Statistical Modelling of Equal Risk Portfolio Optimization with Emphasis on Projection Methods

Master Thesis
June 2015

Sabrina Neumann

Aalborg University
Department of Mathematical Sciences
In collaboration with Department of Business & Management
Autumn 2014/Spring 2015

Department of Mathematical Sciences
Fredrik Bajers Vej 7G
9220 Aalborg Ø
Telephone 99 40 99 40
Fax 9940 3548
http://www.math.aau.dk/

Title:
Statistical Modelling of Equal Risk Portfolio Optimization with Emphasis on Projection Methods

Project period:
September 2, 2014 -
June 10, 2015

Author:
Sabrina Neumann

Supervisors:
Torben Tvedebrink
Lasse Bork

Circulation:
7

Number of pages:
118

Completed:
June 10, 2015

Synopsis:

The objective of this thesis is to investigate the Equal Risk asset allocation strategy that makes use of latent risk factors in a portfolio. This strategy can be considered from different points of view, but in general it aims to equalize the risk contributions of the assets in a portfolio. There is placed much emphasis on projection methods such as principal component analysis and its functional variant, functional principal component analysis, which are used to extract the latent risk factors. For the purpose of investigating the performance of the Equal Risk strategy compared to other allocation strategies, e.g. the Equally-Weighted or Minimum Variance strategy, there is considered a walk-forward backtest with a rolling estimation window on historical data. Data is kindly provided by *Jyske Bank A/S* and includes different types of assets.

Preface

This master thesis is written by Sabrina Neumann during the period from September 2, 2014 to June 10, 2015. The thesis is composed at the Department of Mathematical Sciences at Aalborg University in collaboration with the Department of Business and Management at Aalborg University, and with *Jyske Bank A/S* in Silkeborg.

The thesis deals with investigating the Equal Risk allocation strategy that is based on latent risk factors, where the emphasis is on projection methods. In order to backtest the different allocation strategies, *Jyske Bank A/S* kindly provides data that includes different types of assets and some R code, which are treated confidentially.

Bibliographical references in the thesis are done with [author,year], and these can be found in the bibliography at the end of the thesis before the appendix. References to definitions, theorems, etc. which start with 'A' refer to the appendix. All figures and code are implemented with the software R.

It is expected that the reader has basic knowledge equivalent to a bachelor degree in Mathematics and Statistics from Aalborg University. In addtion, basic knowledge within measure theory is an advantage.

I would like to thank my supervisors, Assistant Professor Torben Tvedebrink, Department of Mathematical Sciences at Aalborg Univeristy, and Assistant Professor Lasse Bork, Department of Business and Management at Aalborg University, for for their enthusiasm and great help during the thesis. I would also like to thank *Jyske Bank A/S* for providing data, R code, and inside knowledge, and in particular Anders Hartelius Haaning for an introduction to the topic, and his great help and commitment.

Sabrina Neumann
Aalborg, June 2015

Resumé

Formålet med denne specialeafhandling er at betragte aktivallokerings strategier, der er baseret på risikofaktorer i stedet for traditionelle strategier, der allokerer ved hjælp af aktivklasser. Motivationen for at betragte risiko-baserede strategier ligger i problemstillingen at strategier der udelukkende er baseret på estimater af middelværdi og varians, som eksempelvis *Markowitz mean-variance* strategien, ikke kan fange den tunge venstre hale i afkastfordelingen i perioder af høj volatilitet, som for eksempel finanskrisen i 2008/2009. For at opnå en bedre forståelse af den økonomiske tankegang bag risiko-baserede modeller, gives der en introduktion til vigtige økonomiske faktor-baserede modeller, såsom *Capital Asset Pricing Model*, *Arbitrage Pricing Theory* og Fama og Frenchs tre faktor model.

For at kunne finde de underliggende risikofaktorer af en portefølje betragtes *principal component analysis*, som er en statistisk metode til at transformere det oprindelig aktivrum i et lavere dimensionalt rum, der netop udgør risikofaktorerne. Denne metode betragtes også fra en funktionel tilgang: *functional principal component analysis*. Her arbejdes med de samme principper som i *principal component analysis*, men det antages, at data har en underliggende funktionel from. Dermed arbejdes der med egenfunktioner, som observeres over tid i stedet for egenvektorer, der bare giver statiske og ikke-temporære estimater. For at kunne arbejde med *functional principal component analysis* skal data transformeres til funktioner, hvilket gøres ved hjælp af B-splines. Samtidig udglattes data ved at bruge en *penalized weighted least squares* metode, hvilket øger *signal-to-noise* ratioen af data. Motivationen for at betragte den funktionelle tilgang ligger i antagelsen om, at nøjagtigheden af allokerings strategierne kan øges. Derudover er det teoretisk set muligt at indkludere aktiver med forskellige samplingfrekvens i en portefølje, hvilket ikke er muligt i en multivariat tilgang, hvor data betragtes som enkelte observationer.

Der fokuseres specielt på *Equal Risk* strategien, der kan betragtes med forskellige tilgange. Enten kan den betragtes som en optimering der har til formål at vælge porteføljevægtene sådan, at risiko bidragene fra hvert enkelt aktiv er lige store. Eller den kan betragtes som en optimering, hvor vægtene af de enkelte aktiver i en portefølje vælges sådan, at en *principal component analysis* på de historiske afkastserier giver så ens som mulig standardafvigelse i principal komponentretningerne. *Equal Risk* strategien sammenlignes med fast allokering, *Equally-Weighted*, og den traditionelle *Minimum Variance* strategi.

For at kunne sammenligne præstationen af de forskellige strategier betragtes en backtest med rullende estimationsvindue på historiske afkastserier. Der undersøges forskellige estimationsvinduer, og det antages, at *short-selling* ikke er tilladt. Backtesten baseres på tre forskellige portføljer, der indeholder flere aktivklasser samt et forskelligt antal af aktiver. Det observeres, at både længden af estimationsvinduet og sammensætningen af porteføljen har en indflydelse på den profit, der kan opnås ved de forskellige allokerings strategier. Der er lagt meget vægt på at undersøge effekten af at udelade nogle principal komponenter, det vil sige at betragte et mindre antal underliggende risikofaktorer. Der findes ikke et entydig resultat for om denne effekt har en positiv eller negativ indflydelse på allokerings strategierne, men det har været muligt at øge profitten i en af porteføljerne ved at betragte et mindre antal principal komponenter. Desuden kan det konkluderes udfra backtesten, at *Equally-Weighted* strategien det meste af tiden har en højere profit end de andre testede strategier, men at den ikke performer godt igennem finanskrisen. Der er det netop nogle af de betragtede risiko-baserede strategier, der performer stabilt, hvilket motiverer brugen af disse strategier. Den funktionelle tilgang kan ikke outperforme de andre risiko-baserede strategier igennem finanskrisen, men har i nogle portefølje sammensætninger en meget aktraktiv afkast i normale tider. Desuden har de risikobaserede strategier en mindre volatil profit end *Equally-Weighted* strategien, hvilket er en vigtig egenskab for de fleste investorer.

Contents

Introduction 1

This thesis deals with studying asset allocation strategies that are based on *risk factors* instead of traditional strategies, that allocate by asset classes. This is motivated by the fact that strategies which are entirely based on the estimates of the mean and variance of assets, e.g. the Markowitz mean-variance strategy, cannot capture the heavy left tail in the distribution of a return that occurs in times of high volatility, e.g. during the recent financial crisis in 2008/2009.

The aim is to establish asset allocation strategies that are based on independent, underlying drivers of assets. These drivers are also called risk factors, which can be classified into observable factors, e.g. growth or inflation, and latent factors that are not observable. The focus in this thesis is on strategies that are based on latent factors such that the risk of a portfolio is spread out to assets that do not all behave similar, especially during financial crises where many assets are likely to crash. The emphasis is on the Equal Risk portfolio optimization strategy that can be considered from different points of view, which will be investigated in this thesis. For the purpose of comparing the performance of the risk-based strategies that will be introduced, these strategies will be compared to the traditional Minimum Variance strategy and the simple Equally-Weighted strategy.

The mathematical tool used to extract the underlying, lower dimensional risk factors is *principal component analysis*, which is a well-known tool in the field of statistics to find linearly uncorrelated, i.e. independent, variables in data. In addition, the functional variant, *functional principal component analysis*, is considered in order to explore possible improvements in the allocation strategies. Note that principal component analysis should not be confused with *factor analysis*, which assumes that there is an underlying model, since principal component analysis just is a dimensionality reduction method.

In the recent years there is conducted some research of risk-based allocation strategies. [S. Maillard, T. Roncalli, J. Teiletche, 2009] introduces the general idea of the Equal Risk strategy and shows relations between the Equal Risk strategy to the traditional Minimum Variance strategy and the simple Equally-Weighted strategy. [Kind, 2013] focuses on the risk-based strategies using principal component analysis, and [Meucci, 2010] introduces a related strategy: Diversified Risk Parity, and extensions are presented in [A. Meucci, A. Santangelo, R. Deguest, 2014]. This thesis presents among others some of the most important results of these articles.

But first in order to achieve a better understanding of the economical terminology, concepts, and models the following sections introduce the basic economical theory for factor models, their motivations and critics, the concept of *diversification*, and *risk premia*.

1.1 Portfolio Theory

This section introduces general concepts within portfolio theory and is inspired by [Luenberger, 2009]. The *rate of return*, r, of an asset over a single period is defined as:

$$r = \frac{X(1) - X(0)}{X(0)},$$

where $X(1)$ is the amount received and $X(0)$ the amount invested. It is often assumed that prices are log-normally distributed, which implies that *log returns*, $\log(R)$, are given by:

$$\log(R) = \log(1 + r) = \log\left(\frac{X(1)}{X(0)}\right) = \log(X(1)) - \log(X(0)).$$

In order to form a portfolio, suppose that there are N different assets available in the market. Let $X_i(t)$ denote the price of asset i at time t, then the log returns, $\log(R_i(t))$, are computed as follows:

$$\log(R_i(t)) = \log(X_i(t)) - \log(X_i(t-1)), \quad \text{for } i = 1, \ldots, N \text{ and } t = 2, \ldots, T, \tag{1.1}$$

where $\log(R_i(1)) = 0$ and T is the number of observations in the sampling periode. Both academics and practitioners often assume that asset log returns follow a normal distribution, which is basis for many fields in finance. However, it can be observed that real financial data often is more heavy tailed than the normal distribution, e.g. under crashes return distributions have a heavy left tail. [Longin, 2005] This can also be observed in Figure 1.1, which shows histograms for the monthly log returns of two indexes describing equities and credits, respectively.

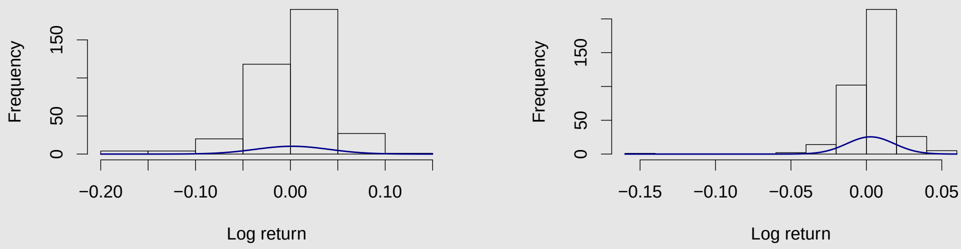


Figure 1.1. Histogram of equities (left panel) and credits (right panel) monthly log returns in the period from January 20, 1999 through September 29, 2014. The solid line indicates the normal distribution with the same mean and standard deviation as the indexes.

When forming a portfolio of N assets, the investor has to select the amounts to be invested in each asset, which can be expressed by *portfolio weights* $\boldsymbol{w} = (w_1, \ldots, w_N)^\top$. These are constrained by:

$$\mathbf{1}^\top \boldsymbol{w} = 1, \tag{1.2}$$

where some of the weights can be negative if short-selling is allowed. Then the overall *rate of expected return* of a portfolio, $r_P(\boldsymbol{w})$, is given by the weighted returns:

$$r_P(\boldsymbol{w}) = \boldsymbol{w}^\top \boldsymbol{r}.$$

Asset prices are not static, but flucate over time. They change due to demand and arrival of new information, i.e. political or economical news. Therefore it is convenient to model asset prices as a stochastic process. In general it is assumed that the price of an asset follows a first order Markov process, which means that the future price only depends on the present price.

Depending on the investment strategy, an investor could want to reblance the portfolio weights periodically in order to maintain the desired asset allocation. This means to buy or sell assets until the disired level of allocation is reached. For instance, consider originally holding 50% stocks and 50% bonds in a portfolio and the investor wants to maintain this strategy. Imagine that stocks perform pretty good during a period such that the weight of stocks increases to 80%. In order to reweight to the original allocation the investor could sell some stocks and buy more bonds until the weights again are 50/50. The time interval of rebalancing is varying, e.g. every month.

Given some weights $\boldsymbol{w}(t) = (w_1(t), \ldots, w_N(t))^\top$ of the assets at time t, where it is assumed that short-selling is not allowed and that the portfolio weights sum to one, the value process, also called *profit*, $P(t)$, of a portfolio consisting of N assets is given by:

$$P(t) = \boldsymbol{w}(t)^\top \boldsymbol{X}(t),$$

where $\boldsymbol{X}(t)$ is a vector of asset prices at time t.

The processes $(w_1(t))_{t\geq 0}\ldots,(w_N(t))_{t\geq 0}$ are adapted, i.e. the weights are chosen according to the information available at time t. The process $(w_1(t)\ldots,w_N(t))_{t\geq 0}$ is called the *portfolio strategy.* [Lesniewski, 2008] The *continuously compounded return profit* using log returns is then computed by:

$$P(t+1) = P(t)\cdot \exp\left(\boldsymbol{w}(t)^\top \log(\boldsymbol{R}(t))\right),$$

where $\log(\boldsymbol{R}(t))$ is a vector of log returns at time t as defined in equation (1.1). [Zivot, 2015] From this the *annulized geometrical mean profit*, MP, is given by:

$$\text{MP} = \left(\frac{P(T)}{P(1)}\right)^{\frac{p}{T}} - 1, \tag{1.3}$$

where p indicates the number of periods per year, e.g. when dealing with monthly observed data $p = 12$. And the *annulized volatility*, Vol, of the profit is given by:

$$\text{Vol} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(P(t) - \frac{1}{T}\sum_{i=1}^{T}P(t)\right)^2}\cdot\sqrt{p}. \tag{1.4}$$

The next section introduces the general definitions of portfolio mean and variance, some basic asset allocation strategies, and other measures of risk for a portfolio.

1.1.1 Portfolio Mean and Variance, and other Measures of Risk

With the purpose of measuring the performance of an asset allocation strategy the portfolio mean and variance often are considered. Portfolio optimization often is based on maximizing the *expected rate of return* of a portfolio, $\mu_P(\boldsymbol{w})$, which is given by:

$$\mu_P(\boldsymbol{w}) = \boldsymbol{w}^\top\boldsymbol{\mu},$$

where $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_N)^\top$ is a vector of expected values of the N assets of a portfolio. Correspondingly, the *variance of the rate of return* of a portfolio, $\sigma_P^2(\boldsymbol{w})$, can be determined by:

$$\begin{aligned}\sigma_P^2(\boldsymbol{w}) &= \mathrm{E}\left[(r_P(\boldsymbol{w}) - \mu_P(\boldsymbol{w}))^2\right]\\ &= \mathrm{E}\left[\boldsymbol{w}^\top(\boldsymbol{r}-\boldsymbol{\mu})(\boldsymbol{r}-\boldsymbol{\mu})^\top\boldsymbol{w}\right]\\ &= \boldsymbol{w}^\top\Sigma\boldsymbol{w},\end{aligned} \tag{1.5}$$

where Σ is the covariance matrix. The standard deviation of a portfolio's rate of return, $\sigma_P(\boldsymbol{w})$, is usually used as a basic measure of risk:

$$\sigma_P(\boldsymbol{w}) = \sqrt{\boldsymbol{w}^\top\Sigma\boldsymbol{w}}. \tag{1.6}$$

A higher standard deviation is associated with higher risk and higher risk is required to get a higher expected return. Thereby each investor is faced by the tradeoff of expected return and risk. Note that in financial context the standard deviation is often called volatility and the two terms will be used interchangeably in this thesis.

The traditional Markowitz Mean-Variance, MV, asset allocation strategy aims to find a portfolio, where the expected rate of return of the portfolio is fixed at some abitrary expected value μ while minimizing the portfolio variance:

$$\begin{aligned} \min \quad & \boldsymbol{w}^\top \Sigma \boldsymbol{w} \\ \text{subject to } & \boldsymbol{w}^\top \boldsymbol{\mu} = \mu \text{ and } \mathbf{1}^\top \boldsymbol{w} = 1. \end{aligned} \tag{1.7}$$

This strategy is based on the assumption that asset returns are normal distributed, a strong assumption that as mentioned above often is violated for real data. Moreover, this allocation strategy assumes stable correlations of assets over time, which again is not met by most assets. Although these assumptions are likely to be violated, the MV strategy is widely used by both academics and the financial industry because of the ease of computations.

However, instead of focusing on the expected rate of return of a portfolio, which often is hard to estimate, other portfolio optimization methods consider the *marginal risk contribution*, $\partial_{w_i}\sigma_P(\boldsymbol{w})$, of the ith asset. It is defined by:

$$\begin{aligned} \partial_{w_i}\sigma_P(\boldsymbol{w}) &= \frac{\partial \sigma_P(\boldsymbol{w})}{\partial w_i} = \frac{(\Sigma \boldsymbol{w})_i}{\sqrt{\boldsymbol{w}^\top \Sigma \boldsymbol{w}}} \\ &= \frac{w_i \sigma_i^2 + \sum_{i \neq j} w_j \sigma_{ij}}{\sigma_P(\boldsymbol{w})}. \end{aligned} \tag{1.8}$$

The marginal risk contributions explain the change in the volatility of a portfolio due to a small increase in the weight of one component. [S. Maillard, T. Roncalli, J. Teiletche, 2009]

One strategy that uses this concept is the Minimum Variance strategy, MVa, which is based on the assumption that the marginal risk contributions should be identical across the assets of a portfolio. [Kind, 2013] The MVa weights can be found by solving:

$$\begin{aligned} \min_{\boldsymbol{w}} \ & \boldsymbol{w}^\top \Sigma \boldsymbol{w} \qquad (1.9) \\ \text{subject to } & \mathbf{1}^\top \boldsymbol{w} = 1. \end{aligned}$$

Hence the MVa strategy is a special case of the MV strategy, when the requirement of the expected return is omitted.

But there are also strategies that assume that the *risk contributions* are equal, which will be introduced in Chapter 3. The risk contribution of the ith asset, $\mathrm{RC}_i(\boldsymbol{w})$, can be expressed as:

$$\mathrm{RC}_i(\boldsymbol{w}) = w_i \cdot \partial_{w_i}\sigma_P(\boldsymbol{w}) = w_i \frac{(\Sigma \boldsymbol{w})_i}{\sqrt{\boldsymbol{w}^\top \Sigma \boldsymbol{w}}}. \tag{1.10}$$

In order to show further properties of the risk contributions $\mathrm{RC}_i(\boldsymbol{w})$ the following definition introduces the concept of a *homogeneous function.*

Definition 1.1 (Homogeneous Function)
A function $f : X \subset \mathbb{R}^N \to \mathbb{R}$ is called a homogeneous function of degree τ if there exists a $\lambda > 0$ and $\boldsymbol{x} \in X$, where $\lambda\boldsymbol{x} \in X$ such that the following holds:

$$f(\lambda\boldsymbol{x}) = \lambda^\tau f(\boldsymbol{x}).$$

[Tasche, 2008]

Furthermore, Theorem 1.2 introduces the *Euler decomposition*, which can be used to show the relationship between the portfolio volatility $\sigma_P(\boldsymbol{w})$ and the risk contributions $\mathrm{RC}_i(\boldsymbol{w})$.

Theorem 1.2 (Euler's Theorem on Homogeneous Functions)
Let $X \subset \mathbb{R}^N$ be an open set and let $f : X \to R$ be a continuously differentiable function. Then f is a homogeneous function of degree τ if and only if it satisfies the following:

$$\tau f(\boldsymbol{x}) = \sum_{i=1}^{N} x_i \frac{\partial f(\boldsymbol{x})}{\partial x_i},$$

where $\boldsymbol{x} \in X$. [Tasche, 2008]

Proof. Omitted. □

It is known that the volatility $\sigma_P(\boldsymbol{w})$ is a homogeneous function of degree one, i.e. $\tau = 1$ in Definition 1.1. Thus it satisfies Euler's decompostion as given in Theorem 1.2:

$$\begin{aligned}
\sum_{i=1}^{N} \mathrm{RC}_i(\boldsymbol{w}) &= \sum_{i=1}^{N} w_i \frac{(\Sigma\boldsymbol{w})_i}{\sqrt{\boldsymbol{w}^\top\Sigma\boldsymbol{w}}} \\
&= \boldsymbol{w}^\top \frac{\Sigma\boldsymbol{w}}{\sqrt{\boldsymbol{w}^\top\Sigma\boldsymbol{w}}} \\
&= \sqrt{\boldsymbol{w}^\top\Sigma\boldsymbol{w}} = \sigma_P(\boldsymbol{w})
\end{aligned}$$

This shows that the volatility of a portfolio can be decomposed into risk contributions of the included assets. After having introduced these basic concepts of portfolio theory, the next section presents different economical theory concerning factor models.

1.2 Factor Models and Risk Factors

First, this section gives a short introduction to the *Capital Asset Pricing Model*, CAPM, which is a one factor model that quantifies the tradeoff between expected return and risk within the mean-variance framework. It was introduced independently by Treynor (1961), Sharpe (1964), Litner (1965), and Mossin (1966). [Ang, 2014] The model involves a linear relationship between the expected return of an asset with the covariance of its return and the return of the market portfolio. [J. Y. Campbell, A. W. Lo, A. C. MacKinlay, 1997]

Furthermore in Section 1.2.2, the one factor CAPM model is expanded to a multi-factor model, the *Arbitrage Pricing Model*, APT, developed by Ross (1976). [Ang, 2014] In this setup it is assumed that any risky asset can be considered as a linear combination of various risk factors that affect the asset return. [T. E. Copeland and J. F. Weston, 1988] There are different types of risk factors, like *macro factors*, *style factors*, and *firm-specific factors*, which will be introduced. Furthermore the Fama-French three factor model is shortly presented.

1.2.1 The Capital Asset Pricing Model

This section introduces the one factor model CAPM and is inspired by [Luenberger, 2009] and [M. Grinblatt and S. Titman, 2002]. In order to understand the CAPM, first the concept of the *market portfolio* is explained. The market portfolio M is a theoretical summation of all available assets of the world financial market, where each asset is weighted by its proportion in the market, i.e. its *market value*. Let v_i denote the market value of asset i, then the market weights, w_i^M, can be expressed by:

$$w_i^M = \frac{v_i}{\sum_{i=1}^N v_i}.$$

Thereby the expected return of the market portfolio reflects the expected return of the market as a whole and it is assumed that the portfolio has the lowest volatility among all portfolio that have the same expected return as the market. This is the same as saying that the market has the highest *Sharpe ratio*, which will be introduced later in this section.

Assuming that all investors use the Markowitz mean-variance framework to determine their portfolio weights, that everyone invests in all available assets in the market, and that there are no transactions costs, then the market portfolio M is said to be the *efficient portfolio* in the market. This means that investors will recalculate their estimates of the portfolio weights until demand matches supply, which drives the market to efficiency. Additionally, the CAPM assumes the existence of a *risk-free asset*, which represents the possibility of an investor to borrow or lend cash at the risk-free rate. Let r_f be the return of a risk-free asset, which means that the return is deterministic. When included in a portfolio, a positive weight corresponds to lending cash, whereas a negative weight means borrowing cash.

The following result states that risk-averse investors will invest in a portfolio consisting of a combination of two portfolios:

Result 1.3 (Two-fund Seperation)
Each investor holds an efficient portfolio which is a combination of the risk-free asset and a portfolio of risky assets, i.e. the market portfolio. [T. E. Copeland and J. F. Weston, 1988]

Now consider a *mean-standard deviation diagram* as shown in Figure 1.2, where each point represents an asset with its expected rate of return μ and standard deviation σ. If one plots the market portfolio M in this diagram, then the efficient set of portfolios consists of a straight line, called the *capital market line*, which starts in a risk-free point r_f and passes through the market portfolio M.

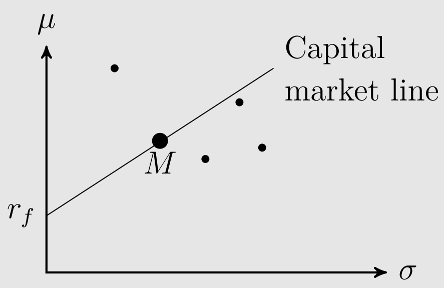


Figure 1.2. Mean-standard deviation diagram with capital market line drawn from a risk-free point r_f and passing through the market portfolio M. The points indicate assets with their expected rate of return μ and standard deviation σ. The figure is inspired by [Luenberger, 2009].

The capital market line describes the relation between the expected rate of return μ of an asset and the risk of return, measured by σ. Let μ_M and σ_M be the expected rate of return and standard deviation of the market rate of return described by the market portfolio, and $\mu_P(\boldsymbol{w})$ and $\sigma_P(\boldsymbol{w})$ be the expected rate of return and standard deviation of an arbitrary efficient portfolio or asset. Then the capital market line describes a portfolio consisting of one risk-free asset r_f and one efficient risky asset r_M such that the expected rate of return of the portfolio is given by:

$$\mu_P(\boldsymbol{w}) = r_f + \frac{\mu_M - r_f}{\sigma_M}\sigma_P(\boldsymbol{w}). \tag{1.11}$$

The numerator in the slope of the capital market line, $\mu_M - r_f$, is called *risk premium*, which is a kind of compensation for an investor that takes extra risk in holding a risky asset compared to a risk-free asset. The slope of the capital market line, $\mu_M - r_f/\sigma_M$, is often called the *price of risk*, since it tells how much the expected rate of return of a portfolio must increase if the standard deviation increases by one unit. It can be used to measure the efficiency of a portfolio by comparing the location of a portfolio in the mean-standard deviation diagram relative to the capital market line, since only portfolios that are on the line are efficient. This introduces the concept of the Sharpe ratio:

$$S = \frac{\mu_P(\boldsymbol{w}) - \mu_b(\boldsymbol{w})}{\sigma_P(\boldsymbol{w})},$$

where $\mu_b(\boldsymbol{w})$ denotes the benchmark expected rate of return, i.e. a reference portfolio which is used to compare the performance of a portfolio. A higher ratio provides better return for the same risk. The Sharpe ratio can be used to compare different investment strategies with each other. The following result states the CAPM.

Result 1.4 (Capital Asset Pricing Model)
Assume that the market portfolio M is efficient, then the expected rate of return μ_i of any asset satisfies the following:

$$\mu_i - r_f = \beta_i(\mu_M - r_f), \tag{1.12}$$

where:

$$\beta_i = \frac{\sigma_{iM}}{\sigma_M^2}. \tag{1.13}$$

[Luenberger, 2009]

Note that the CAPM in equation (1.12) is just a rewrite of equation (1.11). The CAPM describes the risk of an individual asset, measured by the factor exposure β of that asset to the market factor. This means that the higher the exposure of an asset to the market factor, the higher the expected return, which yields a positive risk premium. Consequently, β is exactly the estimate of the slope in a simple linear regression.

In general the β of an asset or portfolio measures the risk arising from general market movements, where the market portfolio has assigned $\beta = 1$. So a portfolio with $\beta > 1$ is predicted to have higher risk than the market portfolio, whereas $\beta < 1$ indicates lower risk. And an asset that is completely uncorrelated with the market has assigned $\beta = 0$. So the CAPM uses β instead of the standard deviation σ as measure of risk of an asset. This implies that an investor prefers assets with a negative β, since during market crashes it will act reversed to the market, and thereby yield higher returns than the market portfolio. The portfolio β, $\beta_P(\boldsymbol{w})$, is the weighted average of the βs of the single assets in the portfolio:

$$\beta_P(\boldsymbol{w}) = \boldsymbol{w}^\top \boldsymbol{\beta}.$$

Inspired by the CAPM in equation (1.12) the random rate of return of asset i can be written as:

$$r_i = r_f + \beta_i(r_M - r_f) + \epsilon_i, \tag{1.14}$$

where ϵ_i is the residual. From equation (1.12) it follows that $\mathrm{E}[\epsilon_i] = 0$. In addition, when taking the covariance of r_i as given in equation (1.14) with the market rate of return r_M, it follows that $\mathrm{Cov}[\epsilon_i, \sigma_M] = 0$. This implies that risk, measured by the variance of an asset, σ_i^2, can be decomposed into *systematic risk* and *specific risk*:

$$\begin{aligned} \sigma_i^2 &= \beta_i^2 \sigma_M^2 + \sigma_{\epsilon_i}^2 \qquad (1.15) \\ \text{Total risk} &= \text{systematic risk} + \text{specific risk}, \end{aligned}$$

where the systematic risk is the risk associated with the market as a whole.

Most investors want a high expected return and at the same time low risk. This introduces the concept of diversification, which is a method to reduce the variance, as given in equation (1.15), by including additional assets in a portfolio. The risk associated with the market, the systematic risk, cannot be reduced by diversification. Whereas specific risk can be diversified due to the uncorrelateness to the systematic risk and the law of large numbers, by including a large number of assets in a portfolio. The following simple example shows how the variance of a portfolio behaves when using the Equally-Weighted strategy and a large number of assets in a portfolio.

Example 1 (Equally weighted portfolio)
Consider an equally weighted portfolio, where $w_i = N^{-1}$. Then the variance of the rate of return of the portfolio can be rewritten as:

$$\begin{aligned} \sigma_P^2(\boldsymbol{w}) &= N^{-2} \sum_{i=1}^{N} \sigma_i^2 + N^{-2} \sum_{i,j}^{\neq} \sigma_{ij} \qquad (1.16) \\ &= N^{-2} N \bar{\sigma}_\bullet^2 + N^{-2} N(N-1) \bar{\sigma}_{\bullet\bullet} \\ &= N^{-1} \bar{\sigma}_\bullet^2 + \frac{N-1}{N} \bar{\sigma}_{\bullet\bullet} \\ &= N^{-1} \bar{\sigma}_\bullet^2 (1-\rho) + \rho \bar{\sigma}_\bullet. \end{aligned}$$

The notation $\sum_{i,j}^{\neq}$ is a short form for $\sum_{i=1}^{N} \sum_{j=1, i\neq j}^{N}$. For a large number of assets N it follows that:

$$\lim_{N\to\infty} \sigma_P^2(\boldsymbol{w}) = \bar{\sigma}_{\bullet\bullet}.$$

Thus, the portfolio variance assymptotically is the average of the covariances between the assets. Hence $\bar{\sigma}_{\bullet\bullet}$ denotes the systematic risk that cannot be diversified, whereas $N^{-1}\bar{\sigma}_\bullet^2(1-\rho)$ is the specific risk that can be diversified by including a large number of assets in a portfolio.
Note that the correlation coefficient ρ has a lower bound, since it has to be ensured

that the portfolio variance $\sigma_p(\boldsymbol{w})$ is positive semi-definite, i.e. $\sigma_P^2(\boldsymbol{w}) \geq 0$. Consider the expression given in equation (1.16) and rewrite it:

$$\begin{aligned}\sigma_P^2(\boldsymbol{w}) =& N^{-2}\left(\sum_{i=1}^{N}\sigma_i^2 + \rho\sum_{i,j}^{\neq}\sigma_i\sigma_j\right) \geq 0\\ \sum_{i=1}^{N}\sigma_i^2 &\geq -\rho\sum_{i,j}^{\neq}\sigma_i\sigma_j\\ \rho &\geq -\frac{\sum_{i=1}^{N}\sigma_i^2}{\sum_{i,j}^{\neq}\sigma_i\sigma_j}.\end{aligned}$$

Considering the simplified case where all assets in a portfolio have variance $\sigma_i = 1$, it follows that:

$$\rho \geq -\frac{N}{N(N-1)} = -(N-1)^{-1}, \tag{1.17}$$

which is the lower bound for ρ. $\triangle$

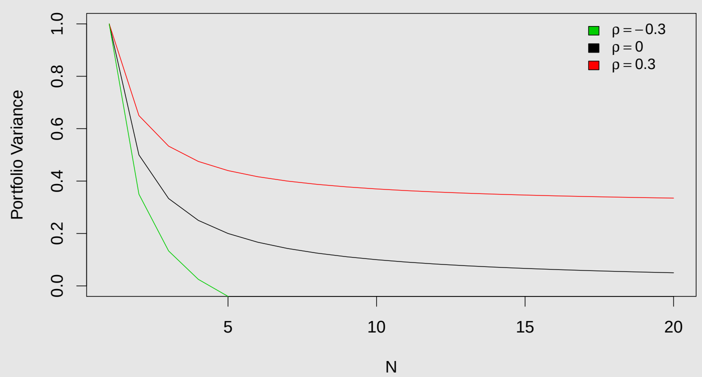


Figure 1.3. The effects of diversification of uncorrelated and correlated ($\rho = \pm 0.3$) assets in an equally weighted portfolio with an increasing number of assets N, cf. Example 1.

Another way to say this is that diversification aims to include assets in a portfolio that do not behave similar, meaning that they are low or negatively correlated with each other. Figure 1.3 shows how the variance of a portfolio behaves when more assets are included. The figure distinguishes between the case of uncorrelated assets, positively, and negatively correlated assets with a correlation of ± 0.3. It can be seen that the benefit of including more assets in a portfolio in order to obtain diversificiation is most significant when holding a few assets and decreases rapidly. Moreover it is observable that in the negatively and uncorrelated case the variance associated with the specific risk decreases much faster than in the correlated case, which confirms the concept of diversification. Note that as shown in Example 1 there is a lower bound for the correlation coefficent ρ that ensures that the portfolio variance does not get negative.

Since equation (1.14) can be considered as a simple linear regression, one can measure the fraction of systematic risk in the variance of the return of the ith asset in a portfolio using the R^2 statistic, which describes the percentage of the total variation in the rate of return that is explained by the regression equation, i.e.:

$$R^2 = \frac{\sigma_i^2 - \sigma_{\epsilon_i}^2}{\sigma_P^2(\boldsymbol{w})},$$

where σ_i^2 is the variance of the ith asset, $\sigma_{\epsilon_i}^2$ the variance of the residual of the ith asset, and $\sigma_P^2(\boldsymbol{w})$ is the portfolio variance. A high R^2 indicates that the variance consists of mostly systematic risk, whereas a low R^2 indicates mostly specific risk.

The concept of diversification is connected with the CAPM by the β of an asset, since it can be rewritten as:

$$\beta = \frac{\rho_{iM}\sigma_i}{\sigma_M},$$

where ρ_{iM} is the correlation between the return of asset i and the market return. So a high β means low diversification benefits. As mentioned earlier, investors want to hold assets with low or negative β such that when the market crashes, they hold assets that do not crash.

One of the drawbacks of the CAPM is that it focuses on the variance and covariances of the asset returns as measure of risk. The variance is a first-moment measure of risk, but most investors also consider higher moment measures of risk, e.g. kurtosis and skewness of asset returns. So indeed including additional assets in a portfolio reduces the variance, but other measures of risk may not be diminished. A portfolio can get more negatively skewed thus it has a higher downside risk, which would not be detected by the variance or β of an asset.

The CAPM has some very strong assumptions, which often are not met in reality. For instance the CAPM assumes that investors have mean-variance utility, but in reality investors have much more complicated utilities. In addition, it is assumed that investors have homogeneous expectations, although they are heterogeneous. And it is disregarding transaction costs, which can vary across assets. Furthermore, the CAPM is a single-period model, which assumes that returns are independently and identically distributed and that they are jointly multivariate normal. [J. Y. Campbell, A. W. Lo, A. C. MacKinlay, 1997]

As mentioned earlier, the market portfolio reflects systematic risk, which effects all assets. The next section extends the model setup to also include style factors like value-growth investing and momentum, or macro factors like inflation and economic growth, that have risk premia based on, e.g. investor characteristics and production capabilities of the economy.

1.2.2 Multifactor Models and the Arbitrage Pricing Theory

Unless otherwise stated, this section is based on [T. E. Copeland and J. F. Weston, 1988], [Overby, 2010], and [Ang, 2014]. The previous section introduced the CAPM, which assumes that there is only one factor in the market, namely the market factor. This section aims to generalize this setup to multifactor models, that describe the underlying drivers of assets by several factors.

One of the first models that used multiple factors was the Arbitrage Pricing Theory, APT, which has less restrictions regarding the assumptions made in the CAPM such as assumptions on the distribution of the returns or the utility function of investors. Moreover it does not require to identify the market portfolio, which can be difficult. The APT is mainly based on the assumption of no arbitrage i.e. the factors cannot be diversified away.

It is assumed that the rate of return, $\boldsymbol{r}$, of the assets can be modelled as a linear function of K unknown factors, $\boldsymbol{f}$, using a multifactor model:

$$\begin{aligned} &\boldsymbol{r} = \boldsymbol{\mu} + B\boldsymbol{f} + \boldsymbol{\epsilon} \\ &\mathrm{E}\left[\boldsymbol{f}\right] = \boldsymbol{0} \\ &\mathrm{E}\left[\boldsymbol{\epsilon}|\boldsymbol{f}\right] = \boldsymbol{0} \\ &\mathrm{E}\left[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top\right] = \Sigma_\epsilon, \end{aligned} \tag{1.18}$$

where $\boldsymbol{\mu}$ is the intercept of the factor model, $B = [\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N]^\top$ is the matrix of factor loadings, and $\boldsymbol{\epsilon}$ is the residual that describes the specific risk. [J. Y. Campbell, A. W. Lo, A. C. MacKinlay, 1997] It is assumed that the factors $\boldsymbol{f}$ only account for common variation, systematic risk, in the asset returns. Note that returns that have similar factor loadings on specific factors are likely to be correlated. The expected return of the assets is given by:

$$\mathrm{E}\left[\boldsymbol{r}\right] = \mathrm{E}\left[\boldsymbol{\mu} + B\boldsymbol{f} + \boldsymbol{\epsilon}\right] = \mathrm{E}\left[\boldsymbol{\mu}\right] + B\mathrm{E}\left[\boldsymbol{f}\right] + \mathrm{E}\left[\boldsymbol{\epsilon}\right] = \boldsymbol{\mu}.$$

By assuming independence of the residuals, the law of large numbers says that in a portfolio consisting of a large number of assets the specific risk vanishes. Consequently, the overall rate of return of a portfolio, $r_P(\boldsymbol{w})$, can be expressed as:

$$r_P(\boldsymbol{w}) = \boldsymbol{w}^\top \boldsymbol{r} = \boldsymbol{w}^\top \boldsymbol{\mu} + \boldsymbol{w}^\top B\boldsymbol{f}, \tag{1.19}$$

where it is assumed that $w_i \approx N^{-1}$, i.e. no asset is overweighted, and $\sum_{i=1}^N w_i\beta_{ij} = \beta_{Pj} = 0$ for each factor f_j where $j = 1, \ldots, k$. The second assumption ensures that there is no exact collinearity between the risk factors and thereby the overall model is factor neutral. This implies that $\boldsymbol{w}^\top B = \boldsymbol{0}^\top$, i.e. $\boldsymbol{w} \in \mathrm{Null}(B)$. Due to this assumptions, the portfolio given in equation (1.19) is called an *arbitrage portfolio*, since both systematic and specific risk are eliminated, and therefore it can be rewritten as:

$$r_P(\boldsymbol{w}) = \boldsymbol{w}^\top \boldsymbol{\mu} = \mu_P(\boldsymbol{w}) = 0. \tag{1.20}$$

The return $r_P(\boldsymbol{w})$ in equation (1.20) has to be equal to zero for otherwise it would be possible to attain an infinite return without risk and capital requirements. The result from equation (1.20) together with $\boldsymbol{w}^\top B = \boldsymbol{0}^\top$ imply that the expected return $\boldsymbol{\mu}$ must be a linear combination of a constant vector $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_k)$ and the coefficient matrix of factor loadings $B = [\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N]^\top$, i.e. it is assumed that:

$$\mathrm{E}\left[\boldsymbol{r}\right] = \boldsymbol{\mu} \approx \boldsymbol{1}\lambda_0 + B\boldsymbol{\lambda}, \tag{1.21}$$

where $\lambda_0, \ldots, \lambda_k$ are risk premia. If there exists a risk-free asset with return r_f then $\lambda_0 = r_f$. Hence equation (1.21) can be rewritten as:

$$\boldsymbol{\mu} - \boldsymbol{1}r_f = B\boldsymbol{\lambda}.$$

Assuming that $\boldsymbol{\lambda} = \bar{\boldsymbol{\delta}} - \boldsymbol{1}r_f$ this can be formulated as:

$$\boldsymbol{\mu} - \boldsymbol{1}r_f = B(\bar{\boldsymbol{\delta}} - \boldsymbol{1}r_f), \tag{1.22}$$

where $\bar{\boldsymbol{\delta}}$ is the expected rate of return on a portfolio with unit loading to the jth factor and zero loading on all other factors. Such a portfolio is called a *pure factor portfolio*. The formulation of $\boldsymbol{\lambda} = \bar{\boldsymbol{\delta}} - \boldsymbol{1}r_f$ is equivalent to the risk premium formulation in the CAPM in equation (1.12) with the difference that the CAPM only considers the market factor whereas the APT considers multiple factors.

Interpreting equation (1.22) as a linear regression equation and assuming that returns are jointly normal distributed and that the factors $\bar{\boldsymbol{\delta}}$ are linearly transformed to be orthonormal, the elements β_{ij} of the matrix B are defined in a similar manner to the CAPM as stated in equation (1.13), this is:

$$\beta_{ij} = \frac{\mathrm{Cov}\left[\mu_i, \bar{\delta}_j\right]}{\mathrm{Var}\left[\bar{\delta}_j\right]}.$$

So the CAPM can be viewed as a special case of the APT.

Let $r_P^{(l)}(\boldsymbol{w})$ denote the return of the lth factor portfolio, where $l = 1, \ldots, k$, which can be expressed by:

$$r_P^{(l)}(\boldsymbol{w}) = \mu_P^{(l)}(\boldsymbol{w}) + \sum_{j=1}^{k} \beta_{Pj}^{(l)}(\boldsymbol{w}) f_j,$$

where:

$$\begin{aligned}
&\mu_{Pj}^{(l)}(\boldsymbol{w}) = \sum_{i=1}^{N} w_i^{(l)} \mu_i, \\
&\beta_{Pj}^{(l)}(\boldsymbol{w}) = \sum_{i=1}^{N} w_i^{(l)} \beta_{ij} \quad \text{for} \quad \beta_{Pj}^{(l)}(\boldsymbol{w}) = \begin{cases} 1 & \text{if } j = l, \\ 0 & \text{if } j \neq l, \end{cases} \\
&\sum_{i=1}^{N} w_i^{(l)} = 1.
\end{aligned} \tag{1.23}$$

In matrix notation equation (1.23) can be expressed as $WB = I_k$, where W is a $N \times k$ matrix of weights, B is a $k \times k$ matrix of factor loadings, and I_k is a $k \times k$ identity matrix. Note that there has to be solved a system of linear equations. Hence it is ensured that there are unique solutions if there are the same number of equations as there are unknowns. Else if there are fewer equations than unknowns, there are infinitely many solutions. The risk premium for the lth factor is then defined by:

$$\lambda_l = \mu_P^{(l)}(\boldsymbol{w}) - r_f.$$

Using this setup, Ross (1976) has shown that in the absence of arbitrage opportunities the APT can be formulated as:

Result 1.5 (Abitrage Pricing Theory)
Consider an investment without specific risk, then the expected rate of return is given by:

$$\boldsymbol{\mu} \approx \mathbf{1}\lambda_0 + B\boldsymbol{\lambda},$$

where λ_0 is the intercept and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_k)^\top$ are risk premia for the K factors. [J. Y. Campbell, A. W. Lo, A. C. MacKinlay, 1997]

The disadvantage of multifactor models is that they do neither specify the number factors nor identify their meaning. As mentioned in [Ang, 2014] risk factors are like nutrients are to food. Such as some types of food are a bundle of nutrients and others contain only one nutrient, assets can consist of one or more risk factors. The intuition behind this is, that the risk factors behind the assets matter, not the assets themselves.

It can be distinguished between macro factors that are common for several assets in a portfolio due to the presence of, e.g. inflation or volatility, in the financial market and style factors like value-growth or size based on firm-characteristics, which will be explained later in this section. These two types of factors are undiversifiable. Specific factors, that for instance only affect a specific firm, can be diversified by including a large number of assets as explained above. Depending on the exposure and type of the underlying, undiversifiable risk factors assets have different risk premia. These premia compensate for low returns during bad times with a premium of high returns in the long run. This means that different risk premia describe different sets of bad times, i.e. bad economical times. Every investor has an individual definition of 'bad time', which among others depends on the investor's income, liabilites, and risk aversion. Depending on the aggregate supply of a factor in the financial markets and type of risk factor, risk premia can be positive, negative, or zero. Assets that have high returns during bad times, e.g. are negatively correlated to market movements, have a high price and thereby a low risk premium. In contrast have assets that are positively correlated with market movements, i.e. crash together with most other assets, a low price and high risk premium to compensate for these losses.

There exists a plenty of different factors, where the most fundamental factor is the market factor described by the CAPM as introduced in the previous section. But in order to use them, one should justify the academic research behind them and they should satisfy the following which is inspired by [Mesomeris, 2013]:

- Risk factors shall be explainable.
- They must be persistent, i.e. continue to exist and not just be a phenomenon for a short time period.
- In isoloation they should have attractive return characteristics.
- A risk factor should have low correlation to traditional market β's and other risk factors considered for a portfolio.
- The risk factor must be accessible.
- Risk factors should be priced.

The following gives a short explaination of macro and style factors.

Macro Factors

Macro factors affect all investors and prices of assets in an economy. For instance affects low growth and high inflation everyone, but to different degree. Many macro factors are lasting, e.g. when inflation is low today, it is likely that it will be low next month. One other well-known macro factor is volatility, since stock returns are negatively correlated to volatility, which is also known as the *leverage effect*. This effect describes the relationship between stock returns and volatilities, since stock prices fall when volatility increases. Volatility can also be viewed as a kind of uncertainty risk factor, since it is highly correlated with the uncertainty of investors to e.g. policy decisions that can affect the economy. There are many other macro factors, e.g. economic growth, interest rate risk, or currency risk.

The following section introduces some of the style factors considered in this thesis.

Style Factors

As mentioned earlier, the assumptions of the CAPM model, introduced in Section 1.2.1, do often not hold in reality. The model has been tested extensively in the 1970s using time series regressions of e.g. the returns of the S&P 500.[1] The tests showed that the assumption of the CAPM that the market factor is the only factor in the market were not satisfied, hence there must be other factors in the market that influcence asset prices. [M. Grinblatt and S. Titman, 2002] This has motivated to introduce models which are based on several factors.

[1]The Standard and Poor's S&P 500 stock market index includes the 500 leading companies in the US. [SP Dow Jones Indices, 2015]

Fama and French (1993) introduced a model that explains assets by three factors. The first factor is the traditional market factor M from the CAPM, and then they had two additional factors to capture a size factor, S, and a value-growth factor, V:

$$\mu_i = r_f + \beta_{i,M}\mathrm{E}\left[r_M - r_f\right] + \beta_{i,S}\mathrm{E}\left[S\right] + \beta_{i,V}\mathrm{E}\left[V\right].$$

The following explains the size and value-growth factors.

Size Factor
This factor describes the market capitalization of stocks. It has been observed that small firms outperform large firms. The effect was discovered in 1981, but since the mid-1980s the existence of this effect is debatable. There are made many studies that argue for the disappearance of the effect, and others can find the effect for special segments. For are discussion of the disappearance of the size effect see e.g. [Crain, 2011].

Value-growth Factor
Stocks that have a low price in relation to their net worth, which is the same as a high book-to-market ratio, are called value stocks. The book-to-market ratio is the book value divided by the market capitalization. This are companies that currently are out of favour in the financial market or newer companies with unknown track records. [J. O. Reilly, S. O. Barba, N. Pavlov, 2003] On the other hand, stocks with low book-to-market ratios are called growth stocks. This are typically companies that had good earnings in the recent years and are expected to continue to yield high profit growth. The investment strategy of going long value stocks and short growth stocks is known as the *Value-Growth strategy.* The reason for this is that value stocks outperform growth stocks, on average. It has been a robust premium for many years.

The factors in the Fama-French model can be constructed by using characteristic-sorted portfolios, which means that the factors are estimated by using portfolios that are formed based on firm characteristics as described for the different style factors. [M. Grinblatt and S. Titman, 2002] Besides the size and value-growth effects there are other effects, as will be explained in the following.

Momentum Factor
Momentum describes the strategy of buying stocks that have increasing returns over the past months and selling stocks with the lowest returns over the same period. The effect is that winner stocks continue to win and losers continue to lose. It is observed that the Momentum strategy yields higher profits than the Value-Growth and Size strategies do, and it is observable in every asset class. The effect can be added to the Fama-French model, where positive momentum βs indicate winner stocks and negative βs indicate loser stocks.

Low Risk Factor
Describes assets that have low volatility and is often called *Low Volatility strategy.*[2]

[2]Source: Internal documentation from Jyske Bank A/S.

Quality Factor
Quality describes the effect of firms facing negative returns in future earnings announcements, when a large proportion of their earnings come from revenue, compared to firms where earnings are based on cash flow.[2]

Strategies like Size, Value-Growth, and Momentum are called *cross sectional strategies*, since they compare one group of stocks with another group, e.g. value stocks with growth stocks. It should be noted that both the CAPM, APT, and Fama-French model assume that the βs are constant, but the exposure of assets to systematic factors vary over time, and often increases during bad times.

This chapter introduced some basic economical concepts regarding portfolio optimization including different measures of risk and some traditional asset allocation strategies. Furthermore, some important economical models for factor models have been introduced in order to get a better understanding of diversification and risk factors. In order to establish risk-based asset allocation strategies in Chapter 3, the next chapter introduces statistical methods to find underlying, lower dimensional factors. These are the projection methods principal component analysis, PCA, and functionional principal component analysis, fPCA.

2 Functional Data Analysis

This chapter introduces the concept of *functional data analysis*, FDA, which is used for data providing information about curves, surfaces, or anything else that varies over a continuum. The focus in this thesis is on asset prices which are observed discretely but in fact follow an underlying stochastic process. This means instead of considering asset prices as individual observations, it is assumed that data has a functional form, i.e. the observations are linked together in some way. Asset prices can be considered as curves whose continuum is time, e.g. consider Figure 2.1, and therefore it is possible to apply FDA techniques. Many of these techniques are conceptually the same as for multivariate data, but instead of vectors there are considered infinite dimensional vector spaces. It is possible to account for unequally spaced observations, i.e. different sampling rates of assets in a portfolio, and FDA also can handle missing values.

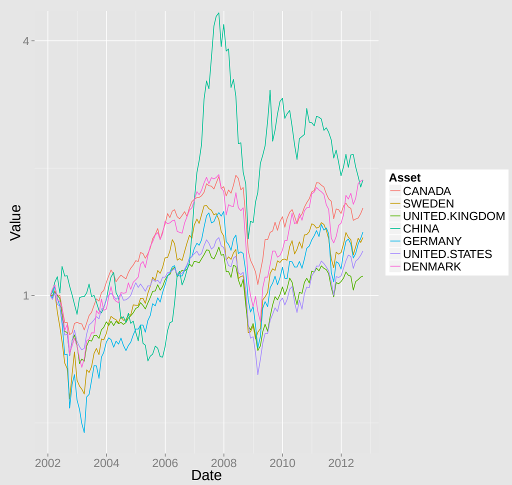


Figure 2.1. Monthly sampled scaled stock indexes from January 1, 2002 through September 24, 2012. Canada (*SPTSX*), Sweden (*OMX*), United Kingdom (*UKX*), China (*SHSZ300*), Germany (*DAX*), United States (*SPX*), Denmark (*KFX*).

Moreover, in order to extract risk factors that lie behind the assets of a portfolio, principal component analysis, PCA, is introduced in Section 2.2 and expanded to functional principal component analysis, fPCA, considering data to be functional instead of multivariate in Section 2.3. The motivation for this is that fPCA is better to capture the variability of the asset returns, which is essential for the asset allocation strategies introduced in Chapter 3. This is due to the fact that in fPCA it is possible to observe the behaviour of the eigenfunctions over time in contrast to PCA that just gives a static, non-temporal estimate of the eigenvectors. In addition, functional data is *smoothed* before performing a fPCA, which improves the signal-to-noise ratio of data and therefore may improve the allocation strategies by specifying the true underlying risk factors.

The following sections give an introduction to FDA, the concept of basis functions, and of smoothing functional data. Furthermore the projection methods PCA and fPCA are introduced, and some important results for these methods are shown.

2.1 Functional Data

This section is based on [J. O. Ramsay , B. W. Silverman, 2005] and [Zhang, 2014]. In functional data analysis, observed data functions are considered to provide information about possible infinite-dimensional curves. The observations often have time as continuum, but could also have other continua such as spatial position.

The reason that functional data is not considered as multivariate data is the assumption of the observations being linked together in some way, meaning that data has a functional form. Consider the observed data (t_j, y_j) for $j = 1, \ldots, n$ where y_j is a response observed at times t_j. The transformation to the functional form has two cases; the discrete values are assumed to be errorless:

$$y(t_j) = y_j = x(t_j),$$

where $x(t_j)$ is a smooth function. This process of converting discrete data into functional data is called *interpolation*. The other case is given by:

$$y(t_j) = y_j(t) = x(t_j) + \epsilon_j. \tag{2.1}$$

The second case may involve smoothing of data in order to reduce the residual ϵ_j. One possible smoothing technique will be introduced in Section 2.1.2. Equation (2.1) can be rewritten to represent all N curves:

$$y_i(t_j) = x_i(t_j) + \epsilon_i(t_j) \quad \text{for } i = 1, \ldots, N \text{ and } j = 1, \ldots, n_i. \tag{2.2}$$

It is assumed that the function $x_i(t)$ can be decomposed into:

$$\bar{x}(t) + \nu_i(t),$$

where $\bar{x}(t)$ is the mean function, which is given by:

$$\bar{x}(t) = N^{-1} \sum_{i=1}^{N} x_i(t), \tag{2.3}$$

and $\nu_i(t)$ is the ith individual variation from $\bar{x}(t)$ with $\sum_{i=1}^{N} v_i(t) = 0$. Thus a functional data set is modelled as independent realizations of an underlying stochastic process:

$$y_i(t_j) = \bar{x}(t_j) + \nu_i(t_j) + \epsilon_i(t_j),$$

where it is assumed that $\nu_i(t)$ and $\epsilon_i(t)$ are independent. Moreover it is assumed that $\nu_i(t)$ and $\epsilon_i(t)$ follow stochastic processes:

$$\nu_i(t) \sim \mathrm{SP}(0, v) \quad \text{and} \quad \epsilon_i(t) \sim \mathrm{SP}(0, v_\epsilon),$$

where $\mathrm{SP}(\bar{x}, v)$ denotes a stochastic process with mean function $\bar{x}(t)$ and covariance function $v(s,t)$, which is given by:

$$v(s,t) = (N-1)^{-1} \sum_{i=1}^{N} \left(x_i(s) - \bar{x}(s)\right)\left(x_i(t) - \bar{x}(t)\right). \tag{2.4}$$

Moreover it is assumed that $v_\epsilon(s,t) = \sigma^2(t)\mathbf{1}[s = t]$, where $\sigma^2(t)$ is the residual variance function that measures the variation of the measurement errors.

Normally the data transfromation takes place seperately for each curve i, but if there is a low signal-to-noise ratio or sparsely sampled data, it can be useful to take information from neighboring or similiar curves into account to get more stable estimates for a curve.

The assumption of a smooth function usually means that the function $x(t)$ has one or more derivatives. Thereby functional data analysis uses information in the features of a curve, i.e. slopes, curvature, crossings, peaks, or valleys. Functional features can be considered as events that are related to a specific value of the argument t. They can be characterized by their location, amplitude, or width, which can be treated as a measure of dimension. For instance, a peak can be considered as being three-dimensional, since location, amplitude, and width have to be known for full information of the peak.

The argument values $t_1, \ldots, t_{n_i}$ can be the same for all curves i, but can also vary from curve to curve. In addition, it is also possible to deal with curves that have missing values, since data is transformed to a continuous structure. It is also important to consider the sampling rate of data, which takes into account the ratio between the argument values t_j relative to the amount of curvature, which is measured by the second order derivative $|D^2x(t)|$ or $[D^2x(t)]^2$. The higher the curvature is, the more points are needed for a good estimation of a curve. [J. O. Ramsay , B. W. Silverman, 2005]

In order to represent the continuous functions, the next section introduces the concept of *basis functions*.

2.1.1 Basis Functions

This section is based on [J. O. Ramsay , B. W. Silverman, 2005]. In functional space basis functions are the counterpart to basis vectors in vector space. This means, that such as every vector can be represented by a linear combination of basis vectors, every continuous function can be represented by a linear combination of basis functions.

Consider the vector-valued function $\boldsymbol{x}$ with components $x_1(t), \ldots, x_N(t)$. Moreover let $\boldsymbol{\phi}$ be the vector-valued function with K independent, real-valued basis function components $\phi_1(t), \ldots, \phi_k(t)$, and let C be a $N \times K$ coefficient matrix, then the simultaneous expansion of all N curves can be expressed by:

$$\widehat{\boldsymbol{x}} = C\boldsymbol{\phi}(t). \tag{2.5}$$

For a single curve $x_i(t)$ this corresponds to:

$$\widehat{x}_i(t) = \boldsymbol{c}_i^\top \boldsymbol{\phi}(t) = \boldsymbol{\phi}(t)^\top \boldsymbol{c}_i.$$

This shows, that basis expansion methods represent possibly infinite functions by finite dimensional vectors.

The number of basis functions K can be seen as a parameter, that is selected according to characteristics in data, often by using a cross validation method. It is preferred to have a low value of K in order to avoid overfitting and to reduce computations. The case $K = n_i$, where n_i is the number of observations of curve i, is known as exact representation or interpolation.

Basis functions ϕ_k are said to be orthogonal over some interval $t \in \mathcal{T}$ if for all m:

$$\int_{\mathcal{T}} \phi_k(t)\phi_m(t) = \begin{cases} 0 & k \neq m, \\ \lambda_m & k = m, \end{cases} \tag{2.6}$$

where $\lambda_m \in \mathbb{R}$ and if $\lambda_m = 1$, then the basis functions are said to be orthonormal. There are different choices of basis functions depending on the structure of data. The most common basis functions are Fourier- , B-splines- , monomials- , and wavelets basis functions. When dealing with periodic data the most widely used basis functions are Fourier functions, whereas in the non-periodic case it is B-splines. [J. O. Ramsay , B. W. Silverman, 2005] In the following B-splines will be introduced.

B-Splines

Spline functions have replaced polynomials, since they have the property of fast computation and a high degree of flexibility. [J. O. Ramsay , B. W. Silverman, 2005] In order to define a spline, consider the interval over which a function is to be approximated. The interval is divided into L subintervals, which are seperated by so-called *breakpoints* τ_l, $l = 1, \ldots, L-1$, i.e. an interval $[a, b]$ is diveded into

$a = \tau_0 < \tau_1 < \cdots < \tau_{L-1} < \tau_L = b$. The breakpoints τ_l for $l = 1, \ldots, L-1$ are called the *interior breakpoints*. Over each subinterval $[\tau_r, \tau_{r+1})$ for $r = 0, \ldots, L-1$, a spline is a polynomial of order m, which are required to join smoothly at the interior breakpoints. This means that the derivatives match up to the order one less than the degree of the polynomial.

The first polynomial has m degrees of freedom, but each subsequent polynomial has only one degree of freedom because of the $m-1$ constraints on its behaviour. This implies that the total number of degrees of freedom in the fit equals the order of the polynomials plus the number of interior breakpoints, $L+m$ degrees of freedom.

The higher the order, the better is the approximation of the function. In addition, greater flexibility in a spline is achieved by increasing the number of breakpoints. The breakpoints do not have to be equally spaced, so it is preferable to have more breakpoints where the function posses the most variation. Note that given an order m and a breakpoint sequence τ, every basis function ϕ_k is itself a spline function. And that any linear combination of basis functions ϕ_k is a spline function.

A B-spline basis function of order m has the *compact support* property. The support of a function is the set of points where the function is not zero-valued, if the function is positive over no more than m intervals, and when these are neighbouring intervals. This property is important for fast computation, since it implies that when there are K B-spline basis functions, the order K matrix of inner products of these functions will be band-structured, with $m-1$ subdiagonals above and below the diagonal containing nonzero values. So the computation of spline functions increases only linearly with K.

A B-spline at time t defined by the breakpoint sequence τ is denoted by $B_k(t, \tau)$, where k refers to the number of the largest breakpoint at or to the left of the value t. Thus a spline function $S(t)$ with interior breakpoints is defined as:

$$S(t) = \sum_{k=1}^{m+L-1} c_k B_k(t, \tau). \tag{2.7}$$

The reason that there are $m + L - 1$ basis functions is that there normally is no information about how the functions behave beyond the interval on which data is collected. Thereby the function may be discontinuous beyond the boundaries, which makes the functions non-differentiable at the boundaries. B-splines avoid this by extending the breakpoint sequence at each end with additional $m-1$ replicates of the boundary breakpoint value. In equation (2.7) only the $m-1$ breakpoints added to the initial breakpoint are also counted.

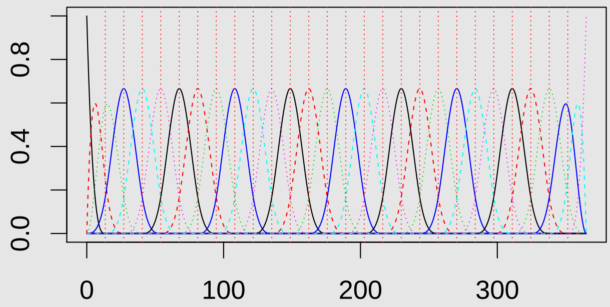


Figure 2.2. B-spline basis with $K = 30$ basis functions.

The position of the interior breakpoints τ_l can be determined by different methods. Often the default method is equal spacing, which is good as long data has a constant sampling rate as mentioned in Section 2.1. An example of an equally spaced B-spline basis with 30 breakpoints can be seen in Figure 2.2. Another possibility is to place a breakpoint at every j'th breakpoint, where j is a fixed number selected in advance. There exists algorithms for breakpoint positioning which are similar to variable selection techniques in multivariate regression. [J. O. Ramsay , B. W. Silverman, 2005]

One possible approach of constructing orthogonal basis functions is functional principal component analysis, fPCA, which will be introduced in Section 2.3. But first the following section explains how data can be smoothed using a *roughness penalty*.

2.1.2 Smoothing Functional Data

This section is inspired by [J. O. Ramsay , B. W. Silverman, 2005] and [J. O. Ramsay, G. Hooker, S. Gaves, 2009]. When data observations may contain errors, data it has to be smoothed in order to obtain a functional form as stated in Section 2.1. There are different techniques available, but the focus here is on smoothing data using a roughness penalty. This method is based on *weighted least squares*, which is given by:

$$\mathrm{SMSSE}(\boldsymbol{y}|\boldsymbol{c}) = (\boldsymbol{y} - \Phi(t)\boldsymbol{c})^\top W(\boldsymbol{y} - \Phi(t)\boldsymbol{c}),$$

where W is a symmetric positive definite weight matrix , $\boldsymbol{y}$ is a response vector, $\boldsymbol{c}$ is a vector of coefficients, and $\Phi(t)$ is a matrix consisting of basis functions as described in Section 2.1.1. The basis functions in Φ could for instance be B-splines.

When the covariance matrix Σ_e for the resiuduals ϵ_j is known, then the weight matrix is given by:

$$W = \Sigma_e^{-1}.$$

To be able to establish smoothing by a roughness penalty, one has to quantify the notion of roughnesss of a function. A natural measure could be the integrated squared mth derivative:

$$\mathrm{PEN}_L(x) = \int [Lx(s)]^2 \, ds,$$

where $L = D^m$ is the mth derivative. A rough function has high curvature, which yields high values for $\mathrm{PEN}_L(x)$. Let $x(\boldsymbol{t})$ be the vector resulting from a function x evaluated at time arguments $\boldsymbol{t}$. Then the *penalized residual sum of squares* is defined by:

$$\mathrm{PENSSE}_\lambda(x|\boldsymbol{y}) = (\boldsymbol{y} - x(\boldsymbol{t}))^\top W(\boldsymbol{y} - x(\boldsymbol{t})) + \lambda \mathrm{PEN}_L(x), \tag{2.8}$$

where $\lambda \geq 0$ is a *smoothing parameter* that measures the tradeoff between goodness of fit of data and the variability of the function x. When $\lambda = 0$ then the penality term has no influence and equation (2.8) is a usual weighted least squares problem. As $\lambda \to \infty$ the penalty term gets more influence so curvature is more penalized. In the field of statistics this is also known as *ridge regression*.

In order to obtain a smoothed functional data set, the aim is to minimize equation (2.8). Substituting the basis expansion $x(\boldsymbol{t}) = \boldsymbol{c}^\top \Phi(t) = \Phi^\top(t)\boldsymbol{c}$ into this equation, yields:

$$\mathrm{PENSSE}_\lambda(x|\boldsymbol{y}) = (\boldsymbol{y} - \boldsymbol{c}^\top \Phi(t))^\top W(\boldsymbol{y} - \boldsymbol{c}^\top \Phi(t)) + \lambda \boldsymbol{c}^\top R \boldsymbol{c}, \tag{2.9}$$

which follows from the following computation:

$$\begin{aligned}
\mathrm{PEN}_L(x) &= \int [Lx(s)]^2 \, ds \\
&= \int \left[L\boldsymbol{c}^\top \Phi(s)\right]^2 ds \\
&= \int \boldsymbol{c}^\top L\Phi(s) L\Phi(s)^\top \boldsymbol{c} \, ds \\
&= \boldsymbol{c}^\top \int L\Phi(s) L\Phi(s)^\top ds \; \boldsymbol{c} \\
&= \boldsymbol{c}^\top R \boldsymbol{c},
\end{aligned}$$

where $R = \int L\Phi(s)L\Phi(s)ds$ is the roughness penalty matrix. It is possible to obtain analytic solutions for some types of basis systems, e.g. for the B-spline basis, but the details for this computation are complicated and omitted in this thesis. Differentiating equation (2.9) with respect to $\boldsymbol{c}$ yields:

$$-2\Phi(t)^\top W \boldsymbol{y} + \Phi(t)^\top W \Phi(t) \boldsymbol{c} + \lambda R \boldsymbol{c} = 0.$$

Hence the estimated coefficient vector is given by:

$$\widehat{\boldsymbol{c}} = \left(\Phi(t)^\top W \Phi(t) + \lambda R\right)^{-1} \Phi(t)^\top W \boldsymbol{y}, \tag{2.10}$$

which is exactly the ridge regression estimate.

In order to determine λ, *generalized cross validation*, GCV, with the following cross validation statistic as criterium is used:

$$\mathrm{GCV}(\lambda) = \frac{n^{-1}\mathrm{SMSSE}}{(n^{-1}\mathrm{trace}(I - H))^2}, \tag{2.11}$$

which was developed by Craven and Wahba (1979), see [Craven and Wahba, 1979]. It is an approximation to the statistic obtained by leave-one-out cross validation, LOOCV, when using a linear regression model which is given by:

$$\text{LOOCV} = n^{-1} \sum_{i=1}^{n} \left(\frac{\text{SMSSE}}{1 - h_i} \right)^2 .$$

This result can for instance be found in [Hyndman, 2014]. LOOCV is a cross validation method where all observations expect from one are used as trainings set and the excluded observation is used as validation set, which is done until all observations have been in the validation set. [J. Friedman, T. Hastie, R. Tibshirani, 2009] When using cross validation to find the smoothing parameter λ one has to pay attention when choosing possible λ values to be validated, since the linear system to be solved has limitations. Define the term in the parenthesis in equation (2.10) as:

$$M(\lambda) = \Phi(t)^\top W \Phi(t) + \lambda R.$$

The matrices $\Phi(t)^\top W \Phi(t)$ and R have elements of completely different size, but $M(\lambda)$ has to be invertible in equation (2.10), which poses some limitations. [J. O. Ramsay , B. W. Silverman, 2005] suggests that the size of λR should not exceed 10^{10} times the size of $\Phi(t)^\top W \Phi(t)$.

It is possible to make λ dimensionless by considering $\log_{10}(\lambda)$, which is in accordance with the fact that $\lambda > 0$ when there is imposed a penality. So the roughness penalty $\text{PEN}_L(x)$ should be multiplied by 10^ν, where $\nu = \log_{10}(\lambda)$.

Before introducing fPCA in Section 2.3, the basic ideas of PCA for multivariate data are explained in the next section.

2.2 Principal Component Analysis

This section is inspired by [Tvedebrink, 2014] and [J. O. Ramsay , B. W. Silverman, 2005]. PCA is a dimensionality reduction method that uses orthogonal transformations to transform a set of variables into a set of linearly uncorrelated variables known as *principal components*. The method is defined in such a way that the first principal component accounts for the largest possible variance in data, and each subsequent component has the highest possible variance subject to the constraint of being orthogonal to the preceding components. The method is often confused with factor analysis, which is a very similar method but assumes that variables can be expressed as a linear combination of underlying factors. Hence factor analysis assumes that there exist underlying factors, whereas PCA just is a dimensionality reduction method.

In the case of multivariate data a matrix X is considered. Without loss of generality, it is assumed that the matrix is centered, which means that $X = X - N^{-1}\mathbf{1}\mathbf{1}^\top X$, where N is the number of observations in data.

In addition, the covariance matrix V can be expressed by:

$$V = N^{-1} X^\top X, \tag{2.12}$$

which is of dimension $p \times p$. It is possible to look at a linear combination of the variable values:

$$f_i = \sum_{j=1}^{p} \beta_j x_{ij} = \boldsymbol{\beta}^\top \boldsymbol{x}_i, \quad i = 1, \ldots, N, \tag{2.13}$$

where $\boldsymbol{\beta}$ is a weighting vector $(\beta_1, \ldots, \beta_p)^\top$ and $\boldsymbol{x}_i$ is an observed vector $(x_{i1}, \ldots, x_{ip})^\top$. PCA is then used to define sets of normalized weights that maximize variation in the f_i's. The first step is to find a vector $\boldsymbol{\xi}_1 = (\xi_{11}, \ldots, \xi_{p1})^\top$ such that the covariance:

$$\begin{aligned} \mathrm{Cov}\left[X\boldsymbol{\xi}_1\right] &= \mathrm{E}\left[(X\boldsymbol{\xi}_1)^\top X\boldsymbol{\xi}_1\right] - \mathrm{E}\left[X\boldsymbol{\xi}_1\right]^\top \mathrm{E}\left[X\boldsymbol{\xi}_1\right] \\ &= \mathrm{E}\left[\boldsymbol{\xi}_1^\top X^\top X\boldsymbol{\xi}_1\right] - \boldsymbol{\xi}_1^\top \mathrm{E}\left[X\right]^\top \mathrm{E}\left[X\right] \boldsymbol{\xi}_1 \\ &= \boldsymbol{\xi}_1^\top \mathrm{Cov}\left[X\right] \boldsymbol{\xi}_1 = \boldsymbol{\xi}_1^\top V \boldsymbol{\xi}_1, \end{aligned} \tag{2.14}$$

is as large as possible. To ensure that the covariance not gets arbitrarily large, the following constraint is introduced:

$$\|\boldsymbol{\xi}_1\|^2 = \boldsymbol{\xi}_1^\top \boldsymbol{\xi}_1 = 1.$$

The problem can also be formulated as maximizing the mean square:

$$\frac{1}{N} \sum_{i=1}^{N} f_{i1}^2 \quad \text{subject to} \quad \|\boldsymbol{\xi}_1\|^2 = 1,$$

where $f_{i1} = \sum_{j=1}^{p} \xi_{j1} x_{ij} = \boldsymbol{\xi}_1 \boldsymbol{x}_i$. The linear combination f_{i1} is also called a *principal component score*. Moreover, the subsequent projections have to be uncorrelated to the previous ones:

$$\sum_{j=1}^{p} \xi_{jk} \xi_{jl} = \boldsymbol{\xi}_k^\top \boldsymbol{\xi}_l = 0, \quad k < m \leq p,$$

where m indicates the number of steps taken, which are limited by the number of variables p.

This means that the projections are orthogonal to each other, which guarantees that they are describing a new underlying feature in data. Based on the introduced constraints, the following maximization problem can be formulated:

$$\max_{\boldsymbol{\xi}_1} \boldsymbol{\xi}_1^\top V \boldsymbol{\xi}_1 - \lambda(\boldsymbol{\xi}_1^\top \boldsymbol{\xi}_1 - 1), \tag{2.15}$$

which is solved by:

$$\begin{aligned} (V - \lambda I)\boldsymbol{\xi}_1 &= 0 \\ V\boldsymbol{\xi}_1 &= \lambda \boldsymbol{\xi}_1. \end{aligned} \tag{2.16}$$

This can be considered as finding the eigenvalues λ and eigenvectors $\boldsymbol{\xi}_j$ of the covariance matrix V. An eigenvector can represent the direction of a component and the corresponding eigenvalue can represent how much variance there is in data in this direction.

Since the covariance matrix V is positive semi-definite, it can be decomposed into an orthogonal matrix U and a diagonal matrix Λ, whose entries are the eigenvalues $\lambda_1 > \cdots > \lambda_p$. The decomposition is then given by:

$$V = U^\top \Lambda U = \sum_{j=1}^{p} \lambda_j \boldsymbol{u}_j \boldsymbol{u}_j^\top. \tag{2.17}$$

Using this expression in equation (2.15) and defining $\tilde{\boldsymbol{\xi}}_1 = U\boldsymbol{\xi}_1$, where:

$$\left\|\tilde{\boldsymbol{\xi}}_1\right\|^2 = \tilde{\boldsymbol{\xi}}_1^\top \tilde{\boldsymbol{\xi}}_1 = (U\boldsymbol{\xi}_1)^\top U\boldsymbol{\xi}_1 = \boldsymbol{\xi}_1^\top U^\top U \boldsymbol{\xi}_1 = \boldsymbol{\xi}_1^\top \boldsymbol{\xi}_1 = \|\boldsymbol{\xi}_1\|^2,$$

the maximization problem from equation (2.15) can be rewritten as:

$$\max_{\tilde{\xi}_1 : \tilde{\xi}_1^\top \tilde{\xi}_1 = 1} \tilde{\boldsymbol{\xi}}_1^\top \Lambda \tilde{\boldsymbol{\xi}}_1 = \max_{\tilde{\xi}_1 : \tilde{\xi}_1^\top \tilde{\xi}_1 = 1} \sum_{j=1}^{p} \tilde{\xi}_{1j}^2 \lambda_j.$$

This is maximal when $\tilde{\xi}_{11} = 1$, so $\tilde{\boldsymbol{\xi}}_1 = \boldsymbol{e}_1$, where $\boldsymbol{e}_1$ is the unit vector with a one in the first entry. As mentioned, the next component has to be uncorrelated with the first component, so:

$$0 = \mathrm{Cov}\left[X\boldsymbol{\xi}_1, X\boldsymbol{\xi}_2\right] = \boldsymbol{\xi}_1^\top \mathrm{Cov}\left[X, X\right] \boldsymbol{\xi}_2 = \boldsymbol{\xi}_1^\top U^\top \Lambda U \boldsymbol{\xi}_2 = \tilde{\boldsymbol{\xi}}_1^\top \Lambda \tilde{\boldsymbol{\xi}}_2. \tag{2.18}$$

Since $\tilde{\boldsymbol{\xi}}_1 = \boldsymbol{e}_1$, equation (2.18) reduces to $0 = \lambda_1 \tilde{\xi}_{21} \Leftrightarrow \tilde{\xi}_{21} = 0$. Hence, the new maximization problem is:

$$\max_{\tilde{\xi}_2 : \tilde{\xi}_2^\top \tilde{\xi}_2 = 1; \tilde{\xi}_{21} = 0} \tilde{\boldsymbol{\xi}}_2^\top \Lambda \tilde{\boldsymbol{\xi}}_2 = \max_{\tilde{\xi}_2 : \tilde{\xi}_2^\top \tilde{\xi}_2 = 1; \tilde{\xi}_{21} = 0} \sum_{j=1}^{N} \tilde{\xi}_{2j}^2 \lambda_j,$$

which implies that $\tilde{\xi}_{22} = 1$, so $\tilde{\boldsymbol{\xi}}_2 = \boldsymbol{e}_2$, where $\boldsymbol{e}_2$ denotes the unit vector with a one in the second entry.

This procedure is repeated until the m principal components are found, where m is at most p. [J. O. Ramsay , B. W. Silverman, 2005]

2.2.1 Singular Value Decomposition and the Smoothing of Data

This section is inspired by [J. Friedman, T. Hastie, R. Tibshirani, 2009]. As mentioned in Section 2.1.2, the estimated coefficient $\widehat{\boldsymbol{c}}$ is similar to a ridge regression estimate. This section shows that ridge regression has a relation to PCA. Consider the case of a weighted matrix as described in Section 2.1.2, where W is a weight matrix. The *singular value decomposition*, SVD, of a weighted and centered matrix Φ is then given by:

$$W^{1/2}\Phi = UDV^\top, \tag{2.19}$$

where U is an $N \times p$ orthogonal matrix whose columns span the column space of Φ and V is an $p \times p$ orthogonal matrix whose columns span the row space of Φ. The matrix D is a diagonal matrix with diagonal elements $d_1 \geq \cdots \geq d_p \geq 0$, which are the so-called *singualar values* of Φ. The matrix Φ is called singular if one or more elements d_j are zero.

Consider the eigendecompostion of $\Phi^\top W \Phi$:

$$\Phi^\top W \Phi = VDU^\top UDV^\top = VD^2V^\top,$$

where the vectors $\boldsymbol{v}_j$ are the eigenvectors of $\Phi^\top W \Phi$. Using equation (2.19), the ridge regression estimate from equation (2.10) can be rewritten as:

$$\begin{aligned}\widehat{\boldsymbol{c}} &= (\Phi^\top W \Phi + \lambda R)^{-1}\Phi^\top W^{1/2}\boldsymbol{y} \\ &= (VD^2V^\top + \lambda R)^{-1}VDU^\top W^{1/2}\boldsymbol{y} \\ &= (V(D^2 + \lambda R)V^\top)^{-1}VDU^\top W^{1/2}\boldsymbol{y} \\ &= V(D^2 + \lambda R)^{-1}DU^\top W^{1/2}\boldsymbol{y}.\end{aligned}$$

Thus it follows that:

$$\widehat{\boldsymbol{y}} = W^{1/2}\Phi\widehat{\boldsymbol{c}} = UD(D^2 + \lambda R)^{-1}DU^\top \boldsymbol{y}. \tag{2.20}$$

Hence, ridge regression as used to smooth functional data finds projections of $\boldsymbol{y}$ on $\boldsymbol{u}_j$ and then shrinks these coordinates by $(D^2 + \lambda R)^{-1}$. So ridge regression shrinks coefficients with least variance most as can be seen in equation (2.20). This is equivalent to the concept of PCA, which finds directions in data that explain most variance and then finds orthogonal directions that describe less variance.

After having introduced the ideas of principal component analysis for multivariate data, the next section introduces the functional variant of this method.

2.3 Functional Principal Component Analysis

The following section is written with inspiration from [J. O. Ramsay , B. W. Silverman, 2005] and [J. O. Ramsay, G. Hooker, S. Gaves, 2009]. Instead of considering variable values, fPCA considers univariate function values $x_i(s)$, where the discrete index j has been replaced by a continuous index s. This implies that the linear combination in equation (2.13) can be expressed as:

$$f_i = \int \beta(s)x_i(s)ds, \quad i = 1, \ldots, N. \tag{2.21}$$

The functions $x(s)$ are assumed to be real-valued and to be elements of a *Hilbert space* $\mathcal{H}$ as defined in Defintion A.3 of Appendix A.1, i.e.:

$$\left(\int x(s)^2 ds\right)^{\frac{1}{2}} < \infty.$$

Considering functions instead of vectors, summations over j are replaced with integrations over s. So the results from Section 2.2 can be formulized to finding a weight function $\xi_j(s)$ that maximizes:

$$\frac{1}{N}\sum_{i=1}^{N} f_{ij}^2 = \frac{1}{N}\sum_{i=1}^{N}\left(\int \xi_j(s)x_i(s)\right)^2 ds,$$

subject to:

$$\int \xi_j(s)^2 ds = 1 \quad \text{and} \quad \int \xi_k(s)\xi_m(s)ds = 0, \quad k < m, \tag{2.22}$$

where $\xi_j(t)$ are eigenfunctions. Similar to the multivariate case, the functions are acquired to be centered, meaning that each function has been subtracted by the mean function as stated in equation (2.3).

The counterpart to the covariance matrix in the multivariate case of a PCA is to consider the variance function in a fPCA. Corresponding to the formulation of the eigendecomposition in equation (2.17) using Mercer's Theorem A.6, as stated in Appendix A.1, the covariance function, $v(s,t)$, given in equation (2.4) can be decomposed into:

$$v(s,t) = \sum_{j=1}^{\infty} \lambda_j \xi_j(s)\xi_j(t), \tag{2.23}$$

where it is assumed that $\int v(t,t)dt < \infty$.

According to Definition A.4 in Appendix A.1, the *covariance operator* V is introduced. It is an integral transform of the weight function $\xi(s)$ and given by:

$$[V\xi](s) = \int v(s,t)\xi(t)dt. \tag{2.24}$$

The following lemma summarizes the properties of the covariance operator.

Lemma 2.1
Let $V : \mathcal{H} \to \mathcal{H}$ be defined as in equation (2.24). Then the following hold:

1. V is compact.
2. V is positive.
3. V is self-adjoint.

[Alexanderian, 2013]

Proof. Omitted. Can be found in [Alexanderian, 2013]. □

Let V be defined as in equation (2.24) and using Lemma 2.1 together with the spectral Theorem A.7 for self-adjoint operators given in Appendix A.1, the maximization problem given in equation (2.16) can be reformulated as:

$$\begin{aligned} \int v(s,t)\xi(t)dt &= \lambda\xi(s) \\ [V\xi](s) &= \lambda\xi(s), \end{aligned} \tag{2.25}$$

for an eigenvalue λ. For complex structured curves it is not possible to solve equation (2.25) exactly, but Section 2.3.1 shows how approximations can be found. In contrast to the multivariate case, where the number of eigenvalue-eigenvector pairs is limited to p, in fPCA the number of function values, thus the number of eigenvalue-eigenfunction pairs, can be infinite. [J. O. Ramsay , B. W. Silverman, 2005]

One can also consider fPCA as finding a set of K orthogonal functions ξ in such a way that the expansion of each curve in terms of these basis functions approximates the curve as closely as possible. Since it is known from the construction of fPCA that these functions are orthonormal, it follows that the expansion is given by:

$$x_i(t) = \sum_{k=1}^{K} f_{ik}\xi_k(t), \tag{2.26}$$

where f_{ik} is the principal component value $\int \xi_k(t)x_i(t)dt$. Equation (2.26) is also known as *Karhunen-Loeve expansion* as stated in Theorem 2.2.

Theorem 2.2 (Karhunen-Loeve Expansion)
Let $x_i : D \times \Omega \to \mathbb{R}$ be a centered mean-squared continuous stochastic process, i.e.:

$$\lim_{k\to\infty} \mathrm{E}\left[\left(x_i(t) - \sum_{k=1}^{K} f_{ik}\xi_k(t)\right)^2\right] = 0.$$

Then there exists a basis expansion ξ_i for $i = 1, \dots, k$ of $\mathcal{H}$ such that for $t \in D$:

$$x_i(t) = \sum_{k=1}^{\infty} f_{ik}\xi_k(t),$$

with convergence in $\mathcal{H}$ and where $f_{ik}(\omega)$ is the principal component value $\int_D \xi_k(t)x_i(t)dt$. Then the following is satisfied:

1. $\mathrm{E}[f_{ik}] = 0$.
2. $\mathrm{E}[f_{ik}f_{ij}] = \delta_{kj}\lambda_j$.
3. $\mathrm{Var}[f_{ik}] = \lambda_j$,

where δ_{kj} is the Kronecker delta. [Alexanderian, 2013]

Proof. Let V be the Hilbert-Schmidt operator as given in equation (2.24), then V has a complete set of eigenvectors $\xi_i \in \mathcal{H}$ and non-negative, increasing eigenvalues $\lambda_i \in \mathcal{H}$.

1. The first allegation is shown by:

$$\begin{aligned}
\mathrm{E}\left[f_{ik}\right] &= \mathrm{E}\left[\int_D x_i(t)\xi_k(t)dt\right] \\
&= \int_\Omega \int_D x_i(t;\omega)\xi_k(t)dtdP(\omega) \\
&= \int_D \int_\Omega x_i(t;\omega)\xi_k(t)dP(\omega)dt \\
&= \int_D \mathrm{E}\left[x_i(t)\right]\xi_k(t)dt = 0,
\end{aligned}$$

where the last equality follows since it is assumed that $x_i(t)$ is a centered proces, i.e. $\mathrm{E}\left[x_i(t)\right] = 0$.

2. The second allegation is verified by:

$$\begin{aligned}
\mathrm{E}\left[f_{ik}f_{ij}\right] &= \mathrm{E}\left[\left(\int_D x_i(s)\xi_k(s)ds\right)\left(\int_D x_i(t)\xi_j(t)dt\right)\right] \\
&= \mathrm{E}\left[\int_D \int_D x_i(s)\xi_k(s)x_i(t)\xi_j(t)dsdt\right] \\
&= \int_D \int_D \mathrm{E}\left[x_i(s)x_i(t)\right]\xi_k(s)\xi_j(s)dsdt \\
&= \int_D \left(\int_D v(s,t)\xi_j(t)dt\right)\xi_k(s)ds \\
&= \int_D [V\xi_j](s)\xi_k(s)ds \\
&= \langle V\xi_j, \xi_k\rangle = \langle \lambda_j\xi_j, \xi_k\rangle = \lambda_j\delta_{kj},
\end{aligned}$$

where δ_{kj} is the Kronecker delta. Then 3. follows from 1. and 2.:

$$\mathrm{Var}\left[f_{ik}\right] = \mathrm{E}\left[\left(f_{ik} - \mathrm{E}\left[f_{ik}\right]\right)^2\right] = \mathrm{E}\left[f_{ik}^2\right] = \lambda_j.$$

It remains to show the convergence in $\mathcal{H}$. Therefore define:

$$\epsilon_k(t) = \mathrm{E}\left[\left(x_i(t) - \sum_{k=1}^{K} f_{ik}\xi_k(t)\right)^2\right], \tag{2.27}$$

where it has to be shown that $\lim_{k\to\infty}\epsilon_k(t) = 0$ uniformely, and hence pointwise in D. Expand equation (2.27):

$$\epsilon_k(t) = \mathrm{E}\left[x_i(t)^2\right] - 2\mathrm{E}\left[x_i(t)\sum_{k=1}^{K} f_{ik}\xi_k(t)\right] + \mathrm{E}\left[\sum_{k=1}^{K}\sum_{j=1}^{K} f_{ik}f_{ij}\xi_k(t)\xi_j(t)\right] \tag{2.28}$$

The first term in equation (2.28) can be expressed as $k(t,t)$, which follows from equation (2.23). The middle term can be rewritten as:

$$\begin{aligned}
\mathrm{E}\left[x_i(t)\sum_{k=1}^{K} f_{ik}\xi_k(t)\right] &= \mathrm{E}\left[x_i(t)\sum_{k=1}^{K}\left(\int_D x_i(s)\xi_k(s)ds\right)\xi_k(t)\right] \\
&= \sum_{k=1}^{K}\left(\int_D \mathrm{E}\left[x_i(t)x_i(s)\right]\xi_k(s)ds\right)\xi_k(t) \\
&= \sum_{k=1}^{K}\lambda_k\xi_k(t)^2 \qquad (2.29)
\end{aligned}$$

And finally the last term can be rewritten as:

$$\begin{aligned}
\mathrm{E}\left[\sum_{k=1}^{K}\sum_{j=1}^{K} f_{ik}f_{ij}\xi_k(t)\xi_j(t)\right] &= \sum_{k=1}^{K}\sum_{j=1}^{K}\lambda_k\delta_{kj}\xi_k(t)\xi_j(t) \\
&= \sum_{k=1}^{K}\lambda_k\xi_k(t)\sum_{j=1}^{K}\delta_{kj}\xi_j(t) \\
&= \sum_{k=1}^{K}\lambda_k\xi_k(t)^2. \qquad (2.30)
\end{aligned}$$

From equations (2.28), (2.29) and (2.30) it then follows:

$$\epsilon_k(t) = v(t,t) - \sum_{k=1}^{K}\lambda_k\xi_k(t)^2.$$

Using Mercer's Theorem A.6 yields:

$$\lim_{k\to\infty}\epsilon_k(t) = 0,$$

which shows the convergence in $\mathcal{H}$ and completes the proof, which is inspired by [Alexanderian, 2013]. □

The basis given in equation (2.26) is referred to as the most efficient basis possible for the selected number of eigenvalues m, which means that it has the lowest total error sum of square, SSE, for a given number of eigenvalues m:

$$\mathrm{SSE} = \sum_{i=1}^{N}\int[x_i(t) - \widehat{x}_i(t)]^2dt.$$

Suppose that the number of eigenvalue-eigenfunctions pairs is truncated to be $N-1$. Then assume that $1 \leq m \leq N-1$, where m is the number of eigenvalues, and using equation (2.26) the SSE can be rewritten as:

$$\begin{aligned}
\mathrm{SSE} &= \sum_{i=1}^{N}\int\left[\bar{x}(t) + \sum_{k=1}^{N-1} f_{ik}\xi_k(t) - \left(\bar{x}(t) + \sum_{k=1}^{m} f_{ij}\xi_k(t)\right)\right]^2 dt \\
&= \sum_{i=1}^{N}\int\left[\sum_{k=1}^{N-1} f_{ik}\xi_k(t) - \sum_{k=1}^{m} f_{ij}\xi_k(t)\right]^2 dt \\
&= \sum_{i=1}^{N}\int\left[\sum_{k=m+1}^{N-1} f_{ik}\xi_j(t)\right]^2 dt \\
&= \sum_{i=1}^{N}\left[\sum_{k=m+1}^{N-1} f_{ik}\int\xi_k(t)^2dtf_{ik} + \sum_{k=m+1}^{N-1}\sum_{l=m+1;l\neq k}^{N-1} f_{ik}\int\xi_k(t)\xi_l(t)dtf_{il}\right],
\end{aligned}$$

where the conditions from equation (2.22) are used. It then follows that:

$$= \sum_{i=1}^{N} \sum_{k=m+1}^{N-1} f_{ik}^2.$$

Using equation (2.21) this can be expressed by:

$$
\begin{aligned}
&= \sum_{i=1}^{N} \sum_{k=m+1}^{N-1} \int (\xi_k(t)x_i(t))^2 \, dt \\
&= \sum_{k=m+1}^{N-1} \sum_{i=1}^{N} \int \xi_k(t)^2 x_i(t)^2 dt \\
&= \sum_{k=m+1}^{N-1} \int \xi_k(t)^2 \sum_{i=1}^{N} x_i(t)^2 dt \\
&= \sum_{k=m+1}^{N-1} \int \xi_k(t)^2 N v(s,t) dt,
\end{aligned}
$$

where the last equation follows from equation (2.23). Hence the relationship between the SSE and the eigenvalues is given by:

$$\text{SSE} = N \sum_{k=m+1}^{N-1} \lambda_k, \tag{2.31}$$

where only the smallest eigenvalues contribute. Hence, the basis expansion found by fPCA has the lowest possible SSE for a given number of eigenvalues. The next section introduces a method to approximate the solution of the eigendecomposition.

2.3.1 Approximate Solution to Eigendecompostion

This section shows one possible way to compute approximate solutions to the functional eigendecomposition in equation (2.25). It is inspired by [J. O. Ramsay , B. W. Silverman, 2005] and [J. Peng and D. Paul, 2012].

In order to estimate the eigenvalues and eigenfunctions of the covariance function, it can be useful to truncate the series on the right side in equation (2.23) at some finite $m \geq 1$ as also mentioned above, which yields the *projected covariance function*:

$$v_{proj}^{(m)}(s,t) = \sum_{j=1}^{m} \lambda_j \xi_j(s) \xi_j(t). \tag{2.32}$$

Since $\left\| v(s,t) - v_{proj}^{(m)}(s,t) \right\| = \sum_{k=m+1}^{\infty} \lambda_k^2$, and as long as the eigenvalues decrease to zero sufficiently fast, the approximation $v_{proj}^{(m)}(s,t)$ of the covariance function results in small bias. Let Λ be a diagonal matrix containing the m eigenvalues, i.e. $\Lambda = \text{diag}\{\lambda_j\}_{j=1}^{m}$. Moreover, assume that the eigenfunctions can be modelled by basis functions:

$$\xi_j(t) = \sum_{k=1}^{K} c_{kj} \phi_k(t) \text{ for } j = 1, \ldots, m \text{ and } K \geq m \geq 1.$$

The eigenfunctions can then defined to be:

$$\boldsymbol{\xi}(t)^\top = (\xi_1(t), \dots, \xi_m(t)) = (\phi_1(t), \dots, \phi_K(t))C,$$

where the $K \times m$ matrix C satisfies:

$$C^\top \left(\int \boldsymbol{\phi}(t)\boldsymbol{\phi}(t)^\top dt \right) C = \int \boldsymbol{\xi}(t)\boldsymbol{\xi}(t)^\top dt = I_m, \tag{2.33}$$

where $\boldsymbol{\phi}(\cdot) = (\phi_1(\cdot), \dots, \phi_K(\cdot))^\top$. This means, that the matrix C lies in a *Stiefel manifold*, which is a space of real-valued matrices with orthonormal columns. The Stiefel manifold $K \times m$ matrices are defined by $\mathcal{S}_{K,m} = \{A \in \mathbb{R}^{K \times m} : A^\top A = I_m\}$.

In order to obtain an approximate solution to the eigendecomposition one converts the continuous functional eigendecomposition to a discrete expression by using the idea of the projected covariance function as given in equation (2.32). Assume that the functions $x_i(t)$ are represented by basis functions as given in equation (2.5), then the covariance function can be expressed by:

$$v(s,t) = N^{-1}\boldsymbol{\phi}(s)^\top C^\top C \boldsymbol{\phi}(t).$$

Define a rank K symmetric matrix $W = \int \boldsymbol{\phi}(t)\boldsymbol{\phi}^\top(t)dt$ and assume that an eigenfunction $\xi(t)$ for equation (2.25) has the basis expansion:

$$\xi(t) = \boldsymbol{\phi}(t)^\top \boldsymbol{c},$$

where $\boldsymbol{c}$ is a coefficient vector. Then the following must apply:

$$\begin{aligned} \int v(s,t)\xi(t)dt &= \int N^{-1}\boldsymbol{\phi}(s)^\top C^\top C \boldsymbol{\phi}(t)\boldsymbol{\phi}(t)^\top \boldsymbol{c} dt \\ &= N^{-1}\boldsymbol{\phi}(s)^\top C^\top C W \boldsymbol{c}. \end{aligned}$$

This implies that the eigenanalysis in equation (2.25) can be expressed by:

$$N^{-1}\boldsymbol{\phi}(s)^\top C^\top C W \boldsymbol{c} = \lambda \boldsymbol{\phi}(s)^\top \boldsymbol{c},$$

which can be rewritten as:

$$N^{-1} C^\top C W \boldsymbol{c} = \lambda \boldsymbol{c},$$

since the eigenanalysis has to hold for all s. Note that the eigenfunctions have to satisfy the normalization constraint $\int \xi(t)^2 dt = 1$ in a fPCA, which implies that:

$$1 = \int \xi(t)^2 dt = \int \boldsymbol{c}^\top \boldsymbol{\phi}(t)\boldsymbol{\phi}(t)^\top \boldsymbol{c} dt = \boldsymbol{c}^\top W \boldsymbol{c}.$$

Additionally, in order to ensure orthogonality of two eigenfunctions $\xi_l(t)$ and $\xi_k(t)$ the coefficient vectors must satisfy $\boldsymbol{c}_k^\top W \boldsymbol{c}_l = 0$.

In order to find principal components, define the eigenvector $\boldsymbol{u} = W^{1/2}\boldsymbol{c}$, which implies that $\boldsymbol{u}^\top \boldsymbol{u} = 1$ in accordance with the normalization constraint in the fPCA.

Then solve the symmetric eigenvalue problem:

$$N^{-1}W^{1/2}C^{\top}CW^{1/2}\boldsymbol{u} = \lambda\boldsymbol{u},$$

which is solved for λ and $\boldsymbol{u}$. The coefficient vector $\boldsymbol{c}$ is then computed by $\boldsymbol{c} = W^{-1/2}\boldsymbol{u}$ and the eigenfunction is given by $\xi(t) = \boldsymbol{\phi}(t)^{\top}W^{-1/2}\boldsymbol{u}$.

The next section considers different possiblities to select the number of principal components.

2.3.2 The Choice of the Number of Principal Components

When working with PCA or fPCA it is known that by construction the principal components are ordered such to describe descending variation in data. Therefore when dealing with high dimensional data it might be appropriate to only consider the first m principal components and thereby ignore the effect of the last $p - m$ components as explained in the previous sections.

In litterature there are studied several methods to determine the number of principal components m. This section is based on the discussion in [Cangelosi and Goriely, 2007] and [K. V. Mardia, J.T. Kent, J. M. Bibby, 1994] on different methods to select the number of principal components. It is a very important parameter, since including too many components to describe data may involve noise, while including too few components may remove valuable information. It can be distinguished between heuristic methods, where there are used subjective criteriums like how much of the variance in data should be explained by the principal components, and statistical methods. This section aims to give a short overview of some heuristic methods, since it is argued in litterature that statistical methods often are computational heavy, overestimate the number of principal components, and some methods require distributional assumptions of data. [Cangelosi and Goriely, 2007]

A possible criterium can be stated such that components whose standard deviations are less than or equal to some predetermined threshold v times the standard deviation of the first component are omitted. That is, the included components satisfy:

$$\sqrt{\lambda_j} \geq v \cdot \sqrt{\lambda_1} \quad j = 1, \ldots, m, \tag{2.34}$$

where $0 \leq v \leq 1$ and $m \leq p$ indicates the number of selected principal components. This criterium is also implemented in the R function `prcomp`, which will be used to perform a PCA in pratice.

Another way to determine the number of principal components is by examining a scree plot, which visualizes the proportion of total variance each factor is accounting for. Hence it is a plot of the eigenvalues λ_j against their indicies j. This means that when there for some value j occurs a sharpe change in the slope of the scree plot then this is chosen to be the number of components m to retain. From equation (2.31) it can be seen that this might imply a small SSE, but the difficulty with this method is that it is not ensured that there will be a clear change or there might be several changes, which makes it difficult to interpret the plot.

Another possibility is to consider the cumulative variance explained by the first j principal components, which is given by:

$$\mathrm{CPV}_j = \frac{\sum_{i=1}^{j} \lambda_i}{\sum_{i=1}^{p} \lambda_i}, \tag{2.35}$$

where $\mathrm{CPV}_1 < \cdots < \mathrm{CPV}_p = 1$, which follows from the property of the eigenvalues. Then there can be selected a certain value of how much of the variance in data that should be explained by the first m principal components, e.g. the components should explain about $70-90\%$ of the variance. As for the other introduced heuristic methods, the drawback of this method is that it is a subjective choice to select how much of the variance that should be explained.

When omitting some principal components, the hypothesis of equal eigenvalues should be checked, i.e. if $\lambda_p = \lambda_{p-1} = \cdots = \lambda_{m+1}$. In general, eigenvalues that are equal to each other should be treated in the same way. This means either all components are omitted or all components are retained. Therefore the standard deviation criterium in equation (2.34) will be used throughout this thesis, since it, in contrast to the scree plot and the cumulative variance explained, incorporates this property.

After having introduced the concept of functional data analysis, principal component analysis, and functional principal component analysis, the next chapter introduces different asset allocation strategies that make use of these methods.

Risk-based Allocation Strategies 3

One of the most used asset allocation strategies in the financial industry is the MV strategy, which attempts to trade off risk and return. However, the method is not very accurate due to estimation errors in the mean estimate. Therefore, this chapter introduces risk-based allocation strategies, that do not have to estimate the expected value, but only the variance, which often is a much more robust estimate than the mean estimate. Furthermore these strategies are more efficient in diversifying a portfolio than the Markowitz approach, since they allocate due to the underlying risk factors of a portfolio instead of only focusing on the overall risk of a portfolio.

In addition to introducing different variants of the Equal Risk, ER, strategy another risk-based allocation strategy will be presented, the Diversified Risk Parity, DRP, strategy. It will be shown that the DRP strategy is related to one of the ER approaches. At the end of this chapter a simplified example shows how the different strategies introduced in this chapter are implemented using the software R and in addition there is considered a simulation study to investigate the similarities and differences of the portfolio weights found by the different strategies.

3.1 Equal Risk Portfolio Optimization

This optimization method can be considered from different points of view:

- A strategy to find a risk-balanced portfolio such that the risk contributions are the same for all assets of a portfolio.
- An optimization, where the weights of the assets are chosen such that the PCA yields equal volatilities in the components.

sam The different approaches are introduced in the following sections. Moreover, there is investigated how the strategy relates to other asset allocation strategies, e.g. EW and the MVa strategies. Also a functional variant of the ER strategy is introduced which uses fPCA instead of PCA.

3.1.1 Risk Contribution Approach

This section is inspired by [S. Maillard, T. Roncalli, J. Teiletche, 2009]. As mentioned above, the idea of an ER portfolio is to find a portfolio in such a way that the risk contributions $\mathrm{RC}_i(\boldsymbol{w})$, as introduced in Section 1.1, are the same for all assets i included in a portfolio. For simplicity it is assumed that short-selling is not allowed, which means that $0 \leq w_i \leq 1$ for $i = 1, \ldots, N$. Then the optimal weights of an ER portfolio can mathematically be described by:

$$\boldsymbol{w}^* = \left\{ \boldsymbol{w} \in [0,1]^N : \sum_{i=1}^{N} w_i = 1; \mathrm{RC}_i(\boldsymbol{w}) = \mathrm{RC}_j(\boldsymbol{w}) \text{ for all } i, j \right\}.$$

In order to get a better understanding of this method, consider the following simple example with a portfolio consisting of two assets.

Example 2 (Two asset case)
A portfolio containing two assets has the weights $\boldsymbol{w} = (w, 1-w)$. It follows from equation (1.10), that the vector of risk contributions is then given by:

$$\begin{pmatrix} \mathrm{RC}_1(w) \\ \mathrm{RC}_2(1-w) \end{pmatrix} = \frac{1}{\sigma_P(\boldsymbol{w})} \begin{pmatrix} w^2\sigma_1^2 + w(1-w)\rho\sigma_1\sigma_2 \\ (1-w)^2\sigma_2^2 + w(1-w)\rho\sigma_1\sigma_2 \end{pmatrix}.$$

It is used that the covariance can be expressed by $\sigma_{12} = \rho\sigma_1\sigma_2$.

The aim is then to find a weight w such that both rows of the risk contribution vector are equal. This means, find a weight w such that $w^2\sigma_1^2 = (1-w)^2\sigma_2^2$ and $0 \leq w \leq 1$:

$$\begin{aligned} w^2\sigma_1^2 &= (1-w)^2\sigma_2^2 \\ w &= \frac{\sigma_2}{\sigma_1 + \sigma_2}, \end{aligned}$$

which yields:

$$\boldsymbol{w}^* = \left(\tfrac{\sigma_1}{\sigma_1+\sigma_2}, \tfrac{\sigma_2}{\sigma_1+\sigma_2} \right).$$

Notice that the optimal solution $\boldsymbol{w}^*$ not depends on the correlation of the assets, but only on their variances. △

This can be generalized to a portfolio containing N assets. There are three different cases to be considered depending on the volatility and correlation structure of the assets in a portfolio.

Different Volatilities, Equal Correlation

The first case is to assume equal correlation across the assets, i.e. $\rho_{ij} = \rho$ for all i, j, where ρ_{ij} denotes the correlation between the return of asset i and the return of asset j. Then the risk contributions are given by:

$$\begin{aligned}
\mathrm{RC}_i(\boldsymbol{w}) &= \frac{w_i^2\sigma_i^2 + \sum_{j\neq i} w_i w_j \rho \sigma_i \sigma_j}{\sigma_P(\boldsymbol{w})} \\
&= \frac{w_i\sigma_i(w_i\sigma_i + \rho\sum_{j\neq i} w_j\sigma_j)}{\sigma_P(\boldsymbol{w})} \\
&= \frac{w_i\sigma_i(w_i\sigma_i - \rho w_i\sigma_i + \rho\sum_{j=1}^N w_j\sigma_j)}{\sigma_P(\boldsymbol{w})} \\
&= \frac{w_i\sigma_i\left((1-\rho)w_i\sigma_i + \rho\sum_{j=1}^N w_j\sigma_j\right)}{\sigma_P(\boldsymbol{w})}.
\end{aligned}$$

From the ER portfolio setup it is known that $\mathrm{RC}_i(\boldsymbol{w}) = \mathrm{RC}_j(\boldsymbol{w})$ and it then follows that $w_i\sigma_i = w_j\sigma_j$. Therefore, together with the constraint $\sum_{i=1}^N w_i = 1$ it is given that:

$$1 = \sum_{j=1}^N w_j = \sum_{j=1}^N \frac{w_i\sigma_i}{\sigma_j} = w_i\sigma_i \sum_{j=1}^N \sigma_j^{-1},$$

which can be rewritten as:

$$\frac{1}{w_i} = \sigma_i \sum_{j=1}^N \sigma_j^{-1}.$$

So the optimal weights for assets with equal correlations are given by:

$$w_i^* = \frac{\sigma_i^{-1}}{\sum_{j=1}^N \sigma_j^{-1}}. \tag{3.1}$$

So the higher the volatility of a component, the lower is the weight in the ER portfolio. Note, that when the assets have equal volatilities, i.e. $\sigma_i = \sigma_j$, then the optimal weights would simply be $w_i^* = N^{-1}$. It will also be shown in Section 3.1.2 that when the ER strategy uses the lower bound of the correlation coefficent ρ, cf. equation (1.17), then it finds similar portfolio weights to the MVa strategy.

Equal Volatilities, Different Correlations

The second case is when correlation across assets differ, but the volatilities are the same, i.e. $\sigma_i = \sigma$ for all i. Then it follows by the same reasoning as above, that the optimal weigths are given by:

$$w_i^* = \frac{\left(\sum_{k=1}^N w_k\rho_{ik}\right)^{-1}}{\sum_{j=1}^N \left(\sum_{k=1}^N w_k\rho_{jk}\right)^{-1}}.$$

Again it can be noted, that assuming equal correlations would yield equal weights, i.e. $w_i^* = N^{-1}$.

Different Volatilites and Correlations

The most general case is when both the volatilites and the correlations of the assets differ. Consider the covariances and risk contributions, respectively:

$$\sigma_{iw} = \mathrm{Cov}\left[r_i, \sum_{j=1}^{N} w_j r_j\right] = \sum_{j=1}^{N} w_j \mathrm{Cov}\left[r_i, r_j\right] = \sum_{j=1}^{N} w_j \sigma_{ij};$$
$$\mathrm{RC}_i(\boldsymbol{w}) = w_i \frac{\sigma_{iw}}{\sigma_P(\boldsymbol{w})}$$

Then from the definition of the $\beta_P(\boldsymbol{w})$ of a portfolio, as stated in equation (1.13), it follows that $\beta_i = \sigma_{iw}/\sigma_P^2(\boldsymbol{w})$. Thereby the risk contributions can be rewritten as $\mathrm{RC}_i(\boldsymbol{w}) = w_i \beta_i \sigma_P(\boldsymbol{w})$. From the ER approach it follows that $\mathrm{RC}_i(\boldsymbol{w}) = \mathrm{RC}_j(\boldsymbol{w})$. Together with the constraint $\sum_{i=1}^{N} w_i = 1$ it follows that the weights are given by:

$$w_i^* = \frac{\beta_i^{-1}}{\sum_{j=1}^{N} \beta_j^{-1}}.$$

Thus, the higher the β an asset, the lower the weight, which implies that assets with high volatility or correlation with other assets will be assigned a lower portfolio weight.

It can be seen that in the last two cases the optimal solution w_i^* is endogenous, since it dependents on itself. This implies that there is no closed-form solution, but the solution has to be found numerically.

One approach is to use *Sequential Quadratic Programming*, SQP, to solve the following optimization problem:

$$\begin{aligned} \boldsymbol{w}^* &= \mathrm{argmin}\, f(\boldsymbol{w}) \\ \text{subject to } \mathbf{1}^\top \boldsymbol{w} &= 1 \text{ and } \mathbf{0} \leq \boldsymbol{w} \leq \mathbf{1}, \end{aligned} \tag{3.2}$$

where:

$$f(\boldsymbol{w}) = \sum_{i=1}^{N} \sum_{j=1}^{N} \left(w_i (\Sigma \boldsymbol{w})_i - w_j (\Sigma \boldsymbol{w})_j\right)^2.$$

The algorithm minimizes the numerator of the risk contributions, i.e. it solves $w_i(\Sigma \boldsymbol{w})_i = w_j(\Sigma \boldsymbol{w})_j$.

In order to summarize this section it is interesting to compare the ER strategy with other asset allocation strategies, mainly the EW and MVa strategies. Mathematically the three strategies make use of the following:

Equally-Weighted (EW):	$w_i = w_j$
Minimum Variance (MVa):	$\partial_{w_i} \sigma_P(\boldsymbol{w}) = \partial_{w_j} \sigma_P(\boldsymbol{w})$
Equal Risk (ER):	$\mathrm{RC}_i(\boldsymbol{w}) = \mathrm{RC}_j(\boldsymbol{w})$.

The next section explains how the three strategies are related.

3.1.2 Properties of the Equal Risk Strategy

This section is based on the work of [S. Maillard, T. Roncalli, J. Teiletche, 2009] and [Stefanovits, 2010] and aims to show that there is a relationship between the ER, EW, and MVa strategies. First it can be shown that the ER strategy is similar to the MVa portfolio for a special structure of the correlation matrix.

Connection to the Minimum Variance Strategy

Let R be a constant correlation matrix such that $R_{ij} = \rho$ and $R_{ii} = 1$, i.e.:

$$R = \rho \mathbf{1}\mathbf{1}^\top + (1-\rho) I_N,$$

where I_N is the $N \times N$ identity matrix. The covariance matrix can then be expressed by $\Sigma = \boldsymbol{\sigma}\boldsymbol{\sigma}^\top R$ and consequently the inverse covariance is given by $\Sigma^{-1} = \Gamma R^{-1}$, where $\Gamma_{ij} = \frac{1}{\sigma_j \sigma_j}$ and:

$$R^{-1} = \frac{\rho \mathbf{1}\mathbf{1}^\top - ((N-1)\rho + 1) I_N}{(N-1)\rho^2 - (N-2)\rho - 1}.$$

The structure of R^{-1} is veryfied by the following calculation:

$$\begin{aligned}
RR^{-1} &= \rho \mathbf{1}\mathbf{1}^\top + (1-\rho) I_N \frac{\rho \mathbf{1}\mathbf{1}^\top - ((N-1)\rho + 1) I_N}{(N-1)\rho^2 - (N-2)\rho - 1} \\
&= \frac{\rho^2 N \mathbf{1}\mathbf{1}^\top - \rho \mathbf{1}\mathbf{1}^\top ((N-1)\rho + 1) + \rho \mathbf{1}\mathbf{1}^\top (1-\rho) - (1-\rho)((N-1)\rho + 1) I_N}{(N-1)\rho^2 - (N-2)\rho - 1} \\
&= \frac{-N\rho I_N - N\rho^2 I_N - 2\rho I_N + \rho^2 I_N - I_N}{(N-1)\rho^2 - (N-2)\rho - 1} \\
&= I_N
\end{aligned}$$

The weights for the MVa strategy are given by:

$$\boldsymbol{w}^*_{MVa} = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}^\top \Sigma^{-1} \mathbf{1}}, \tag{3.3}$$

hence using the constant correlation matrix the weights of the MVa strategy can be rewritten as:

$$w_i^* = \frac{\rho \sum_{k=1}^N (\sigma_i \sigma_k)^{-1} - ((N-1)\rho + 1)\sigma_i^{-2}}{\sum_{k=1}^N \left(\rho \sum_{j=1}^N (\sigma_k \sigma_j)^{-1} - ((N-1)\rho + 1)\sigma_k^{-2} \right)}.$$

Consider the lower bound of the correlation coefficent $\rho = -(N-1)^{-1}$ as shown in equation (1.17), then the MVa weights are:

$$w_i^* = \frac{\sum_{k=1}^N (\sigma_i \sigma_k)^{-1}}{\sum_{j=1}^N \sum_{k=1}^N (\sigma_j \sigma_k)^{-1}} = \frac{\sigma_i^{-1}}{\sum_{j=1}^N \sigma_j^{-1}},$$

which is exactly the solution of the ER strategy with equal correlations and different volatilities as given in equation (3.1). Hence the MVa and ER strategies coincide when having a portfolio consisting of assets that have the same lower bound correlation.

The next section shows that there is a relationship between the volatilities of the ER, EW, and MVa strategies.

The Relationship between the Volatilities of the ER, EW, and MVa Strategies

Before showing the relationship of the volatilities, first it is shown that the ER optimization can be formulated in another way then given in equation (3.2). Consider the following optimization problem:

$$\begin{aligned} \boldsymbol{w}^*(c) &= \operatorname{argmin} \sqrt{\boldsymbol{w}^\top \Sigma \boldsymbol{w}} \\ \text{subject to } &\sum_{i=1}^{N} \ln(w_i) \geq c \;\text{ and } \boldsymbol{w} \geq \mathbf{0}, \end{aligned} \tag{3.4}$$

for any constant $c \in \mathbb{R}$. This optimization problem aims to minimize the volatility of a portfolio subject to the constraint $\sum_{i=1}^{N} \ln(w_i) \geq c$. The Lagrangian for this optimization problem is given by:

$$L = \sqrt{\boldsymbol{w}^\top \Sigma \boldsymbol{w}} - \gamma_1 \left(\sum_{i=1}^{N} \ln(w_i) - c \right) - \boldsymbol{\gamma}_2^\top \boldsymbol{w},$$

where γ_1 and $\boldsymbol{\gamma}_2$ are Lagrange multipliers. Suppose that there exists portfolio weights $\boldsymbol{w}^*$ that are a solution for this optimization problem for an arbitrary constant c, and that the Lagrange multipliers satisfy the first order condition:

$$\nabla L = \frac{\Sigma \boldsymbol{w}^*}{\sqrt{\boldsymbol{w}^{*\top} \Sigma \boldsymbol{w}^*}} - \gamma_1^* \left(\frac{1}{w_1^*}, \ldots, \frac{1}{w_N^*} \right) - \boldsymbol{\gamma}_2^* = \mathbf{0}. \tag{3.5}$$

Moreover the following Kuhn-Tucker conditions are satisfied:

$$\gamma_1^*, (\boldsymbol{\gamma}_2^*)_i \geq 0 \qquad \text{for } i = 1, \ldots, N,$$

$$(\boldsymbol{\gamma}_2^*)_i w_i^* = 0, \tag{3.6}$$

$$\gamma_1^* \left(\sum_{i=1}^{N} \ln(w_i^*) - c \right) = 0. \tag{3.7}$$

Because $\ln(w_i^*)$ is not defined for $w_i^* = 0$ it follows from equation (3.7) that $w_i^* > 0$, hence equation (3.6) implies that $(\boldsymbol{\gamma}_2^*)_i = 0$. Furthermore, assuming that $\gamma_1^* = 0$ must imply that $\Sigma \boldsymbol{w}^* = \mathbf{0}$ from equation (3.5), which leads to $\boldsymbol{w}^* = \mathbf{0}$ which is not allowed. Therefore the second Lagrange multiplier γ_1^* must be greater than zero. These parameter conditions and equation (3.5) imply that the optimal portfolio weights $\boldsymbol{w}^*$ satisfy:

$$\begin{aligned} &\frac{\Sigma \boldsymbol{w}^*}{\sqrt{\boldsymbol{w}^{*\top} \Sigma \boldsymbol{w}^*}} - \gamma_1^* \left(\frac{1}{w_1^*}, \ldots, \frac{1}{w_N^*} \right) = \mathbf{0} \Leftrightarrow \\ &\boldsymbol{w}_i^* \partial_{w_i} \sigma_P(\boldsymbol{w}) = \gamma_1^* \Leftrightarrow \\ &\mathrm{RC}_i(\boldsymbol{w}^*) = \gamma_1^*. \end{aligned}$$

This shows that the optimization problem in equation (3.4) is a ER strategy, since it finds equal risk contributions $\mathrm{RC}_i(\boldsymbol{w}^*)$.

The portfolio weights also have to satisfy $\mathbf{1}^\top \boldsymbol{w}^* = 1$, therefore the portfolio weights are normalized:

$$w_i^* = \frac{w_i^*(c)}{\sum_{i=1}^N w_i^*(c)}. \tag{3.8}$$

Now in order to show the relationship of the volatilities of the MVa, ER, and EW strategies suppose that $c_1 \leq c_2$ then $\sigma_P(\boldsymbol{w}^*(c_1)) \leq \sigma_P(\boldsymbol{w}^*(c_2))$, since $\sum_{i=1}^N \ln(w_i) - c_1 \geq 0$ is less restrictive than $\sum_{i=1}^N \ln(w_i) - c_2 \geq 0$.

In the case of $c = -\infty$ the optimization problem in equation (3.4) yields the MVa portfolio since the inequality constraint disappears. On the other hand, when $c = -N\ln(N)$ then the optimization problem is the EW portfolio, which follows from Jensen's inequality and the constraint $\mathbf{1}^\top \boldsymbol{w} = 1$:

$$\frac{1}{N}\sum_{i=1}^N \ln(w_i) \leq \ln\left(\frac{1}{N}\sum_{i=1}^N w_i\right) = -\ln(N),$$

which implies that $\sum_{i=1}^N \ln(w_i) \leq -N\ln(N)$, hence $\sigma_P(\boldsymbol{w}^*(-\infty)) \leq \sigma_P(\boldsymbol{w}^*(-N\ln(N)))$. For a general $c \in [-\infty, -N\ln(N)]$ it therefore must hold that:

$$\begin{aligned} \sigma_P(\boldsymbol{w}^*(-\infty)) \leq& \sigma_P(\boldsymbol{w}^*(c)) \leq \sigma_P(\boldsymbol{w}^*(-N\ln(N))) \\ \sigma_{\text{MVa}}(\boldsymbol{w}) \leq& \sigma_{\text{ER}}(\boldsymbol{w}) \leq \sigma_{\text{EW}}(\boldsymbol{w}). \end{aligned} \tag{3.9}$$

This means that the MVa strategy is the less volatile strategy, the EW strategy is more volatile, and the ER strategy lies between both. Note that the portfolio variances for the MVa and the EW strategy are given by:

$$\sigma^2_{\text{MVa}}(\boldsymbol{w}) = \frac{1}{\mathbf{1}^\top \Sigma \mathbf{1}} \quad \text{and} \quad \sigma^2_{\text{EW}}(\boldsymbol{w}) = N^{-1}\bar{\sigma}^2_\bullet(1-\rho) + \rho\bar{\sigma}_\bullet, \tag{3.10}$$

where the expression for $\sigma^2_{\text{MVa}}(\boldsymbol{w})$ follows from equations (1.5) and (3.3), and $\sigma^2_{\text{EW}}(\boldsymbol{w})$ is derived in Example 1. There is no closed-form solution for the portfolio variance $\sigma^2_{\text{ER}}(\boldsymbol{w})$, since it depends on the variance and correlation structure of the assets in a portfolio, as discussed in Section 3.1.1. But now it is known that it is bounded below and above.

The following example considers again the case of two assets and shows the relationship of the portfolio weights for the three strategies.

Example 3 (Two asset case - continued)
Example 2 showed that the weights for the ER strategy are given by:

$$\boldsymbol{w}_{ER}^{*} = \left(\frac{\sigma_1}{\sigma_1+\sigma_2}, \frac{\sigma_2}{\sigma_1+\sigma_2}\right).$$

It is known that the portfolio weights for the MVa strategy minimize the portfolio variance, which is given by:

$$\begin{aligned}\sigma_P^2(\boldsymbol{w}) &= (w, 1-w)\begin{bmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{bmatrix}\begin{pmatrix} w\\ 1-w\end{pmatrix}\\ &= \sigma_1 w^2 + 2\rho\sigma_1\sigma_2 w(1-w) + \sigma_2^2(1-w)^2.\end{aligned}$$

So in order to find the the weights that minimize the portfolio variance $\sigma_P^2(\boldsymbol{w})$ the first order derivative with respect to w is determined:

$$\begin{aligned}\frac{d}{dw}\sigma_P^2(\boldsymbol{w}) = 2w(\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2) + 2\rho\sigma_1\sigma_2 - 2\sigma_2^2 \Leftrightarrow\\ w = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}.\end{aligned}$$

In order to verify that these weights find the minimum variance, the second order derivative is determined:

$$\frac{d^2}{dw^2}\sigma_P^2(\boldsymbol{w}) = 2(\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2).$$

Since $\rho \leq 1$ the second order derivative can be rewritten as:

$$2(\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2) \geq 2(\sigma_1^2 - 2\sigma_1\sigma_2 + \sigma_2^2) = 2(\sigma_1 - \sigma_2)^2 \geq 0,$$

which verifies that w minimizes $\sigma_P^2(\boldsymbol{w})$, hence the portfolio weights for the MVa strategy are given by:

$$\boldsymbol{w}_{MVa}^{*} = \frac{1}{\sigma_1 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}\left(\sigma_2^2 - \rho\sigma_1\sigma_2, \sigma_1^2 - \rho\sigma_1\sigma_2\right).$$

For the EW strategy the weights simply are:

$$\boldsymbol{w}_{EW}^{*} = \left(\frac{1}{2}, \frac{1}{2}\right).$$

It can be seen that when $\sigma_1 = \sigma_2$ the weights of the three strategies are the same:

$$\boldsymbol{w}_{MVa}^{*} = \boldsymbol{w}_{ER}^{*} = \boldsymbol{w}_{EW}^{*}.$$

This statement is also shown in Figure 3.1. Furthermore, when $\rho = -1$ then $\boldsymbol{w}_{MVa}^{*} = \boldsymbol{w}_{ER}^{*}$, which corresponds exactly to the case of the lowest possible correlation coefficent as shown in the start of this section. $\triangle$

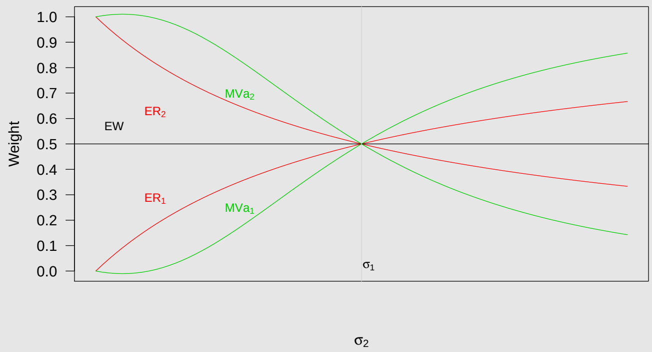


Figure 3.1. The portfolio weights in the two asset case as a function of σ_2. The labels ER_1 and ER_2 correspond to the two weights found by the ER strategy, MVa_1 and MVa_2 correspond to the two weights found by the MVa strategy, and EW indicates the weights found by the EW strategy. As shown in Example 3 the weights of the three strategies are the same when $\sigma_1 = \sigma_2$. In the MVa case the correlation is set to $\rho = 0.2$. This figure is inspired by the work of [Stefanovits, 2010].

From this section it can be learnt that the ER strategy is a tradeoff between the MVa and EW strategies in terms of the portfolio variance $\sigma_P^2(\boldsymbol{w}^*)$, i.e. the risk associated with using the strategy. The ER and MVa strategies are similar when there are equal correlations between the assets of a portfolio. Earlier in this section it has also been shown that the ER and EW strategies coincide when both the correlations and volatilities are equal for all assets of a portfolio.

The next section considers the other ER approaches using PCA and fPCA to find the portfolio weights.

3.1.3 Principal Component Analysis Approach

Principal component analysis as described in Sections 2.2 and 2.3 is a useful statistical tool in many mathematical contexts and it turns out that this is also the case in connection with asset allocation strategies. This section describes how PCA is used to find the weights of an ER portfolio, which is inspired by the work of [Kind, 2013] and [Meucci, 2010].

PCA considers an eigendecomposition of a square matrix, which in financial context either is a covariance matrix or a correlation matrix of asset returns. [Guan, 2013] When modelling asset returns using factor models as given in equation (1.18), it is assumed that the factors may be correlated. Using PCA in portfolio optimization the covariance matrix can be decomposed into lower dimensional, latent factors, which represent independent, i.e. uncorrelated, risk factors. Note that PCA finds a linear combination of the portfolio returns as stated in equation (2.13), i.e. the factors f_{ji} can be expressed as a linear combination of asset returns:

$$f_{1i} = \boldsymbol{\xi}_1^\top \boldsymbol{r}_i, \quad \ldots, \quad f_{pi} = \boldsymbol{\xi}_p^\top \boldsymbol{r}_i, \qquad \text{for } i = 1, \ldots, n, \tag{3.11}$$

where $\boldsymbol{\xi}_j$ is the jth eigenvector, or the so-called *loading*, and $\boldsymbol{r}_i$ is an observed vector of asset returns $(r_{i1}, \ldots, r_{ip})^\top$. Note that if no principal components are omitted, there are as many principal components as there are assets in a portfolio, i.e. $p = N$ where N denotes the number of assets in a portfolio. This is in contrast to factor analysis, where it is assumed that the portfolio is a linear combination of underlying factors. [Seber, 2004] As mentioned in Chapter 1 the factors found by using PCA are unobservable factors, which means that their economical interpretation is not straightforward. Nevertheless, as can be seen in equation (3.11) it is possible to investigate the eigenvectors and thereby determine which asset returns that form the factors. When the original assets are very low correlated it may be possible to deduce the economical interpretation of a factor. [H. Lohre, H. Opfer, G. Orszag, 2013] Alternative, it is possible to use orthogonal rotation techniques to rotate the PCA space into a space that makes it easier to interpret the loadings found by a PCA. A possible technique is *VARIMAX*, which both can be used for multivariate and functional data, which aims to maximize the sum of the squared loadings. Note that the rotated component scores are not uncorrelated anymore, but orthogonality is preserved. The aim of the VARIMAX rotation is to relate each variable to as few as possible factors and thereby ease the interpretation of these factors. But this is beyond the scope of this thesis. [J. O. Ramsay , B. W. Silverman, 2005]

As already stated in equation (2.17) the eigendecomposition of the positive semi-definite covariance matrix is given by:

$$V = U^\top \Lambda U,$$

where Λ is a diagonal matrix consisting of V's eigenvalues λ_j, for $j = 1, \ldots, p$, which are organized in descending order, and the columns in U represent V's loadings $\boldsymbol{\xi}_j$. The loadings $\boldsymbol{\xi}_j$ represent the weight of an asset towards each principal component portfolio, which define a set of p uncorrelated principal portfolios with variances λ_j. The returns and weights in the PCA space, the principal portfolio, can be expressed by:

$$\tilde{\boldsymbol{r}} = U^\top \boldsymbol{r} \text{ and } \tilde{\boldsymbol{w}} = U^\top \boldsymbol{w}, \tag{3.12}$$

where $\boldsymbol{r}$ are asset returns, or log returns, and $\boldsymbol{w}$ are the weights of the original assets.

Since the principal portfolios are uncorrelated, the variance of the principal portfolio is simply given by:

$$\sigma_P^2(\boldsymbol{w}) = \boldsymbol{w}^\top V \boldsymbol{w} = (U\tilde{\boldsymbol{w}})^\top V U \tilde{\boldsymbol{w}} = \tilde{\boldsymbol{w}}^\top U^\top V U \tilde{\boldsymbol{w}} = \tilde{\boldsymbol{w}}^\top \Lambda \tilde{\boldsymbol{w}} = \sum_{i=1}^{m} \tilde{w}_i^2 \lambda_i = \sigma_P^2(\tilde{\boldsymbol{w}}).$$

Hence the portfolio variance $\sigma_P^2(\boldsymbol{w})$ can either be expressed by the quadratic form in the original asset space $\boldsymbol{w}^\top V \boldsymbol{w}$ or in the PCA space $\tilde{\boldsymbol{w}}^\top \Lambda \tilde{\boldsymbol{w}}$. The total number of principal portfolios equals the number of considered principal components m. The marginal risk contributions in the PCA space are given by:

$$\partial_{\tilde{w}_i} \sigma_P(\tilde{\boldsymbol{w}}) = \frac{1}{2\sqrt{\sum_{i=1}^{m} \tilde{w}_i^2 \lambda_i}} 2\tilde{w}_i \lambda_i = \frac{\tilde{w}_i \lambda_i}{\sigma_P(\tilde{\boldsymbol{w}})}.$$

Moreover, since in the PCA space the covariance matrix has covariances that equal zero, the risk contributions $\tilde{\mathrm{RC}}_i$ in the principal space must be equal to:

$$\tilde{\mathrm{RC}}_i = \frac{\tilde{w}_i^2 \lambda_i}{\sqrt{\sum_{i=1}^{m} \tilde{w}_i^2 \lambda_i}}$$

Hence a similar optimization problem to Section 3.1 can be formulated:

$$\begin{gathered} \boldsymbol{w}^* = \operatorname{argmin} f(\boldsymbol{w}) \\ \text{subject to } \mathbf{1}^\top \boldsymbol{w} = 1 \text{ and } \mathbf{0} \leq \boldsymbol{w} \leq \mathbf{1}, \end{gathered} \tag{3.13}$$

where

$$f(\boldsymbol{w}) = \sum_{i=1}^{m} \sum_{j=1}^{m} \left(\tilde{\mathrm{RC}}_i - \tilde{\mathrm{RC}}_j \right)^2.$$

The zero covariances in the PCA space also imply that the weights can be expressed by a closed-form solution similar to equation (3.1). Thus the optimal weights in the PCA space are given by:

$$\tilde{w}_i^* = \frac{(\sqrt{\lambda_i})^{-1}}{\sum_{i=1}^{m} (\sqrt{\lambda_i})^{-1}}. \tag{3.14}$$

Another way of formulating the ER approach is inspired by the work of Jyske Bank, where the principal portfolios found are used to construct a diversified portfolio in such a way that the weights are chosen to yield as equal as possible volatilities in each component. So the aim of the optimization is to have equal risk in all directions of risk. The risk in a direction is defined as the amount of risk of the overall risk that is found in the given direction, where risk is proportional to the variance, i.e. are given by the eigenvalues in each direction. In order to hold a well-diversified portfolio its overall risk should be evenly spread across the principal portfolios. Jyske Bank has established the following objective function:

$$\min_{\boldsymbol{w}} \sum_{i=1}^{m} \mathrm{CPV}_i \quad \text{subject to } \mathbf{1}^\top \boldsymbol{w} = 1 \text{ and } -\mathbf{1} \leq \boldsymbol{w} \leq \mathbf{1}, \tag{3.15}$$

where CPV_i is the ith cumulative proportion of variance from the PCA as defined in equation (2.35). The objective function is the sum of every principal component's cumulative proportion of variance. This means by minimizing equation (3.15), the proportion of variances are made as equal as possible, which corresponds to have equal risk in each component. An example of the cumulative proportion of each component can be seen in Figure 3.2.

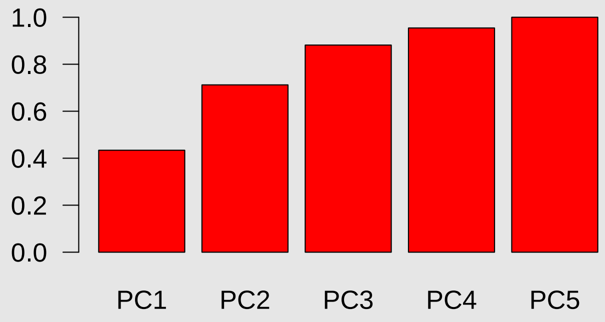


Figure 3.2. Cumulative proportion of variance of a PCA on five assets.

It can be observed that the plot by construction is a convex function that always equals one in the last component, unless some of the principal components are omitted. Thereby, the sum of cumulative proportions can be interpreted as a discrete integral, where the difference between neighbouring bars describes the variance of the component. The minimization problem formulated in equation (3.15) aims to minimize the proportion of the first component and to make the function as little convex as possible, which exactly corresponds to have equal risk in every component. [1]

The optimization algorithm used is the Nelder-Mead algorithm, where in every step of the optimization the input data for the PCA are the log returns multiplied by the portfolio weights. So the PCA is performed on weighted log returns. Since the Nelder-Mead algorithm is unconstrainted, the portfolio weights are normalized as in equation (3.8) in order to ensure that the portfolio weights sum to one.

The question of how many principal components to consider given a number of assets does not have a straight forward answer. In practice, it is often observed that most of the risk of a portfolio is in the first three to five principal components. As discussed in Section 2.3.2 the criterium stated in equation (2.34) is used to select the number of principal components m. To omit some of the principal components makes both from a mathematically and economically perspective good sense. Mathematically it is often known for real data, that the first few principal components describe most of the variance in data, such that only the first few components are needed to describe the whole data. From an economical point of view it also makes sense to assume that there only are a limited number of latent factors in a factor model. For instance, consider a portfolio consisting of 30 assets. When no truncation of the number of considered principal components is conducted, then it would be assumed that there are 30 latent factors. But it seems more realistic to assume that the assets are sensitive to only 2-6 factors, which are the factors with a significant amount of standard deviation in the data.

[1] Source: Internal documentation from Jyske Bank A/S.

The next section expands the ER approach to make use of functional principal component analysis. The reason for considering this approach is, that it is assumed to capture the variability in assets returns better, since it makes use of the underlying functional form the returns, and thereby may find more accurate portfolio weights.

Functional Principal Component Analysis Approach

Functional principal component analysis, fPCA, considers data to have an underlying functional form that can be modelled by univariate functions as explained in Section 2.3. This makes it possible to observe the behaviour of the eigenfunctions over time in contrast to PCA that just gives a static, non-temporal estimate of the eigenvectors. That is, it is possible to see the features characterizing asset returns over a period of time.

In order to obtain an ER portfolio using fPCA, a similar optimization problem as given in equation (3.15) is solved. Hence the ER approach of finding equal volatilities in the components is considered. The reason to only consider this approach is because it is not known how the functional variant will perform in a backtest, so only one of the ER approaches is investigated for possible improvements. The difference to the PCA is that fPCA should be better in taking into account the variation of the asset returns over time.

To make use of fPCA first data has to be transformed to functional data and to be smoothed as explained in Section 2.1.2. The basis functions used for the roughness penalty matrix R are equally spaced B-splines. So one has to determine the number of B-splines K and the smoothing parameter λ.

The number of basis functions is chosen by a three-fold cross validation. This means that data is divided into three equally sized sets, where two of the sets are used to train different values of K and the last set is used to validate the found K values. Simultaneously, there is performed a five-fold cross validation for λ on the trainings sets of the basis functions, again by dividing the training set into five equally spaced sets. The reason that the cross validation for λ is embedded in the cross validation for K is because the smoothing parameter λ depends on the number of basis functions K. The setup is illustrated in Figure 3.3. Hence for every trained K value there is a corresponding λ value. Then the two trained K values with their corresponding λ are tested on the validation set. The criterium used to find the best K and λ is the GCV(λ) as given in equation (2.11). The GCV statistic is used within the performed three- and five-fold cross validation, since it is an approximation to leave-one-out cross validation as explained in Section 2.1.2. Furthermore a rolling window is used, where the first part of data is used as trainings set and the last part of data for validation, which should give better parameter estimates when calculating the portfolio weights today. This also means that this setup might not be completely proper when used in a backtest.

Training set					Training set					Validation set
T	T	T	T	V	T	T	T	T	V	

Figure 3.3. Illustration of the cross validations for finding the number of basis functions K and the smoothing parameter λ. The first two boxes are the trainings sets for K, which are also used to cross validate λ. Then the validation set, the last box, is used to validate K with the corresponding λ value. T and V indicate the training and validation sets for λ, respectively.

After having found values for K and λ, data can be smoothed and a fPCA can be performed. The number of considered principal components m is determined corresponding to the previous section using equation (2.34).

To properly investigate if the fPCA provides better estimates of the portfolio weights and may yield higher profits, Chapter 4 considers different portfolios and investigates the performance of different asset allocation strategies including the ER approach using PCA and fPCA, respectively. But first a related asset allocation to the ER strategy is introduced in the next section.

3.2 Diversified Risk Parity

Another approach to find a well-diversified portfolio is described by [Meucci, 2010], which introduces the concept of the *diversification distribution*.

The Diversified Risk Parity, DRP, strategy normalizes the principal portfolios' contributions by the principal portfolio variance, defined by:

$$p_i = \frac{\tilde{w}_i^2 \lambda_i}{\sum_{i=1}^m \tilde{w}_i^2 \lambda_i}, \quad i = 1, \ldots, m,$$

where $m \leq p$ is the considered number of principal components. Note that the distribution is always non-negative and the elements p_i sum to one. A portfolio is said to be well-diversified when the p_i's are approximately equal such that the diversification distribtion is close to be uniform. Conversely, this means when principal portfolios load on specific factors, the distribution has a peak. For interpretational proposes, the diversification is measured by the exponential of the entropy, H:

$$\mathcal{N}_{\mathrm{H}} = \exp\left(-\sum_{i=1}^m p_i \ln(p_i)\right). \tag{3.16}$$

The interpretation of $\mathcal{N}_{\mathrm{H}}$ can be seen by two extreme cases. The first case is when all risk is due to one single principal portfolio, i.e. $p_i = 1$ for one i and $p_j = 0$ for $j \neq i$, which yields an entropy of zero and thereby $\mathcal{N}_{\mathrm{H}} = 1$. On the other hand if one holds a homogenous portfolio where $p_i = m^{-1}$ for all i, the entropy equals $\ln(m)$, such that the maximium value of $\mathcal{N}_{\mathrm{H}} = m$.

The optimization problem to be solved is given by:

$$\begin{aligned} &\arg\max_{\boldsymbol{w}} \mathcal{N}_{\mathrm{H}} \\ \text{subject to } &\mathbf{1}^\top \boldsymbol{w} = 1 \text{ and } -\mathbf{1} \leq \boldsymbol{w} \leq \mathbf{1}. \end{aligned} \tag{3.17}$$

Hence it is a nonlinear problem with constraints which could be solved using a SQP optimization algorithm as explained in Section 3.1.1.

The difference between the ER and DRP approaches is that they are based on two different measures of risk. The ER approach aims to diversify a portfolio such that the risk contributions are equal. On the other hand, the DRP approach defines a well-diversified portfolio such that the elements p_i of the diversification distribution are uniform distributed. [A. Meucci, A. Santangelo, R. Deguest, 2014] But without any further constraints the solutions to the optimization problems given in equation (3.13) and equation (3.17) coincide. This is due to the fact that the risk contributions in the PCA space $\tilde{\mathrm{RC}}_i$ are almost equal to the diversifications distribution elements p_i, so in this case the ER and DRP approaches optimize a similar problem just with different objective functions. This will also be shown in the simulation study in Section 3.4.2. Table 3.1 summarizes some of the important notations in the asset and principal space, respectively.

Table 3.1. Notions in the asset space and in the PCA space. This table is inspired by [Kind, 2013].

	Asset Space	Principal Space
Weights	w_i	$\tilde{w}_i$
Portfolio Variance	$\sigma_P^2(\boldsymbol{w}) = \sum_{i=1}^N w_i^2\sigma_i^2 + \sum_{i,j}^{\neq} w_i w_j \sigma_{ij}$	$\sigma_P^2(\tilde{\boldsymbol{w}}) = \sum_{i=1}^m \tilde{w}_i^2 \lambda_i$
Marginal Risk Contribution	$\partial_{w_i}\sigma_P(\boldsymbol{w}) = \frac{w_i\sigma_i^2 + \sum_{i\neq j} w_j\sigma_{ij}}{\sigma_P(\boldsymbol{w})}$	$\partial_{\tilde{w}_i}\sigma_P(\tilde{\boldsymbol{w}}) = \frac{\tilde{w}_i\lambda_i}{\sigma_P(\tilde{\boldsymbol{w}})}$
Risk Contribution	$\mathrm{RC}_i = \frac{w_i^2\sigma_i^2 + \sum_{i\neq j} w_i w_j\sigma_{ij}}{\sigma_P(\boldsymbol{w})}$	$\tilde{\mathrm{RC}}_i = \frac{\tilde{w}_i^2\lambda_i}{\sigma_P(\tilde{\boldsymbol{w}})}$

The next section shows that the traditional allocation strategies EW and MVa as introduced in Chapter 1 also can be considered in the PCA space.

3.3 The Equally-Weighted and Minimum Variance Strategies using Principal Portfolios

The strategies EW and MVa can also be performed in the PCA space. This section gives a short introduction to these approaches and is based on the work of [Kind, 2013], but they will not be considered in the backtest performed in Chapter 4. The portfolio weights of the EW strategy used on the principal portfolio are given by:

$$\tilde{w}_i^* = m^{-1} \quad \text{for } i = 1, \dots, m,$$

where m is the number of principal components. In the PCA space the EW strategy probably is a very bad diversification strategy, since the variance of each principal portfolio is given by the eigenvalues λ_i. It is known that eigenvalues are chosen to be variance maximizing and of decreasing order, hence using the EW strategy will give a very unbalanced risk distribution of the principal portfolio. Based on equation (3.3) the portfolio weights of the MVa strategy in the PCA space can be expressed by:

$$\tilde{w}_i^* = \frac{\lambda_i^{-1}}{\sum_{i=1}^m \lambda_i^{-1}}.$$

From equation (3.10) it follows that the portfolio variances in the PCA space for the MVa and the EW strategy are given by:

$$\sigma^2_{\text{MV}}(\tilde{\boldsymbol{w}}) = \frac{1}{\sum_{i=1}^m \lambda_i^{-1}} \quad \text{and} \quad \sigma^2_{\text{EW}}(\tilde{\boldsymbol{w}}) = m^{-1}\bar{\lambda}_{\bullet},$$

where $\bar{\lambda}_{\bullet}$ is the average of the principal portfolio variances. Similar to the property of the portfolio volatilities in equation (3.9), it can be shown that the same property applies for the portfolio volatilities in the PCA space:

$$\sigma_{\text{MV}}(\tilde{\boldsymbol{w}}) \leq \sigma_{\text{ER}}(\tilde{\boldsymbol{w}}) \leq \sigma_{\text{EW}}(\tilde{\boldsymbol{w}}),$$

where the ER strategy using risk contributions in the PCA space, ER- PCA RC, as given in equation (3.13) is considered. For a proof of this statement see [Kind, 2013]. Since it is known that the assets in the PCA space are uncorrelated and may have different volatilites, there is an explicit expression for the portfolio variance in the ER approach. Hence using the expression for the portfolio variance in the PCA space as stated in Table 3.1 with the portfolio weights from equation (3.14), the portfolio variance for the ER- PCA RC strategy is given by:

$$\sigma^2_{\text{ER}}(\tilde{\boldsymbol{w}}) = \frac{m}{\left(\sum_{i=1}^m (\sqrt{\lambda_i})^{-1}\right)^2}.$$

This section shows that the relation between the portfolio volatilities of the three strategies: MVa, ER, and EW is valid both in the original asset space as well as in the PCA space.

The next section considers an example of finding portfolio weights with the different ER strategies and the DRP strategy. Moreover, there is performed a simulation study in order to investigate the similarities and differences of the portfolio weights found by the different risk-based allocation strategies.

3.4 Example

In order to get a better intuition of how the different risk-based asset allocation methods work, consider the following simple portfolio consisting of three indexes describing the asset classes: Equities, Bonds, and Credits. This example is simplified such that reweighting of the portfolio is disregarded thereby the weights are determined based on the whole sampling period.

Data is sampled every month from January 28, 1999 through September 29, 2014. The scaled prices of the three indexes can be seen in Figure 3.4. It can be seen that Bonds are fairly stably increasing throughout the entire sampling period, whereas Equities and Credits behave more volatile with an overall increasing trend, and have a drop in the financial crisis 2008/2009.

Figure 3.4. Scaled prices of the three indexes: Equities, Bonds, and Credits.

Moreover, the correlation matrix for the three assets based on data of the whole sampling period is illustrated below. It can be seen that Bonds are negatively correlated to Equities and Credits. Whereas Equities and Credits are positively correlated to each other:

$$\begin{array}{ccc} \text{Equities} & \text{Bonds} & \text{Credits} \end{array}$$
$$\begin{bmatrix} 1 & -0.03 & 0.39 \\ -0.03 & 1 & -0.21 \\ 0.39 & -0.21 & 1 \end{bmatrix}$$

To get a better overview of how the different strategies work and how they are implemented in `R`, the example considers first the ER and DRP strategies using PCA, after which the ER strategy is considered using fPCA.

3.4.1 Principal Component Analysis

First consider PCA, which can be used to find the orthogonalized representation of the risk factors in this portfolio. In `R` a PCA is performed using the `prcomp` function, which by default centers the variables to have zero mean. Since the input indexes are measured on the same scale and the covariance structure plays an important role in both the ER and DRP strategies, the default setting `scale=FALSE` in `prcomp` is used, which means that the assets are not scaled to have unit variance.

The left panel of Figure 3.5 shows the variances of the three principal components. It can be seen that the first risk factor accounts for the greatest part of the variance in the portfolio, which is consistent with the explained theory in Section 2.2. Considering the scree plot in the right panel of Figure 3.5 it can be seen that the first two components describe about 92 percent of the variance in the portfolio. This suggests to consider a two factor model for modelling the log returns of this portfolio.

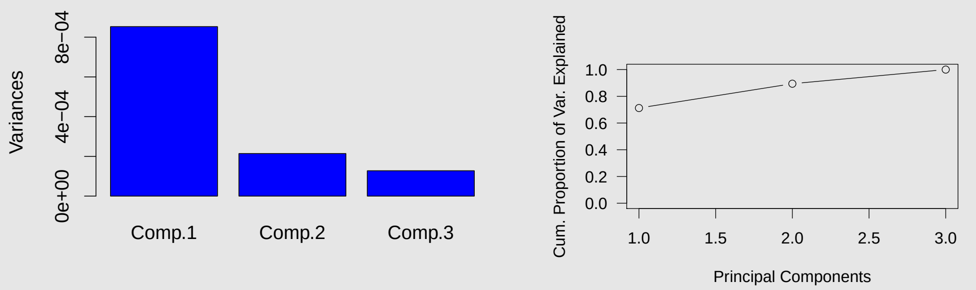


Figure 3.5. Plot of the variances against the number of the principal components in the left panel. And a scree plot of the principal components in the right panel.

Therefore Figure 3.6 shows the loadings of the three asset classes on the first two components, or risk factors. The factors are latent factors, which means that it is not known what they explain. But it is possible to make a guess of the interpretation of the factors by investigating the loadings of the assets on the factors as mentioned in Section 3.1.3. It can be seen that Equities load mostly on the first factor, whereas Bonds and Credits load mostly on the second factor. The reason that Equities are dominated by the first principal component is that Equities have the highest variance of the three indexes. The risk of Equities may be decomposed into equity market risk, i.e. risk of general market movements, industrial membership, and firm-specific risk. And Bonds and Credits may be decomposed into interest rate risk, i.e. risk of interest rate movements, and issuer specific risk, i.e. risk of default. Hence the first risk factor could describe equity market risk, whereas the second risk factor could be interest rate risk.

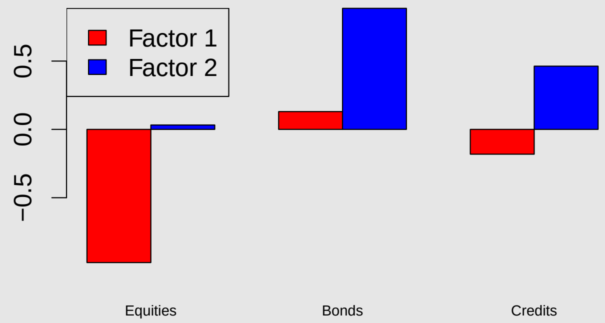


Figure 3.6. Asset loadings for the first two principal components.

In order to find the weights in the ER strategy, there are two possibilities for the PCA approach: the optimization problem in equation (3.13), ER- PCA RC, or the problem given in equation (3.15), ER- PCA.

The optimization method used to solve equation (3.13) is a SQP algorithm implemented in the function `solnp`. In this function it is possible to specify the constraints that the weights have to sum to one and that the weights have to lie in the interval $[0, 1]$, i.e. no short-selling is allowed. The function needs an initial guess of the portfolio weights, which is chosen to be equal weights, i.e. the initial weights are $\boldsymbol{w} = (0.33, 0.33, 0.33)^\top$, for all used strategies in this example.

The optimization problem in equation (3.15) on the other hand is solved using a Nelder-Mead algorithm implemented in the `optim` function. This is an unconstrainted, nonlinear algorithm that aims to minimize CPV_j as described above. In order to ensure that the weights sum to one, the found portfolio weights are normalized as stated in equation (3.8). The reason that the optimization in equation (3.15) is solved using an unconstrainted algorithm is that when using different constrainted algorithms, the algorithms did not converge, i.e. they always choose the initial weights, whereas the Nelder-Mead algorithm converges.

The DRP strategy is solved by the optimization problem given in equation (3.17) also using a SQP algorithm, as in the case for the ER- PCA RC. As described in Section 3.1.3 the optimization procedure for the DRP has to maximize $\mathcal{N}_{\mathrm{H}}$ to yield the number of considered principal components. Note that by default the function *solnp* minimizes a function, thus in order to make it a maximization problem the negative of N_{H} is used.

The found portfolio weights for the ER- RC, ER- PCA RC, ER- PCA, and DRP strategies can be found in Table 3.2, where the weights are displayed when using all principal components and when only using two principal components. For instance the portfolio weights for the ER- PCA strategy are found to be:

$$\boldsymbol{w}^*_{ER} \approx (0.16, 0.43, 0.41)^\top.$$

This means that the ER- PCA strategy suggests to hold 16% Equities, 43% Bonds, and 41% Credits. The next section considers the ER strategy using functional principal component analysis.

3.4.2 Functional Principal Component Analysis

Consider on the other hand the ER strategy using fPCA, where the `R` package `fda` can be used. In order to perform a fPCA, the first step is to transform the discrete observed data into functional data. This can be done as described in Section 2.1.2 using B-splines as basis functions and since the assets included in the portfolio have a constant sampling rate it is appropriate to use equally spaced B-splines. In order to find the number of basis functions K, cross validation is used as described in Section 3.1.3. A three-fold cross validation with a sequence of possible values $K = 10, 15, \ldots, 50$ is considered and at the same time a five-fold cross validation with corresponding λ values is done, where $\lambda = 0.00001, \ldots, 1, \ldots, 10000$ are used. Figure 3.7 shows the chosen B-spline basis, which has 15 basis functions for 365 discrete observations.

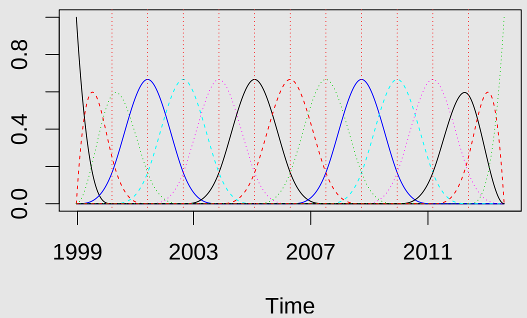


Figure 3.7. B-spline basis with $K = 15$ basis functions.

The corresponding smoothing parameter is found to be $\lambda = 10000$. Then the function `smooth.basis` performs a penalized residual sum of squares to obtain the smoothed curves. Figure 3.8 shows the original log returns and the corresponding smoothed curves. Note that in the ER- fPCA strategy solving equation (3.15) the weighted log returns are smoothed.

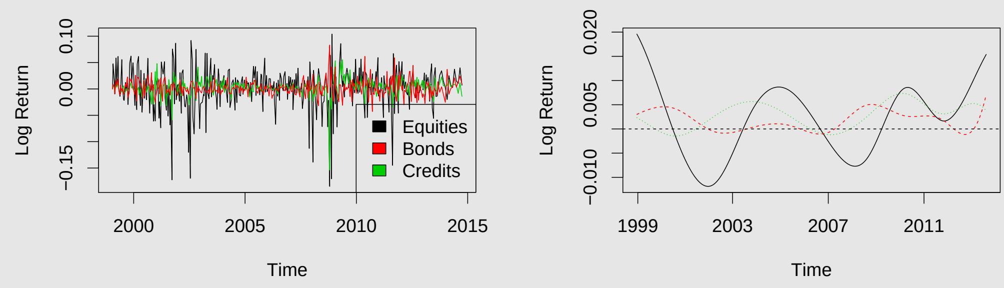


Figure 3.8. Original log returns for the three indexes in the left panel and the smoothed weighted log returns in the right panel.

Then a fPCA can be performed and the weights of the portfolio can be found using the ER- fPCA strategy in accordance with the optimization given in equation (3.15). Again a scree plot is considered in Figure 3.9, which shows that the last principal component does not account for any variation.

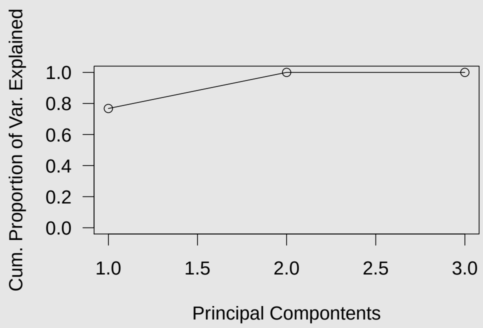


Figure 3.9. Scree plot of the principal components.

Figure 3.10 shows the first two eigenfunctions of the portfolio. It can be be seen that the first eigenfunction accounts for 84.4 percent of the variability of data, whereas the second eigenfunction accounts for 15.6 percent as it also can be seen in Figure 3.9. The solid line represents the mean function $\bar{x}(t)$ and the effects of adding (+) and subtracting (-) a multiple of each principal component. In order to identify which multiple should be used, define a constant C and let T denote the number of observations of an asset, where the root-mean-square difference C^2 between the mean function $\bar{x}(t)$ and its overall time average $\tilde{x}$ is defined as:

$$C^2 = \frac{1}{T} \left\| \bar{x}(t) - \tilde{x} \right\|,$$

where:

$$\tilde{x} = \frac{1}{T} \int \bar{x}(t) dt.$$

Then Figure 3.10 plots the mean function $\bar{x}(t)$ and $\bar{x}(t) \pm C\xi_j(t)$. [J. O. Ramsay , B. W. Silverman, 2005] It can be seen that the first eigenfunction varies from the mean curve. Comparing the first eigenfunction with the smoothed log returns in the right panel of Figure 3.8 it can be observed that the variation of adding the first principal component is very similar to the smoothed Equities log returns, whereas the effect of subtracting is very similar to the Bonds and Credits smoothed log returns. This indicates that the first eigenfunctions describe the overall variation caused by the three indexes. On the other hand the second eigenfunction is very similar to the mean function, it could describe a time shift effect, since adding and subtracting the principal components are shiftet curves compared to mean function.

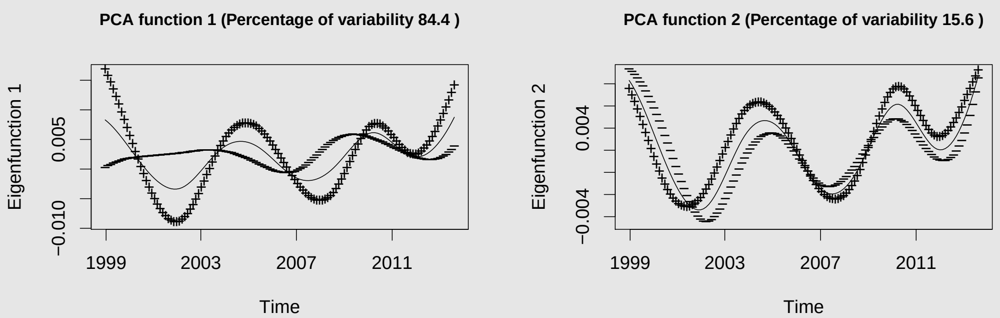


Figure 3.10. The first two eigenfunctions. The third eigenfunction is omitted, since it does not account for any variance of data. The solid line indicates the mean curve, and the effects of adding (+) and subtracting (-) a multiple of each eigenfunction are displayed.

Similar to the ER- PCA strategy in the previous section, a Nelder-Mead algorithm is used to solved equation (3.15) but the CPV_j result from a fPCA instead of a PCA of the weighted log returns. Depending on the used strategy and the number of considered principal components the weights in Table 3.2 are obtained.

Table 3.2. Optimal weights for the different strategies. In the PCA and fPCA strategies the weights are calculated using all principal components, three, and only using two components. The ER- RC strategy is not based on PCA or fPCA, therefore there is only one portfolio weight vector.

	#Components: 2	#Components: 3
Strategy	Weights Equities Bonds Credits	Weights Equities Bonds Credits
ER- RC	$\boldsymbol{w}^* \approx (0.13, 0.45, 0.41)^\top$	
ER- PCA	$\boldsymbol{w}^* \approx (0.16, 0.43, 0.41)^\top$	$\boldsymbol{w}^* \approx (0.16, 0.43, 0.41)^\top$
ER- PCA RC	$\boldsymbol{w}^* \approx (0.07, 0.53, 0.41)^\top$	$\boldsymbol{w}^* \approx (0.11, 0.57, 0.32)^\top$
ER- fPCA	$\boldsymbol{w}^* \approx (0.27, 0.32, 0.41)^\top$	$\boldsymbol{w}^* \approx (0.27, 0.32, 0.41)^\top$
DRP	$\boldsymbol{w}^* \approx (0.19, 0.65, 0.16)^\top$	$\boldsymbol{w}^* \approx (0.11, 0.57, 0.32)^\top$

It can be observed from Table 3.2 that the ER- PCA RC and DRP have the same weights when using three principal components, whereas their weights differ when there only are considered two principal components. It is interesting to observe that the two strategies change the portfolio weights in different directions, i.e. the ER- PCA RC strategy gives less weight to Equities and Bonds, but more to Credits. The DRP strategy behaves exactly reversed. This behaviour is due to the two different optimization problems that have to be solved. When omitting a principal component in the ER- PCA RC strategy, the weight of Credits has to increased in order to maintain the equal risk contributions of the assets, since Equities and Bonds have a larger exposure on the two first principal components than Credits has as can be observed in Figure 3.6. In the DRP strategy the elements of the diversification distribution p_i have to be equal to $p_i = 1/3$ when using all principal components, but by only considering two components the elements have to be equal to $p_i = 1/2$, which explains the increased weights of Equities and Bonds, and the decreased weight of Credits since the DRP solves a maximization problem.
Changing the number of principal components has no effect on the portfolio weights of the ER- PCA and ER- fPCA strategies in this portfolio. Furthermore, it can be seen that the portfolio weights for the ER- RC and ER- PCA strategies are very similar to each other, whereas the weights of the ER- fPCA give more weight to Equities and less to Bonds than the ER- RC and ER- PCA strategies do. Note that all ER strategies give the same weight to Credits when using two principal components.

In order to understand the differences in the portfolio weights when using different strategies, a simulation study is considered in the next section. The aim is to investigate the differences and similarities of the strategies, and the estimation uncertainty of the portfolio weights.

Simulation Study

The study is based on 100 simulations of the log returns of three equally correlated assets drawn from a multidimensional normal distribution with the following correlation matrix:

$$\begin{bmatrix} 2 & 0.3 & 0.3 \\ 0.3 & 4 & 0.3 \\ 0.3 & 0.3 & 6 \end{bmatrix}$$

Since there are considered three assets with equal correlation it is known from Section 3.1.1 that there exists a closed-from solution for the ER- RC strategy as stated in equation (3.1). So the aim of this simulation study is to investigate how accurate the portfolio weights found by the other ER strategies are compared to the closed-form solution. In addition, it is interesting to examine how the portfolio weights found by the ER strategies differ from the portfolio weights found by the DRP strategy, since the example in the previous section showed that there are some variations across the strategies.

Figure 3.11 shows the boxplots of the portfolio weights found by calculating the portfolio weights of the ER- RC strategy using equation (3.2), the ER- PCA RC strategy using equation (3.13), the ER- PCA using equation (3.15) and the corresponding optimization for fPCA, and the DRP strategy using equation (3.17).
It can be seen that the portfolio weights of the ER- RC and ER- PCA strategies are very close to the closed-form solution and the strategies find very accurate weights, i.e. they have a small estimation uncertainty. The DRP and ER- PCA RC strategies find weights that are more widely dispersed compared to the closed-form solution, which means that these strategies are not fully consistent with the basic idea of the ER strategy. The ER- fPCA, ER- PCA RC, and the DRP strategies have a large estimation uncertainty, which is disadvantageous for an asset allocation strategy, since small differences in the portfolio weights can cause big losses.

Note that in the ER- fPCA strategy the number of basis functions is chosen to be 10 and the smoothing parameter is selected to be 100, which is an ad hoc choice in order to reduce computation time in this study.

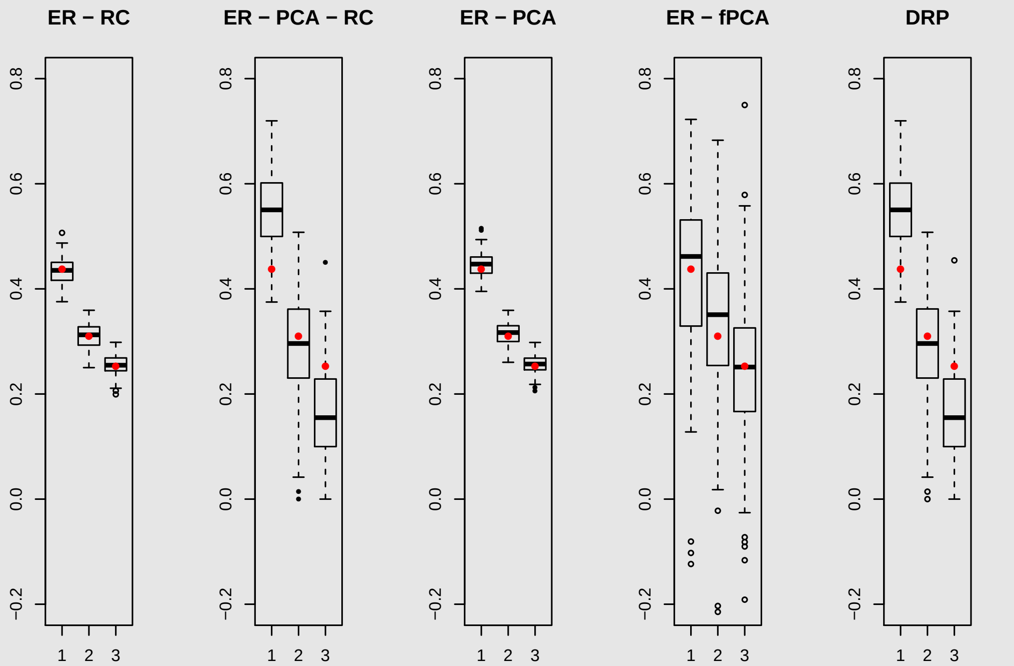


Figure 3.11. Boxplots of weights found by simulating log returns for the different asset allocation strategies. The red dots indicate the weights found with the closed-form solution from equation (3.1) for the case with equal correlations and different volatilities of assets.

In order to check the accuracy of the strategies compared to the closed-form solution, the mean squared error, MSE, for the different strategies can be seen in Table 3.3. The ER- RC and ER- PCA strategies have the lowest MSEs, whereas the ER- PCA RC and DRP have a higher and identical MSE. The ER- fPCA strategy has the highest MSE, which is also due to the large estimation uncertainty in the portfolio weights as seen in Figure 3.11.

Table 3.3. Mean squared error for the different strategies determined by comparing the found weights with the closed-form solution.

Strategy	MSE
ER- RC	0.0016
ER- PCA	0.0015
ER- PCA RC	0.044
ER- fPCA	0.133
DRP	0.044

What can be concluded from the example in Section 3.4 and the simulation study in this section is that the ER- RC and ER- PCA strategies result in very similar portfolio weight, as well do the ER- PCA RC and DRP strategies when using all principal components, but result in different portfolio weights when varying the number of considered principal components. The ER- fPCA is more similar to the ER- RC and ER- PCA strategies than the ER- PCA RC and DRP strategies are, but it has a large estimation uncertainty. The estimation uncertainty might be caused by the choice of the number of basis functions K and the smoothing parameter λ.

As mentioned earlier, the next chapter aims to investigate the different risk-based asset allocation strategies considering different portfolios. The purpose is to figure out which strategy that gives the best mean profit, is least volatile, and hopefully also can perform well throughout a financial crisis like in 2008/2009.

Backtest of Allocation Strategies 4

In order to investigate the performance of the different asset allocation strategies introduced in chapter 3, this chapter considers a backtest on historical data.

Backtesting means to use historical data to test a trading strategy, since it seems reasonable that a strategy that has not been profitable in the past will not be profitable in the future, assuming that there are no major changes in the market. This does not mean that a good backtest guarantees a good performance when trading a strategy in real live, so a backtest may be viewed as a tool to reject strategies, but one has to be careful using it as a validation tool. There is used a walk-forward backtest which in general describes the concept of estimating the parameters of a trading strategy based on a selected set of data, and then roll forward and estimate again. This means that the weights of a portfolio for, e.g. the next month, are determined based on an estimation of data back in time for a specified period, e.g. two years. In order to find the different weights of the portfolio over time one uses a rolling estimation window such as illustrated in Figure 4.1.

$$
\begin{array}{l}
\boxed{1, \quad \ldots \quad , t} \\
\quad \boxed{2, \quad \ldots \quad , t+1} \\
\qquad\qquad \ddots \\
\qquad\qquad\qquad \boxed{n-t, \quad \ldots \quad , n}
\end{array}
$$

Figure 4.1. Rolling estimation window, where t is the specified backtest period and n is the last observation in data.

The next section introduces some portfolios containing different assets, which will be used to backtest the allocation strategies in Section 4.2. In order to benchmark the different strategies the Equally-Weighted, EW, strategy is used as benchmark.

4.1 Portfolios

The aim of this section is to study portfolios consisting of different assets, investigate the correlation between the assets and statistics, such as volatility, of the included assets. Data for this backtest is kindly provided by Jyske Bank and therefore treated confidentially.

4.1.1 Asset Classes

The first portfolio contains four indexes that describe Equities, Bonds, Commodities, and Credits. Data is sampled every month from January 28, 1999 through April 29, 2012. This portfolio is a rather traditional portfolio consisting of indexes describing four different asset classes. The idea of this portfolio construction is to obtain diversification by investing in different asset classes that do not behave similar. But it is questionable if the specific risk that the assets may have will vanish, when the portfolio only consists of four assets.

Figure 4.2 shows the scaled prices of the four considered indexes. It can be seen that Bonds are fairly stable increasing throughout the entire sampling period, whereas Equities, Commodities, and Credits behave more volatile with an overall increasing trend, and have a drop in the financial crisis 2008/2009.

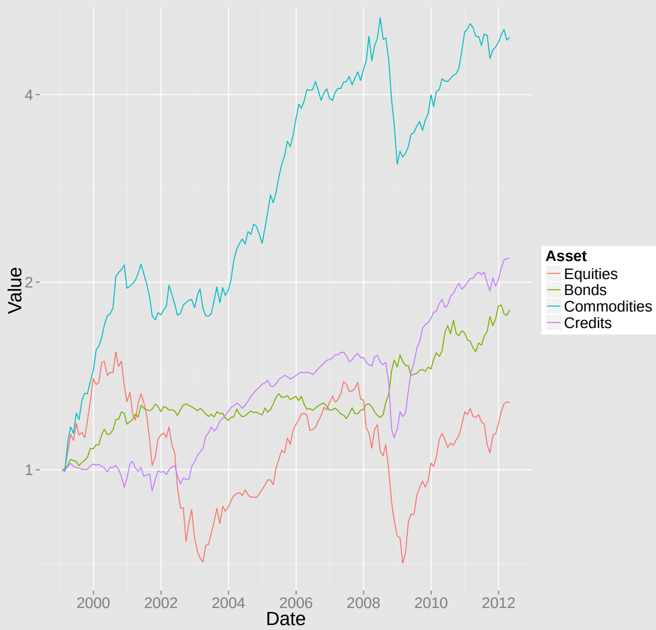


Figure 4.2. Scaled prices of the four indexes.

Moreover, Table 4.1 illustrates the correlation matrix of the four assets based on data of the whole sampling period in the upper half, and the correlation based on approximately four year sampling periods (January, 1999 - Februrary, 2003; March, 2003 - March, 2007; April, 2007 - April, 2012) in the lower half of the table. In the upper half it can be seen that Bonds are only negative correlated to Credits. Commodities are positively correlated to Equities and Credits, and Credits and Equities are positive correlated to each other. The part below the diagonal shows the varying correlations for the approximately four year periods. From this it can be noted that Equities and Bonds are not uncorrelated throughout the whole sampling period, but have a change from positive to negative correlation, which is also true for Bonds and Commodities. For the other assets, the overall correlation from the upper half is a suitable representation of the overall correlation for the 12 considered years. As shown in Figure 1.3 in Section 1.1.1, the less positively correlated the assets of a portfolio are, the more diversification is possible. Consequently, this portfolio sould have a reduced portfolio variance as explained in Section 1.1.1.

Table 4.1. Correlation matrix for the four assets. The part above the diagonal shows the correlation for the whole sample period January 28, 1999 through April 29, 2012. The part below the diagonal shows the correlation for four year periods: January, 1999 - Februrary, 2003; March, 2003 - March, 2007; April, 2007 - April, 2012.

	Equities	Bonds	Commodities	Credits
Equities	1	0	0.5	0.5
Bonds	0.4; 0.2; -0.3	1	0	-0.3
Commodities	0.6; 0.3; 0.5	0.5; 0.2; -0.3	1	0.3
Credits	0.4; 0.2; 0.7	0; 0.1; -0.5	0.2; -0.2; 0.5	1

Table 4.2 shows the volatility of the profit, mean profit, and Sharpe ratio of the portfolio using the EW strategy and the corresponding statistics for the assets themselves. The mean profit and the volatility of the profit are computed according to equations (1.3) and (1.4) as stated in Section 1.1. The reason to consider the EW strategy, is because it is used to benchmark the other introduced allocation strategies later in the backtest. The table illustrates that the profit of Equities and Commodities are much more volatile than that of Bonds and Credits are. However, Commodities have the highest mean profit and actually the highest Sharpe ratio. All in all, using the EW strategy and rebalancing every month, this portfolio has a mean profit of 6.22 and a volatiliy of 8.01, which yields a Sharpe ratio of 0.78. As explained in Section 1.1.1 the Sharpe ratio can be used to compare the performance of the different allocation strategies.

Table 4.2. Statistics for the portfolio.

	Weight	Volatility	Mean Profit	Sharpe Ratio
Portfolio		8.01	6.22	0.78
Equities	0.25	16.07	1.88	0.12
Bonds	0.25	7.11	4.53	0.64
Commodities	0.25	14.9	12.73	0.85
Credits	0.25	8.97	6.04	0.67

The next section describes another portfolio, which is an extension of the portfolio presented in this section.

4.1.2 Asset Classes and Style Factors

It is also interesting to construct a portfolio that consists of both asset classes: Equities, Bonds, Commodities, and Credits; and style factors: LowRisk, Momentum, Quality, Size, and Value. The style factors, as introduced in Section 1.2, are constructed by characteristic-sorted portfolios, which means that the factors are estimated by using portfolios that are formed based on firm characteristics. The intuition behind this construction is that for instance growth firms have similar stock returns, so combining these stocks implies that it is very likely that this portfolio has an exposure to an underlying risk factor, and thereby forms a style factor. [M. Grinblatt and S. Titman, 2002]

The reason to consider such a portfolio in a backtest of risk-based allocation strategies is to see whether it is possible to improve the profit characteristics by including these style factors in the portfolio. The idea of style factors is very similar to pure factor portfolios as introduced in Section 1.2.2, which are portfolios that only load on one factor. It is interesting to see how a PCA-based strategy allocates the style factors, since these strategies are based on finding underlying factors, which style factors are by construction. Furthermore, it will be investigated if risk-based strategies can improve the performance compared to the benchmark strategy.

The style factor indexes can have negative returns in some periods, but in the long run they yield positive returns. This is in accordance with the introduced theory in Section 1.2.2, that states that these indexes are constructed to describe investment styles that compensate for low returns in bad times with high returns in the long run. This implies that including style factors in a portfolio is more advantageous for long-term investors.

Data is sampled every month from January 28, 1999 through April 29, 2012. This portfolio should have lower specific risk than the portfolio considered in the previous section, since there now are included nine instead of only four assets. Considering Figure 4.3 it can be seen that the style factors are fairly stable increasing throughout the sampling period. Expect from Bonds, Quality, and Size the prices of the assets drop in the financial crisis 2008/2009.

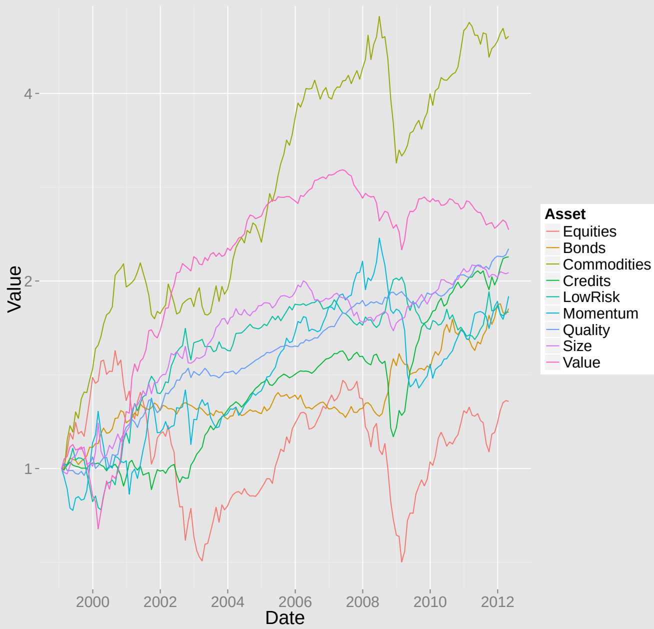


Figure 4.3. Scaled prices of the indexes.

From Table 4.3 it can be seen that especially the style factors are low correlated to each other with exceptions, e.g. are LowRisk and Value highly correlated, $\rho = 0.8$, in the first four years. In general are the correlations very different over time, also with changing sign for almost all assets. This means that the degree of diversification also changes over time. But all in all the low and negative correlations imply that this should be a well-diversified portfolio with a reduced portfolio variance. So including the low and negatively correlated style factors has a positive effect on the diversification properties of the portfolio.

Table 4.3. Correlation matrix for the nine assets. The part above the diagonal shows the correlation for the whole sampling period January 28, 1999 through April 29, 2012. The part below the diagonal shows the correlation for four years: January, 1999 - Februrary, 2003; March, 2003 - March, 2007; April, 2007 - April, 2012.

	Equities	Bonds	Commodities	Credits	LowRisk	Momentum	Quality	Size	Value
Equities	1	0	0.5	0.5	-0.7	-0.1	-0.1	0	-0.1
Bonds	0.4; 0.2; -0.3	1	0	-0.3	0.3	-0.1	0	-0.2	-0.1
Commodities	0.6; 0.3; 0.5	0.5; 0.2; -0.3	1	0.3	-0.3	0.1	-0.2	0.1	-0.1
Credits	0.4; 0.3; 0.7	0; 0.1; -0.5	0.2; -0.2; 0.5	1	-0.4	0	-0.2	0.3	0.1
LowRisk	-0.8; -0.5; -0.7	-0.1; 0.3; 0.6	-0.3; 0; -0.4	-0,4; -0.1; -0.6	1	0.1	0	0.1	0.4
Momentum	-0.4; 0.2; 0.1	0; 0; -0.2	-0.2; 0.3; 0.3	-0.3; 0; 0.1	0.2; 0; -0.1	1	0.3	0.2	-0.2
Quality	-0.2; -0.2; 0	0.1; 0; 0.1	-0.3; 0; -0.1	-0.2; 0; -0.1	0.2; -0.5; -0.2	0.3; 0.3; 0.4	1	-0.2	0.2
Size	-0.3; 0.3; 0.4	-0.2; 0; -0.3	0; 0.2; 0.3	0.2;0.3; 0.5	0.3; 0; -0.3	0.3; 0.3; 0	-0.3; -0.2; 0	1	0
Value	-0.5; -0.2; 0.5	-0.1; 0.2; -0.1	-0.3; -0.1; 0	-0.2; 0.2; 0.5	0.8; 0.1; -0.3	-0.1; -0.2; -0.6	0.5; 0.1; -0.3	-0.1; -0.6; 0.2	1

Table 4.4 shows the volatility, mean profit, and Sharpe ratio of the assets and of the portfolio using the EW strategy. It can be seen that Equities, Commodities, and Momentum have the most volatile profits, whereas Quality has a very low profit volatility. The overall volatility of the profit of the portfolio is only 3.87, hence the portfolio volatility is reduced compared to the portfolio presented in the previous section. This is due to the negative covariances between some of the assets, since the portfolio variance asymptotically equals the average covariance in an equally weighted portfolio as shown in Example 1. And including more assets reduces the contribution of the single assets to the volatility and it may also reduce the specific risk in the portfolio. The mean profit has also declined by adding the style factors. But all in all the Sharpe ratio is increased from 0.78 to 1.52. So for an investor that prefers lower risk and thereby also a slightly lower profit, adding the style factors to the portfolio is a good choice. As mentioned in the start of this section, style factors often first get profitable in the long run, which is also a consideration an investor has to make.

Table 4.4. Statistics for the portfolio.

	Weight	Volatility	Mean Profit	Sharpe Ratio
Portfolio		3.87	5.88	1.52
Equities	0.11	16.07	1.88	0.12
Bonds	0.11	7.11	4.53	0.64
Commodities	0.11	14.9	12.73	0.85
Credits	0.11	8.97	6.04	0.67
LowRisk	0.11	9.68	4.43	0.46
Momentum	0.11	14.47	4.89	0.34
Quality	0.11	5.01	6.28	1.25
Size	0.11	7.94	5.58	0.7
Value	0.11	10.1	6.84	0.68

The next section introduces the final portfolio considered in the backtest later in this chapter.

4.1.3 Large Portfolio - 21 Assets

In the two previous sections there are considered rather small portfolios, but in order to ensure that systematic risk has been diversified and to investigate the behaviour of the asset allocation strategies on a larger portfolio, there is now considered a portfolio consisting of 21 assets. It is especially interesting to observe how many underlying factors the PCA- and fPCA-based strategies suggest using the standard deviation criterium in equation (2.34) compared to the smaller portfolios, since it is difficult to assess the choice of the number of principal components when there maybe are as many underlying factors as there are assets. In this portfolio there are included the same assets as in the previous portfolios: Equities, Bonds, Commodities, Credits, LowRisk, Momentum, Quality, Size, and Value. And then there are included additional equity indexes representing different markets: Europe, USA, Japan, and Emerging Markets (EM); and an additional goverment bond describing the 10 year interest rate in Denmark: BondsDK; and some indexes describing different commodities: LiveStock, Agriculture (Agri), Precious Metal (PrecMetal), Energy, Industrial Metal (IndMetal), Oil, and Gold.

Some of the assets have different sampling rates, i.e. the equity indexes: Europe, USA, Japan, EM, and the commodity index Gold are sampled on all trading days whereas the other assets are sampled on a monthly basis. When combining these assets into one portfolio there are only considered monthly observations such that only the observations that are in accordance with the monthly observations are considered. This is due to technical reason in the backtest setup. As mentioned in Chapter 2 it is possible to include all observation in the functional approach of the Equal Risk, ER, strategy, but this is beyond the scope of this thesis. So data is sampled every month from January 28, 1999 through April 29, 2012. Figure 4.4 shows the scaled prices of the added 12 assets in this portfolio. It is observable that some of the assets behave very similar whereas others act independently of each other. At the same time is the price range of the assets larger in this portfolio compared to the two previous ones.

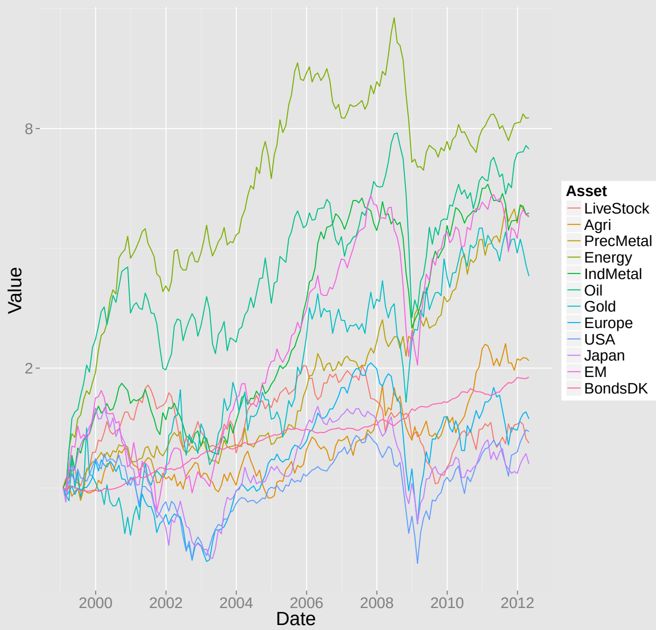


Figure 4.4. The scaled prices of the added 12 assets in the portfolio are shown. The prices of the other assets can be seen in Figure 4.3.

Table 4.5. Statistics for the portfolio.

	Weight	Volatility	Mean Profit	Sharpe Ratio
Portfolio		7.67	7.18	0.94
Equities	0.05	16.07	1.88	0.12
Bonds	0.05	7.11	4.53	0.64
Commodities	0.05	14.9	12.73	0.85
Credits	0.05	8.97	6.04	0.67
LowRisk	0.05	9.68	4.43	0.46
Momentum	0.05	14.47	4.89	0.34
Quality	0.05	5.01	6.28	1.25
Size	0.05	7.94	5.58	0.7
Value	0.05	10.1	6.84	0.68
LiveStock	0.05	17.06	1.96	0.12
Agri	0.05	17.33	5.68	0.33
PrecMetal	0.05	17.24	12.54	0.73
Energy	0.05	25.19	17.43	0.69
IndMetal	0.05	20.43	12.68	0.62
Oil	0.05	30.97	15.85	0.51
Gold	0.05	34.5	9.63	0.28
Europe	0.05	21.11	3.04	0.14
USA	0.05	18.09	2.48	0.14
Japan	0.05	19.76	1.05	0.05
EM	0.05	25.85	12.47	0.48
BondsDK	0.05	3.04	4.93	1.62

Table 4.5 summarizes the portfolio statistics for the single assets and the overall portfolio using the EW strategy with monthly rebalancing. It can be seen that especially the equity and commodity indexes have very volatile profits, and only some of them also have a high mean profit. On the other hand are most of the style factors and the bonds less volatile. All in all the portfolio profit is rather volatile, but has a moderate mean profit. Comparing the Sharpe ratios of this portfolio to the two previous ones it can be seen that the Sharpe ratio is higher than that of the first portfolio, but lower than that of the second portfolio. This portfolio has the highest mean profit of the three considered portfolios when using the EW strategy. Note that the weights are decreased in this portfolio due to the number of included assets, which implies a smaller contribution of the single assets and thereby also has an impact on the profit.

Table 4.6 shows the correlations of the assets for the whole sampling period. In contrast to the other considered portfolios, the four year correlations are disregarded due to lack of space. From the table it can be seen that many assets are low or negatively correlated to each other, which is good for diversifying risk from the portfolio. Some assets, like the equity indexes are highly correlated to each other, which is expected.

Table 4.6. Correlation matrix for the 21 assets. The table shows the correlation for the whole sampling period January 28, 1999 through April 29, 2012.

	E	B	Co	Cr	LR	M	Q	S	V	LS	A	PM	En	IM	O	G	Eu	US	J	EM	DK
E	1	0	0.5	0.5	-0.7	-0.1	-0.1	0	-0.1	0.2	0.3	0.1	0.3	0.6	0.2	0.2	0.6	0.7	0.5	0.7	-0.3
B		1	-0.3	0	0.3	-0.1	0	-0.2	-0.1	0.4	0.1	0.2	-0.1	-0.1	0.1	0	-0.6	-0.3	-0.2	-0.5	0.5
Co			1	0.3	-0.4	0	-0.2	0.3	0.1	-0.2	0.1	0	0.2	0.3	0.1	0.2	0.6	0.5	0.4	0.6	-0.1
Cr				1	-0.3	0.1	-0.2	0.1	-0.1	0.3	0.6	0.3	0.8	0.8	0.6	0.3	0.2	0.2	0.2	0.3	-0.3
LR					1	0.1	0	0.1	0.4	0.1	-0.2	0	-0.1	-0.4	0	-0.1	-0.6	-0.7	-0.5	-0.6	0.4
M						1	0.3	0.2	-0.2	-0.1	0	0.1	0.2	0	0.1	0.1	0	-0.1	0	0	0
Q							1	-0.2	0.2	0	0	-0.1	-0.2	-0.1	-0.1	0	0	0	-0.3	-0.1	0
S								1	0	0	0	0.1	0.2	0.1	0.1	0.2	0.1	-0.1	0	0.2	-0.1
V									1	0	-0.1	-0.2	-0.1	0	0	0	0.1	0	-0.1	-0.1	0.1
LS										1	0.2	0.1	0.1	0.2	0.2	-0.1	-0.3	-0.1	-0.2	-0.2	0
A											1	0.2	0.2	0.3	0.1	0.2	0	0.1	0.1	0.2	-0.1
PM												1	0.2	0.3	0.1	0.6	-0.2	-0.1	0	0	0.1
En													1	0.4	0.7	0.2	0.1	0.1	0.2	0.2	-0.2
IM														1	0.3	0.3	0.3	0.3	0.2	0.4	-0.3
O															1	0.1	0	0	0.2	0.1	-0.2
G																1	0.2	0.2	0.2	0.3	0
Eu																	1	0.9	0.6	0.9	-0.3
US																		1	0.6	0.8	-0.2
J																			1	0.6	-0.1
EM																				1	-0.2
DK																					1

After having introduced the different portfolios, the next section deals with investigating the performance of the different asset allocation strategies on historical data by among others comparing their profits, Sharpe ratios, and turnovers. Furthermore, the selected portfolio weights are investigated over time.

4.2 Backtest

This section aims to examine the performance of the different risk-based allocation strategies introduced in Chapter 3 compared to the benchmark strategy: EW, and the traditional MVa strategy by a walk-forward backtest on the portfolios as described in Section 4.1. The considered asset allocation strategies are:

- *Equally-Weighted strategy* (EW):
$$\boldsymbol{w}^* = N^{-1}.$$
- *Minimum Variance strategy* (MVa):
$$\boldsymbol{w}^* = \min_{\boldsymbol{w}} \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j \sigma_{ij}.$$
- *Equal Risk strategy – risk contribution approach* (ER - RC):
$$\boldsymbol{w}^* = \operatorname*{argmin}_{\boldsymbol{w}} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(w_i(\Sigma \boldsymbol{w})_i - w_j(\Sigma \boldsymbol{w})_j\right)^2.$$
- *Equal Risk strategy – PCA approach* (ER - PCA):
$$\boldsymbol{w}^* = \min_{\boldsymbol{w}} \sum_{j=1}^{m} \mathrm{CPV}_j.$$
- *Equal Risk strategy – fPCA approach* (ER - fPCA):
$$\boldsymbol{w}^* = \min_{\boldsymbol{w}} \sum_{j=1}^{m} \mathrm{CPV}_j.$$
- *Equal Risk strategy – PCA risk contribution approach* (ER - PCA RC):
$$\boldsymbol{w}^* = \operatorname*{argmin}_{\boldsymbol{w}} \sum_{i=1}^{m} \sum_{j=1}^{m} \left(\tilde{\mathrm{RC}}_i - \tilde{\mathrm{RC}}_j\right)^2.$$
- *Diversified Risk Parity strategy* (DRP):
$$\boldsymbol{w}^* = \arg\max_{\boldsymbol{w}} \mathcal{N}_H.$$

The reason that the MVa strategy is tested and not the Markowitz Mean-Variance, MV, strategy is due the fact that the aim of this backtest is to compare the performance of the alternative risk-based strategies with the benchmark strategy and a traditional strategy, but the MV strategy might find completely other portfolio weights due to a high estimation error of the mean. So in order to obtain a meaningful comparison of the different allocation strategies the MVa strategy is considered, which is a special case of the MV strategy, and is exclusively based on the estimation of the covariance matrix Σ, hence it is more similar to the other considered strategies.

Appendix A.2 illustrates how the backtest is set up in R using a flow chart. First the selected portfolio is read from a database, which then is inputted in a function that can perform the backtest. Data is transformed to log returns as described in equation (1.1) and outliers are detected and cut off by considering the standard deviation of the log return of an asset in the following way:

$$|\log(R_i(t))| > 2\sigma_i \quad \Rightarrow \quad \log(R_i(t)) = \pm 2\sigma_i \quad \text{for } i = 1, \dots, N,$$

where $\log(R_i(t))$ is the log return of asset i at time t, σ_i is the standard deviation of asset i, and N is the number of assets in a portfolio. This means that values that are larger in absolute value than two times the standard deviation of the considered asset are cut off and set to be two times the standard deviation. Then this function calls an optimization function, which is dependent on the asset allocation strategy. Before performing the optimization some strategies need to specify some additional parameters, e.g. the PCA-based strategies need the number of principal components, which is determined by the criterium in equation (2.34). Thus the number of principal components m, the number of basis functions K, and the smoothing parameter λ are computed before performing the backtest, such that the used parameters are the same throughout the whole backtest period. Furthermore, the optimization methods need an initial guess for the portfolio weights, which for all strategies are set to equal weights.

Since all three portfolios introduced in Section 4.1 are modified to have a monthly sampling rate, the portfolio weights of the different strategies are rebalanced every month based on an estimation of the past two or five years, respectively. Thereby data from the first two or five years are exclusively used for estimation purposes. It is assumed that short-selling is not allowed, which is due to the theoretical limitations of some of the strategies in this thesis, e.g. the ER- RC strategy. After the first estimation of the weights, the estimated portfolio weights are used recursively as initial weights to find the weights for the next period. This is in accordance with the theory introduced in Section 1.1. Another possibility would be to always use equal weights as initial weights, which might result in a different allocation.

In the backtest setup transaction costs are ignored, but in order to determine how good an asset allocation strategy performs it is also essential to know how much it will cost to use a specific strategy. Therefore the statistic describing the average *turnover* over T periods can be considered:

$$\text{TO} = \frac{1}{T-1} \sum_{t=2}^{T} \sum_{i=1}^{N} |w_i(t) - w_i(t-1)|, \tag{4.1}$$

where N is the number of assets, $w_i(t)$ is the portfolio weight of asset i at time t. The value of TO equals the average amount of buy and sell transactions as a percentage of the portfolio value. [Kind, 2013] Often a low value of TO is preferred, but if a strategy has very high profits, a high turnover might still pay off.

It is also interesting to consider the concentration of the portfolio, i.e. if some assets are weighted much higher than others, using the different strategies over time t. One possible statistic to measure concentration is the so-called *Herfindahl index*, $h(t)$, which is given by:

$$h(t) = \sum_{i=1}^{N} w_i(t)^2,$$

where $w_i(t)$ is the weight of asset i at time t. The index equals one if a portfolio is concentrated, i.e. there is only invested in one asset. On the other hand when there is invested uniformly in all assets then the index equals N^{-1}. Usually the *normalized Herfindahl index*, $\mathrm{NH}(t)$, is considered:

$$\mathrm{NH}(t) = \frac{h(t) - 1/N}{1 - 1/N},$$

where $\mathrm{NH}(t) \in [0, 1]$. In general, a low value of $\mathrm{NH}(t)$ is preferred, since this indicates a balanced portfolio. [S. Maillard, T. Roncalli, J. Teiletche, 2009]

The following sections backtest the introduced asset allocation strategies on the different portfolios introduced in Section 4.1. A summary of the most important results of the backtest can be found in Section 4.3.

4.2.1 Asset Classes

The first backtest of the mentioned asset allocation strategies is based on the portfolio consisting of asset classes as introduced in Section 4.1.1. First a two year estimation window is used and later in this section also a five year estimation window is considered.

For the strategies based on PCA the number of principal components m is determined using the standard deviation criterium from equation (2.34) with $v = 0.1$. This means that only components whose standard deviation is larger than 10% of the first principal component's standard deviation are included in the PCA, which is done to ensure that the included components account for enough variation and thereby represent an underlying risk factor. The choice of the parameter v is ad hoc and will therefore be investigated. This is also done for the ER- fPCA strategy and in addition the smoothing parameter λ and the number of basis functions K are determined using cross validation as explained in Section 3.1.3. Note that since all assets have a constant sampling rate, equally spaced B-splines are used to smooth data in the ER- fPCA strategy. The determined parameters are summarized in Table 4.7.

Table 4.7. The determined number of principal components m for the strategies based on PCA and fPCA for $v = 0.1$. And the chosen number of basis functions K and the smoothing paramter λ for the fPCA approach determined by cross validation.

Strategy	m	K	λ
PCA-based	4	-	-
fPCA-based	3	15	10000

It can be seen from Table 4.7 that for the PCA-based strategies all principal components are used, whereas in the case of ER- fPCA only three principal components are used. This means that the PCA-based strategies suggest that there are four latent factors, whereas the ER- fPCA strategy suggests three factors. The difference in the number of principal components might be caused by smoothing data, which can remove noise and reduces the variability of data, hence the eigenvalues might also decrease and thereby the standard deviation criterium suggests a lower number of principal components. But since this portfolio only consists of four assets, it might be appropriate that there are four underlying factors that drive the assets.

The scaled portfolio profits for the different strategies are displayed in Figure 4.5. The first two years are used as initial backtest period, hence the portfolio weights are first determined from ultimo January 2001. After this period there can be observed differences in the resulting profits of the different strategies, where it can be noted that the ER- PCA RC and DRP outperform the benchmark strategy and that the EW and ER- fPCA strategy drop during the financial crisis, whereas the other strategies perform quite stable.

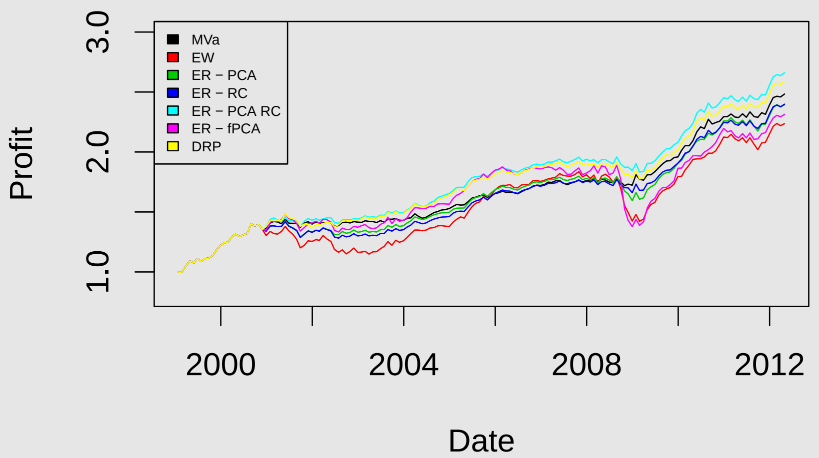


Figure 4.5. Scaled portfolio profits of the asset allocation strategies. The first two years are exclusively used to estimate the first backtest weights.

Figure 4.6 shows the portfolio weights over time for the different strategies. It can be observed that the weights of the ER- RC and ER- PCA approaches are very similar to each other, which is consistent with the results obtained in the simulation study in Section 3.4. As investigated in Section 4.1.1 the correlations of the assets in this portfolio are very similar to each other, expect from the negative correlations of Bonds and Credits. Therefore it is expected from the results in Section 3.1 that the ER- RC strategy finds weights that are inversely related to the volatilites of the single assets. Using the volatilies from Table 4.2 it is assumed that the weights for Equities and Commodities are very similar and small, whereas the weights for Bonds and Credits are larger and on average also are very similar. This can be confirmed by the portfolio weights found in Figure 4.6. The weights of the ER- PCA RC and DRP approaches are also very similar, which also is expected from the results of Section 3.4. The ER- fPCA strategy finds completely other weights and has a very high turnover, which is disadvantageous for an asset allocation strategy since this can imply high transaction costs. The MVa strategy gives most weight to the least volatile assets, Bonds and Credits, which is in accordance with the aim of finding portfolio weights that result in the minimum portfolio variance. As expected from the correlation matrix in Figure 4.1, where it can be seen that Bonds and Credits are negatively correlated, the weights for these behave reversed in all strategies, but in the case of the ER- fPCA strategy it is difficult to see this behaviour due to the high turnover of this strategy.

Figure 4.6. The weights of the different strategies estimated on a rolling window of two years and rebalanced every month. Note that the vertical axes differ for some of the strategies.

Even though the different asset allocation strategies find different portfolio weights it still can be noted that they weight the different assets in a similar way. Hence to a certain extent the allocation strategies agree, but of course can small differences in the portfolio weights cause major differences in the portfolio profit.

Figure 4.7 shows the normalized Herfindahl index for the seven different strategies, which describes how concentrated the portfolio is. It can be observed that the ER- RC and ER- PCA strategies are fairly balanced throughout the sampling period since the index is close to the value 0.25, which indicates a balanced asset allocation. On the other hand are the ER- fPCA, ER- PCA RC, MVa, and the DRP strategies very concentrated in some periods. This is also consistent when comparing the normalized Herfindahl index to the the portfolio weights in Figure 4.6, where these strategies often assign a large weight to one asset.

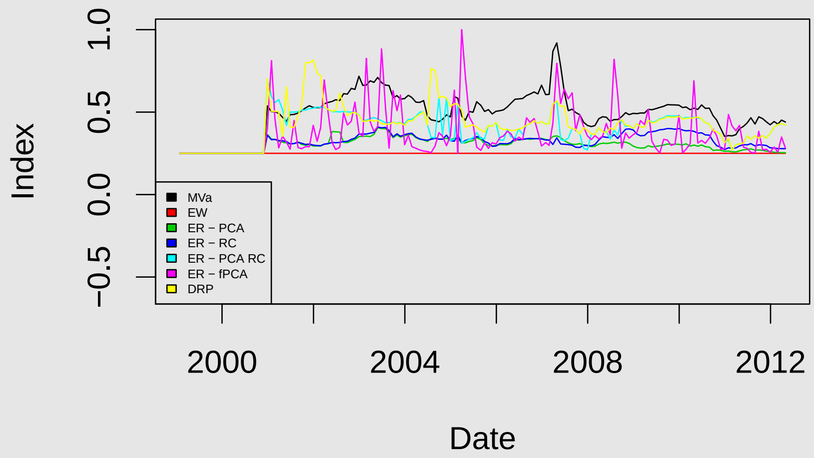


Figure 4.7. Normalized Herfindahl index for the asset allocation strategies. It measures how concentrated the allocations are.

In order to compare the performance of the different strategies, Table 4.8 shows the mean profits, volatilities, the Sharpe ratios, and the turnovers of the portfolio using the different asset allocation strategies.

Table 4.8. Statistics for the optimized portfolio. The bolded numbers indicate the lowest volatility, highest mean profit, highest Sharpe ratio, and the lowest turnover, respectively.

Strategy	Volatility	Mean Profit	Sharpe Ratio	Turnover
Minimum Variance	**5.22**	7.06	1.35	0.07
Equally-Weighted	8.01	6.22	0.78	**0**
Equal Risk - RC	5.74	6.78	1.18	0.053
Equal Risk - PCA	6.03	6.77	1.12	0.046
Equal Risk - fPCA	8.68	6.49	0.75	0.41
Equal Risk - PCA RC	5.59	**7.61**	**1.36**	0.16
Diversified Risk Parity	5.62	7.39	1.31	0.16

As expected from the relation of the portfolio variances shown in Section 3.1.2, the MVa strategy has the least volatile portfolio profit. Moreover, the EW strategy has the lowest turnover, since it always assigns equal weight to every asset. Note that by the definition of turnover as given in equation (4.1) the EW yields a turnover of zero. This is of course not true in practice, since the investor still has to perform trades in order to maintain the strategy as explained in Section 1.1. The ER- PCA RC strategy has the highest profit and also the highest Sharpe ratio. At the same time has this strategy a high turnover, which might cause high transaction costs. So it has to be considered if the higher profit compensates for the high transaction costs that might arise from this strategy. The ER- PCA strategy has also a low turnover and at the same time a moderate mean profit, which might make the strategy preferable compared to the ER- PCA RC strategy. As expected from the plot of the portfolio weights in Figure 4.6 the ER- fPCA has a very high turnover. It also has the highest volatility of all strategies. One way to dim the turnover might be to impose a limit on the replacement of assets in a portfolio, e.g. that only 40% of the portfolio are allowed to be replaced in every rebalancing. This might imply a better performance of this strategy, but is beyond the scope of this thesis.

As mentioned earlier in this section it is interesting to observe the behaviour of the different asset allocation strategies when the estimation window is expanded from two to five years, which will be investigated in the next section.

Backtest based on 60 Months

It can be imagined that some strategies behave different when using a larger estimation window. The estimation of the covariance matrix is now based on 60 instead of 24 observations, which should improve the estimate of the covariance matrix in the ER- RC and MVa strategy. This implies also that the estimates of the eigenvalues should be improved and thereby the risk-based strategies using PCA might be more accurate. Since the sampling period is still the same as in the case of the two year estimation window, the same number of principal components m, smoothing parameter λ, and number of basis functions K are used.

The top panel of Figure 4.8 considers the profit of the different strategies using a 60 month rolling estimation window. It can be seen that the EW strategy most of the time has a higher profit than any other strategy. Moreover, at first glance it could look like the ER- PCA RC performs worse than the other strategies, but it should be noted that the ER- PCA RC performs quite stable throughout the whole backtest, but some of the other strategies yield higher profits in the beginning of the backtest period, which give them an upward shift throughout the whole backtest period. This means that although the other strategies yield higher profits in the backtest it does not mean that they in general perform better. Therefore in order to compare the performance of the different allocation strategies using different estimation windows, it is important to ensure that the estimation starts at the same time in order to avoid misleading shifts in the profit in some periods. Therefore the backtest using a two year estimation window is performed again based on a sampling period from January, 2002 through April, 2012 such that the first estimated portfolio weights of the two and five year estimation are consistent. The changed sampling period implies that the number of principal components m, the smoothing parameter λ, and the number of basis functions K are recalculated. The new parameter estimates are shown in Table 4.9. It can be seen that all parameters are unchanged.

Table 4.9. The chosen number of principal components m for the strategies based on PCA and fPCA for $v = 0.1$. And the chosen number of basis functions K and the smoothing paramter λ for the fPCA approach for the sampling period from January, 2002 through April, 2012.

Strategy	m	K	λ
PCA-based	4	-	-
fPCA-based	3	15	10000

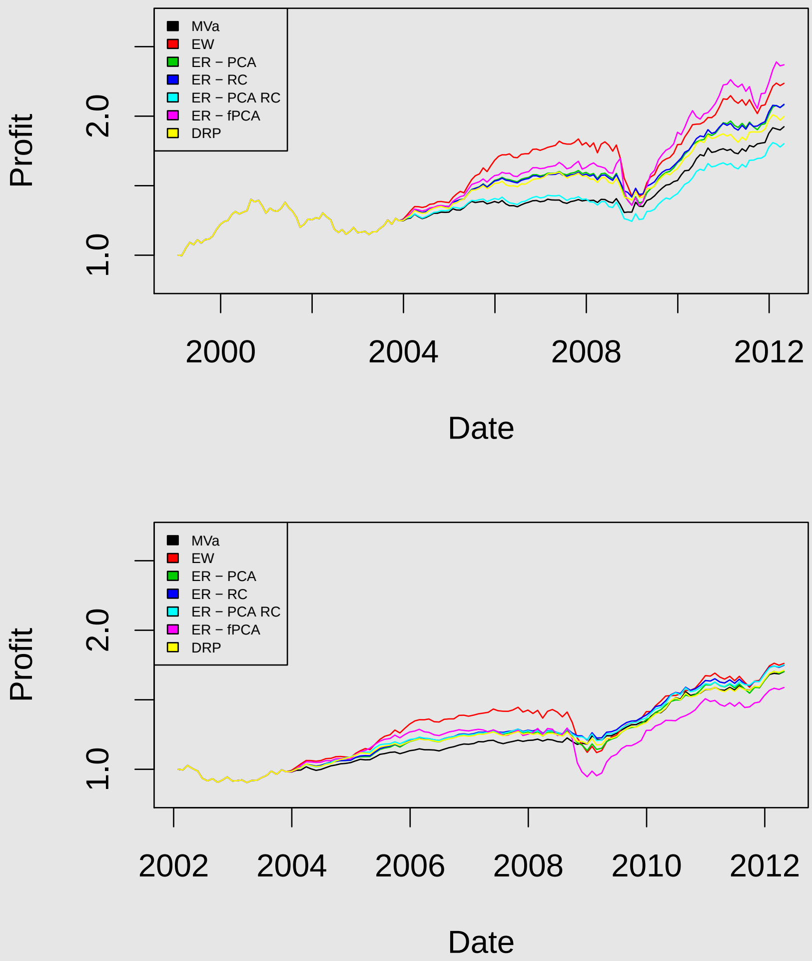


Figure 4.8. Scaled portfolio profits of the different asset allocation strategies. The top panel shows the profit when using a five year estimation window. The bottom panel shows the profit when using a two year estimation window, but where the estimation period starts at the same time as for the five year estimation, i.e. three years later than for the original two year estimation as investigated previously in this section.

The bottom panel of Figure 4.8 shows the profit of the strategies using the modified sampling period for the two year estimation. The performances of the strategies are very similar to each other in the two frameworks using the two year estimation window. Whether to use a two or five year estimation window is dependent on several aspects. The accuracy of the estimation of the covariance matrix and the eigenvalues may be improved by using a larger window. On the other hand, when an event such as the financial crisis in 2008/2009 happens it can be seen from Figure 4.8 that the two year estimation window is better in determining the portfolio weights throughout the crisis. This is due to the fact that a shorter estimation window is better in accounting for the changes that happen right there, whereas a larger window is affected by low volatile times before the crisis, or vice versa. Figure 4.9 shows the normalized Herfindahl index over time for the five year estimation window, where it can be seen that especially the ER- fPCA strategy is very concentrated in some periods.

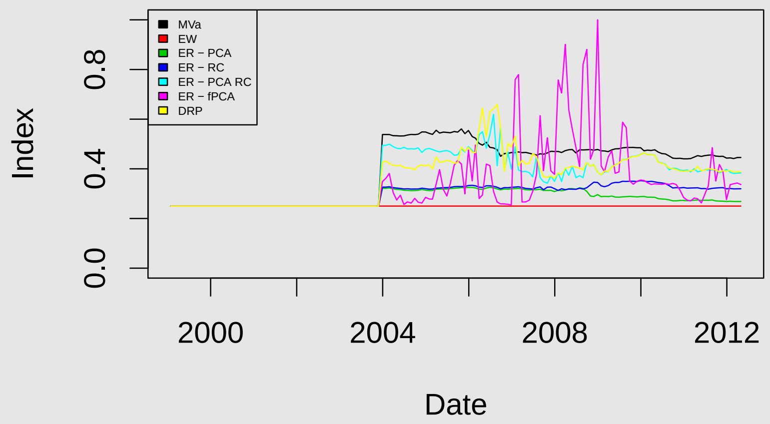


Figure 4.9. Normalized Herfindahl index for the different asset allocation strategies for the five year estimation window. It measures how concentrated the allocations are.

Table 4.10 summarizes the performance of the different strategies using the 60 months rolling estimation window and Table 4.11 shows the result for the 24 month estimation window with a shorter sampling period. Most of the strategies yield higher Sharpe ratios in the two year framework, since the strategies are less volatile during the financial crisis in 2008/2009, but at the same time is the turnover also higher. The ER- PCA RC is the best performing strategy in the two year framework, but one of the worst performing the five year framework. In the five year framework there is not one best performing strategy, but several strategies show preferable characteristics. But the ER- RC yields the highest Sharpe ratio and might therefore be the preferable strategy in this framework.

Table 4.10. Statistics for the optimized portfolio using a five year estimation window. The bolded numbers indicate the lowest volatility, highest mean profit, highest Sharpe ratio, and the lowest turnover, respectively.

Strategy	Volatility	Mean Profit	Sharpe Ratio	Turnover
Minimum Variance	**6.61**	5.03	0.75	0.027
Equally-Weighted	7.58	**6.05**	0.8	**0**
Equal Risk - RC	6.75	5.67	**0.84**	0.017
Equal Risk - PCA	7.05	5.65	0.8	0.014
Equal Risk - fPCA	8.96	6.68	0.75	0.20
Equal Risk - PCA RC	6.82	4.5	0.66	0.11
Diversified Risk Parity	7.12	5.63	0.79	0.13

Table 4.11. Statistics for the optimized portfolio using a modified two year estimation window. The bolded numbers indicate the lowest volatility, highest mean profit, highest Sharpe ratio, and the lowest turnover, respectively.

Strategy	Volatility	Mean Profit	Sharpe Ratio	Turnover
Minimum Variance	**5.33**	5.28	0.99	0.07
Equally-Weighted	7.63	5.63	0.74	**0**
Equal Risk - RC	5.56	5.56	1.00	0.057
Equal Risk - PCA	6.00	5.29	0.88	0.045
Equal Risk - fPCA	8.92	4.58	0.51	0.34
Equal Risk - PCA RC	5.6	**5.58**	**1.00**	0.17
Diversified Risk Parity	5.62	5.34	0.95	0.13

PCA-based Strategies

For the strategies based on PCA: ER- PCA, ER- PCA RC, and DRP, one can consider the cumulative proportion of variance, CPV_j for $j = 1, \ldots, m$ as described in equation (2.35), over time. The CPV_j are determined from a PCA that is performed every month using data from the last two years to extract the latent risk factors embedded in assets of the portfolio.

It can be seen from Figure 4.10 that the cumulative variances increase from about 2008 until 2011 and then drop again, this means that the first principal component accounts for much more variability in this period. Furthermore the plot shows that the first two principal components account for about 75% of the variance most of the time. But throughout the recent financial crisis it can be observed that they actually account for about 90% of the variance. This might question the necessity of considering all principal components as discussed in the beginning of this section.

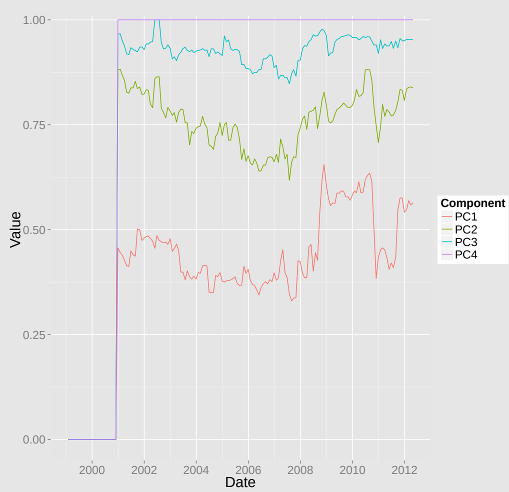


Figure 4.10. Cumulative proportion of variance extracted from the PCAs performed every month based on a two year estimation window.

The question is whether the strategies perform better using a smaller number of principal components? It turns out that first when $v > 0.3$ there only are three principal components considered in the three PCA-based strategies. Therefore a higher value of the parameter v, $v = 0.4$, is chosen such that only three principal components are considered. Then the backtest is performed again for the PCA-based strategies.

Figure 4.11 shows the weights for the three PCA-based strategies using only three principal components. It can be observed that in the case of the ER- PCA strategy the portfolio weights are very similar to the weights found using all principal components as shown in Figure 4.6. Actually, Table 4.12 shows that the ER- PCA strategy yields exactly the same volatility and mean profit as in Table 4.8. This means that in this portfolio the ER- PCA strategy finds the same portfolio weights whether it uses three or four principal components. On the other hand perform the ER- PCA RC and DRP strategy slightly worse, since they are more volatile and yield a lower mean profit when using three instead of four principal components, which can be observed by comparing Tables 4.8 and 4.12. In addition, Figure 4.11 shows that when using three principal components instead of four the portfolio weights for the ER- PCA RC and DRP differ, whereas they were very similar when using four principal components. This conforms with the results obtained in Section 3.4, which made the same observation.

The reason that the ER- PCA yields unchanged portfolio weights, whereas the ER- PCA and DRP change might be due to the different optimization problems for the strategies. In the case of the ER- PCA the sum over the CPV_j is minimized and as can be seen from Figure 4.10 the last principal component does not account for much cumulative variance, hence the optimization yields the same result using three or four principal components. On the other hand incorporate the optimization problems for the ER- PCA RC and DRP strategies the principal portfolio weights $\tilde{\boldsymbol{w}}$ in the minimization of the risk contributions $\tilde{\mathrm{RC}}_i$ and the maximization of the diversification distribution p_i, respectively. The principal portfolio weights give most weight to the component with the lowest variance as stated in equation (3.14), hence when removing one principal component the terms $\tilde{\mathrm{RC}}_i$ and p_i will change, which implies a different asset allocation. Note that since the portfolio considered in this section consists only of four assets, the difference of removing one or more principal components may be larger than removing principal components in a portfolio consisting of a larger number of assets.

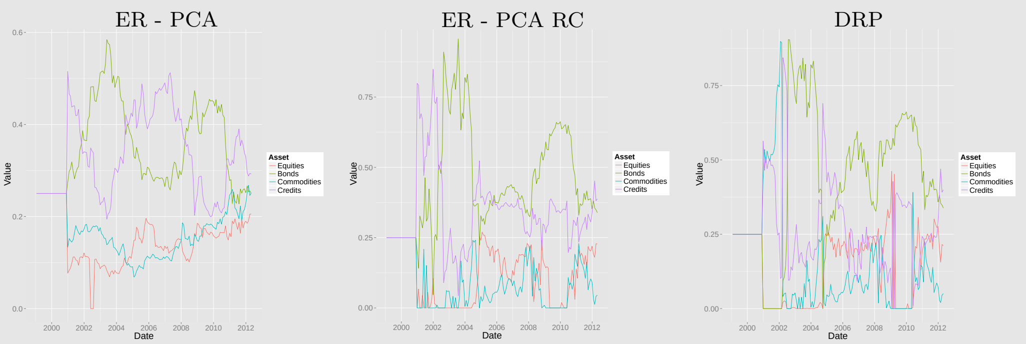


Figure 4.11. The weights of the different strategies estimated on a rolling window of two years and rebalanced every month using only three principal components. Note that the vertical axes differ for the different strategies.

Table 4.12 shows the portfolio statistics for the PCA-based strategies using only three principal components. As mentioned above, the ER- PCA is unchanged, whereas the ER- PCA RC and DRP strategies yield a higher profit volatility and a lower mean profit. From this it is difficult to tell which framework that is most profitable. But since the fPCA also suggests to only consider three components, it might be more appropriate to do that.

Table 4.12. Statistics for the optimized portfolio based on PCA with three principal components.

Strategy	Volatility	Mean Profit	Sharpe Ratio	Turnover
Equal Risk - PCA	6.03	6.77	1.12	0.046
Equal Risk - PCA RC	5.68	7.09	1.25	0.12
Diversified Risk Parity	5.99	6.26	1.05	0.16

After backtesting the different strategies on the portfolio consisting of asset classes it can be concluded that varying the time period of the estimation window changes the performance of the strategies, i.e. mean profit and volatility. In general a five year window is prefered over a two year window in order to improve the estimation of the covariance matrix and the eigenvalues used in the different allocation strategies.

In general it can be noted that the ER- RC and ER- PCA, and the ER- PCA RC and DRP strategies yield very similar portfolio weights when using all principal components, respectively. The weights for the ER- PCA RC and DRP strategies may differ when $m < N$, i.e. the number of principal components is lower than the number of included assets in a portfolio. The weights found by the ER- fPCA strategy differ from the other ER strategies and the strategy has a higher turnover than the other strategies. Therefore it could be considered to impose a limitation on the replacement of the assets, which might improve the overall performance of the ER- fPCA strategy. Nevertheless, the performance of the ER- fPCA strategy is improved by using the five year estimation window. Also the MVa strategy finds other portfolio weights than the other strategies, which is due to the different aim of this strategy, namely to find the mimimum variance portfolio. Furthermore, the ER- fPCA only uses three principal components, while the PCA-based strategies use four principal components when using a low value for the parameter v in equation (2.34). Thereby the ER- fPCA strategy suggests that there are three latent risk factors and the other PCA-based strategies suggest that there are four latent risk factors. As investigated, it might be appropriate to increase the parameter v in order to reduce the number of principal components m.

It should also be noted that the strategies that yield a high profit as shown in Figure 4.5, also have a high turnover. Therefore an investor considering the ER- PCA RC or DRP strategies has to determine if it is worth the transaction costs involved by these strategies. Consequently, a good alternative that trades off turnover and the Sharpe ratio are the ER- RC and ER- PCA strategies. These strategies have a relatively low turnover and at the same time a moderate Sharpe ratio, which means a moderate mean profit and low volatility.

The next section considers an extension of the portfolio considered in this section by including five style factors in addition to the four asset classes as described in Section 4.1.2.

4.2.2 Asset Classes and Style Factors

This section has the objective of backtesting the asset allocation strategies on a portfolio consisting of both asset classes and style factors. Since the style factors are constructed to be very low correlated and to be factors themselves it is interesting to investigate the performance of the PCA-based strategies, that aim to extract independent factors and to see if these strategies can improve the profit compared to the benchmark strategy.

The setup of the backtest in this section is inspired by the results of the previous section. This means that both a two and five year rolling estimation window is considered, but such that the estimation of the portfolio weights of the two different window sizes starts at the same time. This means that the backtest using the two year estimation window is based on the sampling period from January, 2002 through April, 2012, whereas the sampling period for the five year estimation window is from January, 1999 through April, 2012. Furthermore the parameter v in the standard deviation criterium for determining the number of principal components is increased from 0.1 to 0.3. The reason for this choice is to ensure that the considered risk factors are explanatory underlying drivers of the assets and is based on the experiences of the previous section. Nevertheless, the profit statistics for the PCA-based strategies will also be determined using all principal components in order to investigate the differences of omitting components and including all components. But it is maintained that the figures of the scaled profits and the Herfindahl index only are considered for the PCA- and fPCA-based strategies making use of the standard deviation criterium, since it is assumed in this thesis that there have not to be as many underlying risk factors as there are assets in a portfolio. Nevertheless, it is investigated if including the same number of principal components as there are asset makes a difference in the ER- fPCA strategy. The result is that there is no difference in neither portfolios and therefore these results are disregarded.

This implies that the number of principal components m, the number of basis functions K, and the smoothing paramter λ are determined for the two different sampling periods and might be different. The determined parameters can be seen in Table 4.13, where it can be seen that m is identical for the fPCA-based strategy for the two and five year estimation window. On the other hand, m differs in the case of the PCA-based strategies. The number of basis functions differs in the two frameworks, whereas smoothing parameter is identical. So in the two year framework there are used 5 additional basis functions, which is explainable by the variation in data, which is larger considering the modified two sampling period, hence there are needed more basis functions to better cover these variations.

Table 4.13. The chosen number of principal components m for strategies based on PCA and fPCA for $v = 0.3$. And the chosen number of basis functions K and the smoothing paramter λ for the fPCA approach. The parameters are displayed for 'two year estimation winodow/ five year estimation winodow'.

Strategy	m	K	λ
PCA-based	5/6	-	-
fPCA-based	4/4	15/10	10000/10000

Figure 4.12 shows the CPV_j over time for the performed PCAs based on an estimation every month on a two year estimation window, shown in the left panel, and a five year estimation window, shown in the right panel. This figure is especially

interesting in connection with the ER- PCA strategy, which aims to minimize the CPV_j for $j = 1, \ldots, m$. It is observable that the CPV_j's have different paths over time for the two different estimation windows. It is especially interesting to examine the different behaviour during the financial crisis, where in the five year framework there is a drop, which means that there are needed more principal components to describe the same variance, whereas in the two year framework the CPV_j is rising. This means that depending on which estimation window that is used, the exposure of the assets in the portfolio is rising for some risk factors in the five year framework and falling in the two year framework. The shorter estimation window should be better to capture the dynamics, i.e. the latent risk factors, that affect the assets to a specific time since it only examines the dynamics within the past two years. Based on intuition it seems reasonable that during a financial crisis, there might arise additional risk from factors that did not affect the assets that much during normal times, e.g. assets may have a much higher exposure to the volatility risk factor.

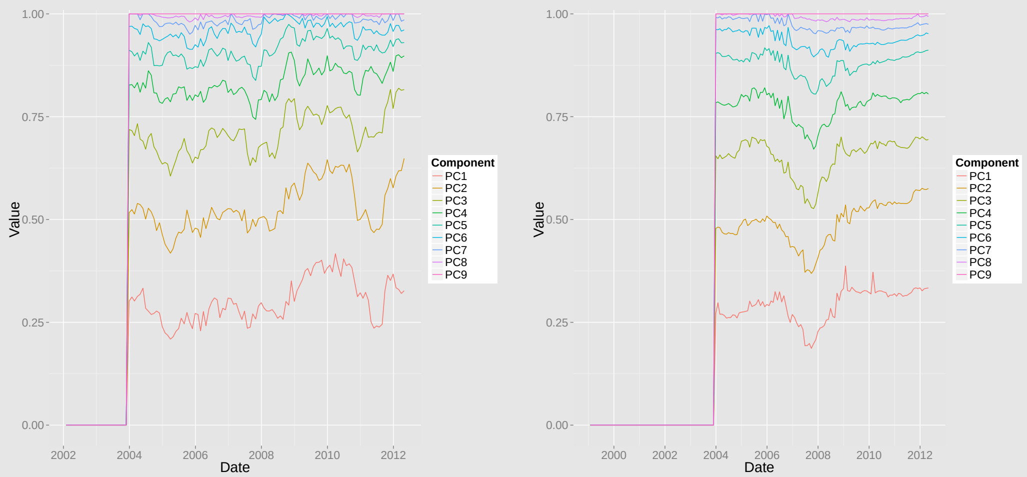


Figure 4.12. Cumulative proportion of variance extracted from the PCAs performed every month on a two year estimation window (left panel) and a five year estimation window (right panel).

The differing exposure to the latent risk factors when using the different estimation windows must imply that the ER- PCA strategy finds different portfolio weights in the two frameworks, which also was the case for the portfolio considered in the previous section since the profits differed. This can be investigated by looking at Figure 4.13, where the left panel shows the weights found in the two year framework and the right panel shows the portfolio weights determined in the five year framework. As expected from the above discussion, the portfolio weights differ.

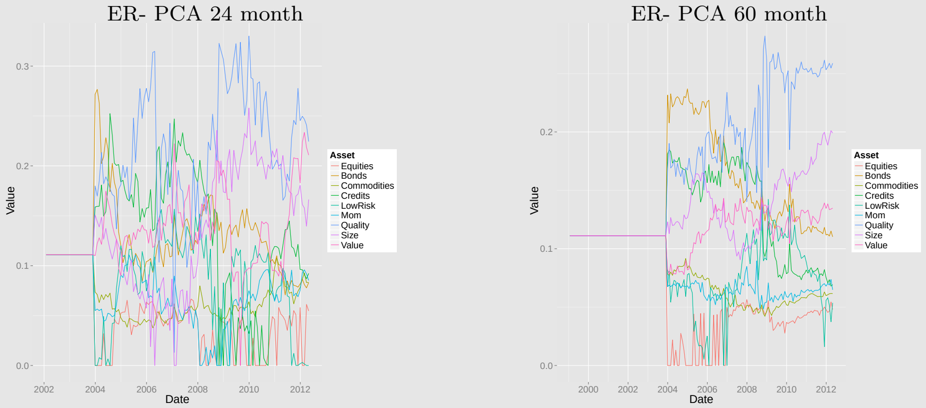


Figure 4.13. The weights of the ER- PCA strategy using a 24 month (left panel) and 60 month (right panel) estimation window, respectively. Note that there are varying vertical axes.

In order to get an idea of how the portfolio weights of the other strategies look like and how diversified the portfolio is during the backtest, the normalized Herfindahl index is inspected. Figure 4.14 shows the index for the two year estimation in the top panel and for the five year estimation window in the bottom panel. For the two year estimation window it is especially the MVa strategy that is concentrated most of the time, but gets less concentrated in the end of the backtest period. The DRP and ER- fPCA strategies are concentrated in some periods, but well-diversified most of the time. In the five year framework it is again the MVa and the DRP strategies that are concentrated, but also the ER- PCA RC strategy is likely to allocate much weight to one asset. It can be concluded that especially the MVa and DRP strategies that often are concentrated in both estimation frameworks and at the same time these strategies have a high turnover, which can be seen in Tables 4.14 and 4.14.

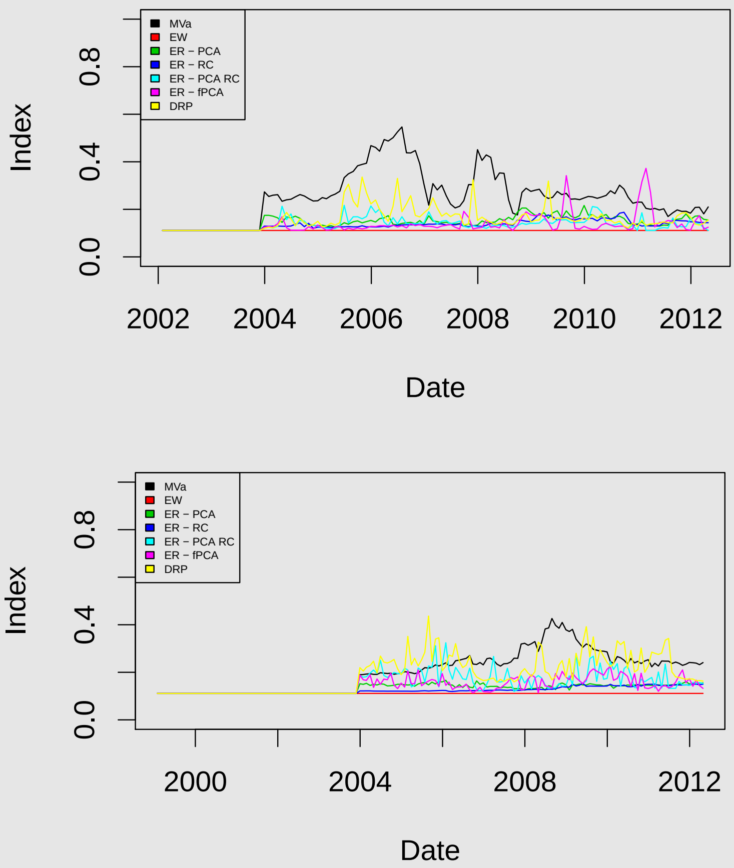


Figure 4.14. Normalized Herfindahl index for the different asset allocation strategies. It measures how concentrated the portfolio is over time. The top panel shows the index for the two year estimation window and the bottom panel shows it for the five year estimation window.

The scaled profits of the different asset allocation strategies using the different estimation windows are shown in Figure 4.15. The top panel shows the profits when using the two year estimation window. It is observable that the MVa strategy performs good and stable throughout the whole backtest period. The ER- RC, ER-PCA, and ER- PCA RC strategies are very similar to each other, whereas the ER-fPCA has a drop during the financial crisis in 2008/2009. The DRP strategies actually beats the EW strategy during and after the financial crisis. On the other hand when looking at the bottom panel of Figure 4.15, which shows the five year framework, it can be seen that the ER- fPCA strategy performs better than the other strategies and in some periods, e.g. after 2009 it beats the benchmark strategy by having higher profits.

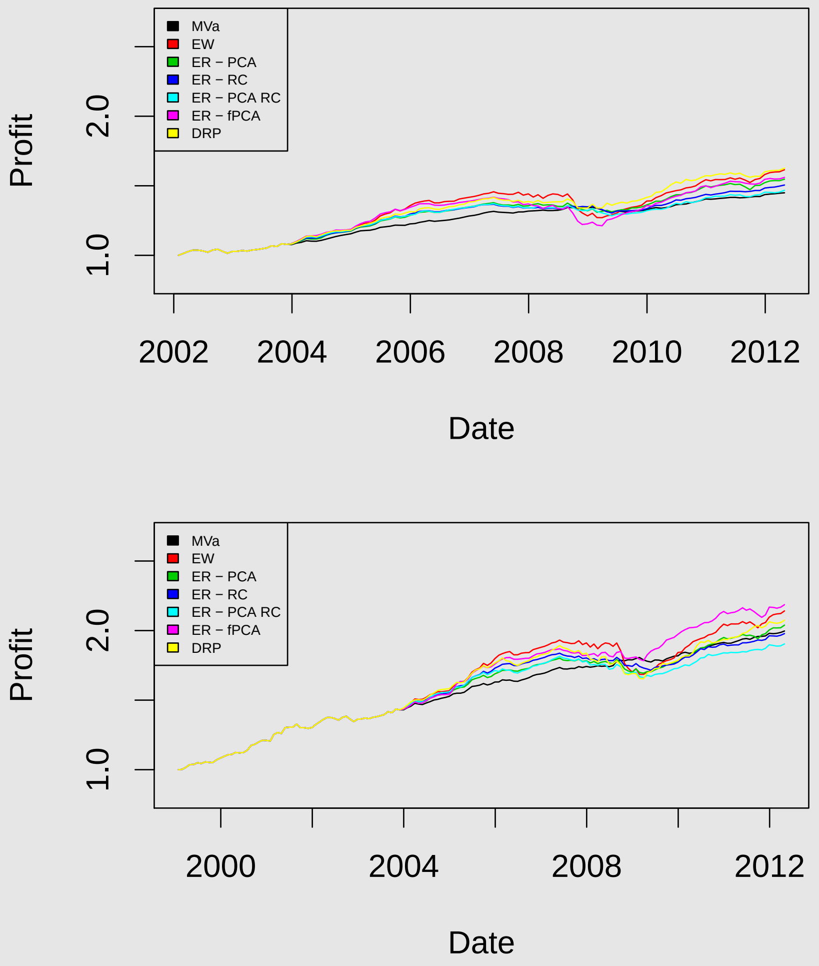


Figure 4.15. Scaled portfolio profits of the different asset allocation strategies. The top panel shows the profit when using a two year estimation window. The bottom panel shows the profit when using a five year estimation window.

Table 4.14 shows the backtest statistics for the different allocation strategies using the two year estimation window. It can be seen that the DRP strategy has the highest mean profit and at the same time the highest turnover. The MVa has the lowest volatility and the highest Sharpe ratio. Using all principal components in the PCA-based strategies lowers the volatility of the profit, but for the ER- PCA RC and DRP strategies it also lowers the mean profit and increases the turnover. Therefore it is concluded that the omitting four principal components is reasonable for the PCA-based strategies.

Table 4.14. Statistics for the optimized portfolio for the two year estimation window. The bolded numbers indicate the lowest volatility, highest mean profit, highest Sharpe ratio, and the lowest turnover, respectively. The numbers in the brackets show the statistics using all principal components.

Strategy	Volatility	Mean Profit	Sharpe Ratio	Turnover
Minimum Variance	**1.88**	3.66	**1.94**	0.14
Equally-Weighted	3.79	4.75	1.25	**0**
Equal Risk - RC	2.26	4.04	1.79	0.07
Equal Risk - PCA	2.7 (2.62)	4.32 (4.45)	1.6 (1.66)	0.13 (0.1)
Equal Risk - fPCA	3.66	4.4	1.2	0.26
Equal Risk - PCA RC	2.54 (2.2)	3.78 (3.56)	1.49 (1.62)	0.19 (0.23)
Diversified Risk Parity	3.17 (2.34)	**4.82** (3.66)	1.52 (1.56)	0.38 (0.43)

On the other hand, Table 4.15 shows the corresponding statistics for the 60 month estimation window. In this setup the ER- fPCA performs very well, since it yields the highest mean profit. It has not the highest volatility as it could be seen in previous setups, but it still has a fairly high turnover. When comparing the results of Tables 4.14 and 4.15 it is observable that the profit gets more volatile for all strategies expect for the ER- fPCA strategy, where the volatility decreases from 3.66 to 3.23, and at the same time the mean profit increases from 3.66 to 5.91, which implies a higher Sharpe ratio. In general all mean profits increase when using the larger estimation window and simultaneously the turnover decreases for all strategies. In this framework there are omitted three principal components in the PCA-based strategies, which also in this case seems reasonable.

Table 4.15. Statistics for the optimized portfolio for the five year estimation window. The bolded numbers indicate the lowest volatility, highest mean profit, highest Sharpe ratio, and the lowest turnover, respectively. The numbers in the brackets show the statistics using all principal components.

Strategy	Volatility	Mean Profit	Sharpe Ratio	Turnover
Minimum Variance	**2.69**	5.52	**1.98**	0.06
Equally-Weighted	3.87	5.88	1.52	**0**
Equal Risk - RC	3.01	5.25	1.75	0.03
Equal Risk - PCA	3.08 (3.07)	5.49 (5.55)	1.78 (1.8)	0.05 (0.04)
Equal Risk - fPCA	3.23	**5.91**	1.83	0.33
Equal Risk - PCA RC	3.18 (3.05)	4.95 (4.78)	1.55 (1.57)	0.2 (0.15)
Diversified Risk Parity	3.7 (2.94)	5.62 (5.18)	1.52 (1.76)	0.31(0.19)

In the 60 month estimation window the ER- fPCA only uses four principal components in contrast to the PCA-based strategies that use six principal components. It could be observed that the ER- fPCA performs very well in the five year framework, which might be explainable by the fact that the ER- fPCA focuses on only four latent risk factors, which seems reasonable when examining the right panel of Figure 4.12, since the first four principal components explain about 80% of the variance almost during the whole backtest period.

As mentioned in the beginning of this section the motive of adding style factors to the portfolio was to investigate the allocation process of the risk-based strategies, since style factors themselves represent risk factors. And indeed there can be observed a different allocation behaviour compared to the portfolio only consisting of asset classes. Mainly, the volatility of the profit is decreased for all strategies. This means that adding the style factors to the portfolio makes the portfolio profits much more stable performing, which is an attractive property for many investors.

4.2.3 Large Portfolio - 21 Assets

The last portfolio considered in the backtest of the different asset allocation strategies consists of 21 assets as introduced in Section 4.1.3. The reason to consider a larger portfolio is in order to investigate the behaviour of especially the PCA- and fPCA-based strategies, since it is interesting to see how many principal components, i.e. how many underlying risk factors, will be considered using a larger portfolio and how these strategies perform compared to the strategies that solely are based on the estimation of the covariance matrix.

There is similar to the previous sections considered a backtest with a two year estimation window, starting from January, 2002, and a five year estimation window using the whole sampling period. Table 4.16 shows the chosen number of principal components using $v = 0.3$, the number of basis functions K, and the smoothing parameter λ for the ER- fPCA strategy for the different estimation windows. It is interesting to see that λ has the same value for both estimation windows, whereas the number of basis functions differs. This was also the case in the previous section, but there were needed more basis functions in the two year framework, whereas here it is in the five year framework. Similar to the previous backtests the fPCA uses less principal components. So in the portfolio consisting of the 21 assets, the PCA-based strategies suggest that there are six or seven latent risk factors driving the assets, whereas the fPCA-based strategy suggests that there only are four latent factors. Note that similar to the previous section the statistics when using all principal components in the PCA-based strategies are displayed in brackets.

Table 4.16. The chosen number of principal components m for strategies based on PCA and fPCA for $v = 0.3$. And the chosen number of basis functions K and the smoothing paramter λ for the fPCA approach. The parameters are displayed for 'two year estimation winodow/ five year estimation winodow'.

Strategy	m	K	λ
PCA-based	7/6	-	-
fPCA-based	4/4	10/25	10000/10000

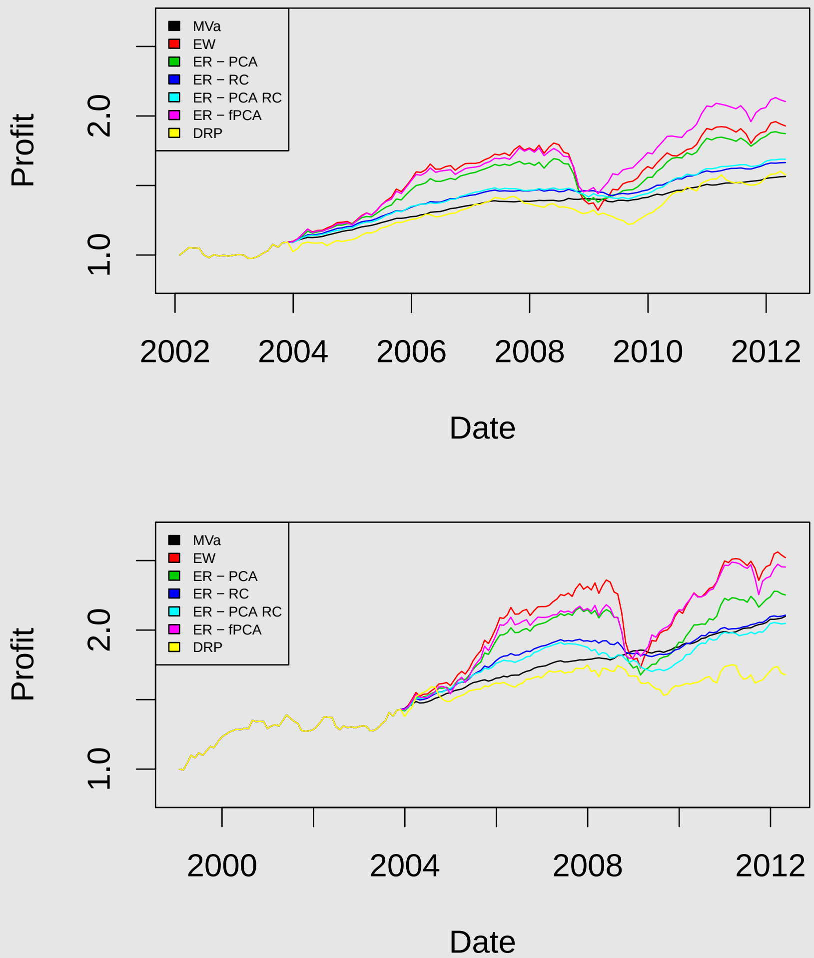


Figure 4.16. Scaled portfolio profits of the different asset allocation strategies. The top panel shows the profit when using a two year estimation window. The bottom panel shows the profit when using a five year estimation window.

Figure 4.16 shows the scaled profits using the two year estimation window in the top panel and the five year estimation window in the bottom panel. Moreover Tables 4.17 and 4.18 show the profit statistics: volatility, mean profit, Sharpe ratio; and the turnover of the different asset allocation strategies. From the figure and the tables it can be seen that the EW, ER- PCA, and ER- fPCA strategies have very similar profits both in the two and five year estimation window. Their profits are increasing until 2008/2009 and then drop during the financial crisis, thereafter the profit increases again. It is notable that although the ER- fPCA and EW stratgies have a large drop in the financial crisis, they manage to outperform the other strategies afterwards. On the other hand perform the MVa, ER- RC, and ER- PCA RC also very similar to each other and have a more moderate, continuously increasing profit with a small drop during the financial crisis. The DRP strategy has a bit lower, but stable profit than the other strategies throughout the whole backtest period in both frameworks.

Table 4.17. Statistics for the optimized portfolio for the two year estimation window. The bolded numbers indicate the lowest volatility, highest mean profit, highest Sharpe ratio, and the lowest turnover, respectively. The numbers in the brackets show the statistics using all principal components.

Strategy	Volatility	Mean Profit	Sharpe Ratio	Turnover
Minimum Variance	**3.23**	4.4	1.36	0.2
Equally-Weighted	7.66	6.56	0.86	**0**
Equal Risk - RC	3.49	4.93	1.41	0.16
Equal Risk - PCA	5.91 (5.8)	6.26 (5.86)	1.06 (1.01)	0.26 (0.25)
Equal Risk - fPCA	6.91	**7.46**	1.08	0.28
Equal Risk - PCA RC	3.78 (3.57)	5.48 (4.62)	**1.45** (1.29)	0.19 (0.21)
Diversified Risk Parity	5.46 (3.67)	4.51 (4.92)	0.83 (1.34)	0.8 (0.8)

Table 4.18. Statistics for the optimized portfolio for the five year estimation window. The bolded numbers indicate the lowest volatility, highest mean profit, highest Sharpe ratio, and the lowest turnover, respectively. The numbers in the brackets show the statistics using all principal components.

Strategy	Volatility	Mean Profit	Sharpe Ratio	Turnover
Minimum Variance	**4.67**	5.74	**1.23**	0.15
Equally-Weighted	7.67	**7.18**	0.94	**0**
Equal Risk - RC	4.96	5.72	1.15	0.08
Equal Risk - PCA	6.61 (6.2)	6.28 (6.63)	0.95 (1.07)	0.2 (0.18)
Equal Risk - fPCA	7.35	6.96	0.95	0.33
Equal Risk - PCA RC	5.13 (4.93)	5.29 (5.73)	1.03 (1.16)	0.09 (0.15)
Diversified Risk Parity	6.38 (5.36)	3.96 (4.52)	0.62 (0.84)	0.4 (0.59)

As also observed in the previous backtests, the ER- RC strategy has a very stable profit, since it has a low volatility. The benchmark strategy, EW, is performing very well for this portfolio, but the ER- fPCA is also performing well and actually to an extend similar to the benchmark strategy. In the two year framework the ER- fPCA also yields higher profits than the EW strategy during and after the financial crisis. Omitting principal components has a negative effect on both the volatility and mean profit, which might indicate that there are omitted too many components.

In addition it can be observed from Figure 4.17 that expect from the DRP and MVa strategy, are the allocation strategies diversified throughout the sampling period. Especially in the five year framework the DRP strategies assigns much weight to specific assets. This may explain the constant, but low profit in the backtest period.

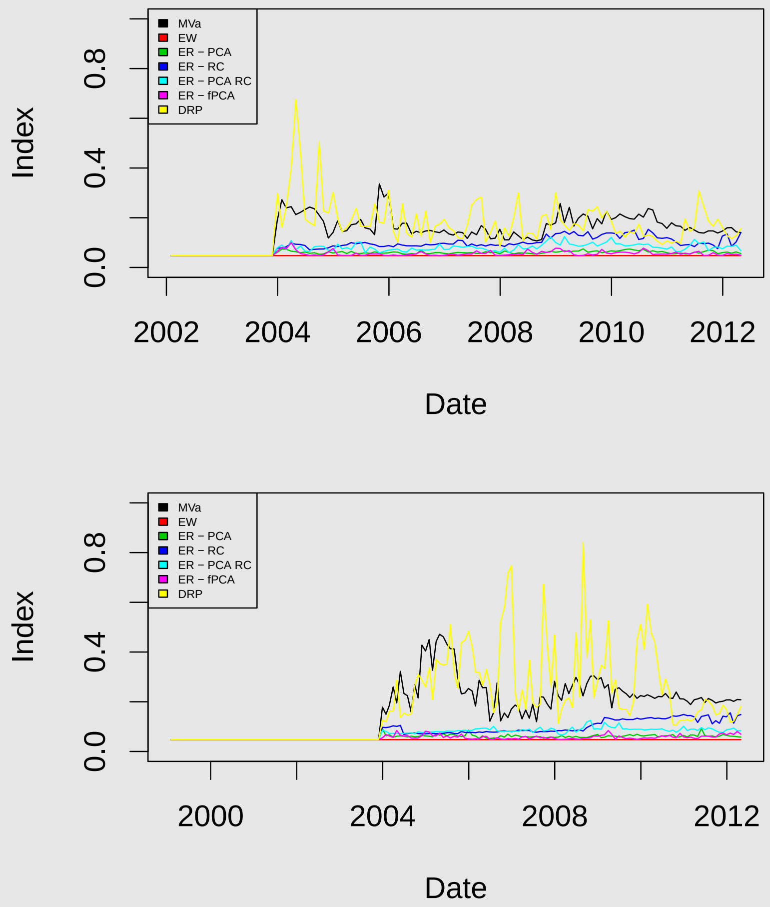


Figure 4.17. Normalized Herfindahl index for the different asset allocation strategies. It measures how concentrated the portfolio is over time. The top panel shows the index for the two year estimation window and the bottom panel shows it for the five year estimation window.

This section shows that the PCA- and fPCA-based strategies only consider a small number of latent risk factors, when using the standard deviation criterium in equation (2.34) with $v = 0.3$, compared to number of assets included in the portfolio. In contrast to the results of the previous section it can be observed in the backtest of this portfolio that including all principal components improves both the volatility and the mean profit of the PCA-based allocation strategies. This may indicate that there are avoided too many principal components and thereby the strategies may perform better when reducing the parameter v. However, there are suggested many different methods for determining the appropriate number of principal components m in litterature, but there is no unique or proofed method given the right number, since most methods are ad hoc and subjective as discussed in Section 2.3.2. But it seems reasonable that the 21 assets are driven by four to seven latent risk factors. Comparing the risk-based allocation strategies it is especially the ER- fPCA strategy that yields high profits, expect from the period during the financial crisis. This could indicate that during the crisis, the ER- fPCA is missing some latent factors that may be included in the PCA-based approaches, since most of these strategies perform more stable during the crisis.

4.3 Review of the Backtest

This section aims to summarize the most important results from the backtest. The focus is on the impact of the size of the estimation window, the effect of omitting principal components, the composition of the portfolios, and the performance of the risk-based strategies compared to the benchmark strategy.

The backtest considered both traditional asset allocation strategies, i.e. the MVa strategy and the simple EW strategy, but also alternative risk-based allocation strategies like the ER and DRP strategies. Three different portfolios were considered: one classical consisting of indexes describing asset classes, one that combines the classical portfolio with style factors, and finally a large portfolio including both classical asset classes, style factors, and additional equity and commoditity indexes.

What can be concluded from the backtest is that the size of the estimation window used to estimate the portfolio weights affects the performance of the asset allocation strategies. A larger window might improve the estimates of the covariance matrix used in the MVa and ER- RC strategies, and thereby also the estimates of the eigenvalues in the PCA-based strategies. But at the same time the allocation in a larger estimation window also is influenced by data that might not capture the dynamics of the current economic situation, i.e. the large estimation window might not response fast enough to the current economical dynamics.
This implies that most of the strategies get more volatile when enlarging the estimation window, since there is a larger drawdown during highly volatile times, which in this backtest period is the financial crisis in 2008/2009. At the same time the strategies yield a higher mean profit in a larger window, hence in normal times a

larger estimation window implies higher profits, but during highly volatile times one should use a shorter window in order to better capture the market dynamics. The effect on the Sharpe ratio is dependent on the composition of the portfolio, i.e. the portfolio consisting of the nine assets has an increased Sharpe ratio when using the larger estimation window, whereas in the other two profits most of the strategies have a lower Sharpe ratio due to the increased volatility. This shows the positive effect of including style factors in a portfolio, which are very low correlated to each other and thereby the portfolio has good diversification properties which imply more stable profits.

Furthermore, differing sampling periods may yield different degree of smoothing in the ER- fPCA strategy. The reason that the smoothing parameter λ may differ is due to the setup of the performed cross validation as described in Section 3.1.3. Since a smaller sampling period implies that the training- and validation sets get smaller in the cross validation, which has a large impact on the choice of the parameters. A different degree of smoothing also means that the number of principal components m differs, since the parameter m is determined based on the smoothed data. The more smoothed data is, the less is the number of principal components, since smoothing removes much of the variability of data and thereby the standard deviations used in equation (2.34) become smaller, hence more principal components are avoided. In the considered backtest the same smoothing parameter is selected for all setups, whereas the number of basis functions differs. This is an interesting observation since the training- and validation sets cover the same time periods, but the portfolios consist of different assets. The selected smoothing parameter is the highest possible value, which is due large flucations in the included assets of the portfolios, in particular the equity and commodity indexes, hence data is more smoothed as explained in Section 2.1.2.
When enlarging the size of the estimation window and the number of assets included in a portfolio, the ER- fPCA yields higher profits during the backtest, expect from the large portfolio, but still has a drawdown during the financial crisis when compared to other risk-based strategies. This shows that the strategy is better in accounting for the variability in data during normal times when considering an appropriate amount of data, i.e. a large enough estimation window. The reason that the strategy not performs that well during the financial crisis in 2008/2009 might be explained by a wrong number of principal components, since this parameter is static throughout the backtest period, but it can be imagined that the number of underlying risk factors increases during a financial crisis. Or data is smoothed too much, which can remove valuable variablility. Therefore it should be considered to make these parameters time-varying.

Excluding principal components that explain little variance in data seems reasonable to do, but the choice of the parameter v that controls the number of principal components is not that straightforward, as discussed in Section 2.3.2, and some of the allocation strategies are rather sensitive to this parameter as will be explained

below.
It could be examined in the backtest of the portfolio consisting of nine assets that omitting components improved the mean profit, but also made the profit more volatile. In the case of the large portfolio consisting of 21 assets it was more profitable to include all principal components than only considering five or six as determined by the criterium in equation (2.34). As mentioned above, this could be caused by the large drawdown in the financial crisis due to missing risk factors that explain the underlying market movements in this period. For the ER- fPCA strategy it can be observed that there is no difference whether to use the number of principal components suggested by the standard deviation criterium or by using as many components as there are assets in the portfolio.
This raises the question of how many principal components that should be included. In order to investigate this, Figure 4.18 shows the Sharpe ratio of the three strategies that are based on PCA using a varying number of principal components by trying different values for the parameter v, i.e. considering $v = 0.1, 0.2, \dots, 0.9$. The reason to consider the Sharpe ratio is to find out which number of principal components that makes the PCA-based strategies most profitable. The results for the portfolio consisting of nine and 21 assets can be seen in the left and right panel of Figure 4.18, respectively.
It can be observed that the three strategies behave very different when changing the number of principal components. The Sharpe ratio of the ER- PCA strategy does not change that much, but decreases a bit when omitting some components. This is caused by the fact that the profit volatility increases while the mean profit decreases in proportion to each other. The ER- PCA RC strategy might yield a higher Sharpe ratio when only including a small number of components, whereas the DRP strategy in the case of the nine assets performs worse by omitting components, but in the case of 21 assets yields the highest Sharpe ratio when considering 11 components. In both strategies the profit gets more volatile and at the same time the mean profit increases, but to different degree.
Studying different methods and thresholds for how many principal components that should be included may improve the performance of the PCA-based strategies, but is beyond the scope of this thesis.

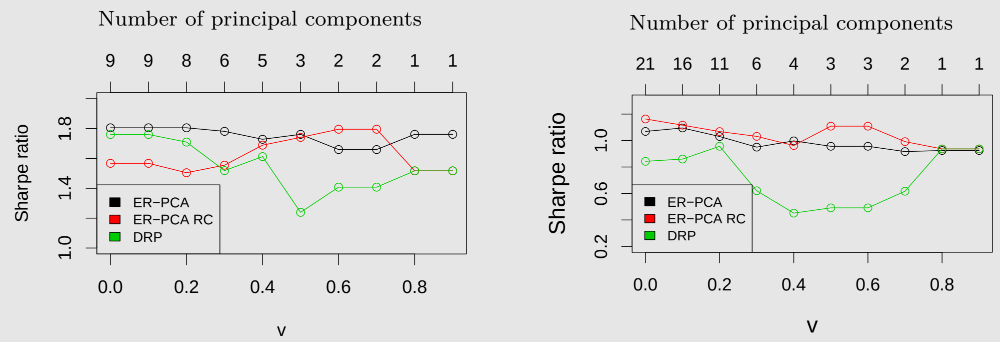


Figure 4.18. Varying the number of principal components by different values of the parameter v and the development in the Sharpe ratio for the PCA-based strategies. The upper axis shows the considered number of principal components.

Table 4.19 summarizes the results of the three different portfolios using the five year estimation window. The reason to consider this estimation window is due to the assumption that the estimates of the covariance matrix and thereby the eigenvalues are improved and that this estimation window yields higher profits compared to the two year estimation window for most of the strategies.

It can be seen that the portfolio consisting of nine assets yields the least volatile profits and the highest Sharpe ratios, which confirms the good diversification properties of adding style factors to the portfolio. The portfolio consisting of 21 assets yields the highest mean profits, expect from the DRP strategy that has the smallest mean profit in this framework, which would be improved by considering all principal components as could be seen in Table 4.18.

Comparing the four different approaches of the ER strategy it can be seen that the ER- RC strategy is the least volatile approach, whereas the ER- fPCA approach has the highest mean profit but at the same time also has the highest turnover of these strategies. Although the four ER strategies are based on the same principal, i.e. that the assets contribute with the same risk in a portfolio, they result in different portfolio weights which is due to their different optimization problems. The two risk contribution approaches, ER- RC and ER- PCA RC, use a sequential quadratic programming, SQP, algorithm to find the optimal portfolio weights, whereas the ER- PCA and ER- fPCA strategies use a Nelder-Mead algorithm. The reason to use the Nelder-Mead algorithm for two of the strategies was caused by a lack of convergence when using constrainted optimization algorithms. Furthermore the PCA in the ER- PCA and ER- fPCA strategies is performed on the weighted log returns, whereas in the ER- PCA RC it is performed on the unweighted log returns.

Table 4.19. Statistics for the optimized portfolio for the five year estimation window for the three different portfolios. The bolded numbers indicate the lowest volatility, highest mean profit, highest Sharpe ratio, and the lowest turnover, respectively.

#Assets	Strategy	Volatility	Mean Profit	Sharpe Ratio	Turnover
	Minimum Variance	**6.61**	5.03	0.75	0.027
	Equally-Weighted	7.58	**6.05**	0.8	**0**
	Equal Risk - RC	6.75	5.67	**0.84**	0.017
4	Equal Risk - PCA	7.05	5.65	0.8	0.014
	Equal Risk - fPCA	8.96	6.68	0.75	0.20
	Equal Risk - PCA RC	6.82	4.5	0.66	0.11
	Diversified Risk Parity	7.12	5.63	0.79	0.13
	Minimum Variance	**2.69**	5.52	**1.98**	0.06
	Equally-Weighted	3.87	5.88	1.52	**0**
	Equal Risk - RC	3.01	5.25	1.75	0.03
9	Equal Risk - PCA	3.08	5.49	1.78	0.05
	Equal Risk - fPCA	3.23	**5.91**	1.83	0.33
	Equal Risk - PCA RC	3.18	4.95	1.55	0.2
	Diversified Risk Parity	3.7	5.62	1.52	0.31
	Minimum Variance	**4.67**	5.74	**1.23**	0.15
	Equally-Weighted	7.67	**7.18**	0.94	**0**
	Equal Risk - RC	4.96	5.72	1.15	0.08
21	Equal Risk - PCA	6.61	6.28	0.95	0.2
	Equal Risk - fPCA	7.35	6.96	0.95	0.33
	Equal Risk - PCA RC	5.13	5.29	1.03	0.09
	Diversified Risk Parity	6.38	3.96	0.62	0.4

Some strategies have a high turnover, which might imply high trading costs when traded in real live. This applies mostly to the ER- fPCA, ER- PCA RC, and the DRP strategies which at the same time also often have a high mean profit. Also here the size of the estimation window has an influence, since a smaller estimation window for most of the strategies implies a higher turnover. The reason for this might be due to larger flucations in the asset returns, which influence the estimates of the covariance matrix and eigenvalues and thereby imply that the portfolio weights have to change more. In order to dim the turnover there could be imposed a limitation on the replacement of the assets when rebalancing the portfolio weights.

The backtest also confirms that the volatility of the ER- RC approach lies between the volatility of the MVa and EW strategies as shown theoretically in Section 3.1.2. But comparing the performance of all strategies to the performance of the EW strategy, it can be seen that no strategy can outperform the EW strategy during the whole backtest period, expect from the ER- PCA RC and DRP strategy using

the portfolio consisting of four classical assets and using a two year sampling period as can be seen in Figure 4.5.
The good performance of the simple benchmark strategy rises the question whether the work of implementing the rather complicated risk-based strategies can pay off? This is difficult to answer, but a great disadvantage of the EW strategy is that it is not possible to control the risk of this strategy. The ER strategies are a good tradeoff between the EW strategy that has a lack of risk monitoring and the MVa strategy that only focuses on risk. Moreover the MVa strategy is not that diversified as measured by the Herfindahl index, which applies in all frameworks and all portfolios, since it gives much weight to low volatile assets.
It should also be noted that during the whole backtest period there has been falling interest rates, which has a positive effect on the returns of bonds. Since bonds are very low volatile the MVa strategy allocates mostly in bonds, i.e. the MVa strategy benefits from the falling interest rates in the backtest period. This should be considered when comparing the different strategies with the MVa strategy, since it is not ensured that the MVa strategy always will perform as good as it does in this selected backtest period.

Besides the performance measured by the profit, turnover, and diversification, of the different asset allocation strategies it is also important to consider the computational effort needed to find the portfolio weights. The computations for estimating the covariance matrix and the eigenvalues in the PCA-based strategies are fairly fast. On other hand, the fPCA-based strategy has to determine additional parameters, i.e. the number of basis functions K and the smoothing parameter λ, which is done by using cross validation. Afterwards data has to be smoothed using these parameters, and then the fPCA can be performed to estimate the eigenvalues needed to find the optimal portfolio weights. This implies that this strategy is more complex and time consuming than the other strategies.

The performance of the risk-based strategies might be significantly improved when estimations errors in the covariance matrix and the eigenvalues are ensured to be minimal. The constraint of long-only portfolio might also have an impact on the performance of the strategies. This is investigated by [Kind, 2013] who reports that allowing long-short portfolios improves the diversification properties of the risk-based allocation strategies, but it does not improve the performance the strategies. It should of course be noted that good diversification of a portfolio does not necessarily imply good performance.
In the backtest it should be considered to include other performance measures besides the volatility of the profit, mean profit, Sharpe ratio, turnover, and Herfindahl index. As can be seen from the above discussion, often more volatile profit can be explained by a large drop in the cumulative return, this could be measured by the drawdown or maximum drawdown.

5 Recapitulation

The main conclusion of this thesis is that most of the introduced risk-based asset allocation strategies outperform the Equally-Weighted strategy for medium-sized portfolios during times of high volatility. I expect the outperformance to be even more significant for larger portfolios, especially for the functional approach of the Equal Risk strategy. Furthermore, I conclude that these strategies perform very stable, which is especially preferable for long-term investors.

The purpose of this thesis has been to establish asset allocation strategies that are based on latent risk factors instead of traditional allocation strategies that allocate by asset classes. The reason to consider latent risk factors is due to the fact that assets consist of latent factors and therefore there should be diversification benefits when allocating in these factors. The motivation to consider alternative, risk-based strategies is among others due to the recent financial crisis in 2008/2009 which showed that many traditional asset allocation strategies that are based on the estimates of mean and variance could not capture the heavy left tail of the return distribution and thereby did not perform well. Therefore it is interesting to consider risk-based strategies, since these strategies allocate by considering the underlying risk factors, hence these strategies might have captured the risks assets where exposed to during the crisis. I have placed much emphasis on projection methods to extract the underlying risk factors, namely principal component analysis and its functional variant, functional principal component analysis.

Principal component analysis is a statistical method to transform the original variable space into a lower dimensional space without losing too much information. The reason to consider a lower dimensional representation of the original variable space is due to the purpose of only considering the important features in data. The aim is to reduce the asset space consisting of assets returns to latent factors, which are represented by the principal components. The principal components are linear combinations of the original asset returns and are constructed such that the first component is variance maximizing and the subsequent components are variance maximizing with the constraint of being orthogonal to the preceding components. Functional principal component analysis is based on the same principles, but considers data to have an underlying functional form that can be modelled by univariate functions. The functional approach accounts for the dependence of the underlying observations over time and thereby models the asset returns' underlying functional process. This implies that the functional data analysis can handle assets with different sampling rates or missing observations, which is a great advantage to multivariate analysis. Thereby eigenfunctions can be modelled over time, which is contrast to principal component analysis, that gives a static, non-temporal estimate of the eigenvectors. Hence it is assumed that the functional framework is better in caputuring the variability in asset returns over time. In order to use functional principal component analysis, data has to be transformed to functional data. The transformation is done using basis functions, which are the counterpart to basis vectors in functional space. Since asset returns are non-periodic, I introduced B-splines as basis functions. Furthermore a technique for smoothing data has been described which uses a roughness penalty and is based on weighted least squares. After studying these methods to extract risk factors, some risk-based asset allocation strategies have been introduced.

The main focus of this thesis was on the Equal Risk strategy which covers over different techniques. One way to consider this strategy is by choosing the portfolio weights of the single assets in a portfolio such that every asset contributes with the same risk to the portfolio. This approach can either be considered in the original asset space or it can be transformed into the principal space by using principal component analysis. Another way to consider this approach is to select the portfolio weights of the single assets in a portfolio such that a principal component analysis performed on the asset returns yields as equal as possible volatility in the principal component directions. This approach was also performed using functional principal component analysis. Note that the Equal Risk strategy can be considered as a tradeoff of the Minimum Variance and Equally-Weighted strategies in terms of the portfolio variance, i.e. the risk associated when using this strategy.

For the purpose of investigating the performance of the different asset allocation strategies I considered a backtest on historical data. A walk-forward backtest with rolling estimation window was used to compute the portfolio weights every month, where different sizes of the estimation window were analyzed. Using the assumption that short-selling is not allowed, three different portfolios were backtested. These portfolios included a different number of assets and different types of assets in order to investigate the behaviour of the different strategies. Moreover, I focused on investigating the impact of omitting some principal components, i.e. to reduce the number of underlying risk factors. Selecting a method and a threshold for deciding on how many principal components that should be omitted is a difficult choice and I found ambivalent results regarding this issue. In some cases the mean profit of a risk-based strategy that uses principal component analysis could be improved by omitting components with the cost of a higher volatility of the profit, which in all cases led to a reduced Sharpe ratio compared with the performance of the strategies that used all principal components. Further research on this topic could improve the performance of the risk-based strategies that are based on principal component analysis.

From the backtest it could also be concluded that the different risk-based strategies perform different when varying the size of the estimation window and the composition of the portfolio. During normal times a larger estimation benefits the risk-based strategies, but during highly volatile periods a shorter estimation window should be used in order to capture the exposure to the risk factors present during this period. The risk-based strategies most of the time had a lower volatile profit than the Equally-Weighted strategy, the benchmark strategy. At the same time the benchmark strategy often had a higher mean profit than the risk-based strategies.

There is still room for improvement of the risk-based strategies. There may be placed more emphasis on the estimation of the covariance matrix, since small estimation errors can cause large deviations in the computation of the risk contributions or in the principal component analysis and thereby in the selected portfolio weights. Furthermore, the process of smoothing data in the functional approach may be investigated more comprehensively since this also can cause a completely different allocation. In the setup of the backtest it should be considered to make the parameters time-varying, i.e. to compute the number of principal components, the number of basis functions, and the smoothing parameter for every estimation. This might improve the principal component analysis-based strategies, since it seems reasonable that the number of underlying risk factors varies over time, and that data has to be smoothed to different degree over time in the functional approach of the Equal Risk strategy. This has been disgraded in this thesis due to lack of time and thereby simplification of the implementation of the backtest. It could also be considered to implement the other principal component analysis-based strategies using the functional approach, since there only is transformed one Equal Risk approach to make use of functional principal component analysis. Furthermore, in the setup of the backtest there still are disregarded observations for assets with different sampling rates in the functional approach for ease of the implementation of the backtest, although it theoretically is possible to include all observations. Hence including all available observations of all assets might improve the functional approch. This thesis focuses on the standard deviation as the measure of risk, but it could be investigated if other measures of risk, e.g. the quantile-based value at risk, VaR, or shortfall-based measures, could be used in the computation of the risk contributions in the Equal Risk strategy. Furthermore, other projections methods could be considered, e.g. independent component analysis.

The usage of risk-based asset allocation strategies is not that widespread yet, but several Nordic investors allocate a part of there assets in alternative, risk-based strategies. To mention some of them: the Danish pension fund PKA, the Swedish pension fund SPK, and the Swedish state pension funds AP2 and AP3. The Danish pension fund PKA changed their equity strategy to also allocate in alternative, risk-based strategies in 2012. [Liinanki, February/March 2015] Also the government pension fund of Norway, SPU, also called 'The Oil Fund' is known to use factor investing. It is one of the world largest pension funds, which at the end of 2012 managed $685 billion. [Ang, 2014] The reason that many investor still are hesitating to use risk-based allocation strategies is that it can be time consuming to understand and implement these allocation strategies. [Liinanki, February/March 2015]

Bibliography

A. Meucci, A. Santangelo, R. Deguest. Measuring Portfolio Diversifcation Based on Optimized Uncorrelated Factors. 2014.

A. Alexanderian. A Brief Note on the Karhunen-Loeve Expansion, 2013.

A. Ang. *Asset Management: A Systematic Approach to Factor Investing*. Oxford University Press, 2014.

T. Björk. *Arbitrage Theory in Continuous Time*. Oxford University Press, 2009.

R. Cangelosi and A. Goriely. Component Retention in Principal Component Analysis with Application to cDNA Microarray Data. *Biology Direct*, 2007.

M. A. Crain. A Literature Review of the Size Effect. 2011.

P. Craven and G. Wahba. Smoothing Noisy Data with Spline Functions. *Numerische Mathematik*, 31:377 – 403, 1979.

M. Guan. Principal Component Analysis in Portfolio Management, 2013. URL `https://systematicedge.wordpress.com/2013/06/02/principal-component-analysis-in-portfolio-management/`.

H. Lohre, H. Opfer, G. Orszag. Diversifying Risk Parity. *JEL Classification*, 2013.

R. J. Hyndman. Fast Computation of Cross-validation in Linear Models, 2014. URL `http://robjhyndman.com/hyndsight/loocv-linear-models/`.

J. Friedman, T. Hastie, R. Tibshirani. *The Elements of Statistical Learning*. Springer-Verlag New York, 2009.

J. O. Ramsay , B. W. Silverman. *Functional Data Analysis*. Springer Science+Business Media, Inc., 2005.

J. O. Ramsay, G. Hooker, S. Gaves. *Functional Data Analysis with R and MATLAB*. Springer Science+Business Media, 2009.

J. O. Reilly, S. O. Barba, N. Pavlov. Growth vs. Value Trading Strategies, 2003. URL `https://faculty.fuqua.duke.edu/~charvey/Teaching/BA453_2003/CPAM/HTML/Index.htm`.

J. Peng and D. Paul. A Geometric Approach to Maximum Likelihood Estimation of the Functional Principal Components From Sparse Longitudinal Data. *Journal of Computational and Graphical Statistics*, 2012.

J. Y. Campbell, A. W. Lo, A. C. MacKinlay. *The Econometrics of Financial Markets*. Princeton University Press, 1997.

K. V. Mardia, J.T. Kent, J. M. Bibby. *Multivariate Analysis*. Academic Press, 1994.

C. Kind. Risk-Based Allocation of Principal Portfolios. 2013.

A. Lesniewski. Interest Rates and FX Models - 3. Girsanov, Numeraires, and All That. 2008.

C. Liinanki. Lessons and Experiences from the Nordic Pioneers. *Nordic Fund Selection Journal*, February/March 2015.

F. Longin. The Choice of the Distribution of Asset Returns: How Extreme Value Theory can Help? *Journal of Banking and Finance*, 29:1017–1035, 2005.

D. G. Luenberger. *Investment Science*. Oxford University Press, 2009.

M. Grinblatt and S. Titman. *Financial Market and Corporate Strategy*. McGraw-Hill Higher Education, 2002.

S. Mesomeris. Risk Premia in Asset Allocation, 2013. URL https://secure.cfauk.org/assets/2891/Spyros_Mesomeris___A_new_asset_allocation_paradigm.pdf.

A. Meucci. Managing Diversification. *Bloomberg Education & Quantitative Research and Education Paper*, 2010.

L. J. Overby. CAPM, Factor Models and APT, 2010. URL http://www.juloverby.dk/courses/cfi/slides/mean_var_3.pdf.

S. Maillard, T. Roncalli, J. Teiletche. On the Porperties of Equally-weighted Risk Contributions Portfolios. 2009.

G. A. F. Seber. *Multivariate Observations*. WILEY, 2004.

SP Dow Jones Indices. SP 500, 2015. URL http://us.spindices.com/indices/equity/sp-500.

D. Stefanovits. *Equal Contributions to Risk and Portfolio Construction*. Master Thesis, ETH Zurich, 2010.

T. E. Copeland and J. F. Weston. *Financial Theory and Corporate Policy*. Addison-Wesley Publishing Company, 1988.

D. Tasche. Capital Allocation to Business Units and Sub-Portfolios: the Euler Principal. 2008.

T. Tvedebrink. Slides from MAT8 Data Mining Course, 2014.

J.-T. Zhang. *Analysis of Variance for Functional Data.* CRC Press, 2014.

E. Zivot. Introduction to Computational Finance and Financial Econometrics - Return Calculations, 2015. URL `http://faculty.washington.edu/ezivot/econ424/returncalculationslides.pdf`.

Appendix A

A.1 Hilbert Spaces

In order to define a Hilbert space, which is an infinite dimensional vector space, firstly there is given a formal definition of an *inner product*.

Definition A.1 (Inner Product)
Consider a real vector space $\mathcal{H}$. A mapping $\langle \cdot, \cdot \rangle : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ is called an inner product on $\mathcal{H}$ if it has the following properties:

- For any $\alpha, \beta \in \mathbb{R}$ and $\boldsymbol{f}, \boldsymbol{g}, \boldsymbol{h} \in \mathcal{H}$

  $$\langle \alpha \boldsymbol{f} + \beta \boldsymbol{g}, \boldsymbol{h} \rangle = \alpha \langle \boldsymbol{f}, \boldsymbol{h} \rangle + \beta \langle \boldsymbol{g}, \boldsymbol{h} \rangle$$

  and

  $$\langle \boldsymbol{h}, \alpha \boldsymbol{f} + \beta \boldsymbol{g} \rangle = \alpha \langle \boldsymbol{h}, \boldsymbol{f} \rangle + \beta \langle \boldsymbol{h}, \boldsymbol{g} \rangle .$$

  This means that an inner product is bilinear.
- In addition, it is symmetric:

  $$\langle \boldsymbol{f}, \boldsymbol{g} \rangle = \langle \boldsymbol{g}, \boldsymbol{f} \rangle , \quad \text{for all } \boldsymbol{f}, \boldsymbol{g} \in \mathcal{H}.$$

- And it is positive definite:

  $$\langle \boldsymbol{f}, \boldsymbol{f} \rangle \geq \boldsymbol{0} \text{ for all } \boldsymbol{f} \in \mathcal{H} \text{ with equality if and only if } \boldsymbol{f} = \boldsymbol{0}.$$

[Björk, 2009]

Definition A.1 shows that the inner product is a generalization of the standard scalar product on $\mathbb{R}^n$. The definition below shows how a norm and the concept of orthogonality also can be generalized to an infinite dimensional vector space.

Definition A.2

- Let $\boldsymbol{f} \in \mathcal{H}$, then for any vector $\boldsymbol{f}$ the norm is denoted by $\|\boldsymbol{f}\|$ and defined by:
$$\|\boldsymbol{f}\| = \sqrt{\langle \boldsymbol{f}, \boldsymbol{f} \rangle}.$$
- Two vectors $\boldsymbol{f}, \boldsymbol{g} \in \mathcal{H}$ are said to be orthogonal if $\langle \boldsymbol{f}, \boldsymbol{g} \rangle = \mathbf{0}$.
- For any linear subspace $M \in \mathcal{H}$ the orthogonal complement is defined to be:
$$M^{\perp} = \{\boldsymbol{f} \in \mathcal{H} : \boldsymbol{f} \perp M\}.$$

[Björk, 2009]

A vector space with an inner product is called an *inner product space*, which leads to the following definition of a Hilbert space.

Definition A.3 (Hilbert Space)
A Hilbert space is an inner product space which is complete, i.e. every Cauchy sequence is convergent, under the induced norm $\|\cdot\|$ in Definition A.2. [Björk, 2009]

A generalization of the idea of matrices in infinite-dimensional spaces are *Hilbert-Schmidt operators* as given in the following definition.

Definition A.4 (Hilbert-Schmidt operator)
Let $D \subset \mathbb{R}^n$ be a bounded domain. The function $k : D \times D \to \mathbb{R}$ is called a Hilbert-Schmidt kernel if:
$$\int_D \int_D |k(x, y)|^2 dxdy < \infty,$$
that is, $k \in \mathcal{H}$. Define the integral operator K on $\mathcal{H}$, $K : u \to Ku$ for $u \in \mathcal{H}$, by:
$$[Ku](x) = \int_D k(x, y)u(y)dy. \tag{A.1}$$
The mapping K is called a Hilbert-Schmidt operator. [Alexanderian, 2013]

The following lemma states that the Hilbert-Schmidt operator is a compact operator, i.e. some theory concerning matrices can be applied to this operator.

Lemma A.5
Let D be a bounded domain in $\mathbb{R}^n$ and let k be a Hilbert-Schmidt kernel. Then the integral operator $K : \mathcal{H} \to \mathcal{H}$ given by equation (A.1) is a compact operator. [Alexanderian, 2013]

Proof. Omitted. □

The following theorem generalizes the result that any positive semi-definite matrix is the Gram matrix of a set of vectors to Hilbert spaces.

Theorem A.6 (Mercer's Theorem)
Suppose that $k(s,t)$ is a symmetric, continuous, and nonnegative definite kernel function on $D \times D$. Suppose further that the corresponding Hilbert-Schmidt operator K is positive. Then there exists an orthonormal set of eigenfunctions $\xi_j(x)$ and eigenvalues λ_j such that:

$$k(s,t) = \sum_{j=1}^{\infty} \lambda_j \xi_j(s) \xi_j(t),$$

where convergence is absolute and uniform on $D \times D$. [Alexanderian, 2013]

Proof. Omitted. □

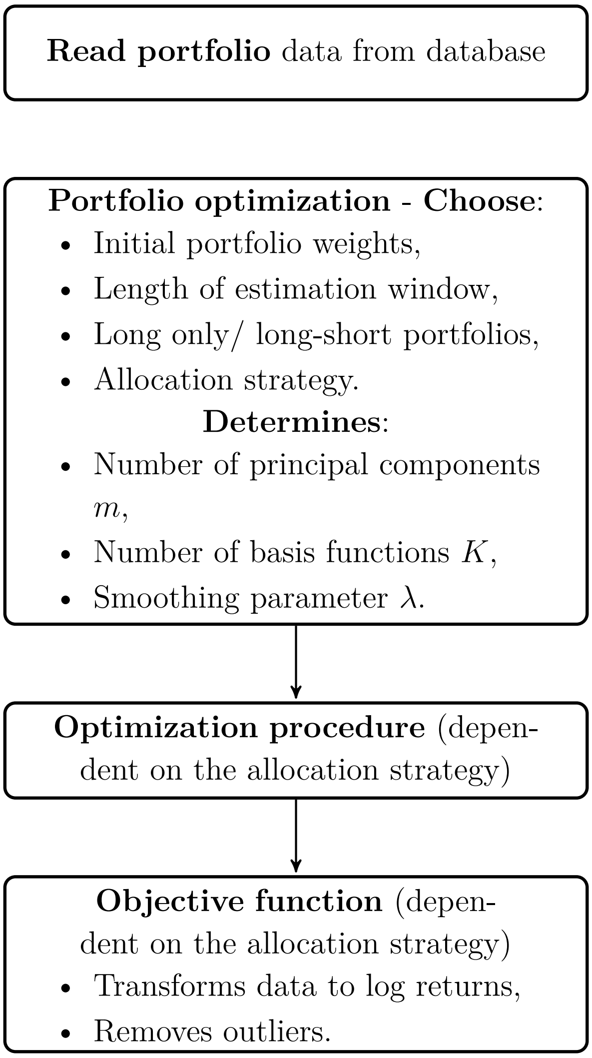
Compact self-adjoint operators on infinite dimensional Hilbert spaces act very similiar to symmetric matrices.

Theorem A.7 (Spectral Theorem)
Let $\mathcal{H}$ be a Hilbert space and let $T : \mathcal{H} \to \mathcal{H}$ be a compact self-adjoint operator. Then, $\mathcal{H}$ has an orthonormal basis $\{\xi_i\}$ of eigenvectors of T corresponding to eigenvalues λ_i. [Alexanderian, 2013]

Proof. Omitted. □

A.2 Flow Chart: R Code

The above flow chart illustrates the programming flow in R.



First a portfolio data is read, where every series constains a date and a price. The next function is used to backtest an asset allocation strategy as described in Section 4. In order to run a backtest one has to select between several options as described in the flow chart. At the same time determines this function parameters that are used in the optimization procedures, e.g. the number of principal components used in the PCA-based strategies.

The **portfolio optimization** function calls the **optimization procedure**, which depends on the chosen allocation strategy. The **objective function** to be optimized is in another function. It is also here that data is transformed to log returns and that outliers are removed.

It is possible to determine portfolio weights based on a specific period of time by skipping the second step, **Portfolio optimization**, and just using the **Optimization procedure**.