DEGREE PROJECT IN MATHEMATICS,
SECOND CYCLE, 30 CREDITS
STOCKHOLM, SWEDEN 2017

Beating the MSCI USA Index by Using Other Weighting Techniques

TROTTE BOMAN

SAMUEL JANGENSTÅL

KTH ROYAL INSTITUTE OF TECHNOLOGY
SCHOOL OF ENGINEERING SCIENCES

Beating the MSCI USA Index by Using Other Weighting Techniques

TROTTE BOMAN

JANGENSTÅL SAMUEL

Degree Projects in Financial Mathematics (30 ECTS credits)
Degree Programme in Applied and Computational Mathematics
KTH Royal Institute of Technology year 2017
Supervisor at Öhman: Filip Boman
Supervisor at KTH: Camilla Landén
Examiner at KTH: Camilla Landén

TRITA-MAT-E 2017:37
ISRN-KTH/MAT/E--17/37--SE

Royal Institute of Technology
School of Engineering Sciences
KTH SCI
SE-100 44 Stockholm, Sweden
URL: www.kth.se/sci

Abstract

In this thesis various portfolio weighting strategies are tested. Their performance is determined by their average annual return, Sharpe ratio, tracking error, information ratio and annual standard deviation. The data used is provided by Öhman from Bloomberg and consists of monthly data between 1996-2016 of all stocks that were in the MSCI USA Index at any time between 2002-2016. For any given month we use the last five years of data as a basis for the analysis. Each time the MSCI USA Index changes portfolio constituents we update which constituents are in our portfolio.

The traditional weighting strategies used in this thesis are market capitalization, equal, risk-adjusted alpha, fundamental and minimum variance weighting. On top of that, the weighting strategies are used in a cluster framework where the clusters are constructed by using K-means clustering on the stocks each month. The clusters are assigned equal weight and then the traditional weighting strategies are applied within each cluster. Additionally, a GARCH-estimated covariance matrix of the clusters is used to determine the minimum variance optimized weights of the clusters where the constituents within each cluster are equally weighted.

We conclude in this thesis that the market capitalization weighting strategy is the one that earns the least of all traditional strategies. From the results we can conclude that there are weighting strategies with higher Sharpe ratio and lower standard deviation. The risk-adjusted alpha in a traditional framework performed best out of all strategies. All cluster weighting strategies with the exception of risk-adjusted alpha outperform their traditional counterpart in terms of return.

Alternativa viktportföljer för att prestera bättre än MSCI USA Index

Sammanfattning

I denna rapport prövas olika viktningsstrategier med målet att prestera bättre i termer av genomsnittlig årlig avkastning, Sharpekvot, aktiv risk, informationskvot och årlig standardavvikelse än det marknadsviktade MSCI USA Index. Rapporten är skriven i samarbete med Öhman och data som används kommer från Bloomberg och består av månadsvis data mellan 1996-2016 av alla aktier som var i MSCI USA Index vid någon tidpunkt mellan 2002-2016. För en given månad används senaste fem åren av historisk data för vår analys. Varje gång som MSCI USA Index ändrar portföljsammansättning så uppdaterar vi vilka värdepapper som ingår i vår portfölj.

De traditionella viktningsstrategierna som används i denna avhandling är marknadviktat, likaviktat, risk-justerad alpha viktat, fundamental viktat och minsta varians viktat. De klusterviktade strategierna som används i denna avhandling är konstruerade genom att använda K-medel klustring på aktierna varje månad, tilldela lika vikt till varje kluster och sedan använda traditionella viktningsstrategier inom varje kluster. Dessutom används en GARCH skattad kovariansmatris av klustrena för att bestämma minsta varians optimerade vikter för varje kluster där varje aktie inom alla kluster är likaviktade.

Vi konstaterar i detta arbete att den marknadsviktade strategin har lägst avkastning av alla viktningsmetoder. Från resultaten kan vi konstatera att det finns viktningsmetoder med högre Sharpekvot och lägre standardavvikelse. Risk-justerad alpha viktning använt på traditionellt vis är den strategi som presterar bäst av alla metoder. Alla klusterviktade strategier med undantag av risk-justerad alpha viktning presterar bättre än deras traditionella motsvarighet i termer av avkastning.

Acknowledgements

We would like to thank Filip Boman and coworkers at Öhman for their support and guidance through this project. We would also like to thank our supervisor at KTH, Camilla Landén, for her much appreciated inputs regarding this thesis.

Contents

1 Introduction

There are many investment strategies to build an equity portfolio, but there is one kind of portfolio which has been central within the finance community for a long period of time - capitalization-weighted portfolios. The prevalence of the cap-weighted portfolio is due to a number of reasons, but most of it can be credited to four factors [1]. One is that it is a passive strategy, which means it requires little active management. Another is that large companies receive greater weights, thus the portfolio consists mainly of highly liquid stocks which reduces expected transaction costs. Another benefit is that the cap-weighted portfolio is automatically rebalanced as the stock prices vary which means the only rebalancing costs are for replacing constituents. Finally, under the most common interpretation of the Capital Asset Pricing Model, a diversified cap-weighted portfolio is Sharpe-ratio maximized, i.e. mean-variance optimized.

The first three factors require no assumptions and are considered factual. However, the fourth only holds when certain, very specific, assumptions are made. It has been shown that very little deviation from these assumptions renders the cap-weighted portfolio sub-optimal [1]. This has created a desire within the financial community to investigate other strategies for constructing portfolios.

Cap-weighted portfolios are based on technical analysis, which means that the weights are constructed using the trading history and the price history of a stock. To simply describe the problem with cap-weighted portfolios, it can be argued that it assigns too large weights to stocks that are overvalued by the market and too small weights to stocks that are undervalued by the market [2]. This can be thought of as that the cap-weighted portfolio takes only the markets view of a stock into account and then misses other factors which could influence the stock's value.

There are other strategies than the cap-weighted strategy that use technical analysis. Many of these strategies are some form of mean-variance optimization. It is often preferable to use only variance-minimization due to the fact that errors in the expected future means of a stock or portfolio lead to more incorrect weighting than errors of the same size in the expected future variances [3]. In other words, the mean needs to be estimated far more accurately than the variances to render realistic results, and both mean estimation and variance estimation is very hard to do.

There is another approach to constructing a portfolio that is based on another form of analysis than technical analysis. Instead, it is based on fundamental analysis. Rather than analyzing the historical prices and the historical trading of a stock, fundamental analysis focuses on the data which describes a company's value without taking the stock market into account. Examples of data which fundamental analysts may look at are a company's book value and revenue.

In addition to investigating portfolios constructed using technical analysis and fundamental analysis, a third approach will be investigated which is based on cluster analysis. The principle of clustering is to group together data points which are similar to each other into one cluster, while data points that are non-similar will be in different clusters. This is a purely mathematical approach which is used within many fields, where some of the more notable are machine learning, pattern recognition and medicine [4]. The method of using cluster analysis within finance is a relatively new approach, but there are reports that suggest that a clustering approach can yield better results than a traditional approach [5]. In the area of portfolio construction the clustering consists of finding which stocks have similar temporal behavior, i.e. when we construct a cluster we try to find stocks which are highly correlated and group them together. Once the stocks are clustered we can first assign weights to the cluster by some method, and then assign weights within the cluster by some other method. This gives much more flexibility and the reasoning is that since stocks with similar behavior are in the same cluster, assigning weights to all clusters should give us a diversified portfolio.

In this thesis we will, with the help of asset management company Öhman, examine alternative weighting strategies to the cap-weighted strategy. Some alternative portfolios will be based on technical analysis, but assign weights according to other methods than the cap-weighted portfolio. Other portfolios will be constructed using other types of analysis. Additionally, we will compare how the weighting strategies perform in the traditional framework to when the same strategies are applied in a cluster framework. The stocks in our portfolio will be based on the constituents of the MSCI USA Index and we will determine the weights from monthly data between 1996-2016. To compare our weighting strategies to that of a cap-weighted portfolio, the determined portfolios will be backtested on the stock market between 2002-2016. By comparing different performance metrics such as return, Sharpe ratio and information ratio we can find the strengths and weaknesses of the different strategies and in which scenarios a certain strategy is suitable and when it is not.

2 Theory

In this section the theory that the thesis relies on is presented. Note that there exist alternative definitions of the concepts presented here - the definitions used in this context are related to portfolio management.

2.1 Notation and Definitions

The following notation and definitions are used throughout the report. All notations used in this report are considered at a given month t if nothing else is stated.

- The number of stocks in the portfolio: N
- Weight:
  w_i - Percentage weight of stock i.
  $\boldsymbol{w} = (w_1, ..., w_N)$ - Vector of stock weights in the portfolio.
- Price: p_i - Closing price per share of stock i.
- Total Return Over Last Month: R_i - The percentage return of stock i over the last month.
- Total Return Over Last Month Of Index: R_B - The percentage return of the index over the last month.
- Total Return Over Last Month Of Portfolio: R_P - The percentage return of the portfolio over the last month.
- Risk-free rate: r_F - Based on the US 3-month Treasury Bill.
- Standard deviation and covariance matrix:
  σ_i - Standard deviation of the return of stock i.
  Σ - Covariance matrix of the stocks, which is an $N \times N$ matrix.

2.2 Construction of Returns

The return for stock i in month t is defined as

$$R_{i,t} = \frac{p_{i,t} - p_{i,t-1} + d_{i,t}}{p_{i,t-1}} \tag{1}$$

where $p_{i,t}$ is the closing price per share of stock i in month t and $d_{i,t}$ is the average monthly dividend per share of stock i over the last year in month t. Adding $d_{i,t}$ when constructing returns means that we reinvest all paid dividends into the stock.

The return for a portfolio $R_{P,t}$ in month t is defined as

$$R_{P,t} = \sum_{i=1}^{N} w_{i,t} \cdot R_{i,t}. \tag{2}$$

The return for the market capitalization index $R_{B,t}$ is constructed in the same way as $R_{P,t}$.

2.3 General Theory

This section describes how the common concepts of beta and alpha have been implemented in this thesis.

2.3.1 Beta

The beta of a portfolio is a measure of how volatile the portfolio is compared to the market as a whole. The beta of a portfolio at time t is calculated using regression analysis of the Capital Asset Pricing Model, which is defined as

$$R^A_{P,j} - r_{F,j} = \alpha_{P,t} + \beta_{P,t} \cdot (R^A_{B,j} - r_{F,j}) + \epsilon_j, \quad j = y-4, ..., y \tag{3}$$

where j denotes which 12-month period is being investigated and y denotes the 12-month period from the currently investigated month (i.e. if we are constructing beta in June 2011, then y is the period from July 2010 to June 2011), $R^A_{P,j}$ is the annual return of the portfolio at time j, $R^A_{B,j}$ is the annual return of the benchmark at time j, $r_{F,j}$ is the average risk-free rate of the 12 months of year j and ϵ_j is the residual at time j [6]. Solving the system of equations in (3) using regression analysis, an estimate of $\beta_{P,t}$ is obtained. If $\beta_{P,t} < 1$ then the portfolio is less volatile than the market and the opposite holds if $\beta_{P,t} > 1$.

2.3.2 Jensen's alpha

Jensen's alpha is the intercept of the regression equation in the Capital Asset Pricing Model (3) and is the excess return adjusted for systematic risk. Ignoring the error term in (3), Jensen's alpha of a portfolio is defined as

$$\alpha_{P,t} = R^A_{P,j} - r_{F,j} - \beta_{P,t} \cdot (R^A_{B,j} - r_{F,j}), \quad j = y-k, ..., y. \tag{4}$$

The value of $\alpha_{P,t}$ indicates how the portfolio has performed when accounting for the risk taken. If $\alpha_{P,t} < 0$ then the portfolio has earned less than expected given the risk taken and if $\alpha_{P,t} > 0$ the portfolio has earned more than expected given the risk taken.

2.4 Traditional Weighting Strategies

2.4.1 Market Capitalization Weighting

In market capitalization weighting the stocks are weighted according to their total market capitalization. This is one of the most common ways to weight index funds and will be the benchmark weighting strategy of this thesis. The total market capitalization is determined by the current market price of a stock multiplied by the number of outstanding shares of a stock [7]. The weight of each stock i is given by

$$w_i = \frac{n_i \cdot p_i}{\sum_{i=1}^{N} n_i \cdot p_i}. \tag{5}$$

where n_i is the number of shares outstanding of stock i. An outstanding share is a share of a stock that has been authorized, issued, purchased and is held by an investor.

2.4.2 Equal Weighting

In an equally weighted portfolio, the weight of each stock i is given by

$$w_i = \frac{1}{N}. \tag{6}$$

This is the simplest possible portfolio to construct and at first glance it should not be able to compete with any worthwhile portfolio strategy. However, research shows that an equally weighted portfolio can outperform the market cap portfolio in terms of a larger average annual return [8].

2.4.3 Risk-Adjusted Alpha Weighting

The risk-adjusted alpha weighting intends to provide large weights to stocks that have large returns and low variance. Jensen's alpha at time t of a stock i is defined according to (4), i.e

$$\alpha_{i,t} = R^A_{i,j} - r_{F,j} - \beta_{i,t} \cdot (R^A_{B,j} - r_{F,j}), \quad j = y-k, ..., y \tag{7}$$

where j denotes year, $R^A_{i,j}$ is the annual return of stock i and k denotes how many years we base our regression on. We will use five years of data so $k = 5$. The risk-adjusted Jensen's alpha is defined as [9]

$$\alpha^{adj}_{i,t} = \frac{\alpha_{i,t}}{\sigma_i}. \tag{8}$$

The weight of stock i is given by

$$w_i = \frac{\alpha^{adj}_{i,t}}{\sum_{i=1}^{N} \alpha^{adj}_{i,t}}. \tag{9}$$

The advantage a of risk-adjusted alpha weighting strategy is that the assignment of stock weights is based on the risk-return trade-off. The risk-adjusted alpha method has been shown to perform well in falling markets [9].

2.4.4 Fundamental Weighting

In a fundamentally weighted index, the weights are based on fundamental criteria such as a company's revenue, dividends, earnings, book value etc. Proponents of the fundamental weighting method claim that this is a more accurate measure of a company's value than the value implied by using market capitalization.

The weight for stock i is defined according to [10] as

$$w_i = \frac{1}{4} \times \left(\frac{r_i}{\sum_{i=1}^{N} r_i} + \frac{c_i}{\sum_{i=1}^{N} c_i} + \frac{d_i}{\sum_{i=1}^{N} d_i} + \frac{b_i}{\sum_{i=1}^{N} b_i}\right) \tag{10}$$

where

- r_i - The five year average of the total revenue from the day-to-day operations for company i.
- c_i - The five year average of the net amount of cash and cash-equivalents moving in and out of company i on a per share basis. Represents the net cash a company produces.
- d_i - The five year average of dividends per share for company i. Based on all dividends that have gone 'ex', i.e all dividends that have been confirmed by the company to be given out to the shareholders.
- b_i - The five year average of reported book values for company i. The book value of a company is the total value of company i's assets that the shareholders would theoretically receive if the company was liquidated.

2.4.5 Minimum Variance Weighting

Many investment strategies are built on some form of mean-variance optimization. However, these constructed portfolios are very sensitive to the mean estimations which means that small errors in the mean estimations create large deviations from the desired portfolio [3]. Minimum variance weighting does not take mean estimations into account, and a portfolio based on mean-variance optimization is about one tenth as sensitive to errors in the estimations of the variances and covariances as it is to errors in the estimations of the means [3].

In the minimum variance portfolio, the weights are determined by finding the linear combination of the assets that gives the smallest standard deviation (risk) of the future portfolio value [11]. This is

equivalent to maximizing the Sharpe ratio of the portfolio (the definition of the Sharpe ratio is found in Section 2.7.1).
The weights are obtained from solving the following optimization problem

$$\underset{\boldsymbol{w}}{\operatorname{argmin}} \sqrt{\boldsymbol{w}^T \Sigma \boldsymbol{w}}$$
$$\textit{such that} \begin{cases} \sum_{i=1}^{N} w_i = 1 \\ w_i \geq 0 \text{ for all } i \\ w_i \leq c \text{ for all } i \end{cases} \tag{11}$$

These three constraints holds for all weighting techniques and are further explained in Section 3.2.

2.5 Cluster Weighting Strategies

Clustering refers to a very broad set of techniques for finding groups, i.e clusters, in a data set. When the stocks in a dataset are clustered, they are partitioned into distinct clusters so that the stocks within each cluster are similar to each other. There exists many different algorithms to find clusters and many different measures of similarity to compare the data. In terms of clustering stocks together, you can use different stock data as a basis for clustering. In this thesis we will cluster stocks together based on four different measures of similarity for the K-means algorithm based on two different types of stock data. The first type of stock data is the stocks historical one-month return.

The second type of data is fundamental data. The fundamental data $\boldsymbol{F}$ we will use is

$$\boldsymbol{F} = \frac{1}{2}(\boldsymbol{r}_i \odot \frac{1}{\boldsymbol{n}_i} + \boldsymbol{\eta}_i \odot \frac{1}{\boldsymbol{n}_i}) \tag{12}$$

where $\boldsymbol{r}_i$ is a vector with 5 years of monthly historical revenues for stock i, $\boldsymbol{n}_i$ is a vector with 5 years of monthly historical data on the number of shares outstanding for stock i and $\boldsymbol{\eta}_i$ is a vector with 5 years of monthly historical net incomes for stock i. The $\odot$ symbol represents element-wise multiplication so that $\boldsymbol{F}$ is a vector of the same size as $\boldsymbol{r}_i$, $\boldsymbol{n}_i$ and $\boldsymbol{\eta}_i$.

2.5.1 K-means clustering

In K-means clustering the data is partitioned into K clusters. Let $C_1, ..., C_K$ denote the sets of all clusters. If the i:th stock is in the j:th cluster, then $i \in C_j$. The clusters satisfy two properties

$$C_1 \cup C_2 \cup ... \cup C_K = \{1, .., N\} \tag{13}$$

$$C_j \cap C_l = \emptyset \quad \text{for all} \quad j \neq l \tag{14}$$

Equation (13) states that each stock belongs to at least one of the K clusters and (14) states that the clusters are non-overlapping, i.e no stock belongs to more than one cluster [12].

There are various K-means algorithms. In this thesis, the K-means++-algorithm is used. At any given time, let $D(\boldsymbol{x}_i, \boldsymbol{c}_j)$ denote the similarity between $\boldsymbol{x}_i$ and $\boldsymbol{c}_j$ according to some similarity measure D, where $\boldsymbol{x}_i$ and $\boldsymbol{c}_j$ are more similar the smaller D is. X is an $N \times 60$ matrix, containing N stocks at a given time and the 5 year historical data of these stocks. The stock $\boldsymbol{x}_i$ is a row in X. The following steps defines the K-means++ algorithm:

1a. Choose an initial centroid $\boldsymbol{c}_1$ as $\boldsymbol{c}_1 = \boldsymbol{x}_i$ uniformly at random from X.
1b. Choose the next centroid $\boldsymbol{c}_2$ as $\boldsymbol{c}_2 = \boldsymbol{x}_m \in X$ with probability $\frac{D(\boldsymbol{x}_m, \boldsymbol{c}_1)^2}{\sum_{\boldsymbol{x}_i \in X} D(\boldsymbol{x}_i, \boldsymbol{c}_1)^2}$.
1c. Repeat step 1b until K centroids are chosen.
2. For each $j \in \{1, ..., K\}$, set the cluster C_j to be the set of points in X that are closer to $\boldsymbol{c}_j$ than they are to $\boldsymbol{c}_l$ for all $l \neq j$.
3. For each $j \in \{1, .., K\}$, set $\boldsymbol{c}_j$ to be the center of mass of all points in C_j.

4. Repeat steps 2 and 3 until the centroids no longer change [13].

When we cluster stocks we consider each stock to be in a 60-dimensional space since we have 60 historical data points. When we state that a stock $\boldsymbol{x}_i$ belongs to the cluster which it is nearest we mean that it belongs to the cluster which has its center of mass nearest to this stock in this 60-dimensional space. The question is then how 'nearest' is defined in this space, and there are various ways of defining the similarity measure $D(\boldsymbol{x}_i, \boldsymbol{c}_j)$. However, it is important to note that these similarity measures are not necessarily distances. In this thesis the following similarity measures will be examined:

- Squared Euclidean Distance

$$D_{SED}(\boldsymbol{x}_i, \boldsymbol{c}_j) = (\boldsymbol{x}_i - \boldsymbol{c}_j)(\boldsymbol{x}_i - \boldsymbol{c}_j)^T \tag{15}$$

  The Euclidean distance is a common measure of distance and often referred to as the L^2-norm. The squared Euclidean distance, however, is not a distance as it does not obey the triangle inequality. This can be easily proven in one dimension, and the same argument can be applied in any number of dimensions. If we walk to $x = 2$ from the origin, equation (15) equals 4. If we walk from the origin to $x = 1$ or from $x = 1$ to $x = 2$, equation (15) equals 1. The total distance to walk from the origin to $x = 2$ is then 2 in the case that we make a stop in $x = 1$, but the total distance is 4 if we do not make a stop. However, the distance is obviously the same and so the squared Euclidean distance can not be a measure of distance. Instead, it is a measure of similarity. For our purposes, this is preferable. Two stocks are considered more similar the smaller D_{SED} is, and the further apart they are the faster the rate of dissimilarity grows, as made obvious by the non-metric example above. The intuitive interpretation of using the squared Euclidean distance is that we assign each point to the cluster with the closest mean.

- City Block Distance

$$D_{CBD}(\boldsymbol{x}_i, \boldsymbol{c}_j) = \sum_{l=1}^{6} 0|\boldsymbol{x}_{i,l} - \boldsymbol{c}_{j,l}| \tag{16}$$

  where l denotes which element of $\boldsymbol{x}_i$ and $\boldsymbol{c}_j$ is being inspected. This is the classic L^1-norm. The interpretation of using city block distance is that we assign each stock to the cluster with the closest median.

- Cosine Similarity

$$D_{Cos}(\boldsymbol{x}_i, \boldsymbol{c}_j) = 1 - \frac{\boldsymbol{x}_i \boldsymbol{c}_j^T}{\sqrt{(\boldsymbol{x}_i \boldsymbol{x}_i^T)(\boldsymbol{c}_j \boldsymbol{c}_j^T)}} \tag{17}$$

  Cosine similarity is not a distance, it is a similarity measure. When we use cosine similarity we think of $\boldsymbol{x}_i$ and $\boldsymbol{c}_j$ not as points, but as the vectors to those points. All vectors in the space, i.e. the vectors pointing to all stocks and centroids, are normalized to unit length. The cosine similarity between a stock and a centroid is then determined from the cosine of the angle between their two normalized vectors. The second term of equation (17) is the cosine angle between the vectors, and since $\cos(x) \to 1$ as $x \to 0$, D_{Cos} tends to 0 as the angle between the two vectors tends to 1.

- Correlation Distance

$$D_{Corr}(\boldsymbol{x}_i, \boldsymbol{c}_j) = 1 - \frac{(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{c}_j - \bar{\boldsymbol{c}})}{\sqrt{(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^T}\sqrt{(\boldsymbol{c}_j - \bar{\boldsymbol{c}})(\boldsymbol{c}_j - \bar{\boldsymbol{c}})^T}} \tag{18}$$

  where $\bar{\boldsymbol{x}}$ and $\bar{\boldsymbol{c}}$ are the mean vectors of $\boldsymbol{x}$ and $\boldsymbol{c}$ for all i and j respectively. The second term is the Pearson correlation coefficient, ρ, which we know has a range between -1 and 1. The more similar $\boldsymbol{x}_i$ and $\boldsymbol{c}_j$ are, the closer ρ is to 1. If they correlate negatively, ρ will instead tend to -1. This means that D_{Corr} ranges between 0 and 2, and the closer to 0 the value is the more similar the stock is to the centroid. We assign each stock to the centroid for which D_{Corr} is the smallest, i.e. to the cluster to which it has the highest positive correlation.

2.5.2 Choosing the number of clusters K

A value of K has to be chosen before running the algorithm. In order to choose a suitable value of K two methods are used, namely the silhouette index and the ratio of between-cluster sum of squares to total sum of squares.

- Silhouette Index (SI)

  The silhouette index, SI, works by measuring how similar an object is to its own cluster compared to other clusters. The index works by assigning a value z_i to each object i, which in our case means that we assign a z_i to each stock i for each month. Then we calculate z_i as

$$z_i = \frac{b - a}{\max(a, b)} \tag{19}$$

  where a is the average length to all other stocks within the same cluster and b is the average length to all stocks within the nearest cluster. The larger the value of z_i is, the better the stock i is matched to its own cluster in comparison to how poorly it is matched to neighboring clusters. The silhouette index SI is then calculated as the average value of z_i for all stocks i in the dataset, i.e

$$SI = \frac{\sum_{i=1}^{N} z_i}{N}. \tag{20}$$

  If most stocks have a high value of z_i, then the clustering configuration is appropriate and the silhouette index will be large too. If many points have a low or negative value, then the clustering configuration may have too many or too few clusters and it will be reflected by a small value of the silhouette index [14].

- Between-Cluster Sum of Squares to Total Sum of Squares Ratio (r)

  The between-cluster sum of squares to total sum of squares ratio reflects how much of the variance between clusters that has been accounted for [15]. We denote the between-cluster sum of squares by $BCSS$ and it is the sum of the squared Euclidean distance for all objects to all centroids of the clusters they do not belong to. The total sum of squares, denoted TSS is the sum of the between-cluster sum of squares and the within-cluster sum of squares. The within-cluster sum of squares is the sum of the squared Euclidean distance for all objects to the centroids of the clusters they do belong to and is denoted $WCSS$. The ratio r then becomes

$$r = \frac{BCSS}{TSS} = \frac{BCSS}{BCSS + WCSS}. \tag{21}$$

In order to choose a suitable value of K we will combine the silhouette index SI and the between-cluster sum of squares to total sum of squares ratio r in Equation (20) and Equation (21), respectively. We run the K-means algorithm 50 times for each month to stabilize across different initializations, and the average value over the 50 runs is used for the silhouette index. For the between-cluster sum of squares to total sum of squares ratio, we note empirically that the mean of r increases with the number of clusters while the variance decreases. This is what we expect. As the number of clusters increase, there will generally be fewer stocks in each cluster. This decreases the within-cluster sum of squares. By the same argument, the number of stocks in other clusters increases with the number of clusters, and then the between-cluster sum of squares increases too. Then we note that when the variance of r is small, the mean value of r is large (≥ 0.99), i.e. the value stabilizes as it approaches 1. By setting a small enough threshold for the variance of r we are then guaranteed that the ratio is large enough for our purposes, i.e. there is a lot more variance between clusters than within clusters, and since the variance is small we know that this ratio is stable across different initializations of the algorithm.

The value of K is then chosen as the K_i for which the Silhouette index SI is as large as possible given that the variance of r_i is over some threshold c and that $K \geq 3$, i.e.

$$\begin{aligned} &K = \text{argmax}\left(SI(K_i)\right) \text{ for which} \\ &\begin{cases} K \geq 3 \\ \text{Var}(r_i) \leq \gamma \end{cases} \end{aligned} \tag{22}$$

By examining a small subsample we find that $\gamma = 0.0001$ is a suitable choice, i.e. that with this limit for the variance of r_i, the mean of $r_i \geq 0.99$ which is sufficiently large to consider the within-cluster variance to be small in comparison to the between-cluster variance, while the variance of r is low enough to consider r stable across different initializations of the algorithm.

2.5.3 GARCH-Estimation of the Covariance Matrix

Generalized autoregressive conditional heteroskedasticity, or GARCH, is a method of estimating the stylized features of a return process $\{Z_t\}$ [16]. The stylized features of a process include things as volatility and tail heaviness. Here, we will focus on the volatility. We will use GARCH for two different cases. In the first case, our aim is to model the covariance matrix between the constructed clusters and then determine the cluster weights using minimum variance optimization. In order to determine a covariance matrix we need to determine the returns of each cluster in each timestep. This can not be done without weighting the constituents in some way beforehand. We will use equally weighted constituents. The end result is that the cluster weights are based on minimum variance optimization of the covariance matrix that has been determined using a multivariate GARCH-approach, and the cluster constituents are equally weighted within each cluster. The second case consists of applying GARCH directly on stocks so that a comparison can be made if GARCH works better in a cluster framework or traditional framework.

The idea behind GARCH-estimation of the covariance is different from normal covariance estimation. The standard way is to take the historical data and see what the covariance is for those historical data points. In GARCH-estimation we look at the historical data points and try to estimate what the covariance should be in the next timestep. This is done by accounting for errors in previous predictions when trying to minimize the prediction error in the ongoing prediction. Robert Engle, who won the Noble prize in 2003 for his work with ARCH/GARCH models, writes in one paper that the regression coefficients for an ordinary least squares regression will have too narrow confidence intervals of the standard errors in the presence of heteroskedasticity, creating a false sense of precision. In ARCH/GARCH-modeling we do not consider this as a problem to be corrected, but rather we treat the heteroskedasticity as a variance to be modeled. This not only solves the problem of deficiencies in the least squares approach, it gives us an estimation of the variance for each error term too, which is of particular interest in finance [17].

In the univariate case, a GARCH(p,q)-process is defined as [16]

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i Z_{t-i}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2 \tag{23}$$

where $\alpha_0 > 0$ and $\alpha_i, \beta_i \geq 0$ for all i. We can see that the volatility at time t is a function of the squared returns and the squared volatilities in the previous steps.
However, in the context of finding the variance and covariance of clusters of stocks we are interested in the multivariate version of a GARCH-model. We let $\sigma_t^2 \to H_t$ where H_t is the covariance matrix. We construct the multivariate version of Equation (23) using the BEKK-formulation. We set $p = q = 1$ so that we have a multivariate GARCH(1,1)-model. It can be written as [18]

$$H_t = CC^T + AZ_{t-1}Z_{t-1}^T A^T + BH_{t-1}B^T \tag{24}$$

where C is a lower triangular matrix and A and B are parameter matrices of the same size as H_t, i.e. $N \times N$-matrices. The multivariate case is basically just the same as the univariate case but with matrices instead of scalars. In our case, the number of instruments to be estimated is quite large, and the computational power needed to solve Equation (24) for each month is simply too large. Therefore we will use simpler verions of the parameter matrices so that $A = \alpha I_N$ and $B = \beta I_N$ where I_N is the identity matrix of size $N \times N$. Since multiplying a matrix with the identity matrix gives the same matrix, this simplifies equation (24) to

$$H_t = CC^T + \alpha^2 Z_{t-1}Z_{t-1}^T + \beta^2 H_{t-1}. \tag{25}$$

Using the iterative form of Equation (25) we can get an estimate of the next month's covariance matrix H_{t+1} of the clusters using our historical data. We can then use a minimum variance optimization on H_{t+1} to weight the clusters and the stocks.

2.6 Overview of Weighting Strategies

Many of the weighting strategies that we will use have been tested before in different articles and scenarios, and some are used by many portfolio managers on a regular basis. Here, we have used the MSCI USA Index as benchmark, which is market capitalization weighted, but MSCI also offers alternative indexes which are weighted using equal weighting, minimum variance optimization and many other methods.

Fundamentally weighted indexes have recently been developed as an alternative to market capitalization weighted indexes and price indexes, both of which were seen as the two most efficient indexation methods [9]. The fundamental weighting strategy introduced by Arnott, Hsu and Moore [19] has challenged this idea and many empirical studies have investigated different fundamental weighting strategies since then. The allocation of the portfolio weights depends on fundamental characteristics of companies, and in this thesis the fundamental weighting strategy used is the one introduced by Arnott, Hsu and Moore, and it is used by index funds such as the RAFI Fundamental Index. However, other fundamental characteristics can be used to construct the portfolio weights. Perhaps the most notable characteristic which we have not used is the number of employees the company in question has.

Another type of recently developed indexes are risk-weighted indexes. These are weighting methods which seek to reduce risk by diversifying. Examples of risk-weighted methods are equal weighting, minimum variance optimization, maximum Sharpe ratio optimization and equally weighted risk contribution [20]. The risk-adjusted alpha strategy introduced by Agarwal [9], which is examined in this thesis, belongs to this category of weighting strategies too. In Agarwal's paper, three different risk-weighted methods were introduced, and all three of them use regression to find alpha and beta. The estimates of alpha and beta are then used to construct the weights. The risk-adjusted alpha strategy outperformed the market cap weighted portfolio and the other two investigated strategies. Due to this fact, we have chosen to investigate the risk-adjusted alpha strategy in this thesis. The other two methods investigated in Agarwal's paper was weighting strategies using Treynor's square ratio and the appraisal ratio.

There are few studies that have investigated the effects of clustering stocks [21]. However, classifying stocks into distinct groups is important for all investors. The primary method of classifying stocks into groups is to form industry groups [22]. The idea of industry groups is that most of the companies within one industry group tend to move as a whole on the market. By knowing the trends in place within the industry group, investors can better understand the investment potential of the companies within that group. One example of a weighting strategy that classifies stocks into distinct groups without using clustering is the risk-cluster equal weight strategy. This method is too naive for some investors because the portfolio allocation weights are dictated largely by the arbitrary choice of which group the stocks belong to [10].

In this thesis we cluster stocks together into groups by two different types of stock data, namely their historical one-month returns and their fundamental data. The clustering is done using the K-means algorithm. We were not able to find previous research on a weighting strategy that cluster stocks based on historical one-month returns by using the K-means algorithm. However there are previous research that cluster stocks based on historical one-month return data by using hierarchical clustering. In our study we tried to cluster stocks by using hierarchical clustering but our data did not show a hierarchical nature. This is line with Marvin's arguments [15], which state that there is no hierarchical nature to stock data. Marvin has performed clustering of stocks based on fundamental data using the K-means algorithm too, and we have used the same fundamental data to create our clusters. In her paper, she only clusters the stocks using the correlation distance measure.

2.7 Portfolio evaluation

In this section different measures to evaluate a portfolios performance are presented.

2.7.1 Sharpe Ratio

Investors are risk averse in general. Given the same return of two portfolios they would prefer the one with less risk. The Sharpe ratio is a way to evaluate portfolios with different returns and different levels of risk. The Sharpe ratio, SR, is defined as

$$SR = \frac{\mathrm{E}[R_P^A - r_F^A]}{\sigma_P} \tag{26}$$

where R_P^A is the annual return, r_F^A is the annualized risk-free rate and σ_P is the standard deviation of the portfolio's annual returns. In practice we calculate the Sharpe ratio as the mean of the annual portfolio return subtracted by the mean risk-free rate for that year and then divide it by the standard deviation of the annual returns of the portfolio. This means that the Sharpe ratio is the average excess return over the risk-free rate per unit of volatility. It is the most commonly used measure of a portfolio's risk-adjusted return. As seen in (26) a higher value of SR is preferable [6].

2.7.2 Tracking Error

Tracking error is the standard deviation of the difference between the returns of a portfolio and its benchmark [23]. The tracking error, TE, is defined as

$$TE = \sqrt{\mathrm{Var}(R_P^A - R_B^A)}. \tag{27}$$

It is a measure of how the portfolio changes in relation to the benchmark. If the tracking error is 0, it means that the new portfolio changes exactly like the benchmark. As the tracking error increases, the portfolio behave less and less like the benchmark. This means that for portfolios that strive to replicate an index fund we want the tracking error to be as close to 0 as possible. The less an investor believes in the optimality of the benchmark portfolio, the less he needs to worry about the tracking error.

2.7.3 Information Ratio

Information ratio is a measure of the excess return over benchmark divided by the tracking error. The information ratio, IR, is defined as

$$IR = \frac{\mathrm{E}[R_P^A - R_B^A]}{TE} \tag{28}$$

where TE is the tracking error defined in equation (27). Just like Sharpe ratio, the information ratio is a measure of risk-adjusted return. The difference is that while the Sharpe ratio attempts to measure the risk-adjusted return in relation to the risk-free rate, the information ratio attempts to measure it in relation to the benchmark. A positive information ratio indicates outperformance in relation to the benchmark and a negative information ratio indicates underperformance. Underperformance with a low tracking error is considered worse than underperformance with a high tracking error [23]. This seems unintuitive at first, but if we have lower returns than the benchmark with a low tracking error it actually means that we will consistently underperform the benchmark no matter how the market moves, while with a high tracking error our portfolio performs differently from the benchmark and we can outperform the benchmark if the market behaves differently.

3 Data and Methodology

3.1 Data

The data we use is provided by Öhman from Bloomberg. The data consists of monthly data between 1996-2016 of all stocks that were in the MSCI USA Index at any time between 2002-2016. For any given month we will use the last five years of data as a basis for the analysis. Each month in the time period 2002-2016 we update our portfolio constituents rebalance our portfolio weights. Since we have chosen an index containing only American stocks, all stocks are traded in US dollars and currency exchanges do not need to be considered. The reason the initial portfolio is constructed five years after the first month of data is because we need to have a sufficient amount of historical data to base the analysis on for the weighting strategies which require historical data.

3.2 Weights and Constituents

All constructed portfolios will have the constraints

$$\sum_i w_i = 1 \tag{29}$$

$$w_i \geq 0 \text{ for all } i \tag{30}$$

$$w_i \leq 0.1 \text{ for all } i. \tag{31}$$

The first constraint states that we must invest all our capital into stocks. The second constraint states that short-selling is not allowed in the construction of the portfolios. This is since the MSCI USA Index does not utilize short-selling when trading. Finally, we impose a constraint that is a simplification of the 5-10-40 rule which most funds follow. The rule states that no more than 10% of the capital can be invested in one single stock (the part of the rule we have used as a constraint) and that all stocks that have a weight larger than 5% summed together should not have a weight larger than 40%.

The constituents of our portfolio will be the stocks which were in the MSCI USA Index at a given time for which there is 60 months of historical data on the closing prices available. This means that the benchmark will not be the return of the actual index but we will need to construct a fictional index by using the market capitalization weighting method in Section 2.4.1. The choice of constructing the fictional index is made so that the covariance matrix and methods built on five-year averages should have enough data to be meaningful. For some stocks there is not 60 months of available data. In some cases this is because the data was missing in Bloomberg, but generally it is because the stock was not publicly listed for the entirety of the 60-month period.

3.3 Sample Covariance Matrix

The sample covariance matrix will be estimated from the last 60 months of historical data. To construct the sample covariance matrix we will use the *Honey, I Shrunk the Sample Covariance Matrix* script [24]. This is since if we try to estimate the sample covariance matrix by standard methods when the amount of data points for each stock (we use 60 months of historical data) is less than the number of stocks we want to investigate (the index has around 500 constituents) we will not necessarily get an invertible matrix which it must be in order to be a covariance matrix. On top of that, the sample covariance matrix will most likely get heavy outliers which would have a huge effect on the calculations made with the covariance matrix, and these outliers are likely to produce unwanted results. The script takes care of both of these problems which should lead to better results [25].

The method employed by the script is a method called shrinking and is described in detail by Ledoit and Wolf [25] but we will outline the idea. We start with the matrix M which is an $N \times 60$-matrix of historical excess returns. We then construct what we will call our naive sample covariance matrix as

$$\hat{\Sigma}_{naive} = \frac{1}{60} MM' \tag{32}$$

which is an $N \times N$-matrix. We calculate the mean variance and mean covariance of $\hat{\Sigma}_{naive}$ and use these to construct another highly structured $N \times N$ sample covariance matrix which we will call $\hat{\Sigma}_{struct}$. The highly structured matrix consists of a vector of the sample variances on the diagonal, and the covariances between the stocks are set as the average correlation between stocks. This means that the covariances between stocks are the same between all stocks in Σ_{Struct}. The sample covariance matrix we seek, denoted $\hat{\Sigma}$, is then acquired from

$$\hat{\Sigma} = \delta\hat{\Sigma}_{struct} + (1-\delta)\hat{\Sigma}_{naive} \tag{33}$$

where $0 \leq \delta \leq 1$ is called the shrinking parameter.

3.4 Weighting Strategy Methods

For some methods presented in Section 2.4 and Section 2.5 the entire methodology was not explained, only the general theory. Here we will present the methodology that is specific to our thesis.

3.4.1 Minimum Variance - Choice of c

In the minimum variance weighting strategy we have a constant c which determines the largest possible weight any stock can receive. As described in Equation (31) we have simplified the 5-10-40 rule so that we set $c = 0.1$ in the minimum variance optimization described in Equation (11).

3.4.2 Risk-Adjusted Alpha & Fundamental Weighting Producing Negative Weights

For the risk-adjusted alpha method, $\alpha_{i,t}$ for stock i at time t is calculated according to Equation (7) each month. For some stocks, the value of $\alpha_{i,t}$ will be less than 0. This would give a negative weight according to equation (9), which means short-selling of that stock. The same problem can arise in the fundamental weighting method described in Equation (10) since we can have large negative cash flows or a large negative revenue, or even a negative book value. When a negative weight occurs for either of the two methods we set that weight to 0 since the constraint (30) means that short-selling is not allowed. The weight vector that now contains only positive weights is then normalized so that the sum of the weights equals 1.

3.4.3 K-Means Clustering

The choice of the upper limit c for the variance has more impact on the squared Euclidean distance and the city block similarity measures. In the case of the cosine and correlation similarity measures, $\mathrm{Var}(r_i)$ is smaller than c (and the mean is larger than 0.99) almost always. However, for these methods the silhouette index suggests a large number of clusters in general and so the choice of c is irrelevant when the minimum number of clusters is $K = 3$. Due to this argument we keep $\gamma = 0.0001$ even when the cosine and correlation similarity measures are used and let the silhouette index determine the number of clusters when these similarity measures are used.

The cluster assignment process and the choice of K is repeated each month. This results in a cluster assignment matrix, where each stock in the index is assigned a cluster between $1, .., , K^{(t)}$ where t is the month currently clustered and $K^{(t)}$ is the number of clusters for that month.

Once the cluster assignment matrices for all similarity measures have been constructed, various weighting techniques are applied to these clusters. Except for when the clusters are GARCH-weighted, discussed in Section 2.5.3, all clusters are equally weighted. The following weighting techniques are used to weight the constituents of each cluster.

- Equal Weighting

  All clusters that contain more than one stock are equally weighted, and the stocks within each cluster are equally weighted according to Section 2.4.2. Clusters that contain only one stock are removed, since they would gain a larger weight than 0.1 if they were kept, which is not allowed according to the constraint in Equation (31).

A manual check of the largest weights in the weight matrices shows that this does not yield any singular stock to have a weight larger than 10% when clustering on historical one-month returns. In the case of clustering on fundamental data we find that the largest weights in some cases exceed 10%. The stocks which have weights larger than 10% are treated as outliers and the clusters they belong to are removed. The remaining clusters are equally weighted and the stocks within each cluster is again equally weighted.

- Fundamental Weighting, Market Capitalization Weighting & Risk-Adjusted Alpha Weighting

  All clusters with more than one stock are equally weighted. However, when we weight the constituents of each cluster there is no easy way to guarantee that any stock will not get a weight larger than 0.1. What has been done to solve this is that each time a constituent has a weight $w_i > 0.1$ we adjust that weight so $w_i = 0.1$ and then distribute the rest of the capital initially assigned to w_i equally on all remaining stocks that have not already been readjusted. This ensures that the largest possible weight is $w_{max} = 0.1$.

- Minimum Variance Weighting

  We solve the minimum variance problem described in Equation (11) for the constituents in each cluster. This means that the total weight when summing over all clusters won't equal to 1, since the total weight for each cluster equals 1. Since we equally weight all clusters this is adjusted by dividing all weights by the number of clusters, which both makes all clusters equally weighted and the sum of all weights to 1. We need to adjust c in Equation (11) so that the largest possible weight is 0.1 after we divide by the number of clusters rather than before we divide by the number of clusters. This is done by changing the upper limit c in the optimization problem as

$$c = \frac{\#\text{ clusters}}{10}. \tag{34}$$

When we then divide all weights by the number of clusters the upper limit is, just as before, $w_{max} = 0.1$.

3.4.4 GARCH-Approach For Weighting the Clusters

In order to weight the clusters using minimum variance optimization of a GARCH-estimated covariance matrix we need a measure of the returns of the clusters in the previous timesteps. In this report we have set each clusters return to the average return of its constituents in each timestep. This is equivalent to equally weighting the constituents and so the cluster constituents must be equally weighted within each cluster using this approach.

In order to find H_{t+1}, the estimated covariance matrix one month after the available data, we have used Kevin Sheppard's MFE Toolbox [26], a toolbox of MATLAB scripts to help with ARIMA and GARCH calculations and simulations.
In the minimum variance optimizer we set

$$c = \tfrac{1.7}{\#\text{ clusters}}.$$

For some key values this means that when we have two clusters we get $c = 0.85$, when we have five clusters we get $c = 0.34$ and when we have ten clusters we get $c = 0.17$. This is to guarantee that not only the cluster with the least variance is weighted.

3.4.5 GARCH-Approach When Comparing GARCH on Clusters and Stocks

Initially, we wanted to investigate the differences of how GARCH works when applied to clusters to when GARCH is applied directly to stocks. However, due to the extensive amount of computational power needed to calculate the GARCH covariance matrices for a large number of data points, one would need access to a computer with GPU's or to some server for an extended period of time. Still, there would be limitations to what could actually be done. The BEKK-formulation used to estimate H_t presented in Equation (24) requires that we have a larger number of time steps than we have stocks, which restricts us to using a maximum of 59 stocks when using five years of historical monthly data.

If we had access to a server to perform the necessary calculations, we would have selected 59 stocks at random in each month and performed GARCH-estimation to construct their covariance matrices. The same 59 stocks would be clustered, and in each month we would use GARCH-estimation to construct the covariance matrices of the clusters. The reason we would choose 59 random stocks each month rather than choosing 59 stocks for the entire period is because there might be a bias towards one method or the other if we only use stocks that are in the index the entire time, since these are typically larger stocks, or stocks that have performed well. The two methods would then both get weights assigned according to minimum variance optimization, where the clusters' constituents would be equally weighted. From this, we could get an idea if it is preferable to use GARCH directly on stocks or in a cluster framework.

4 Results

4.1 Traditional Weighting Strategies

In Table 1 the results of the traditional weighting strategies are presented.

Weighting Technique	**Average Annual Return**	**Excess Return over Benchmark**	**Sharpe Ratio**	**Tracking Error**	**Information Ratio**	**Annual Standard Deviation**
Market Capitalization	4.50%		0.30			15.59%
Equal	6.88%	2.38%	0.41	5.54%	0.50	18.39%
Risk-Adjusted Alpha	18.95%	14.45%	1.39	6.68%	2.10	13.78%
Fundamental	6.44%	1.94%	0.39	4.39%	0.56	18.63%
Minimum Variance	8.73%	4.23%	0.63	4.36%	0.89	13.73%

Table 1: The results of the traditional weighting strategies.

In Table 1 it can be seen that the risk-adjusted alpha strategy has outperformed all other traditional weighting strategies in terms of average annual return, Sharpe ratio and information ratio. The market capitalization portfolio which was used as benchmark has the lowest average annual return.

In Table 2 the Jensen's alpha and beta for the different weighting strategies are presented. The p-values are used to determine if the values are significant at the 1% level.

Weighting Technique	**Alpha**	**Beta**	$p < 0.01$? **Alpha**	$p < 0.01$? **Beta**
Equal	0.02	1.13	No	Yes
Risk-Adjusted Alpha	0.15	0.81	Yes	Yes
Fundamental	0.02	1.17	No	Yes
Minimum Variance	0.05	0.86	Yes	Yes

Table 2: The alphas and betas of the traditional weighting strategies, as well as their significance.

The p-value is the probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis is true. The null hypothesis H_0 states that alpha$= 0$ against the alternative hypothesis H_A which states that alpha$\neq 0$. The same holds for beta. In this thesis the null hypothesis is rejected when the p-value is less than 0.01, i.e. a 1% chance of rejecting the null hypothesis when it is true. With such a value there is only a 1% chance that the results of the regression analysis would have occurred in a random distribution, or, there is a 99% probability that the coefficient is having some effect on the regression model. When the null hypothesis is rejected, the result is said to be statistically significant [6].

As seen in Table 2, the p-values for the parameter estimates of alpha for equal weighting and fundamental weighting are larger than 0.01. This means the null hypothesis is not rejected for these parameter estimates and hence the obtained values of alpha can not be used at this level of risk.

In Table 3 the average annual returns for the traditional strategies over different time periods are presented.

Weighting Technique	2002-2007	2008-2010	2011-2016
Market Capitalization	4.65%	−3.97%	8.85%
Equal	8.07%	1.77%	8.32%
Risk-Adjusted Alpha	19.05%	16.54%	20.07%
Fundamental	6.45%	−0.098%	10.34%
Minimum Variance	9.02%	5.19%	10.26%

Table 3: Average annual returns for the traditional weighting strategies over different time periods.

As seen from Table 3, the equally weighted portfolio, the risk-adjusted alpha portfolio and the minimum variance portfolio produced positive returns in all time periods, including the period between 2008-2010 when there was a financial crisis. The market capitalization portfolio and the fundamental portfolio made a loss during the financial crisis.

4.2 Cluster Weighting Strategies

4.2.1 Cluster Weighting Using Historical Return Data

In Table 4 the results of the K-means cluster weighting strategies using historical return data are presented.

Weighting Within Cluster Clustering Method	Average Annual Return	Excess Return over Benchmark	Sharpe Ratio	Tracking Error	Information Ratio	Annual Standard Deviation
Market Capitalization						
Squared Euclidean	5.11%	0.61%	0.31	6.79%	0.16	18.71%
City Block	4.04%	−0.46%	0.25	6.85%	0.04	19.91%
Cosine	2.53%	−1.97%	0.17	3.63%	−0.64	13.21%
Correlation	1.37%	−3.13%	0.08	5.66%	−0.61	13.51%
Equal						
Squared Euclidean	6.98%	2.48%	0.37	10.78%	0.31	21.70%
City block	6.12%	1.62%	0.34	8.81%	0.29	21.40%
Cosine	3.92%	−0.58%	0.26	5.80%	−0.12	14.94%
Correlation	3.07%	−1.43%	0.20	8.15%	−0.18	15.92%
Risk-Adjusted Alpha						
Squared Euclidean	17.08%	12.58%	0.86	12.20%	1.08	21.32%
City Block	17.19%	12.69%	0.87	11.01%	1.21	21.06%
Cosine	16.29%	11.79%	1.23	5.59%	2.03	13.26%
Correlation	14.59%	10.09%	1.01	8.83%	1.11	14.58%
Fundamental						
Squared Euclidean	7.85%	3.35%	0.42	8.61%	0.50	21.42%
City Block	6.58%	2.08%	0.37	7.67%	0.36	20.24%
Cosine	3.83%	−0.67%	0.26	3.92%	−0.23	14.15%
Correlation	2.21%	−2.29%	0.15	5.45%	−0.43	15.11%
Minimum Variance						
Squared Euclidean	10.95%	6.45%	0.76	7.46%	0.82	14.48%
City Block	9.15%	4.65%	0.58	5.48%	0.85	16.32%
Cosine	8.91%	4.41%	0.90	7.26%	0.49	9.23%
Correlation	8.71%	4.21%	0.83	7.34%	0.47	9.89%

Table 4: Results of the K-means clustering when clustering is done using data of the historical returns.

When equally weighting clusters formed by using squared Euclidean distance as the similarity measure and then applying the traditional weighting strategies on the cluster constituents, we get a higher average annual return for all strategies except the risk-adjusted alpha strategy. The cosine similarity and correlation similarity have produced higher average annual returns while reducing the standard deviation for the minimum variance strategy, and the same is true for the cosine similarity using a risk-adjusted alpha strategy.

In Table 5 the average annual returns over different time periods for K-means clustering using historical return data are presented.

Weighting Technique	**2002-2007**	**2008-2010**	**2011-2016**
Market Capitalization			
Squared Euclidean	7.90%	0.81%	5.41%
City block	4.58%	0.01%	5.58%
Cosine	3.82%	−3.91%	4.59%
Correlation	4.21%	−3.98%	1.31%
Equal			
Squared Euclidean	10.11%	10.15%	2.42%
City block	7.52%	4.70%	5.44%
Cosine	6.82%	0.50%	2.80%
Correlation	7.20%	1.09%	0.05%
Risk-Adjusted Alpha			
Squared Euclidean	18.02%	25.19%	12.32%
City Block	19.10%	19.96%	13.96%
Cosine	16.40%	13.41%	17.65%
Correlation	17.02%	13.48%	12.74%
Fundamental			
Squared Euclidean	10.00%	5.02%	7.15%
City block	8.51%	2.29%	6.86%
Cosine	5.00%	−1.11%	5.20%
Correlation	5.38%	−3.02%	1.78%
Minimum Variance			
Squared Euclidean	12.13%	12.79%	8.88%
City Block	8.03%	7.20%	11.28%
Cosine	7.64%	5.04%	12.22%
Correlation	8.19 %	4.43%	11.43%

Table 5: Average annual returns for cluster weighting strategies based on data of historical returns over different time periods.

In Table 5 we can see that all portfolios created using the clusters that were based on the squared Euclidean distance as similarity measure have produced positive returns during all time periods. All these strategies have a higher return than their traditional counterpart during the financial crisis, but a lower return for the years following the financial crisis.

4.2.2 Cluster Weighting Using GARCH-Estimation of the Covariance Matrix

The results of using minimum variance weighting on covariance matrices of the clusters estimated from GARCH where all the cluster constituents are equally weighted are presented in Table 6.

Clustering Method	Average Annual Return	Excess Return over Benchmark	Sharpe Ratio	Tracking Error	Information Ratio	Annual Standard Deviation
Squared Euclidean	7.90%	3.40%	0.51	8.27%	0.41	16.17%
City Block	7.05%	2.55%	0.50	5.84%	0.39	13.97%
Cosine	1.42%	-3.08%	0.07	8.88%	-0.42	10.77%
Correlation	2.42%	-2.08%	0.17	8.32%	-0.33	11.09%

Table 6: Results when using GARCH to estimate the covariance matrices of the clusters.

Weighting Technique	2002-2007	2008-2010	2011-2016
Squared Euclidean	10.22%	6.48%	6.34%
City Block	9.98%	1.51%	6.99%
Cosine	4.78%	0.71%	-1.48%
Correlation	5.19%	1.06%	0.39%

Table 7: Average annual returns in different time periods when using GARCH to estimate the covariance matrices of the clusters.

We can see that the squared Euclidean similarity measure seems most suitable, as it produces the highest return, best Sharpe ratio and it performs well during the crisis period.

4.2.3 Cluster Weighting Using Fundamental Data

In Table 8 below the results of the K-means cluster weighting strategies using fundamental data are presented.

Weighting Within Cluster Clustering Method	**Average Annual Return**	**Excess Return over Benchmark**	**Sharpe Ratio**	**Tracking Error**	**Information Ratio**	**Annual Standard Deviation**
Market Capitalization						
Squared Euclidean	1.32%	−3.18%	0.10	4.43%	−0.62	17.86%
City Block	2.15%	−2.35%	0.15	3.23%	−0.63	17.06%
Cosine	2.82%	−1.68%	0.19	7.35%	−0.08	20.94%
Correlation	3.94%	-0.56%	0.26	3.17%	−0.20	15.12%
Equal						
Squared Euclidean	3.58%	−0.92%	0.23	6.56%	0.01	20.66%
City block	4.68%	0.19%	0.27	8.28%	0.15	21.58%
Cosine	4.73%	0.23%	0.27	9.74%	0.16	23.18%
Correlation	6.52%	2.02%	0.40	5.46%	0.42	17.69%
Risk-Adjusted Alpha						
Squared Euclidean	15.61%	11.11%	0.79	11.08%	1.06	21.15%
City Block	13.48%	8.98%	0.76	8.66%	1.08	18.85%
Cosine	12.09%	7.59%	0.71	5.52%	1.40	17.72%
Correlation	15.04%	10.54%	1.03	5.96%	1.72	14.75%
Fundamental						
Squared Euclidean	4.03%	−0.47%	0.25	6.01%	0.06	20.06%
City Block	4.89%	0.39%	0.29	9.05%	0.19	22.43%
Cosine	4.00%	−0.50%	0.25	11.55%	0.14	25.31%
Correlation	5.68%	1.18%	0.35	4.23%	0.37	17.64%
Minimum Variance						
Squared Euclidean	1.77%	-2.73%	0.14	6.88%	-0.26	19.74%
City Block	3.81%	-0.69%	0.23	10.02%	0.04	22.22%
Cosine	5.01%	0.51%	0.29	9.10%	0.23	23.38%
Correlation	6.07%	1.57%	0.37	6.86%	0.31	18.65%

Table 8: Results of the K-means clustering when clustering is made on fundamental data.

When clustering on fundamental data the correlation similarity measure generally produces the best results in terms of return and standard deviation.

In Table 9 the average annual returns over different time periods for K-means clustering using fundamental data are presented.

Weighting Technique	**2002-2007**	**2008-2010**	**2011-2016**
Market Capitalization			
Squared Euclidean	2.34%	−8.72%	5.68%
City Block	3.38%	−7.11%	5.84%
Cosine	2.80%	−4.19%	6.54%
Correlation	5.40%	−4.69%	7.04%
Equal			
Squared Euclidean	3.06%	−1.86%	6.96%
City Block	4.22%	1.59%	6.73%
Cosine	5.01%	0.81%	6.47%
Correlation	7.98%	1.44%	7.67%
Risk-Adjusted Alpha			
Squared Euclidean	10.39%	22.92%	17.43%
City Block	11.16%	14.00%	15.59%
Cosine	11.86%	7.69%	14.60%
Correlation	15.11%	14.40%	15.30%
Fundamental			
Squared Euclidean	3.55%	−2.90%	8.17%
City Block	5.44%	−1.05%	7.43%
Cosine	4.84%	−2.07%	6.31%
Correlation	6.86%	−2.98%	9.08%
Minimum Variance			
Squared Euclidean	3.47%	-8.62%	5.62%
City Block	5.47%	1.15%	3.51%
Cosine	7.70%	-4.26%	7.23%
Correlation	8.80%	-0.19%	6.60%

Table 9: Average annual returns for cluster weighting strategies based on fundamental data over different time periods.

In Table 9 we can see that for each weighting strategy, the best performing similarity measure performs worse than the best performing similarity measure in clusters formed from historical return data with three exceptions. These are for the market capitalization portfolio, equal portfolio and fundamental portfolio in the time period 2011-2016.

For the alphas, betas and their corresponding p-values of the cluster strategies, see Appendix A.

5 Discussion

5.1 Jensen's Alpha

In many of the weighting strategies we obtain p-values for the parameter estimates of alpha that are larger than 0.01 (see Table 2 and Table 11, Table 12 and 13 in Appendix A). These parameter values are not significant at the risk level of 1% and therefore the obtained estimates of alpha for these weighting strategies can not be used at this level of risk.

5.2 Traditional Weighting Strategies

5.2.1 The Weak Performance of the Market Capitalization Portfolio

The market capitalization weighting technique is the most common way to invest capital for investors operating on portfolios with many constituents, but our results indicate that it is far from the best strategy. In fact, out of all the traditional strategies it is the one that earns the least. The second worst strategy earns almost 2% more per year on average, while the best earns almost 15% more per year on average. The only argument for the market capitalization portfolio is that it has a low standard deviation, but even here it is outperformed by the minimum variance portfolio and the risk-adjusted alpha portfolio.

Another argument against the market capitalization portfolio is its poor performance in the time period between 2008-2010. One of the key indicators of a good portfolio is that it should be able to perform adequately at times when the market does not. The market capitalization portfolio has a negative average annual return during the crisis period 2008-2010. By the end of 2010 the market capitalization portfolio has lost almost 12% of the capital that it had at the start of 2008. The other traditional weighting strategies have all significantly outperformed the market capitalization portfolio in this time period. The equally weighted portfolio still makes a profit. The minimum variance portfolio and the risk-adjusted alpha portfolio are not particularly affected and still make a larger average profit per year than the market capitalization does per year over the whole period between 2002-2016. The fundamental portfolio does not perform especially well for this period but basically breaks even. It is during this period that the market capitalization portfolio performs worst in comparison to all other portfolios and for an investor this could be considered as the worst time to perform poorly.

Then why is market capitalization weighting so prominent? As mentioned in Section 1, there are four main factors used as an argument for the portfolio strategy. The first three are that it is a passive strategy, it consists of highly liquid stocks and it is automatically rebalanced. The effect of these three factors are that it requires little active management and that the transaction costs are small. In this report, no regards are taken to transaction costs, and if we had, the results of the market capitalization portfolio would improve in comparison to the other portfolios. However, it seems unlikely that they would improve so drastically that the market capitalization portfolio would suddenly be the best portfolio. Moreover, the fourth factor that states that the market capitalization portfolio is Sharpe ratio maximized seems incorrect according to our results. We see that both the minimum variance portfolio and risk-adjusted alpha portfolio have much higher returns and lower standard deviations, both resulting in a higher Sharpe ratio.

However, we have excluded all stocks for which there are not five years of historical data. This is in order to be able to use the same constituents in different portfolios where most other portfolios (all traditional but the equally weighted portfolio) require the five years data to be meaningful. The removal of all stocks without five years data helps the other portfolios while it makes the market capitalization portfolio more limited. One of the strengths of the market capitalization portfolio is that it actually doesn't require historical data but can place weights using only todays price and number of shares outstanding. This means that the market capitalization portfolio can include stocks that have just been publicly listed while the methods based on five years of data can only use stocks that have been listed for five years. If we instead would work in a framework where all stocks are included but the other methods could only include the ones with five years of historical data, the market capitalization portfolio and the equally weighted portfolio could, and probably would, improve, while the other methods would not be influenced by this change.

5.2.2 The Beauty of Simplicity - Equally Weighted Portfolios

The simplest possible portfolio to construct is to equally weight all of the constituents in the portfolio. The initial reaction to such a simple strategy is that it should not perform on an adequate level. However, as is well known, equally weighted portfolios generally perform really well and actually beat the market capitalization portfolio in many scenarios [8].

To try to understand how the equally weighted portfolio can perform so well we must consider what the effects are of equally weighting all stocks in a portfolio. To compare the good performance of the equally weighted portfolio to the market capitalization portfolio, we remember that a market capitalization portfolio puts an emphasis on stocks with a high price and many outstanding shares. This means that in a market capitalization portfolio we invest more in the larger companies which, in general, are less risky since they are more established than smaller companies. On top of that, the percentage changes in a stock's price generally decreases as the stock's price increases, and therefore it is harder to make a large profit on more expensive stocks. The smaller companies' stocks, however, can both increase and decrease at a much faster rate.

In an equally weighted portfolio we have put no emphasis on whether the company is small or large and the effect is that in comparison to the market capitalization portfolio we have more capital invested in the smaller companies and less capital invested in the larger companies. That this leads to a higher return could be explained by something called the 'size effect'. The size effect is a widely discussed theory discovered by Banz in 1981 and it states that smaller companies' stocks have higher average returns than larger companies' stocks over long horizons even when adjusting for risk [27]. New research has suggested, however, that since Banz noted the size effect in 1981 and funds were built on the premise of exploiting it, the size effect has diminished or even vanished in US markets [27]. Other research indicates that the level of outperformance of the equally weighted portfolio in comparison to the market capitalization portfolio decreases when the size of the company and the price and liquidity of the stocks increases, which suggests that there is some size effect in play [8].

Plyakha, Uppal and Vilkov conclude that the higher return of the equally weighted portfolio can be attributed to two factors - that it has a higher exposure to the market and that it has an alpha larger than 0 [8]. In our results, the value for the beta estimate is larger than 1 and the annual standard deviation of the equally-weighted portfolio is rather high compared to the other traditional weighting strategies. These results indicate that the equally weighted portfolio is more volatile than the market, which offers a probability of higher return but also posing more risk, which is in line with the conclusion by Plyakha, Uppal and Vilkov. In our case the alpha estimate is not significant at the 1% level of risk, which means that we can not comment on whether alpha had an influence on the better results.

5.2.3 Minimum Variance Weighting

That a portfolio built on the premise of having the least possible variance can outperform the market capitalization portfolio with a significant margin seems unnatural at first glance. However, it has been shown that over long periods of time stocks with lower volatility actually consistently outperform stocks with higher volatility in terms of return. This is referred to as the low-volatility anomaly and is examined by Baker, Bradley and Wurgler in [28]. They show that by investing a dollar in 1968 and seeing how it has developed from then until 2008 the minimum variance approach yields much better returns and a more stable growth than the market capitalization approach, regardless if risk is defined as volatility or beta.

Another benefit of the minimum variance portfolio and a reason that it has performed so well during our 15 year period compared to the market capitalization portfolio is due to the stability the minimum variance portfolio had during the years of the financial crisis. At times of great financial instability it should be very beneficial to choose stocks which exhibit little variance as they should be least affected by the crisis. We can see from Table 3 that while most weighting techniques give significantly smaller returns during the time of the crisis, the minimum variance portfolio is not affected nearly as much. During this period, it still has a higher average yearly return than the market capitalization portfolio has over the entire period.

However, there are some downsides to using a minimum variance weighting strategy. The most obvious downside is that it is very difficult to estimate a covariance matrix, and even when a covariance matrix has been estimated it is very difficult to determine if the estimated covariance matrix is actually descriptive of the stocks' relationship in the real world. The first of these issues is mostly solved by using the shrinking approach described in Section 3.3 when estimating the covariance matrix. However, whether this accurately portrays the real world is impossible to determine, since if we could determine if the covariance matrix was correct we would not need to estimate it to begin with. All we can say is that the covariance matrix should somewhat accurately describe the relationships between the stocks for the historical time period from which the matrix was estimated and hope that this will be descriptive of the relationship between the stocks for the next investment period.

Another downside of the minimum variance approach is that we need historical data and an optimizer that we trust to solve the numerical optimization problem. For a market capitalization portfolio or an equally weighted portfolio we only need data on the stocks from today to construct portfolio weights, but a covariance estimation needs historical data to be meaningful. This means that we can not include stocks which have missing data in the time period from which we construct our covariance matrix or stocks that have not been publicly listed for the entirety of the period. This means that stocks that are new on the market or stocks that have missing data will need to be excluded in the minimum variance portfolio. In this report we have restricted ourselves to the subset of stocks that we do have five years of historical price data on, meaning that the minimum variance portfolio can include all stocks in our subset. In a real scenario, however, the minimum variance portfolio would have to ignore stocks that had not been publicly listed for five years. This restriction means that great investment opportunities could be missed as some of the stocks which get listed are very attractive on the market from the get-go. It should be noted that this restriction is one of the reasons that the minimum variance is more stable, since the stocks that are recently listed have a tendency to vary more than stocks that have been listed for a longer period of time.

5.2.4 Risk-Adjusted Alpha Weighting

The strategy that has worked the best out of all investigated strategies is the risk-adjusted alpha strategy. The average annual returns are four times greater for the risk-adjusted alpha portfolio than the market capitalization portfolio. On top of that, the annual standard deviation of the risk-adjusted alpha portfolio is lower than for the market capitalization portfolio.

The idea behind the strategy is to invest in stocks with a high average return and a low standard deviation. This principle seems fantastic for constructing a portfolio. Even so, it has performed well above expectation. Agarwal, who introduced the risk-adjusted alpha weighting strategy, got results where the outperformance of the strategy in comparison to the market capitalization strategy were on the same level as in our results, which means that it has substantially outperformed the benchmark index in at least two scenarios [9].

Our results suggests that the risk-adjusted alpha weighting strategy is something worth trying for everyone who distrusts the optimality of the market capitalization portfolio. It has the highest average annual return of all strategies with more than twice as large annual returns as the second best strategy. It is barely affected by the financial crisis. On top of that, it has the second lowest standard deviation of all portfolios, being beaten only by the minimum variance portfolio by 0.05%. This results in the by far best Sharpe ratio. On top of that, the tracking error is not that large and the information ratio is, as a result, by far the best too.

The portfolio is built on an easy idea. Let the portfolio consist of those stocks that we are most certain will perform well in terms of return. It is determined by finding stocks with high average returns and low standard deviation. This is no guarantee that they actually will perform well in the future but in our case it has certainly been the case in general. Whether this strategy can perform as well as it has on our dataset is doubtful, but it is definitely a strategy worth exploring as the method outperforms all other methods by far.

5.2.5 Fundamental Weighting

The fundamentally weighted portfolio has outperformed the market capitalization portfolio both in terms of return and Sharpe ratio, but all other traditional methods have performed better than the fundamental portfolio. It has the second lowest return and the highest standard deviation of all the traditional methods, but the second lowest tracking error.

If we look at how the fundamental portfolio performed in the different time periods we can see that it outperformed the market capitalization portfolio in all time periods. On top of that, in the time period of 2011-2016 where the market has been performing well the fundamental portfolio has produced the second largest returns, being beaten only by the risk-adjusted alpha portfolio. When the market performed poorly in the time period of 2008-2010 the fundamental portfolio basically broke even.

The performance of the portfolio seems highly correlated with the performance of the market in general, just as for the market capitalization portfolio. This seems counter-intuitive as the point of weighting on fundamental criteria is not to have a portfolio based on what the market expects of the stocks. However, the fact that the stocks price is influenced by the company's fundamentals is not a surprise, and with this in mind it seems more intuitive that the two strategies end-results are similar. The difference our results suggest is that the fundamental portfolio yields a higher return but taking a greater risk. For someone who believes that the market capitalization portfolio is not to be trusted the fundamental portfolio seems a good alternative as it presents a way to get a portfolio of similar performance in different time periods built on a different premise.

5.3 Cluster Stocks On Historical Returns

5.3.1 Most Suitable Similarity Measure When Clustering On Historical Returns

In this thesis, four different similarity measures have been used in order to determine the similarity between stocks.

Clustering on correlation has been the absolute worst, and is heavily outperformed by all other similarity measures and traditional weighting methods in terms of average annual return. The only reason to choose correlation as similarity measure when clustering on historical returns is that the annual standard deviation is the lowest among all similarity measures. If a low standard deviation is preferable to higher returns, this similarity measure might be considered but it is still outperformed by the cosine measure, and in all other aspects there seems to be no reason to cluster stocks on correlation. This is not that surprising. In Section 3.3 we argued that the variance and covariance are very hard to estimate and we adjust our data to solve the problem. The Pearson correlation coefficient which is used to cluster here is a function of these measures, and as such it is not a huge surprise that they yield less than satisfactory results.

Clustering on cosine is the second worst and is outperformed by most of the traditional weighting strategies and all similarity measures except correlation in terms of average annual return. When we cluster using a cosine similarity measure we try to cluster stocks that move in the same general direction at the same time, but with no regard to how much they actually move in that direction. The most likely reason as to why the cosine clustering does not perform well is that when trying to understand which stocks have similar temporal behavior we can not only consider whether their prices increase or decrease into consideration, but should also consider the size of the price movements.

Trying to understand this, we can construct a simple three-stock example that showcases the weaknesses of this type of clustering. Consider three stocks A, B and C in three different time-steps. Stock A and B increase with 30% in the first time-step, 2% in the second time-step and then 35% in the third time-step. Stock C on the other hand increases with 1% in the first time-step, 55% in the second time-step and 4% in the third time-step. In this three-dimensional space they would all have the cosine angle 0 and be considered equal in the cosine similarity measure, while it is obvious that their price increases are not positively correlated.

However, the standard deviation of the portfolios constructed from clusters formed by cosine similarity is lower than for any other similarity measure. This suggests that portfolios constructed from clusters using the cosine similarity measure are more diversified than for the other similarity measures, which could be preferable, if not for the low returns.

For the risk-adjusted alpha strategy, though, the returns are high for all similarity measures. Using cosine similarity, the returns are slightly lower than for the squared Euclidean distance similarity and the city block distance, but the standard deviation and tracking error are significantly lower too. The result is that the risk-adjusted alpha weights used within clusters constructed from cosine similarity has the best Sharpe ratio and information ratio out of all cluster strategies.

The city block and squared Euclidean distance similarity measures outperform traditional weighting strategies in terms of average annual return in most of the cases. We can see that the squared Euclidean clustered portfolios outperform the city block portfolios in all cases in terms of average annual return. However, using these similarity measures the standard deviation of the portfolios become quite high in comparison to the traditional strategies.

The performance of our clustering methods in terms of return seems to be correlated with how many clusters there are. In Table 10 the mean and median for the number of clusters are presented for the different similarity measures.

Similarity Measure	**Mean**	**Median**
Squared Euclidean	9.87	8
City Block	13.98	9.5
Cosine	18.49	19
Correlation	18.45	20

Table 10: The means and medians of the number of clusters for the similarity measures.

In Table 10 we can see that the squared Euclidean similarity measure has both the smallest mean and median. Moreover, the squared Euclidean and city block measures have a much smaller median and mean than the correlation and cosine measures which they heavily outperform in terms of average annual return.

That a smaller number of clusters gives better results strengthens the idea that a span of historical data, in our case a span of 5 years, does not fully portray the similarities between two stocks. The fewer the number of clusters are, the more we focus on the general traits of a stock rather than its specifics. There is much randomness in a stock's price, so much that many people argue that it can be described by a random walk [29]. When our examination of the specifics of a stock gets more detailed, the risk that we try to categorize them into clusters from effects caused by random noise gets larger. With this in mind it seems more reasonable to use a smaller number of clusters so that we try to find the more general similarities of the stocks when clustering them.

5.3.2 Most Suitable Weighting Strategy for the Constituents When Clustering On Historical Returns

In terms of average annual return, the risk-adjusted alpha strategy is the one that has the highest annual return when clustering on historical return data. Compared to using the risk-adjusted alpha strategy in the traditional framework, however, we see a decrease in annual return. In the traditional framework, the Sharpe ratio, information ratio and standard deviation is better than in the cluster framework too. We conclude that the risk-adjusted alpha strategy is better when clustering is not used, but out of all weighting strategies, the risk-adjusted alpha strategy performs best both in the traditional framework and the cluster framework. The only other strategy which could interest an investor is the minimum variance weighting which has a much lower standard deviation and still significantly higher returns than the market capitalization benchmark. For a very risk averse investor, this could be preferable.

It could be argued that the risk-adjusted alpha strategy is the most suitable to use in a cluster framework where we cluster on the historical returns since this strategy outperforms the other strategies in terms of most measures. On the other hand, the risk-adjusted alpha strategy performs better in the traditional framework than in the cluster framework. It could be more interesting to examine which strategy improved the most when going from the traditional framework into the cluster framework.

When clustering on historical returns, the clusters formed using the squared Euclidean similarity measure outperforms the traditional weighting strategy in all cases except for the risk-adjusted alpha weighting strategy. But this comes at a price of a lower information ratio, since the tracking error is increased in all cases. In most cases the standard deviation has increased in the cluster framework, but the Sharpe ratio remains basically the same for the market capitalization portfolio, the equal portfolio and the fundamental portfolio. This is since the average annual returns has had a similar relative increase. It seems that for these strategies, we increase the risk which in turn increases our returns. The only strategy which has seen a significant improvement for the Sharpe ratio is the minimum variance portfolio. The standard deviation has increased slightly in the cluster framework, but the returns are significantly higher than in the traditional framework. Therefore, when working in a cluster framework, we suggest using a minimum variance weighting strategy if the clusters are equally weighted.

5.4 Clustering Stocks On Fundamental Data

When we cluster on the fundamental data defined in Equation (12), we can see that the suitability of the similarity measures are basically inverted compared to when we clustered on the historical returns. The correlation similarity measure has given the highest average annual return for four of the five strategies, while producing the lowest standard deviation for all five. We conclude that when we compare two stocks, the correlation of their revenue per asset and net worth per asset is more descriptive of the relationship between the stocks than the correlation of their returns.

The squared Euclidean distance similarity measure which generally performed the best when clustering on historical returns has performed very poorly. It yields the smallest return for most of the weighting techniques and the standard deviation is not notably smaller than for other similarity measures, even being the highest for the risk-adjusted alpha weighting.

However, even if we only examine the similarity measure that has yielded the highest return for each weighting method in the fundamentally clustered portfolio, it is outperformed both by its traditional counterpart and the best performing similarity measure when clustering on historical returns. This is most likely due to that almost all the fundamentally clustered portfolios' poor performance during the crisis period. There are 20 constructed portfolios and 12 of them lose money during the financial crisis period, and another four of the portfolios make a very small gain of less than 2% over the 3-year period. The only portfolios which yield sufficient returns in the period is when we weight the constituents of each cluster with a risk-adjusted alpha weighting technique which we know has worked phenomenally both in a traditional framework and when clustering on historical returns.

The underperformance of the fundamentally clustered portfolios is made clear if we compare them to the clusters formed by historical return data. As noted, the fundamental clustering had 12 portfolios that lost money and 16 portfolios that made a loss or returned less than 2% over the 3-year crisis period, while the portfolios based on historical returns clustering only had 4 portfolios that made a loss and 8 portfolios that made a loss or earned less than 2% during the same period. On top of that, comparing only the portfolio based on the best performing similarity measure for each weighting technique to its traditional counterpart during the crisis period, the traditional counterpart performs better for all weighting techniques except for the risk-adjusted alpha strategy. But even for this strategy, we get a higher return with a lower standard deviation in the traditional framework during the entire period of 15 years.

As discussed earlier, yielding somewhat stable returns during periods of great financial stress is one of the most notable traits of a portfolio that performs well. We conclude that clustering on a stock's revenue per asset and net worth per asset is inferior in comparison to other methods in general and that it is better to use traditional weighting strategies or to cluster using historical return data.

5.5 GARCH-Estimation on Clusters

The method of estimating the covariance between clusters where the constituents are equally weighted using GARCH and then applying minimum variance weighting of the clusters performed well for the squared Euclidean and city block similarity measures. Using cosine or correlation as similarity measures produced poor results. Since the constituents are equally weighted in the GARCH-method, the cluster method which is most similar to it is the strategy with equally weighted clusters with equally weighted constituents.

We can see that for the cosine similarity and correlation similarity the equally weighted portfolio has performed better in terms of return. However, these similarity measures have produced poor results for both portfolios. For the squared Euclidean similarity and the city block similarity, which are of more interest since they have produced better portfolios, the minimum variance weighted clusters outperform the equally weighted clusters. The standard deviation has, as expected, been reduced for all four similarity measures increasing the Sharpe ratio significantly when clustering using a squared Euclidean or city block similarity measure.

In almost all cases, the portfolio created using minimum variance weights constructed from a GARCH-estimated covariance matrix results in a lower standard deviation than all portfolios where the clusters are equally weighted. If we disregard the cosine and correlation similarity measures due to their poor results for GARCH we see that the city block similarity measure has a lower standard deviation using GARCH-weighting of the clusters than any equally weighted cluster, while the squared Euclidean similarity measure has a lower standard deviation for all weighting techniques except when we weight the constituents of an equally weighted cluster using minimum variance optimization. That the city block similarity measure actually has a lower standard deviation when we use minimum variance on the clusters instead of within the cluster is telling of the strength of the GARCH-estimation of the covariance, since minimizing variance within the clusters take hundreds of stocks into account, while there is typically around 10 clusters (see Table 10).

Comparing the returns of the GARCH-weighted clusters for the squared Euclidean and city block similarity measures we see that it has a higher average annual return as well as a higher Sharpe ratio than when the clusters are equally weighted and the constituents are equally weighted, fundamentally weighted or market capitalization weighted. It is outperformed in terms of return and Sharpe ratio when the constituents of equally weighted clusters are weighted using minimum variance optimization or a risk-adjusted alpha strategy.

Comparing the different time periods we see that when we clustered using the city block similarity measure we performed much worse than the clusters based on a squared Euclidean similarity measure. On top of that, the performance of the portfolios which had been formed using the squared Euclidean distance measure were stable during the financial crisis for almost all portfolios. This suggests that the squared Euclidean similarity measure is the one we prefer when we equally weight the constituents of minimum variance optimized clusters where the covariance matrices of the clusters have been estimated using GARCH.

The GARCH method has created a portfolio with relatively high returns and low variance. However, the computational power needed for the GARCH method exceeds the computational power needed for any other method by far. First each stock has to be assigned to a cluster, and then the covariance matrices between the clusters have to be estimated each month. Both of these things require many calculations. Since other methods create similar results using much fewer calculations, we do not think that the GARCH method is necessarily the best to use on clusters. For example, using minimum variance weights on equally weighted clusters produces higher returns with around the same standard deviation and a better Sharpe ratio, and it takes much less time to calculate the weights this way than with GARCH-estimation of the covariance matrix.

5.6 Comparison Between Clustering and Traditional Weighting

The idea behind clustering is that since stocks that are similar will be in the same cluster, while stocks that do not exhibit similar traits or opposite traits are placed into different clusters, our portfolio weighted in a cluster framework will automatically be diversified if we weight the clusters appropriately. Here, all clusters have been weighted equally in almost all cases with the exception of when we use GARCH to estimate the covariance matrix of the clusters and seek to weight the clusters according to minimum variance optimization. All our clustered portfolios should be diversified by construction.

The difference between the traditional framework and the cluster framework is that in the traditional framework we weight the stocks according to some measure relative to all other stocks in the portfolio, but in the cluster framework we only weight the stocks according to that same measure relative to the stocks in the same cluster. The effect is that stocks that were assigned a relatively small weight in the traditional framework could get a rather large weight in the cluster framework, and vice versa.

Think of a simplified example with 10 stocks where 6 of them show similar temporal behavior and are better than the other 4 according to some measure. In the traditional framework these 6 would get the six largest weights. In the cluster framework, however, these 6 would get clustered together into one cluster C_1 and the other 4 would get clustered together as C_2 and each of these clusters would be assigned half of the capital. Then the 6 stocks in C_1 would only be compared to each other, and the stocks in C_2 would only be compared to each other. The stock in C_2 which is best according to our measure might be relatively better to the other stocks in C_2 than the best performing stock in C_1 in comparison to the other stocks in C_1. Then the largest weight in C_2 would be larger than the largest weight in C_1. This would mean that a stock that had the 7th largest weight in the traditional framework suddenly has the largest weight in the cluster framework. Initially, this might seem like something undesired. However, think of the scenario where the 6 stocks in C_1 all show similar temporal behavior due to some factor which has benefited all these 6 stocks historically but might not benefit them in the future. For example the stocks could benefit from warmer weather and the weather has been exceptional for the examined period of time. Suddenly, the weather gets much worse and all these 6 stocks fall in value since they are all affected by it, while the 4 stocks in C_2 are unaffected by this change. The benefit of clustering and how the cluster framework automatically diversifies is then apparent.

By comparing the results in Section 4.1 to the results in Section 4.2 we can get a picture of whether clustering has given a performance boost to our portfolio or not. We can see that all traditional weighting methods applied to clusters with a squared Euclidean similarity measure with the exception of risk-adjusted alpha outperform their traditional counterpart in terms of return. However, the risk seems to increase when we cluster too. This is most likely due to us not taking much consideration to how many stocks are in each cluster or why the stocks belong to a certain cluster. In the cluster framework, it is more likely that a stock with undesirable historical behavior receives a large weight as it might be assigned to a cluster with few constituents. This is solved with the minimum variance approach, both when applying the minimum variance on the constituents or on the clusters directly when the covariance matrix of the clusters is estimated using GARCH.

A comparison between Table 3 and Table 5 shows that all traditional strategies applied in a cluster framework perform better in terms of return than in the traditional framework during the financial crisis, but worse in the period after the crisis, i.e. between 2011-2016. Additionally, all the traditional strategies applied to clusters with the exception of risk-adjusted alpha has outperformed their traditional counterpart in the years before the financial crisis. The cluster weighting strategies seem to perform better in terms of return before and during the financial crisis than in the years after it. This could be because in the time frame 2002-2007, only a very small part of the historical data is from an unstable market. The last market crash before 2008 was in 1999 when the IT bubble popped, but data from this period is only relevant for the portfolios constructed in 2002 and 2003. The rest of the portfolio weights constructed in the pre-crisis period are formed from more stable data. When all data the clustering is based on is stable, the clustering seems to perform the best. The reason the cluster framework outperforms the traditional framework during the crisis period of 2008-2010 could be because clustering causes a diversification. It is hard to predict which stocks from which groups perform well in a crisis period, but with a clustered portfolio we should have many stocks with different traits which could be beneficial at stressful times. After the crisis period the clustering performs worse. This could be since the data used in the K-means algorithm is constructed by five years of historical returns. The clusters formed by the algorithm during 2011-2014 might be ill-formed since it might be difficult to assign stocks into groups in a time period where the data span into the financial crisis years and therefore is highly dependent on the stocks performance during these years. Consider clustering of stocks in the year 2011. At that point, historical data from 2007-2011 is used to form the clusters, and most of the data that the clustering is based on are data from the years of the financial crisis. Since then, the market has shifted into a new era where the data of the financial crisis might not be very descriptive of the market today, since the market behaves very differently under times of stress.

The clustering of stocks according to stock data has an advantage over grouping stocks into industry groups, as is most common today. The idea of industry groups is that most of the companies within one industry group tend to move as a whole on the market and by knowing the trends in place within an industry group, investors can better understand the investment potential of the companies within that group. This is an important concept for investors, but it could be argued that they should take clustering into account when forming these groups. By using clusters, the investor can be certain that the stocks have actually shown similar temporal behavior and are not only related by the nature of the products they produce or sell. However, the disadvantages of clustering is that it can not be categorized as easily as industry groups. The reason why certain stocks are clustered together can not be explained by more than the fact that they have behaved similarly in the past. On top of that, the clusters need to be updated on a regular basis. Stocks that belong to a certain cluster a certain month are not necessarily in the same cluster the next month. This creates instability, and it could even vary for different initializations of the used clustering algorithm. This means that different investors could get different clusters, while industry groups are constant both in time and for different investors. What we suggest is that instead of using clustering as a replacement for industry groups, they can be used together. An investor can find stocks that are both in the same cluster and the same industry group, and then use many such pairs from different clusters and industry groups to create a portfolio which is definitely diversified according to the historical data and the nature of the stocks.

6 Conclusion

We conclude in this thesis that the market capitalization weighting strategy is not optimal. Out of all strategies implemented in the traditional framework, the market capitalization portfolio is the one that earns the least. On top of that, it is outperformed by all traditional strategies during the years of the financial crisis. The results show that other strategies produce higher returns, better Sharpe ratio and lower standard deviation. The risk-adjusted alpha strategy in the traditional framework performed the best out of all investigated strategies.

To cluster stocks on data of historical returns with the squared Euclidean distance as similarity measure and then applying the traditional weighting strategies resulted in higher returns for all traditional weighting strategies except for risk-adjusted alpha. However, this comes at a price of an increased tracking error in all cases and an increased standard deviation in most cases. The Sharpe ratio is basically the same for most strategies, but has improved for the minimum variance strategy.

GARCH-estimation of covariance matrices between clusters can be used by an investor who is interested in creating a portfolio with very low risk. However, constructing the weights requires significantly more computational power than any other method in this thesis. There are other methods that produce results with similar risk and higher returns to that of the GARCH method, and on top of that the other methods require much less computational power.

6.1 Further research

We encourage readers of this thesis to make further investigations of the risk-weighted alpha strategy. Our results agree with Agarwal, the founder of the risk-adjusted alpha strategy [9], that the strategy outperforms the market capitalization strategy in terms of both return and standard deviation. Due to the level of outperformance, it is of great interest of those who do not believe the market capitalization strategy to be optimal to do further research regarding risk-adjusted alpha weighting. In both our study and in Agarwal's, the method was only tested on one index for quite a short period of time. This strategy should be tested on portfolios constructed from the constituents of different indexes in order to see if the method performs well regardless of the selected index. Additionally, more historical data should be used in order to investigate this strategy over a longer period of time to see if the strategy performs well when the market behaves in different ways.

When clustering the stocks together and performing the weighting strategies within the clusters, which are equally weighted, the returns are higher for the clustered data than for the non-clustered data for all strategies except the risk-adjusted alpha strategy. This indicates that it could be beneficial to explore clustering within portfolio selection further. Further research should focus on if there are different types of data to use as a basis for clustering than historical returns and the fundamental data used in this thesis. Whether there are more suitable clustering algorithms is another topic of interest.

As mentioned in Section 3.4.5, it would be interesting to compare if GARCH performs better when applied directly on stocks. Additionally, this would allow us to compare GARCH-estimation of the covariance matrix to the shrinking methodology used in this report to estimate the covariance matrix. We encourage interested readers with access to great computational power to investigate this further.

7 References

[1] Hsu JC. Cap-Weighted Portfolios are Sub-Optimal Portfolios. Journal of Investment Management. 2006;4(3).

[2] Perold AF. Fundamentally Flawed Indexing. Financial Analysts Journal. 2007;63(6).

[3] Chopra VK, Ziemba WT. The Effect of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice. The Journal of Portfolio Management. 1993;19(2).

[4] Kaufman L, Rousseeuw PJ. Finding Groups in Data - An Introduction to Cluster Analysis. John Wiley and Sons; 1990.

[5] Zhang J, Maringer D. Asset Allocation under Hierarchical Clustering. COMISEF WORKING PAPERS SERIES; 2010.

[6] Darbyshire P, Hampton D. Hedge Fund Modelling and Analysis Using MATLAB. Wiley Finance; 2014.

[7] Downes J, Goodman JE. Dictionary of Finance and Investment Terms. vol. 8. Barron's Educational Series Inc.; 2010.

[8] Plyakha Y, Uppal R, Vilkov G. Why Does an Equal-Weighted Portfolio Outperform Value- and Price-Weighted Portfolios. EDHEC Business School; 2012.

[9] Agarwal N. Risk weighted alpha index - Analysis of the ASX50 index. Journal of Economonic Financial Studies. 2013;1(1).

[10] Chow T, Hsu J, Kalesnik V, Little B. A Survey of Alternative Equity Index Strategies. Financial Analysts Journal. 2011;67(5).

[11] Hult H, Lindskog F, Hammarlid O, Rehn CJ. Risk and Portfolio Analysis Principles and Methods. Springer; 2010.

[12] James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning. 6th ed. Springer; 2013.

[13] Arthur D, Vassilvitskii S. k-means++:The Advantages of Careful Seeding. Stanford University; 2007.

[14] Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Computational and Applied Mathematics. 1986;.

[15] Marvin K. Creating Diversified Portfolios Using Cluster Analysis. Princeton University; 2015.

[16] Brockwell PJ, Davis RA. Introduction to Time Series and Forecasting. 3rd ed. Springer; 2016.

[17] Engle R. GARCH 101: The use of ARCH/GARCH models in applied econometrics. The Journal of Economic Perspectives. 2001;15(4):157–168.

[18] Karanasos M, Kim J. The Inflation-Output Variability Relationship in the G3: A Bivariate GARCH (BEKK) Approach. Brunel University UK; 2005.

[19] Arnott RD, Hsu J, Moore P. Fundamental Indexation. Financial Analysts Journal. 2005;61(2).

[20] Demey P, Maillard S, Roncalli T. Risk-Based Indexation. Lyxor Asset Management; 2010.

[21] Momeni M, Mohseni M, Soofi M. CLUSTERING STOCK MARKET COMPANIES VIA K-MEANS ALGORITHM. Kuwait Chapter of Arabian Journal of Business and Management Review. 2015;4(5).

[22] Fodor A, Jorgensen RD, Stowe JD. Forming Stock Groups with a Cluster Analysis of Common Size Statements; 2015.

[23] Bacon C. Practical Portfolio Performance Measurement and Attribution. 2nd ed. Wiley Finance; 2008.

[24] Ledoit O, Wolf M. Honey I Shrunk the Sample Covariance Matrix - the actual script;. [Online; accessed 3-March-2017]. Available from: https://github.com/brian-lau/highdim/blob/master/%2Butils/covCor.m.

[25] Ledoit O, Wolf M. Honey, I Shrunk the Sample Covariance Matrix; 2003.

[26] Sheppard K. Kevin Sheppard's MFE Toolbox for MATLAB;. [Online; accessed 25-April-2017]. Available from: https://www.kevinsheppard.com/MFE_Toolbox.

[27] Crain MA. A literature review of the size effect. University of Manchester; 2011.

[28] Baker M, Bradley B, Wurgler J. Benchmarks as Limits to Arbitrage: Understanding the Low Volatility Anomaly. Harvard Business School and NYU Stern School of Business; 2010.

[29] Fama EF. Random Walks in Stock-Market Prices. University of Chicago; 1965.

A Appendix

In Table 11 the Jensen's alpha and beta for the K-means clustering weighting strategies with returns as similarity are presented as well as as significance measure of these values obtained reported as p-value.

Weighting Technique	**Alpha**	**Beta**	$p < 0.01$? **Alpha**	$p < 0.01$? **Beta**
Equal				
Squared Euclidean	0.02	1.21	No	Yes
City block	0.01	1.26	No	Yes
Cosine	-0.002	0.88	No	Yes
Correlation	-0.01	0.87	No	Yes
Fundamental				
Squared Euclidean	0.03	1.25	No	Yes
City block	0.02	1.21	No	Yes
Cosine	-0.004	0.88	No	Yes
Correlation	0.02	0.90	No	Yes
Market Capitalization				
Squared Euclidean	0.01	1.11	No	Yes
City block	-0.01	1.21	No	Yes
Cosine	-0.02	0.83	No	Yes
Correlation	-0.03	0.80	No	Yes
Minimum Variance				
Squared Euclidean	0.07	0.81	Yes	Yes
City Block	0.05	0.99	No	Yes
Cosine	0.05	0.58	Yes	Yes
Correlation	0.05	0.60	Yes	Yes
Risk-Adjusted Alpha				
Squared Euclidean	0.13	1.11	Yes	Yes
City Block	0.13	1.14	Yes	Yes
Cosine	0.12	0.80	Yes	Yes
Correlation	0.11	0.77	Yes	Yes

Table 11: Risk measures: K-means clustering based on data of historical returns.

In Table 11 it can be seen that none of the estimated alpha values are significant at the 1% level of risk. This means that the null hypothesis is not rejected for these parameter values and hence the obtained values of alpha can not be used at this level of risk.

In Table 12 the Jensen's alpha and beta for the K-means clustering weighting strategies using fundamental data are presented as well as a significance measure of these values obtained reported as p-value.

Weighting Technique	**Alpha**	**Beta**	$p < 0.01$? **Alpha**	$p < 0.01$? **Beta**
Equal				
Squared Euclidean	-0.01	1.28	No	Yes
City Block	-0.002	1.31	No	Yes
Cosine	-0.002	1.39	No	Yes
Correlation	0.02	1.08	No	Yes
Fundamental				
Squared Euclidean	-0.01	1.25	No	Yes
City Block	0.001	1.35	No	Yes
Cosine	-0.01	1.52	No	Yes
Correlation	0.01	1.10	No	Yes
Market Capitalization				
Squared Euclidean	-0.03	1.11	No	Yes
City Block	-0.02	1.07	No	Yes
Cosine	-0.02	1.28	No	Yes
Correlation	-0.004	0.95	No	Yes
Minimum Variance				
Squared Euclidean	-0.03	1.19	No	Yes
City Block	-0.008	1.29	No	Yes
Cosine	0.002	1.43	No	Yes
Correlation	0.02	1.11	No	Yes
Risk-Adjusted Alpha				
Squared Euclidean	0.11	1.17	Yes	Yes
City Block	0.09	1.07	Yes	Yes
Cosine	0.07	1.08	Yes	Yes
Correlation	0.11	0.88	Yes	Yes

Table 12: Risk measures: K-means clustering based on fundamental data

In Table 12 it can be seen that none of the estimated alpha values are significant at the 1% level of risk. This means that the null hypothesis is not rejected for these parameter values and hence the obtained values of alpha can not be used at this level of risk.

Weighting Technique	**Alpha**	**Beta**	$p < 0.01$? **Alpha**	$p < 0.01$? **Beta**
Squared Euclidean	0.04	0.89	No	Yes
City Block	0.04	0.89	No	Yes
Cosine	-0.02	0.56	No	Yes
Correlation	-0.009	0.60	No	Yes

Table 13: Risk measures for the GARCH approach.

TRITA -MAT-E 2017:37
ISRN -KTH/MAT/E--17/37--SE