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

14. December 2020

Abstract

Markowitz optimization requires estimates of the expected return, standard deviation, and correlations of individual stocks, all of which are subject to estimation errors. 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 asset grouping, shrinkage estimators for historical returns and correlations, and bootstrapping. The goal is to provide a recommended methodology for optimizing a portfolio that includes all stocks from the Swiss Market Index. In summary, two key observations are made. First, asset grouping leads to better portfolio performance relative to individual assets. Similarly, shrinkage leads to increased performance for assets with noisy historical data.

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. [1] [2] **min. halbe Seite lang**

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 shown in table 2 are calculated with eqs. (1) and (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. 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 describes the risk of an asset and is a statistical measure of the dispersion of 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 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 is the standardized covariance and ranges in an interval of $[-1, 1]$ with the diagonal values 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.

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

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

$x = return\ of\ asset\ x$

$y = return\ of\ asset\ y$

$n = sample\ size$

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

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

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

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

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

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 a rather high correlation between them. The correlation between companies from the same industry is also higher than between companies from different industries. It can be concluded that economic similarities are reflected in a higher correlation between stocks. A correlation significance test is performed in chapter 2.4.1.
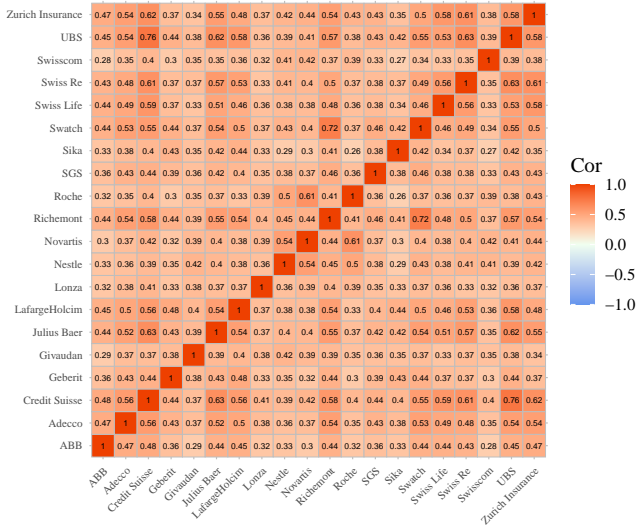

Figure 1: Correlation between SMI constituents

2.4 Statistical key numbers

2.4.1 T-statistic

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

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

$t = t\text{-}statistic$

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

$n = sample\ size$

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

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

The obtained result 18.951, when compared to the t-distribution in figure 2, has a p-value clearly below the threshold of $\alpha = 0.05$. Therefore, the null hypothesis can be rejected meaning that the true correlation is significantly different from zero.
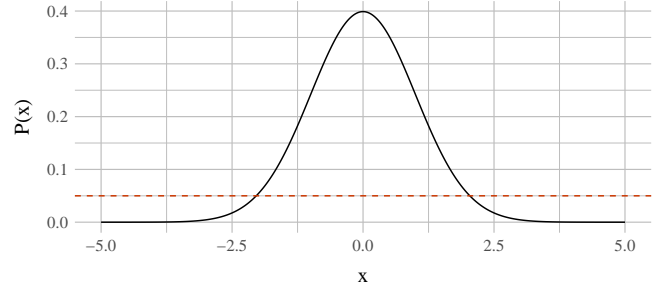

Figure 2: t-distribution with 4815-2 degrees of freedom

2.4.2 Standard error of 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.

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

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

$\sigma = standard\ deviation\ of\ sample$

$n = sample\ size$

Figure 3 shows the logarithmic mean return with the standard errors for each stock. It is easy to see that stocks such as *Swisscom* and *Nestle* with low volatility (table 2) also have a lower standard error. As well *ABB* and *Credit Suisse* have a high volatility and accordingly show higher values for the standard error.
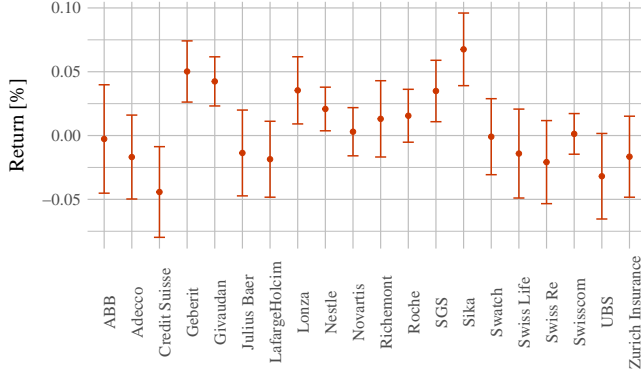
Figure 3: Expected mean return with standard error

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]

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

$\sigma_\sigma = standard\ error$

$\sigma = standard\ deviation$

$n = sample\ size$

As shown in table 3 the values of the standard error of volatility are much lower as the ones of the logarithmic mean returns. This behavior will be clearly visible later in chapter 4.2. Furthermore, we will neglect the values for any further calculations since these values can be estimated nearly 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 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]

It is also assumed that 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, in practice negative weightings are defined as short sales. [9]

$$minimize: \frac{1}{2}\vec{w}^\mathsf{T}\Sigma\vec{w} \qquad (9)$$

With the following two constraints:

I.

$$1 = \vec{w}^\mathsf{T}\vec{1} \qquad (10)$$

II.

$$R_p = \vec{w}^\mathsf{T}\vec{R} \qquad (11)$$

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

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

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 eqs. of the dimension $n + 2$ with $n + 2$ variables. The solution is a known standard problem from linear algebra.

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

$\vec{w} = weight\ vector$

$\Sigma = covariance\ matrix$

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

$\vec{R} = return$

$R_P = total\ portfolio\ return$

3.1 Problem

In order to apply the optimization, some assumptions have to be made at the beginning. As already mentioned before, the returns, risk and correlation coefficients must be known. It is also important that the risk is described by the standard deviation. This leads to the conclusion that a normal distribution is irrelevant, because the standard deviation is used as risk. Another condition applies to the linear relationship of the stocks. The log returns must not have a perfect correlation among themselves, neither negative nor positive. This is described by a correlation coefficient of -1 or 1. If the determinant of the correlation matrix is zero, the system cannot be solved. This situation occurs when stocks have a correlation coefficient of 1 or -1. The system must be uniquely solvable to form the inverse of the covariance matrix Σ^{-1}.

3.2 Minimum variance portfolio

The minimum variance portfolio, short MVP, describes the portfolio of all possible weightings with the minimum volatility. Since only the volatility is minimized, constraint *II* (11) is not included in the eq.. Negative returns do not necessarily have to result in a negative weighting, since in the calculation of the MVP the return is not directly included in the calculation but only the covariance matrix. Therefore, the correlation coefficients are the important factor. The MVP weights are shown in table 4.

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

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

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

Σ = *covariance matrix*

Table 4: Weights of minimum variance portfolio

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 tangent of the capital market line and the efficient frontier, which will be shown more detailed 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 volatility. 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.

The tangency portfolio is often referred to in the literature as the market portfolio. In the equation for the capital market line, shown as eq. (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:

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

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

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

Σ = *covariance matrix*

$\vec{R}$ = *return*

Table 5: Weights of tangency portfolio

Stock	Weight
ABB	-0.001
Adecco	-0.190
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 (figure 4) between the minimum variance portfolio at the left edge of the reachable area and the tangency portfolio. All possible portfolio weights above the MVP 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

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

R_P = total portfolio return

$\vec{w}$ = weights

$\vec{R}$ = mean return

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

σ_P = total portfolio volatility

$\vec{w}$ = weights

Σ = covariance matrix
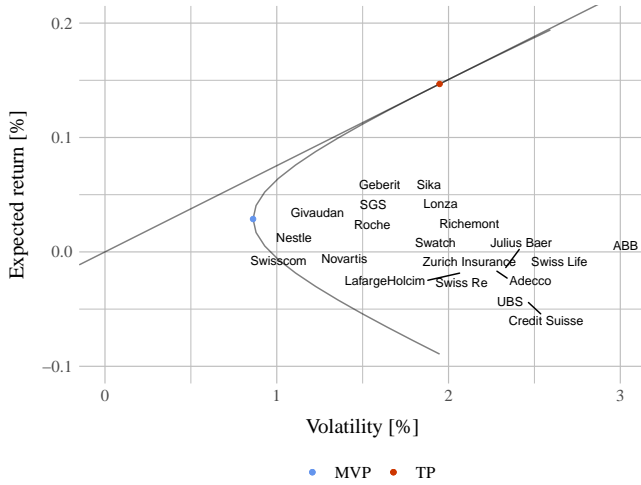
Figure 4: Return-volatility diagram with efficient frontier

In figure 4 it is clearly visible that the efficient frontier covers the range in which all possible portfolio compositions can be located. Through the optimization, the TP is created, which achieves a significantly higher return than the individual stocks included in the SMI. Also, the MVP is at the extreme left of the achievable area on the efficient frontier, which means that there is no portfolio that have a lower volatility.

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 portfolio. 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 eq. (1). Note that the Sharpe ratio is annualized, which means that it is multiplied by $\sqrt{252}$. [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

4 Methodology

4.1 Grouping

Grouping of stocks is an approach to reduce the noise given by individual assets and therefore increase the Sharpe ratio. A more qualitative and a mathematical approach are applied.

4.1.1 Grouping by industry

An intuitive approach is grouping stocks by industry. In the case of the SMI the 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 matrix can be calculated as shown in table 6.

Table 6: Mean returns and Volatility of groups

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. The formula is illustrated below.

$$d_{i,j} = \sqrt{\vec{1}^{\mathsf{T}}(\vec{x_i} - \vec{x_j})^2} \tag{21}$$

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

$\vec{x_i}$ = *correlation coefficients of stock i (column in corr. matrix)*

$\vec{x_j}$ = *correlation coefficients of stock j (column in corr. matrix)*

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

When eq. @ref{eq:eucledean} is applied to the correlation matrix shown in figure @ref{fig:corplot}, 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 stocks with the smallest distance value.

Step 2 Compare the created group with individual stocks or other created groups. The distance between a group (i, j) and a individual stock k is the largest possible distance between a stock within the group and the individual stock.
$d\langle\{i, j\}, k\rangle = max(d\langle i, k\rangle, d\langle j, k\rangle)$.
For distances between groups, the largest possible distance is taken from the distances between each stock of a group with a stock of another group. For distance between group (i, j) and (k, l) with two stocks 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 procedure with the new created distance matrix until only two groups remain. The outcome is a dendogram which is also part of the illustration in figure 5. It can be seen that the first step of grouping is the connection which is drawn lowest. The color scaling describes the different distances. [10]

Figure 5: Grouped by Euclidean distance of correlation coefficients

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 cluster analysis, 20 stocks can now be divided into three groups. The dendrogram in figure 5 clearly shows the allocation. This division makes mathematical sense, but is hardly ever used in practice. First, the groups are different in terms of number of shares. *Group 1* has eleven stocks, but *Group 3* has only three. This would mean that if all shares within a group are weighted equally, the shares of *Group 3* would be weighted much higher than those of *Group 1*. Second, you would have to explain to an investor why a stock from the industrial sector is grouped with a stock from the financial sector. For example ABB is grouped with UBS but not with Geberit, which is intuitively very contradictory.

Nevertheless, figure 6 can be used to provide some clarity. This plot underlines the division and makes clear how the groups are formed. The groups contain stocks with similar volatility and return, so it makes sense that ABB is grouped with UBS because of its high risk and not with Geberit.
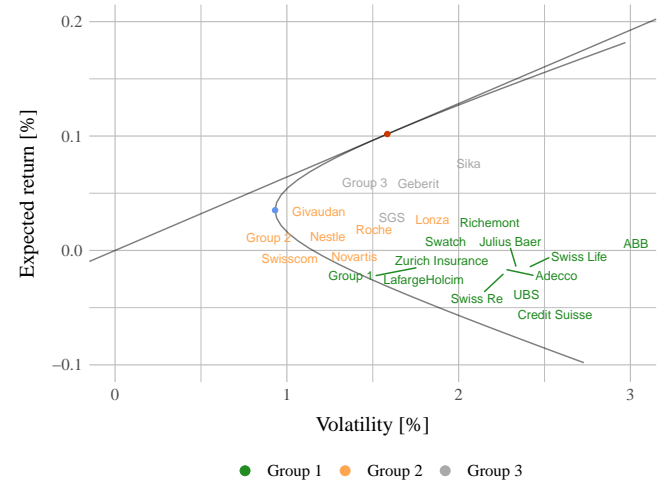

Figure 6: Efficient frontier with stocks

Table 7 shows the returns and volatility of the three groups. It is clearly visible that the values have improved on average compared to table 6. Further methods like k-means or PAM etc. could be considered but are not discussed in this paper.

Table 7: groups with expected return and volatility

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 statistical method of resampling. It allows to create a new sample with drawing from the existing sample. With 100 or more resampled data, each mean and variance of the samples can be analyzed. This method can be used to draw conclusion about a population from its sample.

In this method a bootstrap process is applied to analyze the robustness of the tangency portfolio weights when resampling all historical returns.

Step 1 The dates of the historical data are sampled (using `sample()`) in the same dimension, with replacing included. In other words, the row numbers of the historical data are sampled. With the argument `replace = TRUE` ensures, that certain dates, respectively rows can occur multiple times. If replacing would not be included in the sampling method, the correlation matrix would remain constant to the one of the original data.

Step 2 A new data set is created with the new sorted dates or row numbers. From each sampled date, the corresponding returns of each stock are added to the new data set, consequently maintaining the daily differences between stocks. This new data set is considered as a bootstrap sample.

Step 3 From the bootstrap sample, new mean return of each stock and the corresponding covariance matrix is being calculated. Those two variables are now used as input of the optimization function to calculate minimum variance portfolio and tangency portfolio. The weights of this two portfolios are now being saved in a mvp-weights-matrix and a tp-weights-matrix.

Step 4 Repeat the steps 1-3 100 times. This results in two matrices, each with 100 rows of weights. Those MVP and TP weights can now be compared and analyzed. Additionally, the standard deviation of the resulting weights over those number of bootstrap samples can be investigated. The standard deviation is computed with the 84-quantile minus the 50-quantile to achieve a more robust result, because this calculation is less vulnerable to outliers.

In figure 7 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.
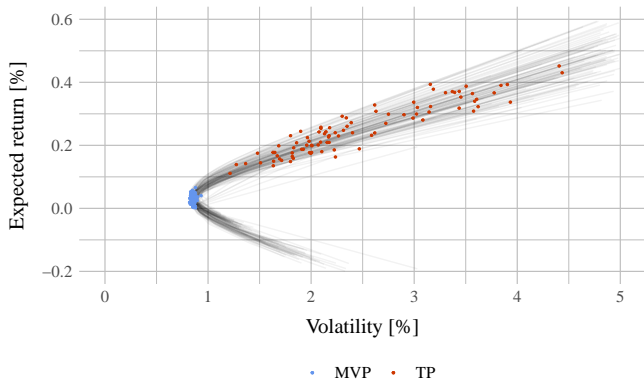

Figure 7: Bootstrap samples efficiency frontier

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).

$$R_{i,w} = \vec{w}^{\mathsf{T}} * \vec{R_i} \tag{22}$$

$R_{i,w}$ = 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).

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

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

$\lambda = $ *shrinkage factor*

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

$\bar{\vec{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).

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

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).

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

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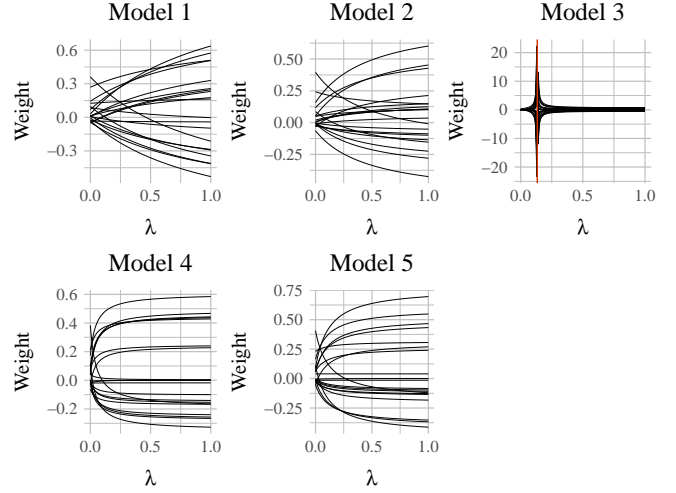
Figure 10: Tangency portfolio weights by model as a function of λ

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.

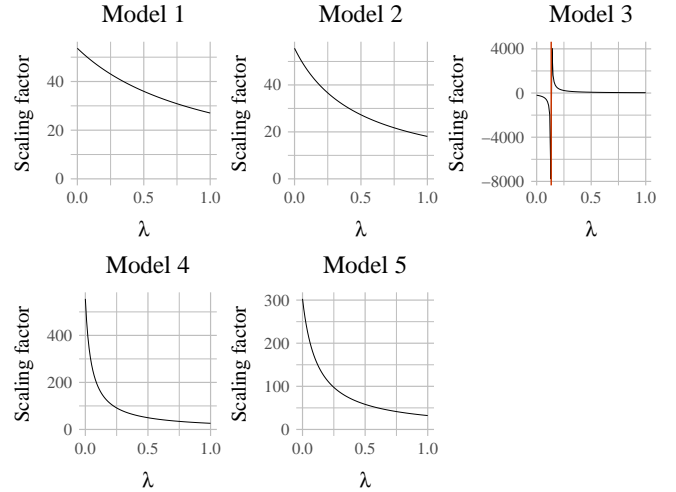The scaling factor as a function of λ is shown in figure 11.

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 Results of individual SMI constituents

Table 8: Sharpe ratio

	In sample	Out of sample
SMI constituents	1.196	0.489

Table 8 depicts the results of the cross validation (4.3) method for 20 stocks. In in sample, the training and test set are the same and therefore the Sharpe ratio is higher than out of sample. Contrary to the out of sample where the testing set is not included in the training set. If a stock behaves differently in this one test set compared to the 4 training sets flowing into the optimization, this stock can easily be over- or undervalued. This leads to a worse Sharpe ratio than the in sample one. The consequences 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.

In the following two illustrations, bootstrapped weights of the SMI constituents can be seen. In the graphics the mean weights and the standard deviation over the 100 samples are shown. Figure 12 shows the MVP weights of the 20 stocks and figure 13 the TP weights.
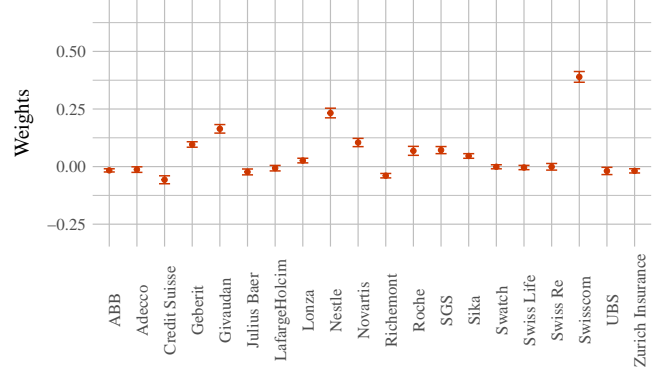

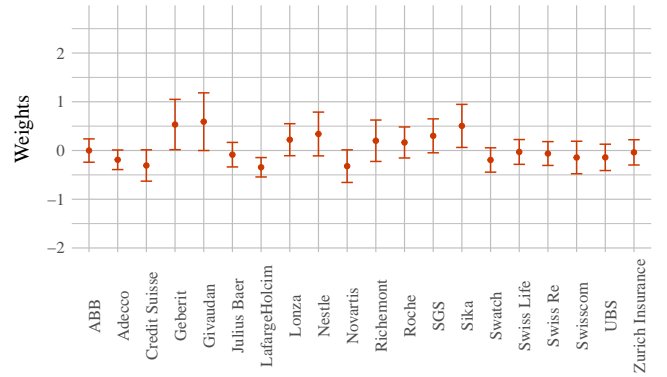Figure 12: MVP weights with standard error

Figure 13: tangency portfolio weights with standard error

5.2 Results of grouped SMI constituents and bootstrapping method

Table 9: Sharpe ratio

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

In table 9, the Sharpe ratio of in and out of sample of the formed groups are presented. The first observation is, that the in sample Sharp ratio decreased after the grouping is applied. 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 with the groups. Nevertheless, the out of sample Sharpe ratio has increased. Due to grouping, the returns are less noisy than the individual stocks. This leads to a more reliable model which performs better out of sample than it would with noisier data.

In chapter 4.1.2, cluster analysis is used to optimize the group composition. Hence, higher correlating stocks are grouped together to have better diversified portfolio than is achieved with grouping by industry. Consequently, the correlation between the groups by correlation shrunk in comparison to the correlation between the groups by industry.

In the following two illustrations, bootstrapped weights of the groups by industry can be seen. As mentioned in section 5.1, the same is shown for the groups. Figure 14 illustrates the MVP weights and figure 15 the TP weights.
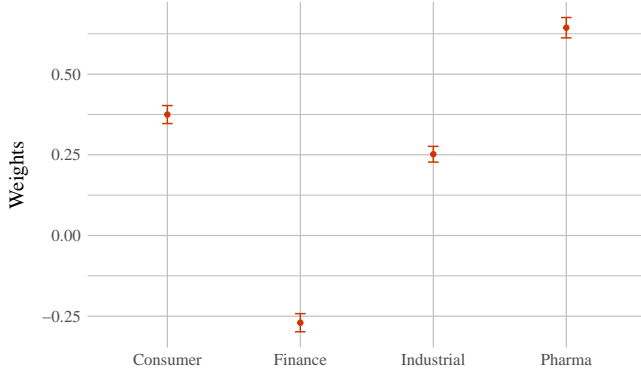
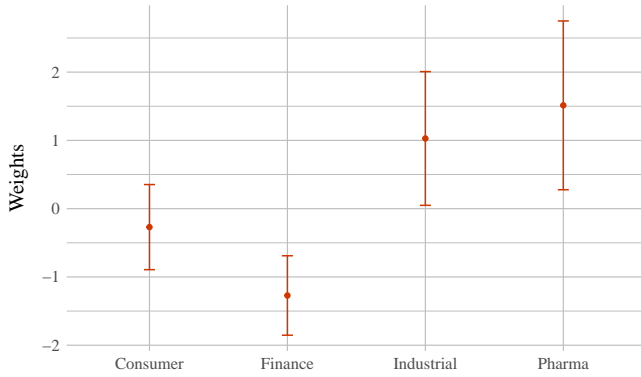Figure 14: MVP weights with standard error

Figure 15: Tangency portfolio weights with standard error by group

It is obvious to see in figure 14 and 15 and also visualized before in figure 7, 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. If the eq. (14) of MVP and the eq. (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 bootstrap sampling of the returns does not influence the covariance matrix as much as the returns itself.

5.3 Results of shrinkage method

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 can be seen that the highest Sharpe ratio is achieved with a shrinkage factor of about 0.05. If past returns are taken too seriously (higher shrinkage factor) or not considered at all (shrinkage factor 0), the out of sample Sharpe ratio drops. They can be trusted to about 5%. 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 non-shrunk mean returns. The groups by correlation do show a significantly higher Sharpe ratio than the grouping by industry and it is also higher than the one of the SMI constituents.
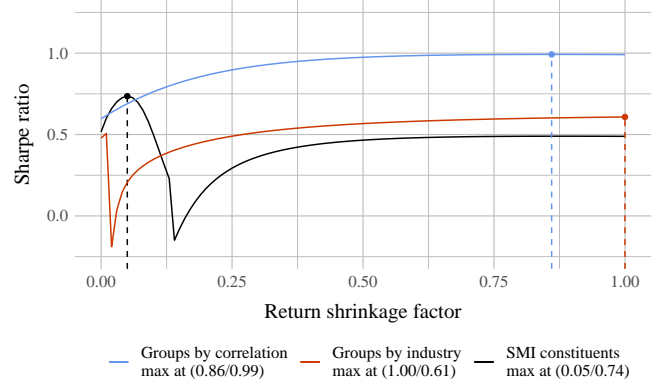

Figure 16: Sharpe ratio as a function of return shrinking factor

Figure 17 shows the same plot for the correlation shrinkage factor. There is a clear maximum for correlation shrinkage factor 0.4. If past correlations are taken too seriously (higher shrinkage factor), the out of sample Sharpe drops. If they are taken less seriously (shrinkage factor 0), it also drops. This result is similar to the one obtained in figure 16 as for the groups shrinkage does not have a great impact. The highest Sharpe ratio is again achieved by the groups by correlation.
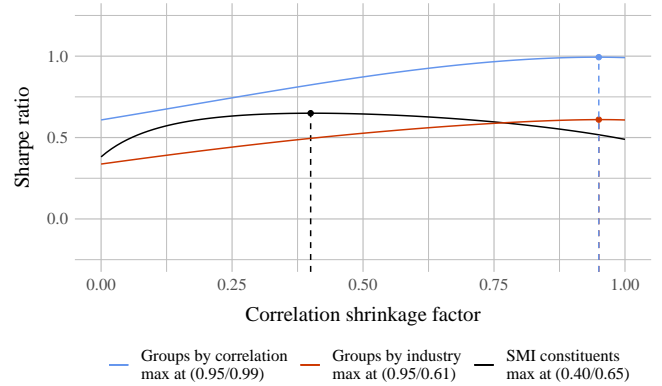

Figure 17: Sharpe ratio as a function of correlation shrinking factor

5.3.2 Shrinkage factors analyzed simultaneously

A further visualization is a three dimensional 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. In the lower 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 about 0.05 and a correlation shrinkage factor of close to one. This matches the result of the previous plots where the mean return shrinkage is better at improving the Sharpe ratio.
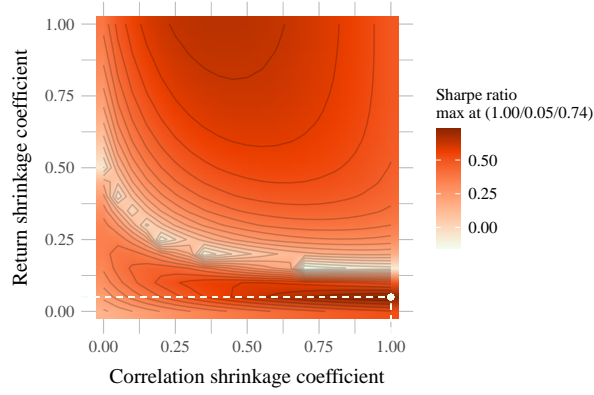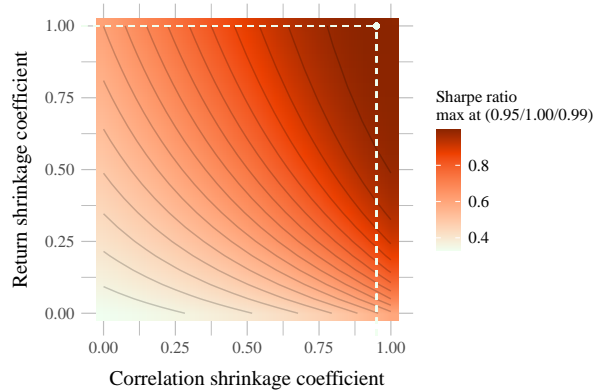
Figure 18: SMI constituents Sharpe ratio as a function of return and correlation shrinking factor

Table 10 provides an overview over all Sharpe investigated Sharpe ratios, namely in sample (IS), out of sample (OS) and out of sample with applied Shrinkage (OS Shrinkage).

Table 10: Sharpe ratio

	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

optimale gewichtungen

6 Conclusion

It can be concluded, that the best improvement of Sharpe ratio is achieved by grouping by correlation with the help of cluster analysis methods. Shrinkage can also improve the Sharpe ratio slightly on already optimized groups and more significantly on ungrouped assets as in the case of the SMI.

Kapitel noch ausbauen, aber Kernaussage ist jene von oben

Figure 19 shows the three dimensional 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 drop in Sharpe ratio in the lower portion while the groups by correlation does not. Overall, latter do show a Sharpe ratio which is about 0.3 greater than the groups by industry.

simultaneous shrinking: if returns are shrunk to 5%, then correlation shrinking does not add more value
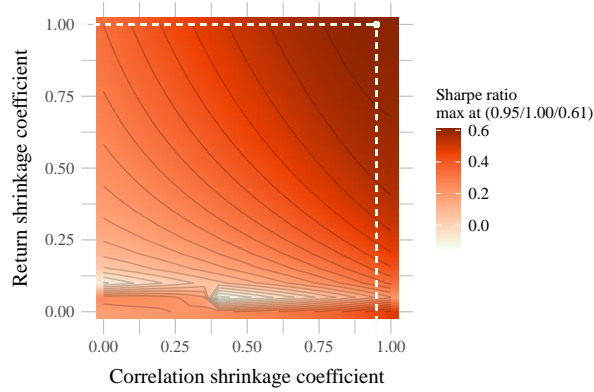
Figure 19: Groups by industry Sharpe ratio as a function of return and correlation shrinking factor

Figure 20: Groups by correlation Sharpe ratio as a function of return and correlation shrinking factor

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 Instintute 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.

[10] A. Ruckstuhl (2020). Statistisches Data Mining. School of Engineering, Zurich University of Applied Science

8 References

[1] R Core Team, *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing, 2019 [Online]. Available: https://www.R-project.org

[2] "Yahoo finance." [Online]. Available: https://finance.yahoo.com

List of Figures

List of Tables