Delft University of Technology
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft Institute of Applied Mathematics

Model Selection in Portfolio Management

Thesis submitted to the
Delft Institute of Applied Mathematics
as part of obtaining

the degree of

BACHELOR OF SCIENCE
in
APPLIED MATHEMATICS

by

E. Hoefkens

Delft, Nederland
July 2018

BSc Report APPLIED MATHEMATICS

“Model Selection in Portfolio Management”

(Dutch Title: “Model Selectie in Portfolio Management”)

Emiel Hoefkens

Delft University of Technology

Supervisor

Dr. J. Söhl

Other thesis committee members

Prof. dr. F.H.J. Redig

Drs. E.M. van Elderen

Delft

July 2018

Preface

You are currently reading the Bachelor thesis that is written in order to obtain the degree of Bachelor of Science at the Delft University of Technology. This project is carried out at the statistics department of Delft Institute of Applied Mathematics under the supervision of J. Söhl.

During the past period, I have spend my time doing research in model selection, Sharpe ratios, Sharpe Ratio Information Criteria (SRIC) and portfolio strategies as all of this gets combined in this thesis.

I would like to thank J. Söhl for the guidance and help during this project and I would also like to thank F.H.J. Redig and E.M. van Elderen for taking place in my thesis committee.

E. Hoefkens
Delft, July 2018

Abstract

Model selection starts with a dataset and a number of candidate models that can explain that data. The AIC and BIC criteria prevents choosing the best fitting model by penalizing for the number of parameters in a model and instead selects the model that performs best when assessed to unseen data. Their performance depends on the sample size and the noise in the data.

In portfolio management, it is common to find a combination of financial products such that an objective is optimized. A common risk measure criterion that gets maximized is the Sharpe ratio. Since portfolio management is also done based on historic data, but wanted to be optimized for unseen data, model selection can be applied to portfolio management as well.

Finding the optimal weights in a portfolio is done by solving a linear system of equations. Applying this to subsets of stocks which are contained in the AEX index, leads to higher in-sample Sharpe ratios than using equal weights or just following the AEX index. The out-of-sample Sharpe ratio gets overestimated by noise fit and estimation error. The Sharpe Ratio Information Criterion (SRIC) corrects for this. This criterion gives an unbiased estimate for the out-of-sample Sharpe ratio and can be used for model selection in portfolio management. Using a trend following strategy, investing proportionally to the returns, also increases you expected out-of-sample Sharpe ratio.

Contents

Introduction

Model selection starts with a dataset and a number of candidate models that can explain that data. Different selection criteria prevents choosing the best fitting model by penalizing for the number of parameters in a model and instead selects the model that performs best when assessed to unseen data. Two common model selection criteria, the AIC and BIC, will be analyzed and we will have a look at their performance by means of simulation.

In portfolio management, a combination of financial products gets chosen in such a way that an objective is optimized. A common criterion that gets maximized is the Sharpe ratio, which is a risk measure of return. Since portfolio management is also done based on historic data, but wanted to be optimized for unseen data, model selection can be applied to portfolio management as well.

The aim of this bachelor thesis is to apply an unbiased closed form estimator for the out-of-sample Sharpe ratio when the in-sample Sharpe ratio is obtained by optimizing over k parameters in a portfolio. This estimator is called the Sharpe Ratio Information Criterion (SRIC). We test two different perspectives of portfolio management on the set of stocks that are currently used to compile the AEX index. To efficiently calculate all the maximum Sharpe ratios and SRIC values, we will be programming a script in R. This script will load all the necessary data and contain the functions we use.

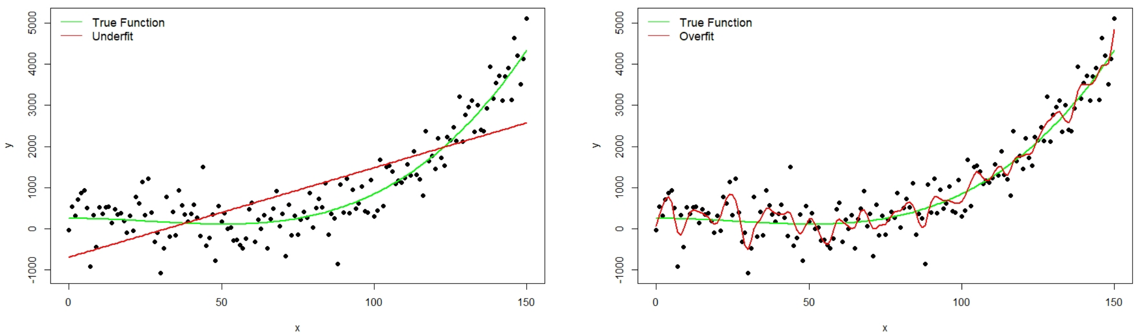
1

Introduction to Model Selection

1.1 Overfitting

True models do not exists, since full reality cannot be captured in a model. There are however often good approximating models. Actually, there may be multiple models that have the ability to give possible explanations of observed data. Thus we seek a good model to approximate the effects or factors supported by empirical data. When choosing such a model out a set of candidate models, it is easy to declare the best fitting model as the best model for prediction. Nevertheless the best fitting model in general does not behave the best on yet unknown data. Model selection techniques are needed to not select the best fitting model, but select the model that performs best when assessed to unseen data.

In general, a complex model fits the data better than a simple model as it has a greater ability to adapt to the data. By adding more and more parameters to a model, therefore making it more complex, a fit can even reach 100% - the model contains all data points in the outcome. The opposite of course also occurs. If the model is too simple, it is not flexible enough to take all changes in the data into account; underfitting arises. See figure 1.1 for displaying both phenomena[(1)].



(a) (b)

Figure 1.1: Data is created by a 3 degree polynomial with added noise. A linear model can't adapt to the decrease and quick increase in the data which ensures a poor model;underfitting has occured (1.1a) . Trying to fit a spline interpolation with a degree of 50 leads to a non realistic view of the data. The function reacts to heavily on small changes (noise) in the data and causes overfitting (1.1b). It may have a good fit on this data, but generalizes poorly.

(1) Spline (or polynomial) interpolation is used to show overfitting. By doing this, local changes do not behave the global behaviour of the function. Otherwise, local changes would have an effect on the whole function which is not really desirable.

1.2 Akaike Information Criterion

The Kullback-Leibler information is a measure of how one probability distribution f diverges from another probability distribution g. First, let f and g be two known simple probability distributions. The K-L information $I(f,g)$ can be interpreted in two different ways: the distance from f to g or the information lost when g is used to approximate f. If f and g are continuous distributions:

$$I(f,g) = \int f(x)\log\left(\frac{f(x)}{g(x|\theta)}\right)dx.$$

If f and g are discrete distributions:

$$I(f,g) = \sum_i p_i \log\left(\frac{p_i}{q_i}\right),$$

where p_i and q_i are corresponding to respectively f and g [1]. The K-L information can also be considered as the expected value of the logarithm of the ratio of two distributions. Keep in mind that we can interpret the K-L information as a distance. This is however a tricky designation, since in general the K-L distance from f to g is not the same as from g to f. It is therefore called a oriented or directed distance.

We first assumed that both f and g are known functions. In our case however, we don't know the true function f or if it even exist - in general there is no true model that fully explains the data. There is only a set of data that we wish to approximiate by a good predicting model. Now let's have a look at relative distance. $I(f,g)$ can also be written as:

$$\begin{aligned} I(f,g) &= \int f(x)\log(f(x))dx - \int f(x)\log(g(x|\theta))dx \\ &= \mathbb{E}_f[\log(f(x))] - \mathbb{E}_f[\log(g(x|\theta))]. \end{aligned}$$

The first expectation, $\mathbb{E}_f[\log(f(x))]$, only depends on the unknown distribution f. Although we can not know the outcome of this expectation, we can consider it as a constant and therefore a measure of relative oriented distance is possible. Define $C = \mathbb{E}_f[\log(f(x))]$. Then $(I(f,g) - C)$ is the relative oriented distance between f and g and leaves $\mathbb{E}_f[log(g(x|\theta)]$ as the only quantity of interest. [2, p. 58]

The statistician Hirotugu Akaike proposed the use of the Kullback-Leibler information as a fundamental basis for model selection criteria. However since we have no full knowledge of f, we can not compute the K-L distance. Akaike found a way to estimate this distance based on the empirical log-likelihood function at its maximum point. He showed that the critical issue for getting an applied K-L model selection criteria was to estimate

$$\mathbb{E}_y\mathbb{E}_x[\log(g(x|\hat{\theta}(y)))],$$

where x and y are independent random samples from the same distribution and expectations taken with respect to the true f [2, p. 60]. Rather than having a measure for minimizing the distance between two models, one has instead an *expected estimated* distance over the set of considered models. This estimate is called the Akaike Information Criterion (AIC). The criterion deals with the trade-off between the goodness of fit of the model and the simplicity of the model. By simplicity, we mean the numbers of parameters of a particular model. The AIC of a model is defined as:

$$AIC = 2k - 2\log(\mathscr{L}),$$

at which k is the number of parameters of the model and $\mathscr{L}$ the maximum value of the likelihood function for the model [2, p. 61]. The lower the AIC value, the better the model is expected to perform on unseen data compared to the other candidate models.

It is not the individual value of the AIC that is important, since this cannot be interpreted due to the unknown constant. AIC is only relative, therefore also only comparative, to other AIC values in the model set. The differences

$$\Delta_i = AIC_i - AIC_{min}$$

are useful and important values. It allows a quick comparison and as a result a ranking of the models.

1.3 Bayesian Information Criterion

The derivation of AIC denies the existence of a true model. There are also criteria developed based on the assumptions that a true model does exist. One of the candidate models is considered the truth and the goal is to identify the model with the highest probability of being the true model based on the data. This means to select the model with the highest value of $\mathbb{P}(M_j|X)$.
For identification of the true model, consistency is key. This means that the probability of selecting the true model in a set of candidates approaches 1 if sample size increases. Criteria satisfying this are also called 'dimension-consistent' since the true parameter value remains fixed. In general, such criteria can only be consistent if its penalty term for the number of parameters is a fast enough increasing function of n. AIC for example is not consistent, as it always has some probability of selecting models that are too large.

The best known dimension consistent selection criterion is the Bayesion Information Criterion (BIC) developed by Gideon E. Schwarz. He used the Bayes' theorem to estimate the highest probability of being the true model. BIC is thus not an estimator of relative K-L information.

Theorem 1 (Bayes' Theorem)**.** *Let A and B events and $\mathbb{P}(B) \neq 0$. Then*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A)\mathbb{P}(B|A)}{\mathbb{P}(B)},$$

where $\mathbb{P}(A|B)$ is the conditional probability of event A occurring given that B is true, $\mathbb{P}(B|A)$ the conditional probability of event B occurring given that A is true and $\mathbb{P}(A)$ and $\mathbb{P}(B)$ the probabilities of observing A and B independently of each other; also known as the marginal probabilities.

Using Bayes' theorem for our situation results in:

$$\begin{aligned} \mathbb{P}(M_j|X) &= \frac{\mathbb{P}(M_j)\mathbb{P}(X|M_j)}{\mathbb{P}(X)} \\ &\propto \mathbb{P}(M_j)\mathbb{P}(X|M_j) \\ &\propto \mathbb{P}(M_j)\int \mathbb{P}(X|\theta_j, M_j)\mathbb{P}(\theta_j|M_j)d\theta_j \quad [3] \end{aligned} \tag{1.1}$$

Schwarz found a model selection criterion by approximating $\log(\mathbb{P}(M_j|X))$ using (1.1):

$$BIC = \log(n)k - 2\log(\mathscr{L}),$$

with n the size of the data sample and again k the number of parameters in the model and $\mathscr{L}$ the maximum value of the likelihood function for the model.

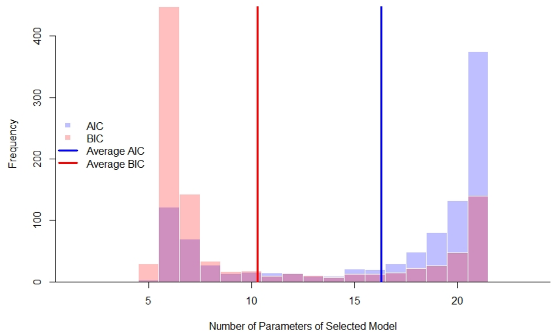
1.4 The Performance of AIC and BIC

To get a feeling of the performances of both selection criteria, we run an example. Data is generated from a 10 degree polynomial with added noise originating from a Gaussian distribution. Different polynomial models, models with a degree from 1 to 20, are fitted to the data with maximum likelihood. Both AIC and BIC are determined for all models. We select the models with the lowest AIC and BIC (this can be 2 different models). We repeat this process a 1000 times to get a good view of their performances. Figure 1.2 shows the results of which models got selected for 4 different sample sizes.

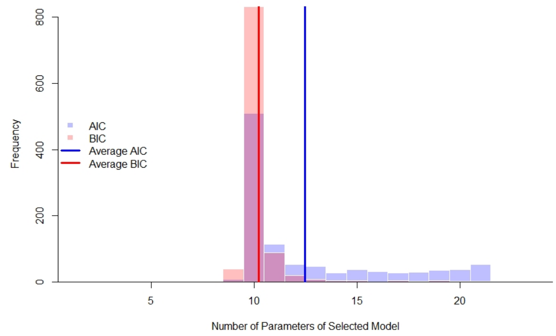
Remark 1.1. *While testing the performances of both AIC and BIC, we found that that is very dependent on the added noise and the size of the in-sample data. In this case they are both chosen in such a way that we get a good view of their selections.*

Remark 1.2. *One may be noticing that we have a 10 degree polynomial, but the models with 11 parameters gets selected most often when sample size increases. This is because the variance of a model is also seen as a parameter. So the model with 11 parameters is considered the true model.*

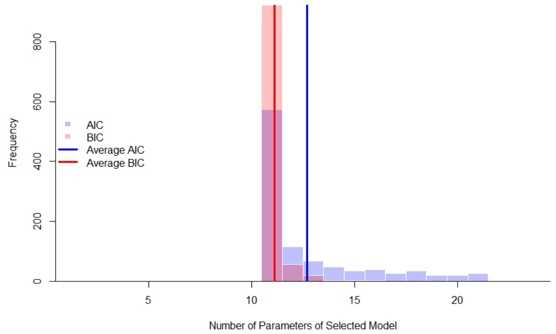
Looking at the 4 histograms in Figure 1.2, many noticeable things arise. When sample size is small, both AIC and BIC behave poorly in finding the true model (which in this case does exist) as can be seen in Figure 1.2a. AIC often overfits and BIC often underfits the data. It is commonly known that AIC does not return good results when sample size is relatively small compared to the complexity of the model. There exists a corrected



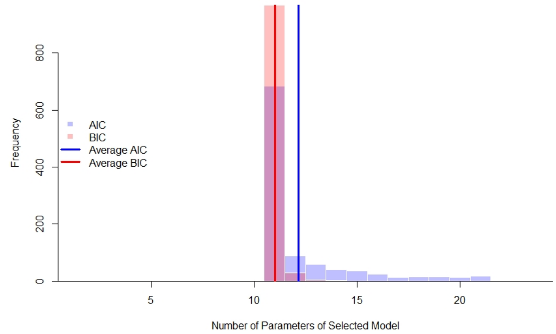
(a) $\#X = 25$



(b) $\#X = 50$



(c) $\#X = 75$



(d) $\#X = 150$

Figure 1.2: Four histograms showing the selection performance of AIC and BIC for different sample sizes of data ($\#X$).

AIC for this case, namely the AICc - indeed *AICcorrected*. However since this is not interesting for our further research, we will leave it at this.
Increasing the sample size leads to significantly better results. Both the criteria see the complexity of the true model and the previous selected underfitting models dissolve from the outcomes.

1.5 AIC and BIC on Portfolio Management

So far we have only looked at model selection for finding the corresponding polynomial model to the data. Doing model selection in portfolio management would be compiling (selecting) an optimal portfolio to reach a certain goal. The AIC can be used for finding a portfolio which would satisfy a best fitting of out-of-sample returns in terms of log-likelihood. The BIC can be used to try and find the underlying portfolio when knowing the returns (which in practice would be a weird situation).
Something the AIC and BIC do not take into account is the exposure of the compiled portfolio. A portfolio that the AIC selects may have high estimated out-of-sample returns, but can also come with a lot of risk. Therefore we are going to introduce a risk measure, called the Sharpe ratio, in the next chapter and use that for compiling our optimal portfolio.

2

Sharpe Ratio

Investing can be very rewarding compared to saving, but there is also the possibility that you lose money by taking a risky position. Economist Harry Markowitz had the idea that owning different kinds of financial assets is less risky than owning only one type. Rather than assessing an asset's risk and return by itself, one must find how it contributes to the portfolio's overall risk and return. See Figure 2.1 for an illustrative example of Markowtiz's idea [(1)].

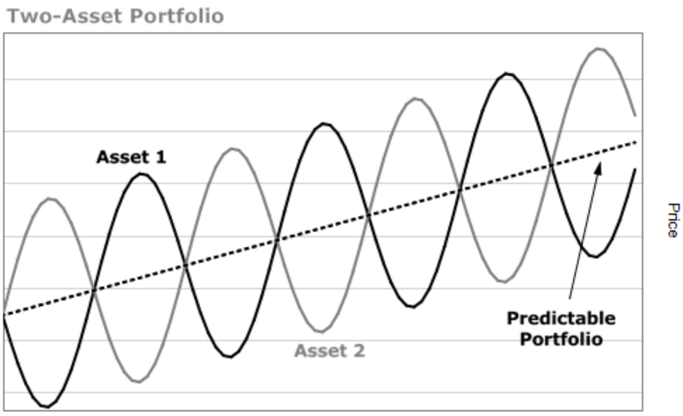


Figure 2.1: Illustrating the idea of Harry Markowitz. The two assets are risky positions on itself, but together they form a perfectly balanced portfolio with rising value. Such a perfect combination is of course not possible in practice, but can be approached (using derivatives of the asset for example). Covering or reducing the exposure of a portfolio is called hedging.

When comparing different portfolios, you would like to invest in the one which is the best. You could consider the portfolio with the highest expected return to be the best, but what if that portfolio also comes with a large(r) risk factor. Is it still the best to choose?

A way to examine the performance of an investment by adjusting for its risk is the Sharpe ratio. This is a risk adjustment measure of return used to evaluate a portfolios performance. It is based on the idea that given two portfolios that offer the same expected return, investors will prefer the less risky one. A portfolio's future performance is often estimated using probabilistic models which has random variables as outcomes. Therefore we can use the standard deviation of a portfolio's return as a proxy for the risk we would take. The Sharpe ratio of a portfolio P is defined as:

$$\rho_P = \frac{\mathbb{E}[R_P - r]}{\sigma_P},$$

where R_P is the rate of return of P, r the best available rate of return on a risk-free asset and σ_P the standard deviation of R_P[4]. Using this ratio lets you see how much additional return you are getting for the added

[(1)] Image: https://healthandwealthbulletin.com/collect-safe-income-with-the-only-free-lunch-in-the-market/

volatility of holding a risky position over a risk-free security. It allows to determine whether a portfolio's returns are due to smart investment decisions or just a higher level of risk. The investment with the highest Sharpe ratio is thus considered the best. Because it is a risk measure, it is used for a relative long term investments. A rule of thumb is that an investment with a Sharpe ratio greater than 1 is seen as a good investment. See Figure 2.2 for an illustrative example of the Sharpe ratio[(2)].

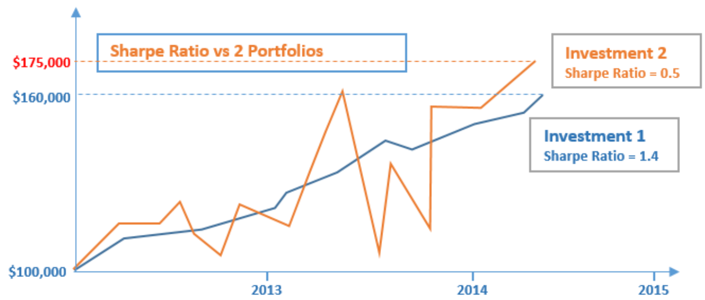


Figure 2.2: Showing the development of the value of two portfolios. When only looking at the return over the past few years, Investment 2 would have been the better investment (higher return). However looking at the big changes in value, investing in this portfolio does seem very risky as this can probably can happen in the future as well. The Sharpe ratio penalizes this by taking in account the standard deviation of the (monthly) returns. Therefore Investment 1 gives you more return per risk you take and also considered the better investment.

2.1 Out-Of-Sample Sharpe Ratio

Notice that up till now we have been discussing Sharpe ratios that are calculated over in-sample data. This helps to validate are choices afterwards, but what about the out-of-sample Sharpe ratio. If we choose a fixed strategy in advance, the expected out-of-sample Sharpe ratio will be the same as the in-sample Sharpe ratio. However, it is not so common to choose a fixed strategy. Often, the in-sample Sharpe ratio of a portfolio gets optimized over a set of n assets. You want to choose the weights of these assets in such a way that you maximize the in-sample Sharpe ratio. What can we expect for the out-of-sample Sharpe ratio when the in-sample Sharpe ratio is optimized over k parameters?

Remark 2.1. *When we are talking about n assets and k parameters, it holds $n = k+1$. With k we mean the number of parameters that influence the Sharpe ratio of a portfolio. This means that the leverage of a portfolio is not counted - the Sharpe ratio is independent of the invested volume. Therefore we say that $k+1$ assets possesses k parameters in a portfolio.*

2.1.1 Decomposition of the Out-Of-Sample Sharpe Ratio

We denote the in-sample Sharpe ratio with ρ and the out-of-sample Sharpe ratio with τ. There are k parameters in a portfolio, but we use θ to denote the vector that contain the weights for the different stocks. So the length of θ is $k+1$. Since we are talking about combining multiple assets, we also need to take into account the (possible) correlation among the different stocks. In such a multidimensional setting we use the covariance matrix Σ of the returns. Define $\hat{\mu}$ as the vector of the in-sample means of the excess returns for the different stocks. Then $\hat{\mu}$ is a noisy observation of the true unkown mean μ for in- and out-of-sample data with $\mathbb{E}[\hat{\mu}] = \mu$. Then the in-sample Sharpe ratio is

$$\rho(\theta) = \frac{\hat{\mu}^T \theta}{\sqrt{\theta^T \Sigma \theta}}.$$

Removing the noise term from $\hat{\mu}$ results in the (true and unobserved) out-of-sample Sharpe ratio:

$$\tau(\theta) = \frac{\mu^T \theta}{\sqrt{\theta^T \Sigma \theta}}.$$

Notice that ρ can always be calculated for any θ using in-sample data. The true mean μ however isn't known and therefore τ can't be determined perfectly. The expectation of τ though, will be equal to ρ when using a fixed strategy (as mentioned before).

(2) Image: https://www.assetmacro.com/wp-content/uploads/2015/06/Investment-Guide.png

Remark 2.2. *We are talking about $\hat{\mu}$ as a noisy observation of μ, but yet nothing is mentioned about the precision of Σ. We assume that Σ is known and therefore there is no difference between $\hat{\Sigma}$ and Σ. In reality however, Σ is also not known and needs to be estimated as well. However, it can be estimated much better than μ by estimating it with higher-frequency data. By doing this, Σ gets more accurate. This does not count for μ and is therefore considered the main estimation error.*

We are mainly interested in the θ that maximizes the Sharpe ratio, so we define:

$$\hat{\theta} \in \arg\max_{\theta \in \Theta} \rho(\theta)$$
$$\theta^* \in \arg\max_{\theta \in \Theta} \tau(\theta)$$

where Θ is the $(k+1)$-dimensional parameter space over which θ gets optimized. See Table 2.1 for all four combinations of optimal θ and in- and out-of-sample Sharpe ratio.

Symbol	Value	Parameter	Description
$\hat{\rho}$	$\rho(\hat{\theta})$	$\hat{\theta}$	In-sample Sharpe ratio of optimal parameter applied to in-sample data
ρ^*	$\rho(\theta^*)$	θ^*	In-sample Sharpe ratio of optimal parameter applied to out-of-sample data
$\hat{\tau}$	$\tau(\hat{\theta})$	$\hat{\theta}$	Out-of-sample Sharpe ratio of optimal parameter applied to in-sample data
τ^*	$\tau(\theta^*)$	θ^*	Ouf-of-sample Sharpe ratio of optimal parameter applied to out-of-sample data

Table 2.1: Defining all four combinations of in- and out-of-sample Sharpe ratio with optimal parameters applied to in- and out-of-sample data.

Remark 2.3. *In Table 2.1 we have defined four different values for the Sharpe ratio, but we can only calculate $\hat{\rho}$ as we do not know anything about the out-of-sample data.*

Remember we said that we are interested in what we can expect for the out-of-sample Sharpe ratio when the in-sample Sharpe ratio is optimized over k parameters. So we are interested in the value of $\hat{\tau}$. Using the notation from Table 2.1, we can decompose $\hat{\tau}$ into

$$\hat{\tau} = \hat{\rho} - (\hat{\rho} - \rho^*) - (\tau^* - \hat{\tau}) + (\tau^* - \rho^*).$$

This decomposition gives us more insight on how the difference between the in- and out-of-sample Sharpe ratio arises when the parameters are optimized. It says that the in-sample Sharpe ratio minus 3 terms equals the out-of-sample Sharpe ratio. These 3 terms are set to be noise fit, estimation error and (general) noise:

- **Noise Fit:** $(\hat{\rho} - \rho^*)$
  The difference between $\hat{\rho}$ and ρ^* can be interpreted as the noise fit on the in-sample Sharpe ratio. The parameter $\hat{\theta}$ gets maximized over in-sample data which (in general) contains noise and gets tuned towards this noise. It makes use of the noise to reduce risk or enhance return. Therefore it diverges from the optimal out-of-sample parameter θ^* and causes a difference.

- **Estimation Error:** $(\tau^* - \hat{\tau})$
  The difference between τ^* and $\hat{\tau}$ can be interpreted as the estimation error of the out-of-sample Sharpe ratio. As we have seen, $\hat{\theta}$ and θ^* are (in general) not equal. So applying these to an out-of-sample data set leads to an estimation error on the true maximum out-of-sample Sharpe ratio.

- **Noise:** $(\tau^* - \rho^*)$
  The difference between τ^* and ρ^* is due to noise. This causes uncertainty in the predictions.

There is that the error for noise fit and estimation error have the same expectation on out-of-sample data, namely

$$\mathbb{E}[\hat{\rho} - \rho^*] = \mathbb{E}[\tau^* - \hat{\tau}] = \frac{-k}{2T\hat{\rho}}.$$

The expected value of the noise $(\tau^* - \rho^*)$ is 0. [5]

2.2 Sharpe Ratio Information Criterion

It turns out that the optimized in-sample Sharpe ratio overestimates the Sharpe ratio that can be expected on out-of-sample data because of tuning a parameter space on a noisy data set. The Sharpe Ratio Information Criterion (SRIC) corrects for both noise fit and estimation error. It is an unbiased closed form estimator for the out-of-sample Sharpe ratio when the in-sample Sharpe ratio is obtained by fitting k parameters in a portfolio. The SRIC is defined as

$$SRIC = \hat{\rho} - \frac{k}{T\hat{\rho}}, \tag{2.1}$$

where $\hat{\rho}$ is the in-sample Sharpe ratio maximized over a k-dimensional parameter space and T time of in-sample data [5]. The unit of T is dependent on which kind of returns are used; yearly returns leads to T in years, monthly returns leads to T in months, etc. This criterion does come with some assumptions:

- We spoke about $\hat{\mu}$ as a noisy observation of true μ. We assume that

$$\hat{\mu} = \mu + \nu, \text{ where } \nu \sim \mathcal{N}\left(0, \frac{1}{T}\Sigma\right).$$

- Also we assume that the Σ is of full rank.

Under these assumptions, the SRIC can be used in two ways:

1. When having obtained the optimal parameters for the maximum in-sample Sharpe ratio, SRIC will give you an estimation for the expected out-of-sample Sharpe ratio. In particular, it even holds:

$$\mathbb{E}[\hat{\tau}] = \mathbb{E}\left[\hat{\rho} - \frac{k}{T\hat{\rho}}\right].$$

   It is dependent of T over which period you could expect that out-of-sample Sharpe ratio. When T is in months, it would be for over a month. If T is in years it is for over one year, etc.

2. When comparing different sizes of parameter spaces and data sizes, SRIC will give you the portfolio with the highest expected return per risk in the future. Therefore it can also be applied as a portfolio selection criterion.

3

Portfolio Management on the AEX Index

Thousands of different stocks (and even much more other securities) get traded every day on exchanges all around the world. If you want to know how the financial market is performing, it would take a lot of time to track every single security that trades on an exchange. Therefore you can take a look at a portfolio sample of the market that is representative for the whole market. This sample is called an index. It is a statistical measure of overall market sentiment. If an index goes up, it means that the price of the stock for (most) companies went up as well. Therefore indices are often used as a benchmark when evaluating portfolios, industries and some are even to be considered as benchmarks for the overall economy of a country; like the S& P 500 for the U.S.

The AEX index is a stock market index of 25 Dutch companies who are listed on the Euronext Amsterdam exchange (formerly known as the Amsterdam Stock Exchange). A few times a year, a lot of rules determine which companies should be taken in the AEX index and which not. Also not every company has equal weight in the index. The weighting is dependent on the number of shares of the companies, the price of the shares and a couple of other factors. See Figure 3.1 for the historical price movement of the AEX.

Remark 3.1. *Indices are great tools for giving a quick overview of the performance of the market, but we cannot buy them. After all, they are statistical measures. It is of course possible to buy all the stocks that make up an index according to the right weightings. In that case the value of your portfolio will follow the index. When doing so however, you need to pay to much transaction costs for buying all the stocks that it isn't profitable for at least a very long time. So instead of buying all those stocks individually, you can also invest in what is called an index fund. That is fund based on an index and that mirrors its performance and there exists one for nearly every index out there.*

To have a look on how different portfolios perform in terms of Sharpe ratio, that consists of a subset of the stocks of the AEX, we divide the *current* 25 AEX stocks in 5 groups with each 5 different stocks(1). The reason why we look at subsets first will become clear in Section 3.3 with remark 3.3. We order these 5 groups in alphabetical order of the AEX stocks with $G1, \dots, G5$. See Table 3.1 for an overview. Therefore we do not have any influence on the grouping of certain stocks that might give very good results. Group G will contain every stock of the AEX index. One of the companies that is listed at the AEX index is 'ASR Nederland'. This insurance company is very new to the financial market as it only went to the exchange in June 2016. To keep equality between all the groups, we will therefore only look at the data from the 20th of June 2016 to the 20th of June 2018 (this results in $T = 24$ months for every SRIC calculation).
The Sharpe ratio measures the excess return per risk. So when calculating Sharpe ratios ourselves, we need to substract the risk free rate from the returns. For this rate, we take the Effective Federal Funds Rate(2). This rate get managed by the Federal Open Market Committee (FOMC) and has influence on the American price stability and economic growth.

(1) All the stock market data used in this research comes from: https://finance.yahoo.com/.
(2) https://fred.stlouisfed.org/series/FEDFUNDS

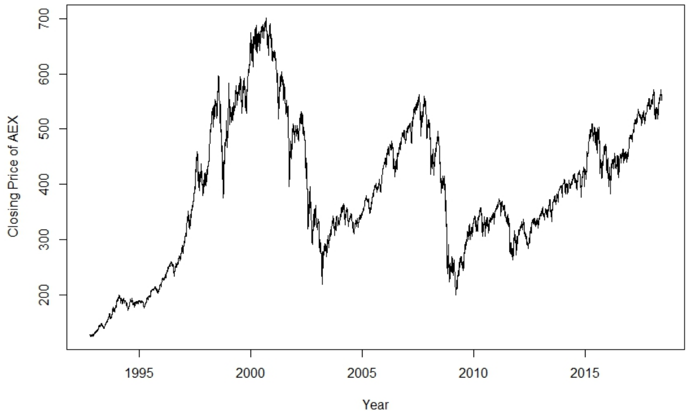


Figure 3.1: Displaying the historical price movemetn of the AEX. The AEX suffered hard from the crisis we had to encounter, but has been rising (with a lot of volatility though) since the end of 2009.

Portfolio	**Included Stocks**
$G1$	Aalbers (AALB), ABN AMRO (ABN), AEGON (AGN), Ahold Delhaize (AD), AkzoNobel (AKZA)
$G2$	Altice (ATC), Arcelor Mittal (MT), ASM Litho (ASML), ASR Nederland (ASRNL), DSM (DSM)
$G3$	Galápagos (GLPG), Gemalto (GTO), Heineken (HEIA) ING (INGA), KPN (KPN)
$G4$	NN Group (NN), Philips (PHIA), Randstad (RAND) Royal Dutch Shell-A (RDSA), RELX Group (REN)
$G5$	Signify (Light), Unibail WFD (UL), Unilever (UNA), Vopak (VPK), Wolters Kluwer (WKL)

Table 3.1: Overview of which companies are in which groups. The abbreviations behind the name of the company is the abbreviation used on the Euronext Amsterdam exchange to denote that company. Some are very logical, some can differ a lot from the full name.

3.1 Equal weight in the Portfolio

Before finding the optimal parameters, we shall start with giving each stock equal weight. This gives us a basis for the in-sample Sharpe ratio from where we need to improve. Remember that this is a fixed strategy, so the in-sample Sharpe ratio will also be an estimation for the expected out-of-sample Sharpe ratio. The in-sample Sharpe ratios can be found in Table 3.2 and an overview of the portfolio values in Figure 3.2 to see how the Sharpe ratios relates to that. The Sharpe ratio of the AEX index is also included in Table 3.2 for comparison.

Portfolio	**In-Sample Sharpe Ratio**
$G1_{eq}$	0.036
$G2_{eq}$	0.332
$G3_{eq}$	-0.0509
$G4_{eq}$	0.092
$G5_{eq}$	-0.18
G_{eq}	0.176
AEX	0.058

Table 3.2: The in-sample Sharpe ratio’s of all groups when all stocks have equal weight and the AEX index.

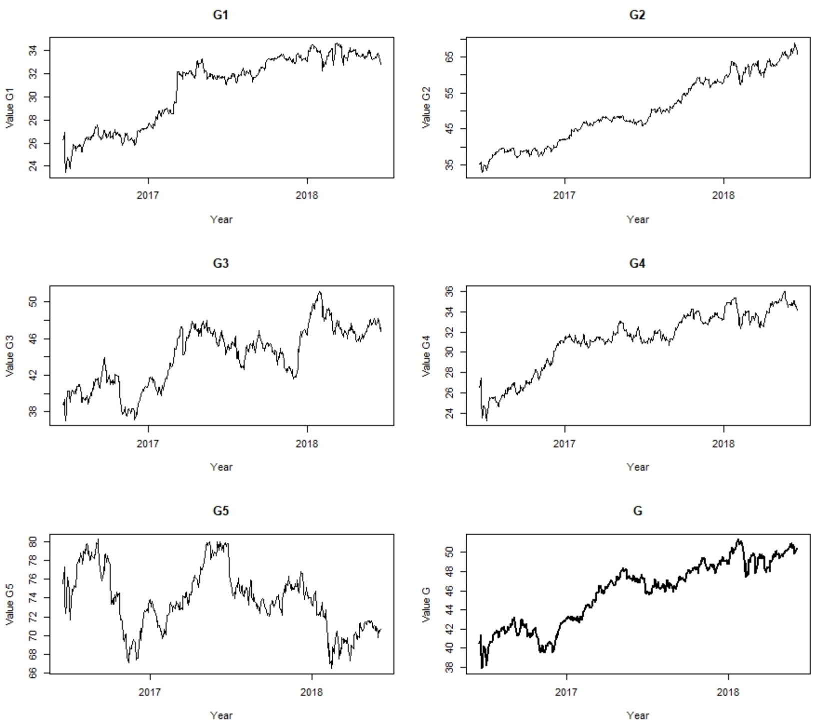


Figure 3.2: The development of the portfolio values of all groups when all stocks have equal weight.

3.2 Optimal In-Sample Sharpe Ratio

Now we are interested in an optimal combinations of stocks which provides us the highest in-sample Sharpe ratio. So we are interested in finding $\hat{\theta}$. We want to find this $\hat{\theta}$ for different k within each group and apply the SRIC as model selection criterion afterwards. Finding $\hat{\theta}$ is maximizing $\rho(\theta)$ over the parameter space Θ. For finding a maximum in a function, we need find the θ where the derivative of $\rho(\theta)$ is zero. Since the log() function is an increasing function, there is that maximizing over $\log(\rho(\theta))$ gives us the same $\hat{\theta}$ as when we would maximize over $\rho(\theta)$ itself. Doing this gets us:

$$\log(\rho(\theta)) = \log\left(\frac{\hat{\mu}^T\theta}{\sqrt{\theta^T\Sigma\theta}}\right) = \log(\hat{\mu}^T\theta) - \log(\sqrt{\theta^T\Sigma\theta})$$

Taking the derivative of this leads to:

$$\begin{aligned}\frac{d\log(\rho(\theta))}{d\theta} &= \frac{1}{\hat{\mu}^T\theta}\hat{\mu} - \frac{1}{\sqrt{\theta^T\Sigma\theta}}\frac{1}{2}(\theta^T\Sigma\theta)^{-\frac{1}{2}}2\Sigma\theta \\ &= \frac{\hat{\mu}}{\hat{\mu}^T\theta} - \frac{\Sigma\theta}{\theta^T\Sigma\theta}\end{aligned}$$

Plugging in $\hat{\mu} = \Sigma\theta$ sets this equation to 0. We get $\hat{\mu}$ and Σ from the in-sample data, so we can solve this system of linear equations to find $\hat{\theta}$. Notice from the later results that this gives indeed a maximum for the Sharpe ratio and not a minimum. See Table 3.3 for all the results on the in-sample data. The R script that is used to get these values can be found in Appendix A.1.

Remark 3.2. *In some optimal cases, stocks get assigned a negative weight. This can be interpreted as going short on that stock. 'Going short' is a way of getting a positive return if the price of a stock goes down. You sell the stock at the beginning and buy it back in a later stadium. The difference is your profit. At a stock exchange, you can sell stocks without first owning them and even have negative amount of securities in your portfolio.*

G1 - In-Sample Sharpe Ratios

		Optimal Weights $(\hat{\theta})$				
k	$\rho(\hat{\theta})$	AALB	ABN	AGN	AD	AKZA
0	0.13455171	1	0	0	0	0
1	0.167841097	2.071833911	0	0	-1.071833911	0
2	0.174470152	6.506259117	0	0	-3.46953388	-2.036725237
3	0.178022907	3.400532904	0	0.475829152	-1.850822179	-1.025539876
4	0.178517803	3.366749318	0.316817143	0.377564607	-1.98004053	-1.081090537

G2 - In-Sample Sharpe Ratios

		Optimal Weights $(\hat{\theta})$				
k	$\rho(\hat{\theta})$	ATC	MT	ASML	ASRNL	DSM
0	0.369909313	0	0	1	0	0
1	0.468548896	0	0	0.566734178	0.433265822	0
2	0.543903769	0.050819711	0	0.477410178	0.471770111	0
3	0.667629	0.044206896	0	0.283381204	0.290701096	0.381710804
4	0.669250297	0.043841703	-0.019392212	0.290342332	0.299760115	0.385448062

G3 - In-Sample Sharpe Ratios

		Optimal Weights $(\hat{\theta})$				
k	$\rho(\hat{\theta})$	GLPG	GTO	HEIA	INGA	KPN
0	0.512046762	0	0	0	0	-1
1	0.574176339	0.263305222	0	0	0	-1.263305222
2	0.605019474	0.323310054	0	0	0.479827104	-1.803137158
3	0.605027886	0.320692525	0	-0.011213983	0.476798969	-1.786277512
4	0.605035055	0.320541864	0.002741901	-0.015077207	0.476663847	-1.784870404

G4 - In-Sample Sharpe Ratios

		Optimal Weights $(\hat{\theta})$				
k	$\rho(\hat{\theta})$	NN	PHIA	RAND	RDA	REN
0	0.22001969	0	1	0	0	0
1	0.243404435	0	1.960984794	0	0	-0.960984794
2	0.277212307	0	22.66108533	-8.654072248	0	-15.00701309
3	0.280275801	5.435233212	40.71024234	-17.73747154	0	-27.40800401
4	0.280346194	6.384163959	50.40039526	-22.09861782	0.974146498	-34.6600879

G5 - In-Sample Sharpe Ratios

		Optimal Weights $(\hat{\theta})$				
k	$\rho(\hat{\theta})$	Light	UL	UNA	VPK	WKL
0	0.454937917	0	-1	0	0	0
1	0.685492604	0	-2.918596614	0	0	1.918596614
2	0.771895704	0	-1.550897668	0	-0.508931975	1.059829644
3	0.839969511	0.38311979	-2.291611761	0	-0.564810125	1.473302096
4	0.841951576	0.379215872	-2.398392062	0.133763186	-0.576221966	1.46163497

Table 3.3: The optimal weights for different portfolio sizes to achieve maximum Sharpe ratio in all 5 groups.

Looking at Table 3.3 gets us that there is quite some difference between the performance of the groups. An optimal combination of stocks in group $G1$ only gets you a maximum Sharpe ratio of 0.179 while in group $G5$ it even passes the 0.8. $G1$ also does not improve that much when we add stocks to the portfolio while the in-sample Sharpe ratio of $G5$ almost doubles when increasing the portfolio size. By looking back at the Sharpe ratios we would just get with equal weight, see Table 3.2, we find that there is quite a lot of improvement. This definitely tells us that using equal weight in a portfolio is not a good idea and there should always be looked at the correlation among all the stocks.

3.3 Applying SRIC

Now that we have obtained the maximum in-sample Sharpe ratios with optimal parameters applied to in-sample data, we can apply the SRIC criterion to estimate the out-of-sample Sharpe ratio with (2.1). This criterion gives the portfolio for each group which is expected to perform best on out-of-sample data (while keeping the same weights). The results of the SRIC selection can be found in Table 3.4.

G1 / G5 - Estimated Out-Of-Sample Sharpe Ratios

	SRIC Values				
k	G1	G2	G3	G4	G5
0	**0.13455171**	0.369909313	**0.512046762**	**0.22001969**	0.454937917
1	-0.080409583	0.379621855	0.501608621	0.072221578	0.624709065
2	-0.303166466	0.390690392	0.467282861	-0.023399648	0.663936386
3	-0.524133922	**0.480399266**	0.398425839	-0.165713469	**0.691154586**
4	-0.755095897	0.420215418	0.329568922	-0.314156853	0.643998783

Table 3.4: The SRIC values that are obtained by using the optimal in-sample Sharpe ratio's from Table 3.1.

The bold numbers are the highest SRIC value's per group. When doing portfolio selection one a single group, the portfolio represented by the bold number is thus the one we should choose. In 3 of the 5 groups (G1, G3 and G4), this would just be one single stock. The other stocks in the group can not contribute enough to be worth the risk of an extra risky position in a portfolio.

We have done portfolio selection to obtain the highest estimated out-of-sample Sharpe ratio within the groups, but can we improve this by combining the selected portfolios of the different groups. So we take the stocks that are selected by the SRIC within the groups, the stocks corresponding to the bold numbers, into consideration. The optimal weights for all portfolio sizes can be found in Table 3.5 and the corresponding SRIC values in Table 3.6.

Remark 3.3. *By combining the best stocks of the different groups based on SRIC values, does not mean that this gives the overall best combination for all the stocks in the AEX index. Some stocks that are not included may have high (negative) correlation with other stocks that can possibly increase the in- or estimated out-of-sample Sharpe ratio. The reason for not looking for the overall best combination of all the stocks in the AEX index is due to a lack of computational power. Although, the outcome is for a relative long term strategy, for this research it is not really feasible. For example, for finding the best combination of 5 out of 25 stocks, you would need to compare over 50.000 combinations and for 12 stocks even more than 5 million which all need to be calculated by carrying out several functions.*

By looking at subsets first and combining the selected portfolios, we found a portfolio of 8 different stocks which, with optimal weights, has an in-sample Sharpe ratio of 1.55 and an estimated out-of-sample Sharpe ratio of 1.37. For comparison, the AEX itself had an in-sample Sharpe ratio of just 0.058 over the same period.

Best of G1 / G5 Combined - Optimal In-Sample Sharpe Ratios

		Optimal Weights $(\hat{\theta})$				
k	$(\hat{\theta})$	AALB	ATC	ASML	ASRNL	DSM
0	0.512046762	0	0	0	0	0
1	0.958319886	0	0	0	3.261986964	0
2	1.164272616	0	0	0	2.189504192	1.602521465
3	1.272954599	0	0	0.414238656	1.219321086	0.89507334
4	1.329803434	0	-0.28921403	0	1.923489811	1.153541285
5	1.439500026	0	-0.2891288	0	1.375566105	0.677150729
6	1.489641556	0	-0.32943976	0	1.462235443	1.014107797
7	1.553898783	0	-0.315274	0	1.38089776	0.624485255
8	1.559464102	-0.19800032	-0.33747164	0	1.460123195	0.615831304
9	1.566209101	-0.24606672	-0.29387707	0.118887578	1.338923461	0.606205575
10	1.566279032	-0.26074785	-0.29708547	0.119933261	1.363326027	0.621106359

	Optimal Weights $(\hat{\theta})$					
k	KPN	PHIA	Light	UL	VPK	WKL
0	-1	0	0	0	0	0
1	-4.261986964	0	0	0	0	0
2	-2.792025657	0	0	0	0	0
3	-1.528633081	0	0	0	0	0
4	-2.275409654	0	0.487592592	0	0	0
5	-1.669983852	0.546471686	0.359924129	0	0	0
6	-1.647621387	0.664712012	0.51177327	-0.675767373	0	0
7	-1.451390479	0.556440323	0.532149288	-0.979789177	0	0.652481031
8	-1.496265634	0.65031881	0.565394165	-0.974146472	0	0.714216589
9	-1.38484434	0.578737962	0.494879044	-0.757095242	0	0.544249748
10	-1.407168264	0.593516842	0.502807907	-0.7764342	-0.01771412	0.558459509

Table 3.5: The optimal weights for different portfolio sizes when combining the selected portfolios of group $G1$ to $G5$.

k	SRIC Values
0	0.512046762
1	0.914841015
2	1.092697169
3	1.174757852
4	1.204471627
5	1.294773851
6	1.321815948
7	**1.36619887**
8	1.34571546
9	1.326777469
10	1.300255764

Table 3.6: The SRIC values that are obtained by using the optimal in-sample Sharpe ratio’s from the combined selected groups.

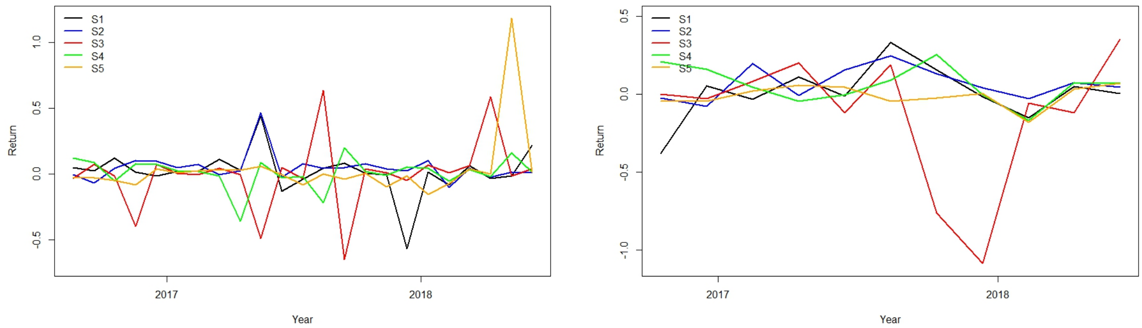
4

Portfolio Management with a Trend Following Strategy

So far we have only looked at combining the price movement of stocks to find higher Sharpe ratio's, but have not made use of the combined returns yet. Now we are going to introduce a trend following strategy that does this. This strategy says that we are going to invest proportional to the returns of the underlying stocks. We develop the strategies per group for groups $G1$ to $G5$. The strategy is fixed in the following way:

1. Set the a number of days for the lookback (lb) you want to take.
2. Determine the returns of all the stocks in the strategy over the past period as long as the lookback. Define the weights by dividing these returns by the sum of it.
3. Invest in the stocks of the strategy proportionally to the weights that are set.
4. Determine the return made by the strategy at the end of the period. Repeat step 1 to 3 to create a list of returns of the strategy.

We perform this strategy on the same groups as we have done so far. So we call $S1$ to $S5$ the stragies of group $G1$ to $G5$ respectively. Such a strategy can be considered as a new financial product. Since we looked at the monthly returns in chapter 3, we are starting with taking a lookback of 30 days. After this we also view the results when the lookback is set to 60 days. The choices of length for this lookback can of be taken differently by everyone, but we will stick to this. A reason for not taking a longer lookback is due to the little number of datapoints you get, which gives unreliable outcomes. The returns of strategies $S1$ to $S5$ for two different lookback periods, $lb = 30$ and $lb = 60$, are showed in Figure 4.1.



(a) $lb = 30$

(b) $lb = 60$

Figure 4.1: Two plots showing the returns of the different strategies with different lookbacks; $lb = 30$ and $lb = 60$.

In the same way we have done before for the single stocks within a group, we are going to find an optimal combination of strategies to maximize our in-sample Sharpe ratio for two different lookbacks; $lb = 30$ and $lb = 60$. Also we use the same time frame as we have always done so far, so the data from the 20th of June 2016 till the 20th of June 2018. The results for the optimal in-sample Sharpe ratio can be found in Table 4.1 and the corresponding SRIC values in Table 4.2.

lb = 30 - Optimal In-Sample Sharpe Ratios

		Optimal Weights $(\hat{\theta})$				
k	$(\hat{\theta})$	S1	S2	S3	S4	S5
0	0.4576939	0	1	0	0	0
1	0.516732018	-0.4019656	1.4019656	0	0	0
2	0.564925869	-0.268729378	1.114544939	0.154184438	0	0
3	0.583306559	-0.246687519	1.021963741	0.140903751	0	0.083820027
4	0.587975330	-0.211837024	0.901888507	0.146964396	0.100761464	0.062222656

lb = 60 - Optimal In-Sample Sharpe Ratios

		Optimal Weights $(\hat{\theta})$				
k	$(\hat{\theta})$	S1	S2	S3	S4	S5
0	0.6757611	0	1	0	0	0
1	0.808377746	-0.456948301	1.456948301	0	0	0
2	0.929134522	-0.304583541	0.911509822	0	0.39307372	0
3	0.993587634	-0.448680957	1.612160448	0	0.733412989	-0.89689248
4	1.009531141	-0.482964891	1.716189345	-0.076465332	0.732538078	-0.8892972

Table 4.1: Optimal combinations of five different strategies to obtain a maximum in-sample Sharpe ratio for two different lookbacks.

Estimated Out-Of-Sample Sharpe Ratios

	SRIC Values	
k	$lb = 30$	$lb = 60$
0	**0.457694**	0.6757611
1	0.432591	0.69591907
2	0.411	**0.73344899**
3	0.359694	0.719100248
4	0.292192	0.649327924

Table 4.2: The SRIC values that are obtained by using the optimal in-sample Sharpe ratio's from Table 4.1 of the trend following strategy.

When updating your portfolio every 30 days proportionally to the returns over that period will give you an estimated out-of-sample Sharpe ratio of 0.46. Using a strategy with a 60 day lookback gives you an estimated Sharpe ratio of 0.73 when applying it to out-of-sample data. These two values are lower than we would combine the optimum of groups $G1$ to $G5$.

There is also the possibility to combine the two strategies, but this comes with a disadvantage. The covariance matrix between the returns is tricky to determine. Because with a 30 lookback we have twice as much as returndates than with a 60 day lookback. Therefore we only take into account the dates on which both strategies are at the end of their period.
So we are going to combine $S2$ with a 30 day lookback with $S1$, $S2$ and $S4$ with a 60 day lookback. De denote the strategies as follows: $S1_{60}$, $S2_{30}$, $S2_{60}$ and $S4_{60}$ to make the difference clear. The optimal in-sample Sharpe ratios can be found in Table 4.3 and the corresponding SRIC values in Table 4.4. The R script that is used to get these values can be found in Appendix A.2.

k	$(\hat{\theta})$	Optimal Weights $(\hat{\theta})$			
		$S1_{60}$	$S2_{30}$	$S2_{60}$	$S4_{60}$
0	0.6757611	0	0	1	0
1	0.82415746	0	0.451868	0	0.548132
2	0.929134522	-0.30458	0	0.91151	0.393074
3	1.081323458	-0.21187	0.324073	0.492774	0.395023

Best of lb = 30 and 60 Combined - Optimal In-Sample Sharpe Ratio

Table 4.3: The optimal weights for different portfolio sizes when combining the selected portfolios of strategy $S1$ to $S5$.

k	**SRIC values**
0	0.6757611
1	0.713851972
2	0.733448992
3	**0.829107278**

Table 4.4: The SRIC values that are obtained by using the optimal in-sample Sharpe ratio's from the combined selected strategies.

With a SRIC value of 0.83 it has a higher estimated out-of-sample Sharpe ratio than not combining strategies with different lookbacks. However is it also lower when finding an optimal combinations of stocks like we did in the previous chapter. A trend following strategy returns good (out-of-sample) Sharpe ratios, but it looks like combining the values of stocks performs better than the trend following strategy.

Conclusion & Discussion

Conclusion

We started with analyzing the AIC and BIC model selection criteria. Although these two criteria come from totally different perspectives on model selection, their formulas look very similar with only a difference on how much complexity gets penalized. The Kullback-Leibler distance, the oriented distance between two probability functions, is the basis for deriving the AIC criterion. The BIC criterion is build on Bayes' theorem which assumes a true model. When looking at the performance of both models, we find that this is very dependent on the ratio between the size of the data sample and the complexity of the model. The sample size must be big enough compared to the complexity of the considered models to give good results.

A maximum Sharpe ratio is a common objective in portfolio management. There is however a difference between the in- and out-of-sample Sharpe ratio when you optimize your portfolio over k parameters. The out-of-sample Sharpe ratio gets overestimated by noise fit and estimation error. The Sharpe Ratio Information Criterion (SRIC) corrects for this and is unbiased closed form estimator for the out-of-sample Sharpe ratio when the in-sample Sharpe ratio is obtained by optimizing over k parameters. Since it is dependent on k and T (time period of data), it can be used as a portfolio selection criterion as well.

We applied the SRIC to sets of stocks which are currently used to compile the AEX index. We optimize over 5 groups of stocks, $G1$ to $G5$, first. The Sharpe ratios of all these portfolios with different k for all groups are calculated by writing a script in R. The SRIC then selects per group among the portfolios of different sizes the one with the highest estimated out-of-sample Sharpe ratio. These selected portfolios already have a much improved (estimated out-of-sample) Sharpe ratio than when we give all stocks just equal weight (which we did first). Combining the stocks from the selected portfolios to find an overall optimal portfolio results in a big improvement over the performance over the AEX.

Instead of combining the value of the stocks, you can also look at combining the returns. This is called a trend following strategy. You update the weights in your portfolio at the end of every certain period proportionally to the returns of the stocks over that same period. Such a period is called a lookback. We look at 5 strategies $S1$ to $S5$ respectively to group $G1$ to $G5$ with 2 different lookbacks. Finding the optimal combinations between those strategies lead to a bit better results than when only looking at the groups. Combining the 2 selected strategies with different lookbacks is a bit tricky, but can lead to a higher estimated out-of-sample Sharpe ratio than with a fixed lookback.

Discussion

There are a couple of things in this research for where there is room for improvement of extension. One thing that would be an improvement for this research is have a look at datasets with a longer timeframe. The stock market data of ASRNL goes only back to June 2016 and to keep equality among the groups, we had this at the starting date of our data. Although it is believed that it is a long enough timeframe when looking at monthly returns, it would be interesting to see what happens when we take a longer timeframe. Does there arise a significant difference in terms of (out-of-sample) Sharpe ratio or not? This bigger dataset can then also be used for evaluating the performance of the SRIC. This is something that is not been done in this research (also because of a lack on time).

Due to a lack of high computational power, it was not feasible to find the best overall combinations of stocks for all portfolio sizes. By first looking at subsets, the total computing time is much and much less and also gives good results. For personal use this if more than satisfying, but for professional use it might be worth to upgrade. It would be interesting so see what the difference is in terms of Sharpe ratio to evaluate the method of working with subsets first. We sorted the groups on alphabetical order so we did not have any influence, but maybe there is a better procedure to sort the groups differently which can improve the results.

Something we have just mentioned once before is transaction costs; when you want a position that follows the price of the AEX, you buy all the stocks that are contained in the index with a lot transaction costs or the corresponding index fund with less transaction costs. These costs do not make a comeback in the research that followed. However for both ways of compiling a portfolio, optimal combination of stocks or trend following strategy, also come transactions costs. These are not taken into account when selecting a portfolio with the SRIC. This could be an extra penalty which may lead to different results.

The Sharpe ratio uses the standard deviation (or covariance matrix) as its proxy for risk, which assumes that returns are normally distributed. It has been shown before that the distribution of returns for financial assets tend to deviate from a normal distribution. This causes the results to be misleading. The difference in distribution is in the tails. The distribution of returns tends to have bigger tails. Because of this, in-sample Sharpe ratios can be overestimated.

A

R Scripts

A.1 Optimal In-Sample Sharpe Ratio

```
File <- function(name){
  filename = paste(c(name, '.AS'), collapse = '')
  filename = paste(c(filename, '.csv'), collapse = '')
  file <- read.table(filename, sep = ',', header = TRUE)
  file$Date = as.Date(strptime(file$Date, "%Y-%m-%d"))
  file$Close = as.numeric(as.character(file$Close))
  file = na.omit(file)
  file = subset(file, select = c(Date, Close))
  return(file)
}

AEXnames = c('AALB', 'ABN',  'AGN', 'AD', 'AKZA',  'ATC', 'MT', 'ASML',
             'ASRNL', 'DSM', 'GLPG', 'GTO', 'HEIA', 'INGA', 'KPN', 'NN',
             'PHIA', 'RAND', 'RDSA', 'REN', 'Light', 'UL', 'UNA', 'VPK',
             'WKL')

FED <- File('FEDFUNDS')

dates_m = seq(as.Date('2016-06-20'), as.Date('2018-06-20'), 'month')
dates_d = seq(as.Date('2016-06-20'), as.Date('2018-06-20'), 'day')

# Functions ---------------------------------------------------------------

StockReturns <- function(name){

  assign(name, File(name))

  Closeprices <- NULL
  Returns <- NULL

  returndates = dates_m

  for(i in 1:length(returndates)){
    location = which.min(abs(returndates[i]-get(name)$Date))
    Closeprices = rbind(Closeprices, get(name)$Close[location])
  }
  for(i in 2:length(returndates)){
    Returns = rbind(Returns,
```

```
                    (Closeprices[i] - Closeprices[i-1])/Closeprices[i-1])
  }
  return(Returns)
}

SharpeTheta <- function(names, ratio = rep(1,length(names))){
  Date = dates_m[2:length(dates_m)]
  returns <- NULL
  for(i in 1:length(names)){
    returns = rbind(returns, t(StockReturns(names[i])))
  }
  returns = t(returns)
  for(i in 1:length(names)){
    returns[,i] = returns[,i] - FED[(nrow(FED)-(nrow(returns)-1)):(nrow(FED)),2]/100
  }

  mu <- NULL
  for(i in 1:length(names)){
    mu = rbind(mu, mean(returns[,i]))
  }

  if(length(names)>1){
    CovarianceReturn = cov(returns)
    theta = solve(CovarianceReturn, mu)
    theta = theta / abs(sum(theta))
    sharpe = (t(theta)%*%mu)/(sqrt(t(theta)%*%CovarianceReturn%*%theta))
    return(c(sharpe, theta))
  }
}

OptimalSharpe <- function(names, k = length(names)){
  max <- rep(-1,(2*k+1))
  comb = combn(names, k, simplify = TRUE)
  for(j in 1:ncol(comb)){
    res = SharpeTheta(comb[,j])
    if(max[1] < res[1]){
      max[1:(k+1)] = res
      max[(k+2):(2*k+1)] = comb[,j]
    }
  }
  return(max)
}

for(k in 2:5){print(OptimalSharpe(AEXnames[1:5], k))}
for(k in 2:5){print(OptimalSharpe(AEXnames[6:10], k))}
for(k in 2:5){print(OptimalSharpe(AEXnames[11:15], k))}
for(k in 2:5){print(OptimalSharpe(AEXnames[16:20], k))}
for(k in 2:5){print(OptimalSharpe(AEXnames[21:25], k))}
```

A.2 Optimal In-Sample Sharpe Ratio - Trend Following Strategy

```
File <- function(name){
  filename = paste(c(name, '.AS'), collapse = '')
  filename = paste(c(filename, '.csv'), collapse = '')
  file <- read.table(filename, sep = ',', header = TRUE)
  file$Date = as.Date(strptime(file$Date, "%Y-%m-%d"))
  file$Close = as.numeric(as.character(file$Close))
  file = na.omit(file)
  file = subset(file, select = c(Date, Close))
  return(file)
}

AEXnames = c('AALB', 'ABN',  'AGN', 'AD', 'AKZA',  'ATC', 'MT', 'ASML',
             'ASRNL', 'DSM', 'GLPG', 'GTO', 'HEIA', 'INGA', 'KPN', 'NN',
             'PHIA', 'RAND', 'RDSA', 'REN', 'Light', 'UL', 'UNA', 'VPK',
             'WKL')

FED <- File('FEDFUNDS')

lb = 30

dates_m = seq(as.Date('2016-06-20'), as.Date('2018-06-20'), 'month')
dates_d = seq(as.Date('2016-06-20'), as.Date('2018-06-20'), 'day')


Trend <- function(name){

  assign(name, File(name))

  returndates = dates_d
  Closeprices <- NULL

  for(i in 1:length(returndates)){
    location = which.min(abs(returndates[i]-get(name)$Date))
    Closeprices = rbind(Closeprices, get(name)$Close[location])
  }

  R <- NULL
  for(i in (lb+1):length(returndates)){
    R = rbind(R, (Closeprices[i] - Closeprices[i-lb])/
                Closeprices[i-lb])
  }
  return(R)
}


Strategie <- function(names){
  s <- dates_d[(lb+1):(length(dates_d))]

  for(i in 1:length(names)){
    s = cbind(s, Trend(names[i]))
  }
  sum <- NULL
  for(i in 1:nrow(s)){
    sum = rbind(sum, sum(s[i,2:(length(names)+1)]))
  }
```

```
  s = cbind(s,sum)

  ratio <- NULL
  for(j in 2:(length(names)+1)){
    weight <- NULL
    dates <- NULL
    for(i in 0:round(nrow(s)/lb)){
      weight = rbind(weight, s[(i*lb + 1),j]/s[(i*lb + 1),(length(names)+2)])
      dates = rbind(dates, s[(i*lb + 1),1])
    }
    ratio = cbind(ratio, weight)
  }
  return(data.frame(dates,ratio))
}

PortfolioValue <- function(names, ratio = rep(1,length(names))){
  timeline = dates_d[(lb+1):length(dates_d)]
  value = data.frame(timeline, rep(0, length(timeline)))
  colnames(value) <- c('Date', 'Total')
  for(i in 1:length(names)){
    assign(names[i], File(names[i]))
    value = merge(value, get(names[i]), by = 'Date')
  }
  for(i in 1:length(names)){
    value[,2] = value[,2] + ratio[,i]*value[,(2+i)]
  }
  return(value[,1:2])
}



PortfolioReturn <- function(names){
  strategie = Strategie(names)
  timeline = strategie[2:(nrow(strategie)),1]
  frame = data.frame(as.Date(timeline), rep(0, length(timeline)))
  colnames(frame) <- c('Date', 'Return')

  for(i in 1:length(timeline)){
    value = PortfolioValue(names, strategie[i,2:6])

    date_n = as.Date(dates_m[1]) + (i+1)*lb
    date_o = as.Date(dates_m[1]) + i*lb
    location_n = which.min(abs(date_n - value[,1]))
    location_o = which.min(abs(date_o - value[,1]))

    value_n = value[location_n,2]
    value_o = value[location_o,2]
    return = (value_n - value_o) / value_o
    frame[i,2] = frame[i,2] + return
  }
  return(frame)
}

SharpeRatioTrend <- function(returns, ratio = rep(1,length(returns))){
  ExpectedReturn = 0
  for(i in 1:length(names)){
```

```
    ExpectedReturn = ExpectedReturn +  ratio[i]*mean(returns)
  }
  SdExpectedReturn = sd(returns - ExpectedReturn)
  return(ExpectedReturn / SdExpectedReturn)
}

SharpeThetaTrend <- function(returnframe){
  Date = dates_m[3:length(dates_m)]
  returns = returnframe
  for(i in 1:length(names)){
    returns[,i] = returns[,i] - FED[(nrow(FED)-(nrow(returns)-1)):(nrow(FED)),2]/100
  }
  mu <- NULL
  for(i in 1:ncol(returns)){
    mu = rbind(mu, mean(returns[,i]))
  }

  if(ncol(returns)>1){
    CovarianceReturn = cov(returns)
    theta = solve(CovarianceReturn, mu)
    theta = theta / abs(sum(theta))
    sharpe = (t(theta)%*%mu)/(sqrt(t(theta)%*%CovarianceReturn%*%theta))
    return(c(sharpe, theta))
  }
}

OptimalSharpeTrend <- function(names, k){
  max <- rep(0,(2*k+1))
  comb = combn(names, k, simplify = TRUE)
  for(j in 1:ncol(comb)){
    frame <- NULL
    for(i in 1:length(comb[,j])){
      frame =  cbind(frame, get(comb[i,j])[,2])
    }
    res = SharpeThetaTrend(frame)
    if(max[1] < res[1]){
      max[1:(k+1)] = res
      max[(k+2):(2*k+1)] = comb[,j]
    }
  }
  return(max)
}

S1 = PortfolioReturn(AEXnames[1:5])
S2 = PortfolioReturn(AEXnames[6:10])
S3 = PortfolioReturn(AEXnames[11:15])
S4 = PortfolioReturn(AEXnames[16:20])
S5 = PortfolioReturn(AEXnames[21:25])

returnnames = c('S1','S2','S3','S4', 'S5')

for(k in 2:5){print(OptimalSharpeTrend(returnnames, k))}
```

Bibliography

[1] Wikipedia. Kullback–leibler divergence, 2018. URL https://en.wikipedia.org/wiki/Kullback%E2%80%93Leibler_divergence.

[2] Kennteth P. Burnham David R. Anderson. *Model Selection and Multimodel Inference*. Springer, Colorado State Univerisity, 2002.

[3] Linear Algebra. Analysis of discrete data: Model selection, akaike and bayesian information criterion, 2016. URL https://www.youtube.com/watch?v=75BOMuXBSP.

[4] Investopedia. Sharpe ratio. URL https://www.investopedia.com/terms/s/sharperatio.asp.

[5] Dirk Paulsen Jakob Söhl. Noise fit, estimation error and a sharpe information criterion: Linear case, 2017. ArXiv:1602.06186.