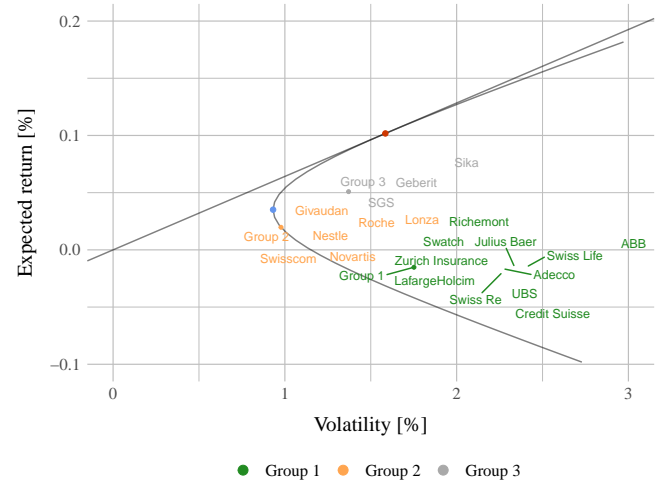
Figure 6: Return-volatility diagram with efficient frontier of groups by correlation

Table 7 shows return and volatility of the three groups. It is visible that the values have improved overall compared to table 6. Further cluster analysis methods such as k-means or k-medoids could be considered but are not discussed in this paper.

Table 7: Mean returns and volatility of groups by correlation

Group	Return [%]	Volatility [%]
Group 1	-0.015	1.752
Group 2	0.020	0.978
Group 3	0.051	1.371

4.2 Bootstrap

Bootstrapping is a method for resampling. It creates a new sample by drawing from an existing sample. With 100 or more samples, statistical measures such as mean and variance can be calculated across all samples. This is useful to draw conclusions about a population from just one sample.

In this case, bootstrapping is applied to analyze the standard deviation of the MVP and TP weights by resampling all historical returns.

Step 1 The historical returns have the structure of a matrix with the rows being each day and the columns each stock. This matrix is sampled row-wise using `sample()` with `replace = T`. In other words, one row, which corresponds to one day, is drawn randomly from the initial matrix as many times as the initial matrix has rows. As `replace = F`, one row/day can be drawn multiple times. If `replace = F`, the correlation matrix of the sample would be equal to the one of the initial matrix which is not desired. The result of this step is a sample returns matrix equaling the dimensions of the initial historical returns matrix.

Step 2 From the sample returns matrix, the mean return and covariance matrix are calculated to obtain the MVP and TP weights.

Step 3 Repeat step one and two 100 times to obtain 100 sets of MVP and TP weights. The standard deviation across these 100 sets of weights is calculated per stock. It is computed with the 84-quantile minus the 50-quantile to achieve a more robust result.

In figure 7 all bootstrap samples are visualized with their MVP and TP as well as efficient frontier. Noticeable is the variance of the TP compared to the MVP. This is due to the high standard errors of the returns. A more detailed explanation is given in the results 5.1.
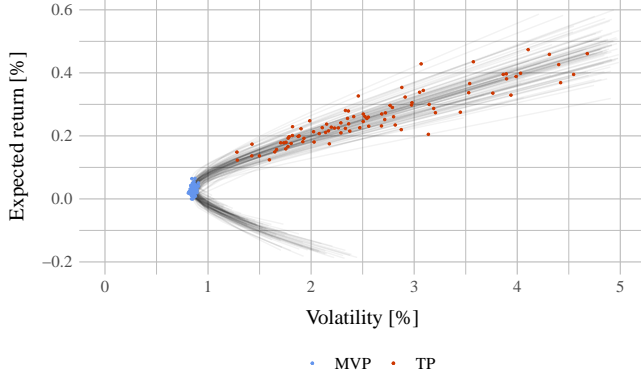
Figure 7: Bootstrap samples efficiency frontier

4.3 Cross validation

Cross validation is a model validation technique for assessing how a model will generalize to an independent data set, i.e., in another time period. 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 historical data is split into five sections and consequentially five models are trained as depicted in figure 8. The gray sections are the training sets and the red ones the test sets.
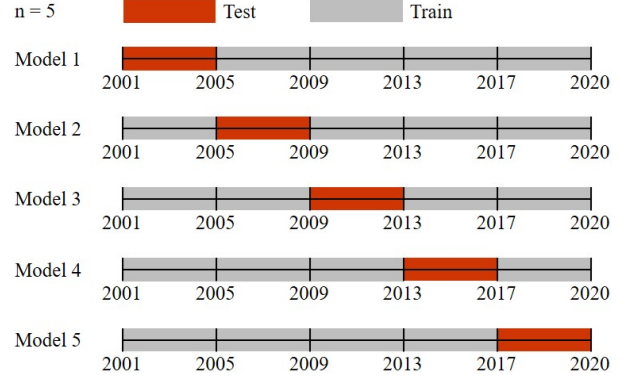
Figure 8: Cross validation training and test sets

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 eq. (22).

$$R_{i,w} = \vec{w}^\intercal * \vec{R}_i \qquad (22)$$

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

$\vec{w} = asset\ weights$

$\vec{R}_i = asset\ returns\ day\ i$

At the end of the cross validation, all five test sets are combined into one time series consisting daily weighted returns and equaling the length of the historical data. This time series is entirely out of sample and is used to calculate the Sharpe ratio and hence measure the performance of the model. The implicit assumption behind this procedure is that the returns at different points in time are independent random variables (and, in particular, do not "remember" the past).

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 two in figure 8 includes this time period as part of the test set and model four of the training set. This results in fluctuating asset weights across the five models which are illustrated in figure 9.
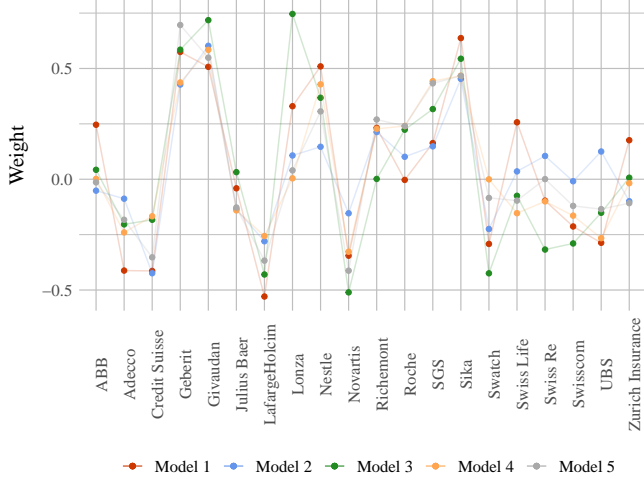
Figure 9: Tangency portfolio weights by model

4.4 Shrinkage

Shrinkage is an approach with the objective 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 shrunk mean returns are calculated as in eq. (23).

$$\vec{R}(\lambda) = \lambda * \vec{R} + (1 - \lambda) * \vec{\bar{R}} \tag{23}$$

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

$\lambda = $ *shrinkage factor*

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

$\vec{\bar{R}} = $ *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 eq. (16) obtains the same asset weights as the minimum variance portfolio eq. (14).

There is an important constraint in mean returns shrinkage. Take the scaling factor of the tangency portfolio eq. (16) and replace $\vec{R}$ with eq. (23) as shown in eq. (24).

$$\vec{w}_{tp,sf} = \frac{1}{\vec{1}^\intercal \Sigma^{-1} (\lambda * \vec{R} + (1 - \lambda) * \vec{\bar{R}})} \tag{24}$$

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 eq. (25).

$$\lambda_0 = \frac{-\vec{1}^\intercal \Sigma^{-1} \vec{\bar{R}}}{\vec{1}^\intercal \Sigma^{-1} \vec{R} - \vec{1}^\intercal \Sigma^{-1} \vec{\bar{R}}} \tag{25}$$

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 in eq. (24) becomes very large which affects the weights. To illustrate this effect, the weights of the five cross validation models from figure 8 are plotted as a function of λ in figure 10.
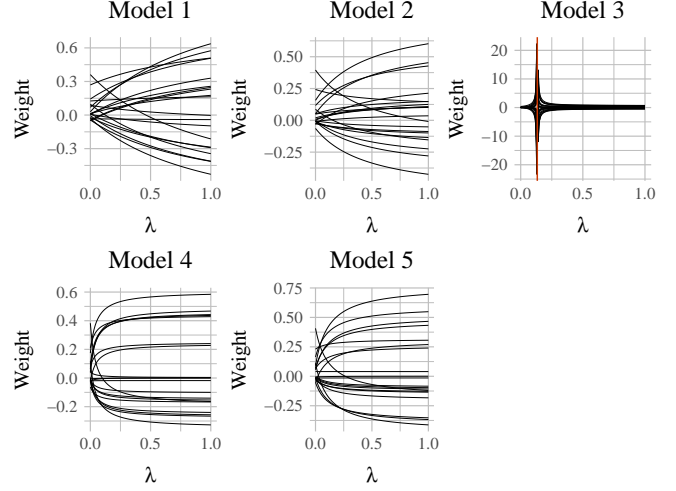
Figure 10: Tangency portfolio weights by model as a function of λ

Model three in figure 10 shows such extreme asset weights ranging from approximately -20 to 20. The asymptote is visualized as a vertical red line. It is not optimal when such extreme weights are applied to returns as done in the cross validation. This can negatively influence further calculations and lead to undesired results. The other four models do show asset weights in a normal scale.

The scaling factor as a function of λ is shown in figure 11.
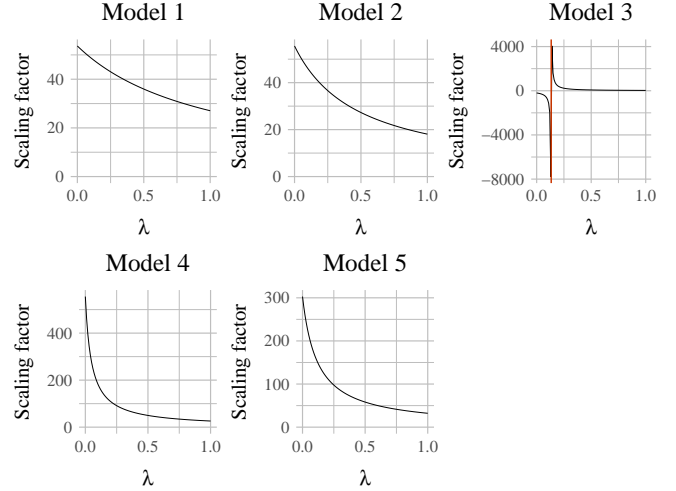
Figure 11: Tangency portfolio weights scaling factor by model as a function of λ

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

4.4.2 Shrinkage of correlation matrix

The shrunk correlation matrix is calculated as in eq. (26).

$$\rho_{i,j}(\epsilon) = I_{i,j} + \epsilon * \tilde{\rho}_{i,j} \qquad (26)$$

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

$\epsilon =$ shrinkage factor

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

$\tilde{\rho}_{i,j} =$ correlation matrix with diagonal zero at i,j

When the shrinkage factor ϵ is set to one, the correlation matrix remains unchanged. When it is decreased from one towards zero, the correlation coefficients converge towards zero, except for the diagonal which always remains one. As the tangency portfolio eq. (16) requires the covariance matrix, it is first standardized to the correlation matrix according to eq. (4), then shrunk with eq. (26) and scaled up to the covariance matrix again according to eq. (4).

There can also arise the case, where for some shrinkage factor ϵ the denominator of the tangency portfolio eq. (16) becomes zero or close to zero. However, this is not further elaborated on.

5   Results

The results of the different approaches described in chapter 4 are presented and compared by their Sharpe ratio.

5.1   SMI constituents

Before the results of the optimization for all individual stocks are shown, the standard deviation of the bootstrapped MVP and TP weights are compared. Figure 12 illustrates the mean weights and their standard deviation of the MVP and figure 13 of the TP.
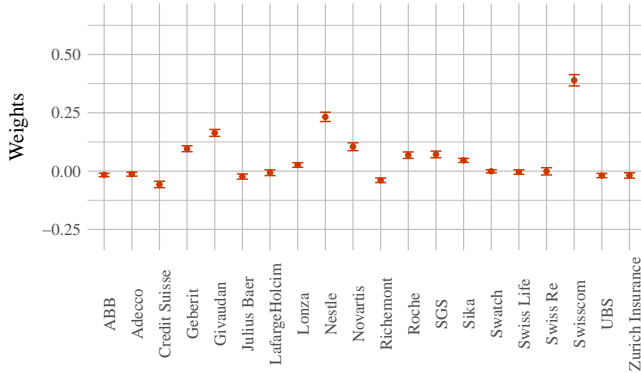
Figure 12: MVP weights with standard deviation
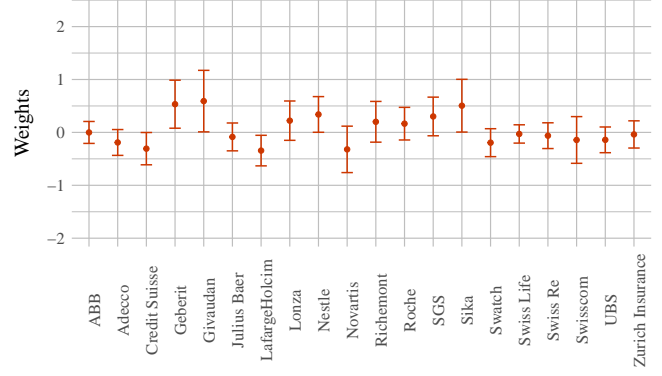
Figure 13: TP weights with standard deviation

It is seen clearly in figure 12 and 13 and also visualized before in figure 7 in chapter 4.2, that TP weights have a higher standard deviation than MVP weights. When considering the MVP eq. (14) and TP eq. (16), it is evident that latter directly depends on the returns which have high standard errors. This carries over to the standard deviation of the TP weights. Although the returns also impact the covariance, it only varies to a small degree which is seen in the small standard deviation of the MVP.

Table 8 compares the in and out of sample Sharpe ratio for all 20 stocks.

Table 8: Sharpe ratio of SMI constituents

	In sample	Out of sample
SMI constituents	1.196	0.489

In sample, training and test set are equal and therefore its Sharpe ratio is higher than out of sample. This is because the model performs best on the data which built it. Out of sample, training set and test set differ. If a stock behaves differently in a test set compared to the training set used to build the model, this stock can easily be over- or undervalued. This leads to a worse performance of the model and therefore lower Sharpe ratio. The conclusion that can be drawn from this insight is, that the in sample Sharpe ratio should always be higher than the out of sample Sharpe ratio.

5.2 Groups by industry and correlation

Figure 14 illustrates the mean weights and their standard deviation of the MVP and figure 15 of the TP.
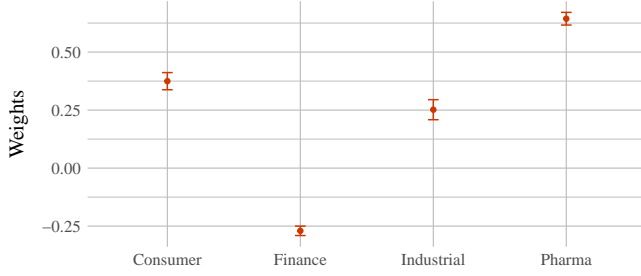

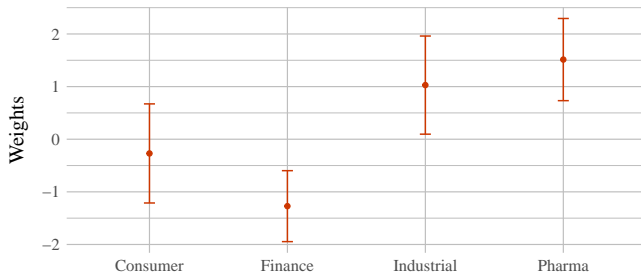Figure 14: MVP weights with standard deviation by group

Figure 15: TP weights with standard deviation by group

The same conclusion can be drawn for the grouped stocks as for the individual stocks in chapter 5.1. The TP weights standard deviation is also much larger than the one of the MVP.

In table 9, the in and out of sample Sharpe ratios for both groups is shown.

Table 9: Sharpe ratio of groups

	In sample	Out of sample
Groups by industry	0.811	0.608
Groups by correlation	1.019	0.991

The first observation is, that the in sample Sharp ratio decreased after the grouping is applied compared to the individual stocks in table 8. The reason being, that with 20 parameters a better fit to the data is achieved than with three or four parameters as it is the case for the groups. Nevertheless, the out of sample Sharpe ratio has increased. Due to grouping, the returns are less noisy than the ones of the individual stocks. This leads to a more reliable model which performs better out of sample.

In chapter 4.1.2, cluster analysis is used to optimize the group composition. Hence, higher correlating stocks are grouped together to obtain a better diversified portfolio than is achieved with grouping by industry. This leads to a higher in and out of sample Sharpe ratio.

5.3 Optimal shrinkage factors

5.3.1 Shrinkage factors analyzed separately

The Sharpe ratio can be visualized as a function of the shrinkage factor. Figure 16 shows such a plot for the mean returns shrinkage factor.

When looking at the SMI constituents curve, it is seen that the highest Sharpe ratio is achieved with a shrinkage factor of 0.05. If past returns are taken too seriously (higher shrinkage factor) or not considered at all (shrinkage factor zero), the out of sample Sharpe ratio drops. So, the returns can be trusted only five percent. This result is plausible, as the returns are very noisy. Further can be seen that the Sharpe ratio has a sharp bend and drop below zero. This is caused by the close to zero denominator of the tangency portfolio eq. (16) as described in chapter 4.4.

For the two group curves the result is inverted. As the grouping of the stocks already reduces the noise significantly, shrinkage does show almost no improvement over unshrunk mean returns. The groups by correlation do show a significantly higher Sharpe ratio than the groups by industry and it is also higher than the one of the SMI constituents.
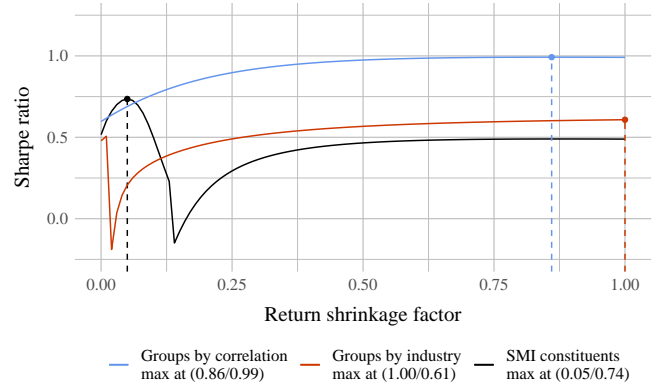

Figure 16: Sharpe ratio as a function of return shrinking factor

Figure 17 shows the plot for the correlation shrinkage factor. For the SMI constituents, the same observation can be made as in figure 16. The out of sample Sharpe ratio drops, if past correlations are taken too seriously (higher shrinkage factor) or less seriously (shrinkage factor zero). The optimum is found at 0.4. Also, the result for groups is similar to the previous one in the sense that shrinkage does not have a great impact. The highest Sharpe ratio is again achieved by the groups by correlation.
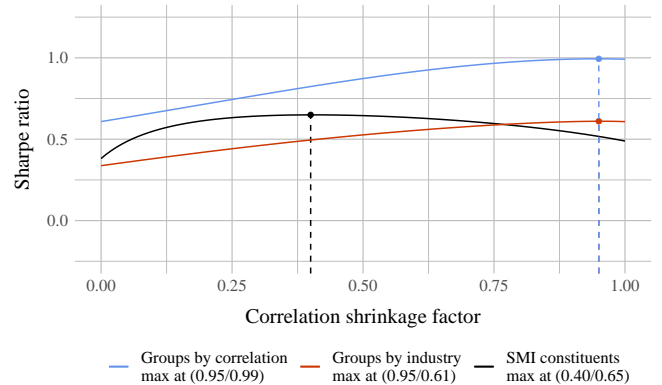

Figure 17: Sharpe ratio as a function of correlation shrinking factor

5.3.2 Shrinkage factors analyzed simultaneously

A further visualization is a heatmap plot where the Sharpe ratio is plotted as a function of both shrinkage factors. The darker the area, the higher the Sharpe ratio.

Figure 18 shows such a plot for the SMI constituents. In the bottom part and left part of the plot the contour lines are very chaotic which is caused by the Sharpe ratio dropping quickly and below zero. The highest Sharpe ratio is achieved with a mean returns shrinkage factor of 0.05 and a correlation shrinkage factor of one. This matches the result of the figure 16 and 17 where mean return shrinkage is better at improving the Sharpe ratio than correlation shrinkage, which does not matter in this case.
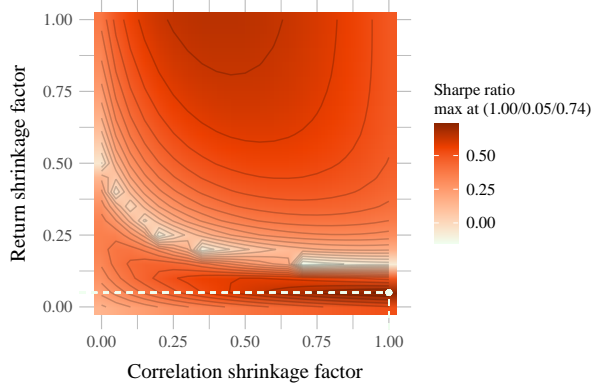
Figure 18: SMI constituents Sharpe ratio as a function of return and correlation shrinkage factor

Figure 19 shows the plot for the groups by industry and figure 20 for the groups by correlation. Both plots are very similar in their general structure. The groups by industry also show a sudden decrease in Sharpe ratio in the lower portion while the groups by correlation do not. Overall, latter does show a Sharpe ratio which is 0.38 higher than the one of the groups by industry.
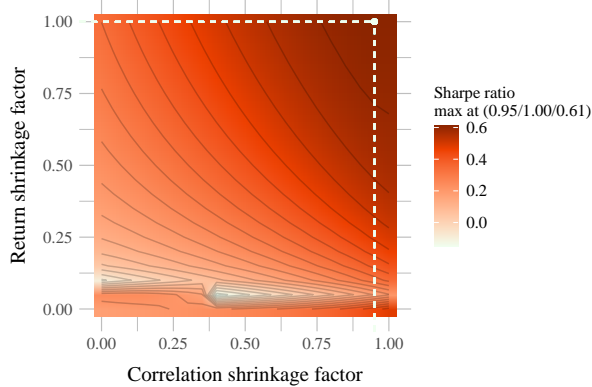
Figure 19: Groups by industry Sharpe ratio as a function of return and correlation shrinkage factor
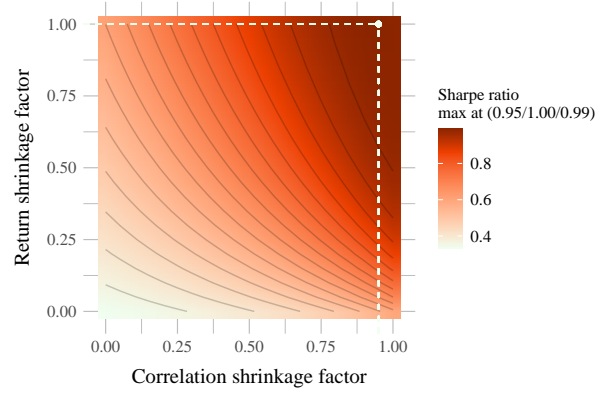
Figure 20: Groups by correlation Sharpe ratio as a function of return and correlation shrinkage factor

Table 10 sums up all investigated Sharpe ratios, namely in sample (IS), out of sample (OS) and out of sample with applied shrinkage (OS shrinkage).

Table 10: Sharpe ratio overview

	IS	OS	OS shrinkage
SMI constituents	1.196	0.489	0.735
Groups by industry	0.811	0.608	0.610
Groups by correlation	1.019	0.991	0.994

To give further insights, the TP weights for unshrunk and shrunk mean returns and correlations are compared in tables 11, 12 and 13.

The weights of unshrunk mean returns and correlations are the ones previously shown in table 5 in chapter 3. They are also the ones which obtain the in sample Sharpe ratio, where all historical data is used to train the model. The shrinkage factors utilized for the shrunk mean returns and correlations are the ones which obtained the highest out of sample Sharpe ratio in figure 18, 19 and 20.

Table 11: Tangency portfolio weights

Stock	Before shrinkage	After shrinkage
ABB	-0.001	-0.013
Adecco	-0.190	-0.053
Credit Suisse	-0.307	-0.114
Geberit	0.533	0.195
Givaudan	0.592	0.260
Julius Baer	-0.085	-0.037
LafargeHolcim	-0.344	-0.083
Lonza	0.222	0.070
Nestle	0.339	0.257
Novartis	-0.321	0.009
Richemont	0.200	0.015
Roche	0.165	0.090
SGS	0.301	0.123
Sika	0.505	0.150
Swatch	-0.194	-0.044
Swiss Life	-0.029	-0.010
Swiss Re	-0.062	-0.015
Swisscom	-0.143	0.269
UBS	-0.141	-0.046
Zurich Insurance	-0.038	-0.023

When comparing the two sets of weights, a significant difference can be seen. Overall, the weights after shrinkage are closer around zero as they are before. This is caused by the fact that the returns are taken into account only five percent. Therefore, the weights are much closer to the MVP weights as seen in table 4 in chapter 3.

Compared to the weights of the individual stocks already shown in table 5, the weights of the groups in table 9 and 10 are significantly higher or lower. To recall, the weights must sum up to one according to constraint (10). If the groups perform differently, like it is the case for *Finance* (negative mean return) and *Pharma* (positive mean return) in the groups by industry, the weights drift apart. This becomes even more prominent in the groups by correlation between *Group 1* and *Group 3*. As mentioned in chapter 3, negative weights correspond to short sales. The difference between the weights before and after shrinkage is insignificant. This is attributed to the shrinkage factor of the mean returns being one and the one for the correlation being 0.95. So there is almost no shrinkage applied.

Table 12: Groups by industry in sample tangency portfolio weights

Group	Before shrinkage	After shrinkage
Consumer	-0.270	-0.268
Finance	-1.272	-1.197
Industrial	1.028	0.950
Pharma	1.513	1.515

Table 13: Groups by correlation in sample tangency portfolio weights

Group	Before shrinkage	After shrinkage
Group 1	-1.155	-1.080
Group 2	0.818	0.776
Group 3	1.337	1.304

6   Conclusion

Considered in isolation, the highest increase of Sharpe ratio is achieved by grouping by correlation using cluster analysis. Nevertheless the highest Sharpe ratio is obtained with a combination of grouping and shrinkage. It should be noted that when one method is applied after the other, its influence on the Sharpe ratio is significantly lower. This has to do with the fact that the noise can only be reduced to a certain degree. In this case, the first method grouping, has a larger increase in Sharpe ratio, while shrinkage, as the second method, has only a small additional increase. However, it is well apparent that for the SMI constituents where only shrinkage is applied, the rise in Sharpe ratio is similarly high compared with grouping as the first method.

With these insights, the best procedure for optimizing multiple assets can be derived. In a first step, one would group by correlation to reduce noise by a significant amount. The second step would be applying shrinkage to mean returns and correlation. Examining the groups implies that returns remain unshrunk, while correlation is shrunk by a factor of 0.95. While this is not generally valid, overall it can be said that shrinkage offers only a small gain in addition to the previous grouping. But if grouping is omitted, it is possible that a higher Sharpe ratio can be achieved with the shrinkage approach, especially with a high number of assets.

7   References

[1] Yahoo Finance, "SMI pr (ˆSSMI)." [Online]. Available: https://finance.yahoo.com/quote/%5ESSMI/components?p=%5ESSMI. [Accessed: 17.10.2020]

[2] SIX, "SMI® – the blue-chip index." [Online]. Available: https://www.six-group.com/exchanges/indices/data_centre/shares/smi_en.html. [Accessed: 15.09.2020]

[3] SIX, "Swiss reference rates (saron)." [Online]. Available: https://www.six-group.com/exchanges/indices/data_centre/swiss_reference_rates/reference_rates_en.html. [Accessed: 06.10.2020]

[4] global-rates.com, "CHF libor interest rate - swiss franc libor." [Online]. Available: https://www.global-rates.com/en/interest-rates/libor/swiss-franc/swiss-franc.aspx. [Accessed: 06.10.2020]

[5] S. Ahn and J. Fessler, "Standard errors of mean, variance, and standard deviation estimators," Jan. 2003.

[6] H. Markowitz, "Portfolio selection," *The Journal of Finance*, vol. 7, p. 77, Mar. 1952, doi: 10.2307/2975974.

[7] W. Sharpe, "Capital asset prices: A theory of market equilibrium under conditions of risk," *The Journal of Finance*, vol. 19, pp. 425–442, Sep. 1964, doi: 10.1111/j.1540-6261.1964.tb02865.x.

[8] A. Ruckstuhl, "Statistisches data mining," pp. 37–47, 65–70, 2020.

List of Tables

List of Figures