Utilizing Machine Learning to Address Noise in Covariance and Correlation Matrices

An Application and Modification of Enhanced Portfolio Optimisation

Master Thesis

16th of May 2021

38 CBS pages

85.290 characters

Copenhagen Business School

M.Sc. Applied Economics and Finance

Spring 2021

Supervisor

Theis Ingerslev Jensen

Author

Bjarne Timm (133274)

Abstract

Portfolio Optimisation is at the core of Asset Management since the invention of the Mean Variance Portfolio Optimisation by Harry Markowitz (1952). Its theoretical usefulness as well as its practical flaws have been studied by several academics. The recent developments within the Machine Learning environment enabled researchers to use modern technologies to find solutions on how to make Mean Variance Optimisation work. Enhanced Portfolio Optimisation has been coined and termed by Pedersen et al. (2021). It makes use of Machine Learning through the unsupervised algorithm Principal Component Analysis to detect noise and structure in the underlying correlation matrices of portfolios. By shrinking the correlation matrix towards the identity matrix, they realise substantially higher Sharpe ratios than their benchmarks. In their study, they are able to effectively address the problem of estimation noise. This finding cannot be confirmed by this thesis. Instead, it shows that their strategy yields inferior Sharpe ratios than the classical Mean Variance Optimisation. A modified version of the Enhanced Portfolio Optimisation is proposed by shrinking towards the average correlations instead of the identity matrix. This approach appears to be superior to the original approach. However, the Mean Variance Portfolio as well as the equally weighted portfolio are tough benchmarks to beat. The main finding is displayed by the dependence of the shrinkage parameter on the prevailing economic cycle, as well as the dependence of Enhance Portfolio Optimisation on the estimation of the correlation matrix.

Keywords: Enhanced Portfolio Optimisation, Mean Variance Optimisation, Machine Learning, Principal Component Analysis, Shrinkage

Acknowledgement

This project represents a 30 ECTS Master Thesis in quantitative finance at Copenhagen Business School. First and foremost, I would like to thank my supervisor Theis Ingerslev Jensen for motivating me to take on this topic and supporting me throughout the whole process. Last, I am grateful for all the support and motivation from my friends and family.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AMEX	American Stock Exchange
ARMA	Autoregressive moving average
CAPM	Capital Asset Pricing Model
CCC	Constant Conditional Correlations
CML	Capital Market Line
DCC	Dynamic Conditional Correlations
EPO	Enhanced Portfolio Optimisation
MVO	Mean Variance Optimisation
NASDAQ	National Association of Securities Dealers Automated Quotations
NCF	Nonlinear Common Factor
NYSE	New York Stock Exchange
PCA	Principal Component Analysis

1. Introduction

Harry Markowitz has been awarded the 1990 Nobel Memorial Prize in Economic Sciences for having developed the theory of portfolio choice in 1952 (Riksbank, 1990). In said theory, Markowitz studies how wealth should be invested when assets differ in terms of risk and expected return. The theoretical contribution is vast, still influencing modern portfolio theory (Berk & DeMarzo, 2014). Opposite to the academic value of Markowitz study, the practical usefulness has not been proven yet. Michaud (1989) was the first to ask why practitioners do not rely on the optimisation techniques proposed by Markowitz. Ang (2012) and several others before him argue that the answer can be found in the high reliance on input parameters to *Mean Variance Optimisation* (*MVO*).

Several approaches have been undertaken to make *MVO* functional, working either on the estimated inputs or on the technicalities proposed by Markowitz. Jagannathan et al. (2003) state that the solution to solving *MVO* is always found in the covariance matrix. Random Matrix Theory and, more recently, Machine Learning helped to gain new insights and develop new approaches to portfolio optimisation. The latest attempt has been undertaken by Pedersen et al. (2021) who established *Enhanced Portfolio Optimisation* (*EPO*). Pedersen et al. and López de Prado (2020) make use of Machine Learning through *Principal Component Analysis*, an unsupervised algorithm to differentiate between random structures and signal contained in the underlying data.

Pedersen et al. (2021) further show that the same result obtained with *Principal Component Analysis* can be achieved by shrinking the correlation matrix towards the identity matrix. By doing so, they argue that the impact of noise in the covariance matrix as well as the expected returns can be significantly reduced. In their simulations, their *EPO* realised higher Sharpe ratios than the equally weighted portfolio and the classical Mean Variance Portfolio. Pedersen et al. (2021) do not motivate why they shrink towards the identity matrix. Depending on the dataset used to simulate the performance of the *EPO*, the underlying true correlation will also differ and thus require a different amount of shrinkage.

This thesis focuses on modifying the *EPO* provided by Pedersen et al. (2021) by shrinking the correlation matrix towards the average correlation of all assets. This approach enables investors to bypass the simulation to figure out the optimal amount of shrinkage needed. This approach is found to be more efficient in predicting Sharpe ratios, yet it produces lower realised Sharpe ratios than the original *EPO*.

The underlying research question is to show if and how Machine Learning technologies can help addressing the problem of noise in covariance matrices when optimising portfolios. First, this thesis presents the theoretical foundations of portfolio choice theory, *Principal Component Analysis*, and estimation noise. Second, *EPO* is introduced. Third, the specific methodology of this thesis is discussed, specifying how *EPO* can be modified to potentially achieve greater estimation accuracy. Next, the results of the analysis and testing are outlined. Lastly, the results are discussed, and a conclusion is drawn.

2. Theoretical Foundations

The fundamental cornerstone of modern portfolio optimisation is the *MVO* as formulated by Markowitz (1952). Its basic idea is to make use of the co-movements of different assets to create diversification effects within a portfolio and thereby enable an investor to balance the rate of return and corresponding risk of the portfolio in any desired way. Before outlining the details of *MVO*, the basics of portfolio theory are explained.

2.1 Portfolio Theory

Two fundamental assumptions are required to explain portfolio theory. First, investors are assumed to prefer high expected returns over low expected returns, ceteris paribus. Second, investors always prefer a low variance over high variance, ceteris paribus. In short, investors are assumed to be greedy and risk averse. Consequently, the investor will always choose the portfolio with the lowest variance amongst all portfolios offering the same expected return. Similarly, investors will always choose the portfolio with the highest expected return amongst all portfolios with the same level of variance.

The Sharpe ratio combines both assumptions by expressing an expected rate of excess return for each unit of risk. A rational investor will always prefer any asset with a higher Sharpe ratio over an asset with a lower Sharpe ratio because each unit of risk is rewarded with a higher rate of excess return by the asset with a higher Sharpe ratio. The Sharpe ratio of an asset can increase by either lowering the risk or an increase of the expected excess return. However, such a scenario is unlikely as the investor is usually compensated for each unit of risk incurred. Therefore, it is more likely that the expected return increases once the variance of a portfolio increases to account for the higher risk of the portfolio. Similarly, investors can expect to obtain lower returns in case of a lower total variance of the portfolio.

Once further assets are included into the investor's portfolio, the relationship of fluctuations of the returns of all different assets needs to be taken into consideration. The co-movements of assets are captured by the covariance and the correlation of assets. The covariance describes the direction of the relationship of the return of assets. If the covariance is positive, one can generally assume the returns of the assets move into the same direction. In contrast, with a negative covariance one can assume that asset returns usually move in opposing directions. Correlation also measures how asset returns are related to each other but is bounded by -1 and 1. A correlation of 1 means that the returns of both assets always move into the same direction. With a correlation of -1 both assets have exactly opposing returns. A perfect negative or positive correlation does not mean that both assets have the same level of return. Instead, it means that the sign of the return is always the same, or always the opposite. A correlation coefficient of 0 can be interpreted as there being no detected relationship between the two assets, their returns are therefore considered to be independent of each other.

The formulas to calculate volatilities, correlations, and covariances are presented in matrix notation where the superscripts $^{\boldsymbol{T}}$ and $^{\mathbf{-1}}$ indicate the transpose and the inverse of a vector or matrix. Matrices and vectors are printed in bold throughout the thesis. The correlation of assets x and y can generally be defined as:

$$\rho = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma(x_i - \bar{x})^2 \Sigma(y_i - \bar{y})^2}} \qquad (01),$$

with x_i and y_i being the individual returns at day i of both assets, and $\bar{x}$ and $\bar{y}$ referring to the means of each series of returns. The pairwise correlations of all assets can then be transformed into matrix form through the correlation matrix $\boldsymbol{\Omega}$.

The variance of each individual asset is calculated as:

$$\sigma^2 = \frac{\sum_{t=1}^{n}(x_i - \bar{x})^2}{n-1} \qquad (02),$$

with n being the number of observations in the sample. The standard deviation or volatility is calculated through the square root of the variance $\sigma = \sqrt{\sigma^2}$. Lastly, the covariance matrix $\boldsymbol{\Sigma}$ showing the pairwise covariances on the off diagonals and the individual variance on the diagonals is calculated as:

$$\boldsymbol{\Sigma} = \boldsymbol{\sigma}^{\boldsymbol{T}} \boldsymbol{\Omega} \boldsymbol{\sigma} \qquad (03).$$

Investors can benefit from the co-movements of assets through diversification of risk. This is best illustrated with an example: An investor has the choice between two assets. Asset A has an expected return of 6% and a volatility of 15%. Asset B also has an expected return of 6% and a volatility of 15%. If the investor chooses to invest into a single asset only, the expected return of the portfolio is 6% with a volatility of 15%. However, if the investor chooses to invest in both assets the co-movement of both assets needs to be considered. For the sake of simplicity of this illustration both assets are assumed to be perfectly negatively correlated. In that case, whenever one asset has a positive return, the other asset has a negative return. The expected return of any portfolio with both assets included is always 6%, irrespective of how the weights are divided. The risk, however, can be entirely eliminated due to the perfect negative correlation of both assets. With an even split of both assets, it is possible to achieve a risk of 0% while maintaining the expected return of 6%. As described above, this portfolio is always preferred by investors as they achieve the same expected return with less risk.

Nevertheless, perfectly negatively correlated assets are rarely existing making the above example impractical. Yet, diversification effects can still be achieved with any other correlation of assets. For example, a correlation of -0.5 would decrease the portfolio variance in the above example to 7.5%. The volatility of a portfolio consisting of two assets can generally be calculated as:

$$\sigma_p = \sqrt{w_1^2 * \sigma_1^2 + w_2^2 * \sigma_2^2 + 2 * w_1 * w_2 * \rho_{1,2} * \sigma_1 * \sigma_2} \qquad (04),$$

with the correlation between the two assets given by $\rho_{1,2}$, the weights of each asset w_i, and the volatilities of the assets σ_i. The formula used to compute expected return of portfolios r_p is:

$$r_p = \sum_{i=1}^{N} r_i * w_i \qquad (05),$$

where the weights w_i of each asset are multiplied with the corresponding expected returns r_i.

While the return formula is applicable for any number N of assets within a portfolio, the formula used to calculate the risk needs to be adjusted. Each asset has a different correlation to all other assets. Hence, each pair of assets is assigned a specific correlation. The formula for the risk of a portfolio is therefore:

$$\sigma_p^2 = \boldsymbol{w}^{\boldsymbol{T}} \boldsymbol{\Sigma} \boldsymbol{w} \qquad (06),$$

where $\boldsymbol{w}$ is the vector of weights and $\boldsymbol{\Sigma}$ is the variance-covariance matrix.

Using the formulas for portfolio risk and return, investors can combine assets and calculate the expected return and risk of the portfolios. The resulting set of risk and return can be plotted as illustrated in Figure 1. By combining the different assets, it is possible to achieve any combination of risk and return on the plotted line. However, one can differentiate between efficient and inefficient portfolios. Efficient portfolios are those on the solid blue line, while those on the dashed orange line are considered as inefficient. For a given level of risk, any point on the orange part has a counterpart on the blue part of the line, which means that a portfolio with the same level of risk but a higher expected return exists. These portfolios are always preferrable to those on the orange line. The blue line is referred to as the efficient frontier as it represents the efficient set of weight allocations within the portfolio.

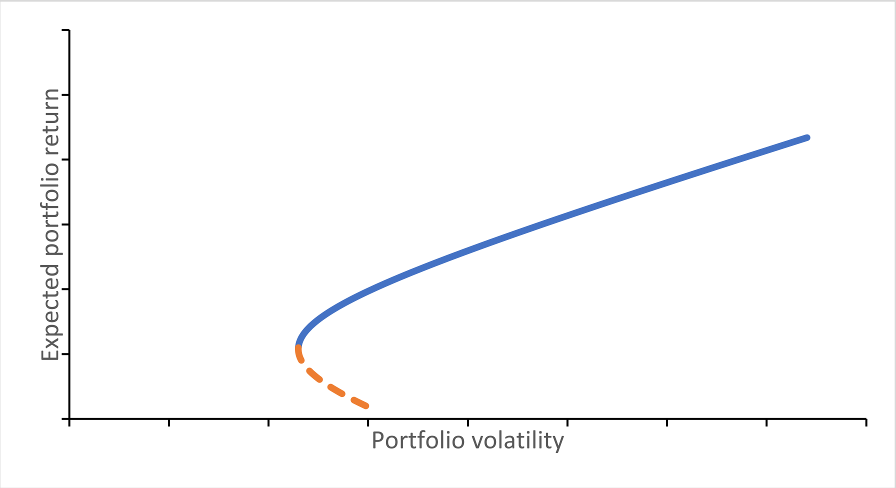

Figure 1 Efficient and inefficient portfolio allocations (own illustration)

2.2 Markowitz' Mean Variance Optimisation

Two portfolios that are of particular interest in portfolio optimisation are the minimum variance portfolio and the maximum slope portfolio. The minimum variance portfolio can be observed in Figure 1 as the point the furthest to the left, where the blue and the orange lines meet. The maximum slope portfolio is not as easily detected. Before plotting the maximum slope portfolio, the intuition is explained. The slope of the graph in Figure 1 can be translated as total expected portfolio return (y-axis) divided by portfolio volatility (x-axis). The slope hence describes how much return is achieved per unit of risk. Identifying the portfolio with the highest slope therefore corresponds to identifying the portfolio offering the highest expected return per unit of risk. One important note on the maximum slope portfolio is that this portfolio does not automatically have the maximum Sharpe ratio of all portfolios. The Sharpe ratio measures excess return per unit of variance. The maximum slope portfolio instead shows the portfolio with the highest share of total expected return per unit of risk.

Mathematically, the minimum variance portfolio is found by minimizing the following objective function:

$$\min_{w} \boldsymbol{w}^{\boldsymbol{T}} \boldsymbol{\Sigma} \boldsymbol{w} \qquad (07),$$

subject to all weights summing up to one:

$$\mathbf{1}^{\boldsymbol{T}} \boldsymbol{w} = 1 \qquad (08).$$

This can be achieved by using the Lagrangian approach and the Lagrangian multiplier λ:

$$L = \boldsymbol{w}^{\boldsymbol{T}} \boldsymbol{\Sigma} \boldsymbol{w} + \lambda * (1 - \mathbf{1}^{\mathbf{T}} \boldsymbol{w}) \qquad (09)$$

The next step is to set the first derivative of L with respect to w equal to 0 and solving the equations for λ. After inserting the found λ back into equation L, the equation can be solved for w:

$$w_{min} = \frac{\boldsymbol{\Sigma}^{-1} \mathbf{1}}{\mathbf{1}^{\boldsymbol{T}} \boldsymbol{\Sigma}^{-1} \mathbf{1}} \qquad (10),$$

which is the portfolio with the lowest achievable risk amongst all set of portfolios. Shaw et al. (2008) show the detailed steps of the computations above. This portfolio is shown in Figure 2 as the most left point on the efficient frontier, indicated by the blue triangle on the frontier. Furthermore, the maximum slope is shown as the yellow diamond. The maximum slope portfolio is the portfolio where the dashed black line touches the frontier. The mathematical derivation on how to find the maximum slope portfolio is outlined in the following.

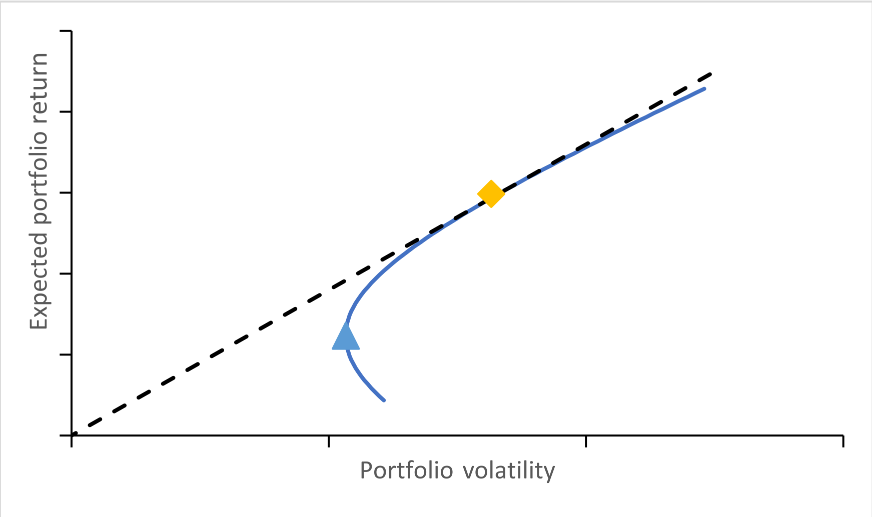

Figure 2: Minimum variance and maximum slope portfolio mapped on the efficient frontier.

The most striking difference between the formula on how to calculate the minimum variance portfolio and the maximum slope portfolio is seen in the inputs. While the objective function of the minimum variance portfolio contains the weight vector and the variance-covariance matrix as inputs, the objective function of the maximum slope portfolio also includes the return variable $\boldsymbol{\mu}$. $\boldsymbol{\mu}$ is a vector of expected returns of all assets. The objective function for the maximum slope portfolio looks as follows:

$$\max_{w} \frac{\boldsymbol{w}^T \boldsymbol{\mu}}{\sqrt{\boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w}}} \qquad (11),$$

subject to all weights summing up to one:

$$\boldsymbol{1}^T \boldsymbol{w} = 1 \qquad (12).$$

Again, by setting the first derivative of the Lagrangian function with respect to the weights $\boldsymbol{w}$ equal to 0, one obtains a function which can be solved for $\boldsymbol{w}$. The ultimate formula which solves for the weights of the maximum slope portfolio is:

$$w_{max} = \frac{\boldsymbol{\Sigma}^{-1} \mu}{\boldsymbol{1}^T * \boldsymbol{\Sigma}^{-1} \mu} \qquad (13).$$

Now, both formulas have been derived to calculate the portfolio with the lowest variance and the portfolio with the steepest slope.

As already pointed out above, the maximum slope portfolio does not take excess returns into account. Overall, interest rates have so far been excluded from the theoretic derivation of the portfolio choice.

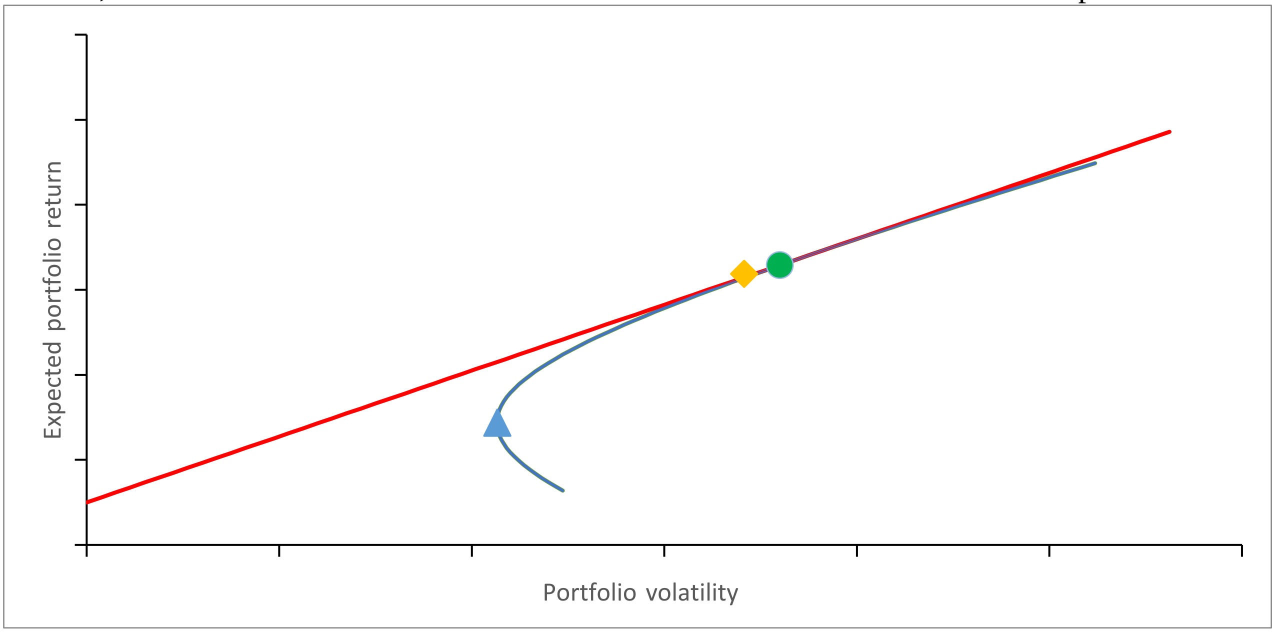

Figure 3Adding the risk-free rate and tangency portfolio to the model (own illustration).

By including the interest rate to the available set of assets, investors are confronted with new possibilities in terms of weight allocation, risk and return trade-offs. The interest rate is referred to as the risk-free rate, enabling investors to achieve a level of return without having to incur any risk at all. Investors are therefore able to construct their portfolios with assets and interest rate instruments such as bonds. The implications of including interest rates as risk-free assets to the portfolio choice theory are that the efficient frontier in Figure 1 and 2 can be further adjusted. The red line in Figure

3 replaced the dashed black line which used to go from the origin through the maximum slope portfolio in Figure 2. The red line starts at the expected return of the interest free rate. Furthermore, the portfolio, which is the tangent point of the red line and the efficient frontier, is shown as the green dot. The red line shows a new set of attainable portfolios by combining the tangent portfolio with the risk-free rate. The maximum slope portfolio, for example, is now considered to be inefficient a portfolio slightly above on the red line with a higher expected return exists. The red line is called the Capital Market Line (CML). The slope of the CML depicts the Sharpe ratio of the tangent portfolio. The derivation of the tangency portfolio is similar to the derivation of the maximum slope portfolio. However, instead of maximising the total return for each unit of risk, the expected excess return for each unit of risk is maximised. Since the tangency portfolio is the only portfolio consisting of only risky assets that lies on the CML, rational investors should ignore all other portfolios from the efficient frontier. The risk aversion should further dictate how much weight is put on the tangency portfolio and how much on the interest rate.

The invention of Markowitz' portfolio theory explained above still has a major influence on modern finance and is taught at universities as the way to optimize portfolios. However, investors face difficulties when trying to implement *MVO* in practice. One of the problems is that *MVO* typically tells investors to build highly leveraged positions on those portfolios with a low estimated variance. The reason for that leverage is that those portfolios are expected to achieve higher returns per unit of risk than the other assets. As explained above, investors are compensated for the amount of risk they assume. This is captured by the Capital Asset Pricing Model (*CAPM*) developed by William Sharpe in 1964. The *CAPM* states that the return of an asset depends on its underlying risk. It should therefore not be possible for the *MVO* to identify assets which yield higher returns per unit of risk than others. The underlying problem of these weight allocations is found in estimation errors. Estimation errors occur when estimations are based on samples from an underlying population. The most important difficulty of *MVO* is known to be estimation errors in the variance-covariance matrix and expected returns. This notion constitutes the focus of this thesis and will be explained in greater detail in the subsequent sections.

2.3 Principal Component Analysis

Principal Component Analysis (*PCA*) is one of the most commonly used unsupervised machine learning algorithms. It further serves as the foundation towards understanding dimensionality reduction (Shlens, 2014). Explaining how *PCA* works, is best done through an example. A correlation matrix of ten assets is a 10×10 matrix has $\frac{10*10}{2} = 50$ correlations to estimate. When applying *PCA* to that matrix, *PCA* decomposes that matrix into a new $10\ x\ 10$ matrix, where each column is now called an eigenvector. This process is called *Feature Extraction* (Abdi et al., 2010). Each of the newly created eigenvectors consists of a combination of weights from the old variables in the initial correlation matrix. The eigenvectors are created in such a way that each eigenvector has a correlation of zero to all other eigenvectors. Further, each eigenvector is assigned a specific eigenvalue which shows how much of the total variance is explained by each eigenvector (Ringnér, 2008). Eigenvectors can be ranked by their eigenvalues, so that the first eigenvector is the eigenvector explaining the largest part of the variance within the underlying data. The eigenvalue can be translated into a percentage by dividing an eigenvalue by the sum of all eigenvalues. The resulting percentage shows how much of the variance can be explained by the corresponding eigenvector. *PCA* therefore provides a way to extract the features explaining the largest share of the variance without deleting any variables but instead reshuffling the weights of the underlying variables (López de Prado, 2020).

In more detail, eigenvectors represent directions, just like the best-fitting line in a basic regression analysis. An eigenvector shows a particular direction in a scatterplot of data, while eigenvalues represent the magnitude or importance of the eigenvectors. The larger the eigenvalue, the higher the importance of the direction of the corresponding eigenvector. The reason why *PCA* is of relevance for this thesis is its ability to split variance into signal and noise and differentiate between signal and noise through the eigenvalues. Noise is the term used in machine learning environments to describe estimation errors. Usually, a lot of variance within one direction indicates an underlying signal or feature that can be detected and used for estimation purposes (Wold et al., 1987). Low variance instead indicates that the underlying feature is random and composed of noise. This is highly relevant as *MVO* takes high leveraged positions in the low estimated variance portfolios as they are expected to be low risk. However, if that variance is estimation noise, one can safely assume the realised volatility to be a lot higher than expected. This is why *MVO* is generally not used in practice.

2.4 Estimation Errors (Noise)

The terms estimation errors and noise are used interchangeably throughout this thesis. Noise is the difference between an estimation and the true underlying parameters of the population. The parameters which are estimated in *MVO* are the expected returns, correlations, as well as variances and covariances. Hence, all these estimated parameters are exposed to estimation noise. Generally, noise can be reduced by increasing the sample size. With infinite data available for an estimation, the estimation converges to the underlying true population values, which has been proven through the Central Limit Theorem (Heyde, 2014).

However, the problem with using return data is that not many data points are available. One data entry per day does not suffice to assume that the estimation based on daily returns converges to the true parameters. That problem could potentially be avoided by taking more years of data into account. Nevertheless, the underlying true parameters may also change over time. Hence, when taking ten years of data into account for an estimation, there is a chance that the underlying parameter changed and is not the same at the end of the estimation period as in the beginning. Longin et al. (1995) prove this in their study on correlation coefficients of international equity returns. Ball et al. (2000) further confirm that finding in their study. This trade-off is addressed in section 5.

To demonstrate the origin of noise, a simulation with three assets is performed. In that simulation the true underlying population values of expected returns, volatilities, correlations, and the corresponding covariances are defined. Based on those values, a random multivariate simulation is performed with the mean, correlation, variance, and covariance inputs taken from the population. This enables to compare the estimations to the true underlying values. The population parameters are defined as:

$$population\ mean = \begin{bmatrix} 0.1 & 0.05 & 0.075 \end{bmatrix}$$

$$population\ volatility = \begin{bmatrix} 0.3 & 0.19 & 0.25 \end{bmatrix}$$

$$population\ correlation = \begin{bmatrix} 1 & 0.2 & 0.8 \\ 0.2 & 1 & 0.4 \\ 0.8 & 0.4 & 1 \end{bmatrix}$$

The true covariance matrix can then be calculated as:

$$population\ covariance = (population\ volatility)^T\ (population\ correlation)\ (population\ volatility).$$

Resulting in

$$population\ covariance = \begin{bmatrix} 0.09 & 0.0114 & 0.06 \\ 0.0114 & 0.0361 & 0.019 \\ 0.06 & 0.019 & 0.0625 \end{bmatrix}.$$

The simulation is a multivariate normal estimation based on the parameters from above, simulating 20 yearly returns. It is performed using numpy in Python 3.7.6.

$$simulation = numpy.random.mutivariate\_normal(population\ mean, population\ covariance, 20)$$

The sample statistics are then calculated based on the simulated returns (rounded to fourth digit):

$$sample\ mean = [0.1238 \quad 0.0366 \quad 0.0813]$$

$$sample\ volatility = [0.291 \quad 0.1349 \quad 0.2306]$$

$$sample\ covariance = \begin{bmatrix} 0.0847 & 0.0065 & 0.0493 \\ 0.0065 & 0.0182 & 0.0092 \\ 0.0493 & 0.0092 & 0.0531 \end{bmatrix}$$

$$sample\ correlation = \begin{bmatrix} 1 & 0.1646 & 0.7346 \\ 0.1646 & 1 & 0.2968 \\ 0.7346 & 0.2968 & 1 \end{bmatrix}$$

One can see the differences between the sample statistics and the population statistics. This difference is calculated below for each parameter:

$$(population\ mean - sample\ mean) = [-0.0238 \quad 0.0134 \quad -0.0063]$$

$$(population\ volatiltiy - sample\ volatiltiy) = [0.0090 \quad 0.0551 \quad 0.0194]$$

$$(population\ covariance - sample\ covariance) = \begin{bmatrix} 0.0053 & 0.0049 & 0.0107 \\ 0.0049 & 0.0179 & 0.0098 \\ 0.0107 & 0.0098 & 0.0093 \end{bmatrix}$$

$$(population\ correlation - sample\ correlation) = \begin{bmatrix} 0 & 0.0035 & 0.0065 \\ 0.0035 & 0 & 0.0103 \\ 0.0065 & 0.0103 & 0 \end{bmatrix}$$

The differences between the sample estimations to the underlying values of the population should, in the absence of noise, all be zero. Instead, they show that noise, stemming from the estimation, exists. Estimation noise in the context of portfolio optimisation has been addressed and researched by several academics. Among others, Michaud (1989) asks why *MVO* is not used by practitioners, which Jorion (1992) answers with the inability of *MVO* to recognize estimation risk. The question arising next is how to differentiate between the signal contained in the estimations and the noise, reasoning why the Marčenko-Pastur distribution is explained next.

2.5 The Marčenko-Pastur Distribution

The Marčenko-Pastur distribution is part of the random matrix theory and describes the distribution of eigenvalues of a completely random matrix (Marčenko & Pastur, 1967). Its mathematical derivation and proof are beyond the scope of this thesis. Instead, the distribution of eigenvalues is

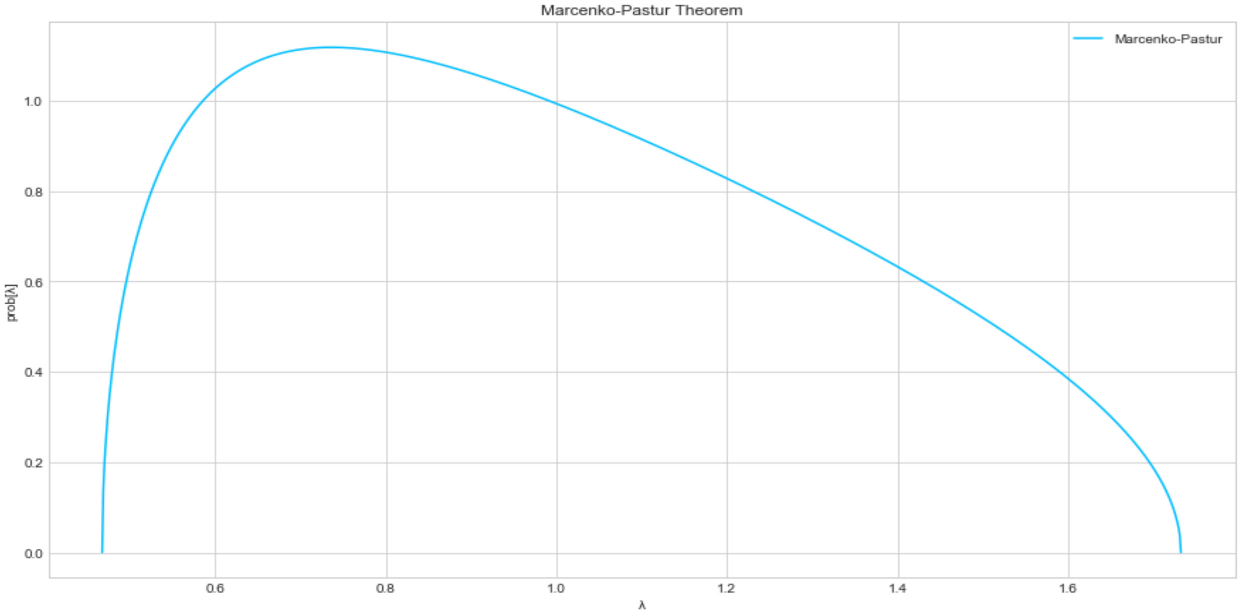

Figure 4 The Marčenko-Pastur distribution of eigenvalues (own illustration)

used to differentiate between those eigenvalues of random eigenvectors and those eigenvalues of non-random eigenvectors. The probability density function of the distribution can be seen Figure 4 below, where the eigenvalues are labelled as λ on the x-axis and the probabilities on the y-axis. Figure 4 displays how eigenvalues are distributed if the underlying matrix on which *PCA* has been performed is entirely made of noise and thus random. The graph in Figure 4 is therefore used as a blueprint to identify those principal component portfolios that are not entirely random. Those portfolios not matching the Marčenko-Pastur distribution hence possess some underlying structure that is able to explain the variance of the distribution.

3. *Related Literature*

3.1 How to make Mean Variance Optimisation work

The theoretical value of *MVO* is unquestioned. Its usefulness, however, is just as clearly rejected due to the impact of estimation noise. The challenge on how to make *MVO* work in practice has been of great concern for many academics. The approaches put forward by academics and researchers thus far can be classified into two categories. One part takes the input criteria as given and tries to work on the mechanisms of *MVO*. The second category looks at the input values of *MVO* and attempts to reduce the impact of noise by reducing the noise in the data itself before calculating the different weights of the assets. Even though the approaches differ in where they tackle the problem of estimation errors, Jagannathan and Ma (2003) find that all approaches can be interpreted as fixing the variance-covariance matrix. Recently, modern technologies enabled researchers to look for other opportunities to solve the problems of *MVO*. Machine learning technologies were used by some scholars to optimize the input datasets and achieve a superior result of *MVO*.

Mechanisms used to adjust the input variables before performing *MVO* include random matrix theory, shrinkage, and resampling. Further, Black and Litterman propose to merge the individual views of investors into the *MVO* calculations. Random matrix theory has been applied by Laloux et al. (1999), who looked at the correlation matrix of stocks in the S&P 500. They apply random matrix theory to extract those eigenvalues which carry most of the information needed to estimate the correlation matrix. This approach is picked up by Pedersen et al.'s (2021) Enhanced Portfolio Optimisation which is explained separately. The resampling approach has been coined by Richard and Robert Michaud (1998) producing several estimates of risks and returns around the initial estimates. Those newly resampled estimates are then used to perform *MVO*. Michaud (1989) further takes the average of all the different outcomes of the *MVO* performed on the different resampled estimates. However, Becker et al. (2009) find that classical *MVO* outperforms Michaud's resampling method within their simulations.

The approach of utilizing shrinkage has been suggested by Ledoit and Wolf (2004), who estimate the covariance matrix of stock returns through a weighted average of the sample covariance matrix and a single-index covariance matrix. This approach is merged with Pedersen et al.'s (2021) Enhanced Portfolio Optimisation in this thesis. Lastly, Black and Litterman (1992) compute a set of neutral weights by using the *CAPM* enabling investors to merge their own personal expected returns with the weights returned by the *CAPM*. Pedersen et al. (2021) incorporate this in two ways: For one, the simple *EPO* a vector of signals can be included. For another, the anchored *EPO* an anchor portfolio can be specified.

The focus of studies looking at resolving the problems of *MVO* by taking the input data for granted lies at setting up bounds of weights, regularizations, or penalisation of objective functions. Setting up bounds for weights signifies to limit the values each weight can take on. Roncalli (2010) examines the impact of weight constraints on portfolio theory. He shows that weight constraints may modify the covariance matrix substantially. This refers to Jagannathan and Ma's (2003) finding that restrictions on weights ultimately imply the same as modifying the covariance matrix itself. Regularization as well as penalisation objectives go into similar directions by restricting the weights in becoming too extreme. However, as argued by Bruder et al. (2013), all these approaches ultimately point towards changing the covariance matrix.

Machine learning algorithms enabled academics and researchers to pursue new opportunities on finding solutions to reducing noise in the covariance matrix. López de Prado (2020) introduces machine learning as means to building powerful financial theories as well as to better understanding existing theories. López de Prado (2020) further classifies machine learning in finance as a separate sub-category of machine learning due to the low signal-to-noise ratio. The main advantage of machine learning applications in finance is its ability to work with unstructured data, which represents 80% of all available data. The first chapter of his book is devoted to denoising and detoning of covariance matrices.

López de Prado (2020) presents two techniques enabling asset managers to work with the correlation matrix: denoising and detoning. Denoising replaces the eigenvalues of the eigenvectors classified as random by Marčenko-Pastur with a constant eigenvalue. This technique leads to an elimination of the noise contained in the correlation matrix while preserving the signal included. The author highlights the key difference between denoising and shrinkage as the ability of denoising to preserve even the smallest signal. Shrinkage instead, as argued by López de Prado (2020), eliminates some noise but also a part of the signal, which is prohibitively dangerous considering the small signal-to-noise ratio of financial return data. This finding is further confirmed by Zakamulin (2014). Once the signal has been extracted and the noisy eigenvectors reduced, López de Prado (2020) proceeds with the detoning of the eigenvectors. Detoning is based on the observation that financial correlation matrices always incorporate the general market factor. This market factor is found in the *PCA* as the eigenvector with the highest eigenvalue. López de Prado (2020) suggests removing that eigenvector to focus on other signals within the correlation matrix. Portfolio optimisation can then be performed on the denoised and detoned eigenvectors. The weights of the original assets can be reversely calculated from the weights of the eigenvectors.

The performance of this approach has been tested in a Monte-Carlo simulation with a minimum variance portfolio. The out-of-sample result for the denoised approach shows an improvement of 60%

compared to the original *MVO*. It was also compared to the Ledoit and Wolf (2004) shrinkage approach, which showed a worse performance compared to the denoising. For a maximum Sharpe ratio portfolio, a similar simulation has been undertaken showing a stronger performance for the denoised approach compared to the original maximum Sharpe ratio portfolio as well as the shrinkage approach (López de Prado, 2020).

Finally, Pedersen et al.'s (2021) Enhanced Portfolio Optimisation is introduced. It takes into account all the theoretical foundations outlined so far. Additionally, it integrates the several approaches made by academics to tackle the estimation noise in covariance and return estimations. López de Prado (2020) laid the foundation of using machine learning to enable investors to use *MVO*. Pedersen et al. (2021), however, take it one step further by showing in a simplistic way, that what is needed to make *MVO* work is to reduce the correlation matrix towards the identity matrix.

3.2 Enhanced Portfolio Optimisation

Enhanced Portfolio Optimisation (*EPO*) is developed by Pedersen et al. (2021) as a solution on how to adjust the correlation matrix in a way that reduces the impact of estimation noise. *EPO* calculates weights for portfolios based on the structure contained in the historic return data. They thereby enable investors to obtain more reliable portfolio weights than from *MVO*. Pedersen et al. (2021) present two use cases of *EPO*: one simple application, and an anchored approach allowing investors to anchor the enhanced portfolio to any desired benchmark portfolio. The simple *EPO* is the approach of interest for this thesis as it shows how machine learning techniques help address noise in covariance matrices and solve that problem for *MVO*.

The simple *EPO* starts by applying *PCA* to the original assets to differentiate between noise and structure. Each eigenvector is treated as a separate portfolio, where the elements of an eigenvector are used as weights. The common statistics such as variance, return, or Sharpe ratio are calculated for the returns of those principal component portfolios. These calculations are performed to show the expected statistics against the realised ones. Figure 5 shows that the portfolios with the lowest eigenvalues are those with higher expected returns compared to the realised returns as well as lower expected volatility than realised (Pedersen et al., 2021). This Figure is used to prove the point that *MVO* tends to leverage on those portfolios which are wrongly estimated through noise.

Figure 5.A shows the estimated volatilities of the principal component portfolios versus the realised volatilities per principal component portfolio. The two main conclusions from that graph are that those portfolios with higher eigenvalues overestimate the risk and tend to realise lower volatilities

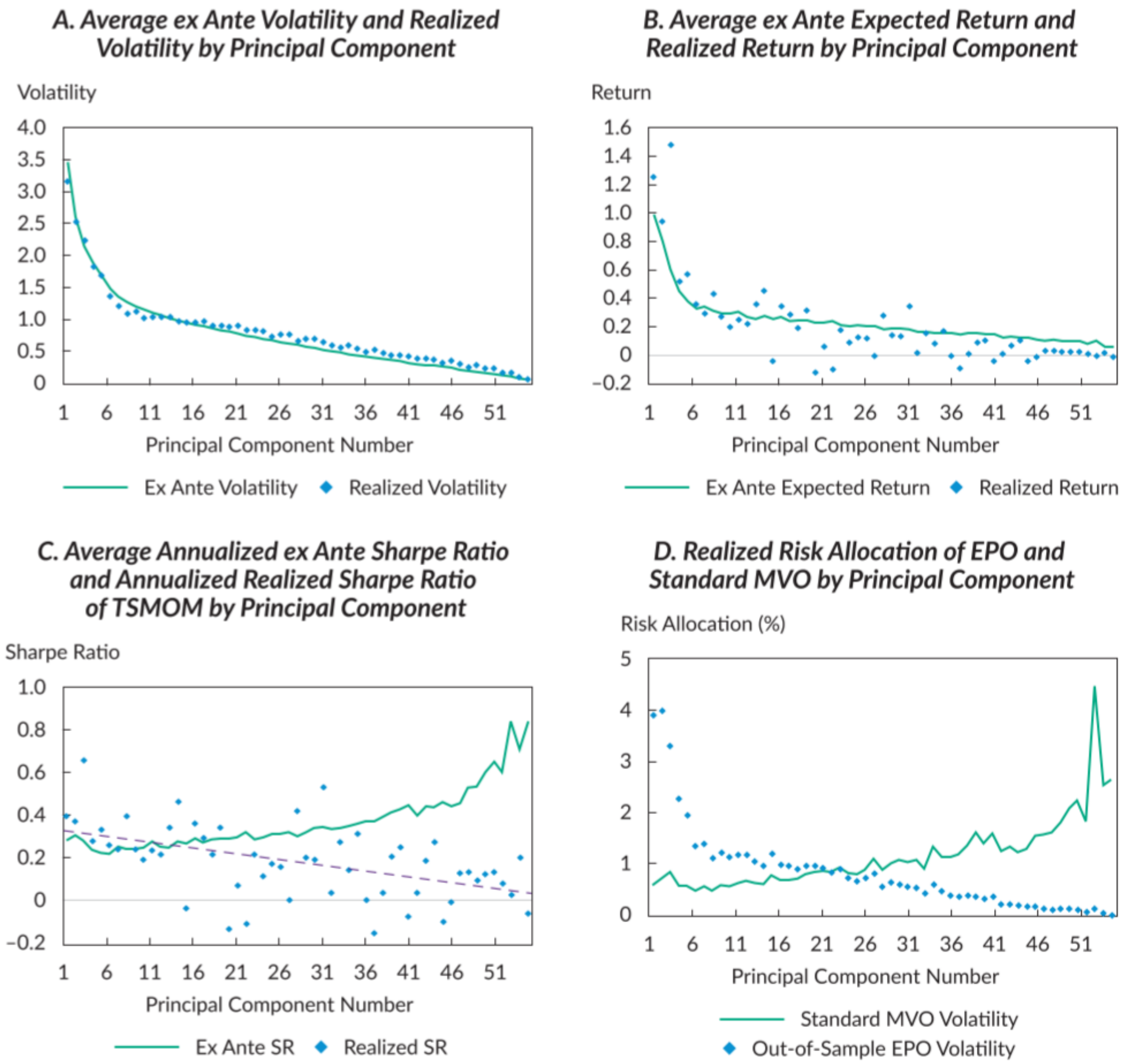

Figure 5 Extracted from Pedersen et al. (2021).

than estimated. Instead, those portfolios on the right end, with low eigenvalues (per definition of *PCA* mostly consisting of noise) realise a higher volatility than estimated. Figure 5.B is similarly built as Figure 5.A but shows expected returns versus realised returns for each principal component portfolio. Again, a striking difference can be noted between those principal component portfolios with high eigenvalues and those with lower eigenvalues. Realised returns tend to be higher than the expected returns for the left half of the portfolios, while the right half realised lower returns than expected.

Figure 5.C merges the two findings from Figure 5.A and 5.B by showing the realised and expected Sharpe ratios per principal component portfolio. Overall, expected Sharpe ratios are higher for those portfolios with low eigenvalues than the portfolios with higher eigenvalues. However, the realised Sharpe ratios are higher for those portfolios with high eigenvalues than those with low eigenvalues. Further, the difference between expected and realised Sharpe ratios is larger for the portfolios with low eigenvalues. This is in line with the previous findings and highlights the randomness of the portfolios with low eigenvalues. Lastly, Figure 5.D shows that *MVO* puts a large part of the portfolio weight into those portfolios which are essentially random and noisy. The danger with *MVO* is therefore that the outcome of the *MVO* portfolio is random and not optimising the portfolio in any way. Rather, it places weights on the sub-optimally estimated portfolios and exposes the investor to the opposite of what it was supposed to: portfolios with high realised volatilities and low realised returns. Michaud framed it in 1989 as:

> *"The unintuitive character of many optimized portfolios can be traced to the fact that MV optimizers are, in a fundamental sense, estimation error maximisers. Risk and return estimates are inevitably subject to estimation error. MV optimization significantly overweights (underweights) those securities that have large (small) estimated returns, negative (positive) correlations and small (large) variances. These securities are, of course, the ones most likely to have large estimation errors"* (Michaud, 1989, p. 33).

Having the flaws of *MVO* established, Pedersen et al. (2021) show how to address them effectively. The basic idea is to shrink the correlation matrix towards the identity matrix and thereby reduce the correlations towards zero. It must be noted that this is not done in the space of principal component portfolios but in the space of the original assets available to the investor. This is the main difference between Pedersen et al. (2021) and López de Prado (2020). The amount of shrinkage needed is an empirical question, which will be addressed further below. The simple *EPO* is built similarly to the minimum variance portfolio. Instead of using the original covariance matrix, a new covariance matrix is computed with the shrunk correlation matrix. Further, risk aversion enters the formula to adjust the portfolios to the specific risk appetite of the investors. Additionally, investors can include their own expectations of future returns through a signal variable.

Before showing the specific mathematical steps, the intuition behind shrinkage is explained and why it solves the problem of *MVO*. Relating back to Figure 5 and moving into the space of principal component portfolios, the intuitive solution to *MVO* is to increase the volatility estimates of the principal component portfolios. Pedersen et al. (2021) show that shrinking the correlation matrix of the original assets towards the identity matrix is the same as changing the volatilities at the level of the principal components. They even go one step further by showing that the shrinking of correlations also addresses noise in the expected return estimations. However, noise in expected returns is not the focus of this thesis and hence not pursued any further.

The reason why a shrunk correlation matrix at the level of the original assets leads to an overall higher estimated variance of the portfolio can be found in the choice of weight allocations in *MVO*. *MVO* places high importance on assets with high correlations towards the other assets by placing large positive or negative weights on those assets. By, for example, shorting an asset with high positive correlations to the other available assets that short position ensures a rather low overall estimated volatility of the portfolio. With pairwise correlations shrunk towards zero, this technique becomes inherently difficult, and *MVO* is forced to place more importance on the volatilities (recall Formula 03 for calculation of the covariance matrix) of the individual assets instead. This, in turn leads to an overall higher expectation of the volatility of the portfolio. Even though it seems counter-intuitive that shrinking correlations leads to an overall higher expected risk, the key to understanding this step is to realise how the weights are allocated using the original *MVO*.

In mathematical terms, shrinking the correlation matrix towards the identity matrix looks as follows:

$$\widetilde{\boldsymbol{\Omega}} = (1-\theta)\boldsymbol{\Omega} + \theta \boldsymbol{I} \qquad (14),$$

where $\boldsymbol{\Omega}$ is the original correlation matrix, $\boldsymbol{I}$ the identity matrix, and θ the shrinkage parameter. The shrunk correlation matrix can further be used, together with the initially estimated volatilities, to calculate a new variance-covariance matrix $\widetilde{\boldsymbol{\Sigma}}$:

$$\widetilde{\boldsymbol{\Sigma}} = \boldsymbol{\sigma}\widetilde{\boldsymbol{\Omega}}\boldsymbol{\sigma} \qquad (15),$$

with the volatilities given by $\boldsymbol{\sigma}$. The weights allocated to the simple *EPO* portfolio are calculated as follows:

$$EPO_s = \frac{1}{\gamma}\widetilde{\boldsymbol{\Sigma}}^{-\mathbf{1}}\boldsymbol{s} \qquad (16).$$

The risk aversion of each investor enters through the parameter γ, while $\boldsymbol{s}$ is the vector of signals through which an investor can incorporate own expectations of future returns into the weight allocations.

Pedersen et al. (2021) compare their method to several benchmarks such as the equally weighted portfolio and *MVO* portfolio with numerous differently composed datasets and portfolios. They find their portfolios to always outperform the benchmarks in terms of gross realised out-of-sample Sharpe ratios.

As impressive as the presented results are, the question of how much shrinkage to use is still unanswered. Pedersen et al. (2021) choose the optimal shrinkage parameter as the shrinkage parameter that would have previously yielded the highest possible Sharpe ratio. On average, the optimal level of shrinkage is reported as 0.75. However, Pedersen et al. (2021) also show that the optimal shrinkage parameter depends on the underlying set of available assets. Pedersen et al. (2021) do not explicitly state that the underlying correlation within the dataset influences the optimal shrinkage parameter. However, it can be drawn from the results presented, and understood intuitively. Also with estimation noise blurring the truly underlying correlation of a dataset of local equities, the probability that these correlations are closer to one than to zero is quite high. Hence, this thesis proposes to shrink the correlation matrix towards the average correlation among all assets instead of the identity matrix.

4. Research Proposal

The research question this thesis intends to answer is how machine learning technologies help addressing the problem of noise in covariance matrix estimations when optimizing portfolios. The needed theoretical background has been established as well as the different approaches proposed by academics on how to deal with noise and how to make *MVO* work in practice are outlined. Pedersen et al. (2021) and López de Prado (2020) are the most recent ones who "fix" *MVO* with the help of machine learning. Both make use of machine learning algorithms when assessing the principal components of the portfolios. The focus of this thesis lies on Pedersen et al.'s (2021) simple *EPO* instead of López de Prado's (2020) denoising approach because it does not solely focus on noise in risk but also noise in expected returns. Even though noise in the expected returns is not the focus of the thesis, it is still better to take it into account instead of blindly accepting it. Ultimately, Pedersen et al. (2021) do not require the help of machine learning anymore, showing that shrinking the correlation matrix yields the same outcome.

This thesis compares estimated risk and returns with the realised risk and return for different portfolio optimisation strategies. Additionally, a mix of Pedersen et al. (2021) and Ledoit and Wolf (2003) will be applied to another strategy of portfolio optimisation. This portfolio is calculated by shrinking the correlation matrix towards the average correlation of all assets instead of shrinking it towards the identity matrix. The motivation for that is found in the portfolios Pedersen et al. (2021) use and the optimal shrinkage parameters they associate with the different portfolios.

In Table 2 Pedersen et al. (2021) report their model to perform best with a shrinkage parameter of $\theta = 0.75$. The underlying portfolio in that example is composed of global equities, bonds, currencies, and commodities. As it includes different types of assets and covers all geographies, the average correlation among this portfolio can be assumed to be comparatively low. In Table 5 instead, they show that the optimal shrinkage parameters are rather low ranging from 10% to 50%. The portfolios considered here only contain equities on industries in the United States. The average correlation among these assets can be assumed to be rather high. This thesis therefore tries to establish a link between the shrinkage parameter and the underlying correlations. Consequently, the approach is to shrink towards the average correlations instead of the identity matrix.

The research undertaken in this thesis adds value on top of the existing literature through the added approach of *Enhanced Portfolio Optimisation*. If the new approach of shrinking towards the average correlations of all assets proves to be performing better than the original *EPO*, an additional portfolio optimizing method has been created. On the other hand, if the new approach does not turn out to be more efficient, the *EPO* gains considerable strength and is shown to be a very strong method to optimise portfolios.

5. Methodology

Saunders (2019) defines a study that focuses on the testing and adjusting of existing theories as an abductive approach. Taking the *MVO* as a starting point and then applying a few selected changes to that method to check if the changes have a positive effect on the performance hence qualifies this thesis as an abductive study. Abductive studies usually first describe a surprising fact about an existing theory, followed by the testing section and a section discussing the results and performance of the modified models. The role *MVO* plays in finance since its invention in 1952 is exceptional and its theoretical usefulness unquestioned. However, the practical usefulness has been questioned by several scholars as shown in the literature review. It is therefore rather surprising that that *MVO* is rarely used in practice, despite its unquestioned theoretical value.

The fundamental problem underlying the purpose of this thesis is the prevalence of estimation noise in data on return and risk of stocks and portfolios. The literature review referred to several papers neglecting the practical usefulness of Markowitz' (1952) *MVO* due to noise. Nevertheless, Pedersen et al.'s (2021) *EPO* makes use of *MVO* and finds a way to deal with noise through the shrinkage of the correlation matrix. To discuss the possibilities of machine learning in addressing the problem of noise in data, this thesis uses different strategies to calculate the weights of several assets within a portfolio and compares the performance over time.

5.1 Dataset

The data needed to perform such tests are high-dimensional quantitative stock- or portfolio returns. Data on individual stocks have a disadvantage, as companies enter and leave the stock market resulting in the number of daily returns available for each stock being not very high. Data on portfolio returns instead have the advantage that they are available for a longer time as portfolios are at any point in time composed of those companies which are at that time publicly traded. One common source for such portfolio returns is the Kenneth French data library.

For the analysis of this thesis two different datasets have been used: ten value weighted industry portfolios and 100 value weighted portfolios on market and book value. Both datasets are retrieved from the Kenneth French data library. The industry portfolio dataset contains daily returns from 1926 until and including 2021. The dataset includes the following industries: consumer nondurables, consumer durables, manufacturing, energy, high-tech, telecommunication, shops, health, utilities, and other. Companies that are listed on the New York Stock Exchange (NYSE), American Stock Exchange (AMEX), or the National Association of Securities Dealers Automated Quotations (NASDAQ) are classified each year to belong to one of the aforementioned industries. The classification is done as per the four-digit SIC codes. The advantage of using this dataset is that it contains daily returns for almost one hundred years. This is extraordinarily important when estimating the covariance and correlation matrices.

As high dimensionality amplifies the problem of noise in estimations of covariances and expected returns (Liu et al., 2015), a dataset with a dimensionality of at least one hundred is needed (Negahban & Wainwright, 2011). The 100 portfolios formed on size and book-to-market dataset includes daily returns from 1926 until and including 2021 and is also retrieved from the French data library. The portfolios include all stocks listed on the NYSE, AMEX, or NASDAQ. This dataset is built in a similar way as the industry portfolios. Instead of looking at the industry code, in this dataset French classifies companies according to their market capitalisation and book-to-market ratio. As the dataset is composed of one hundred portfolios, the dataset matches the high dimensionality criterion. Both datasets will be used in the analysis and testing. One key difference between the two datasets is that the average correlation among the different portfolios is higher for the industry portfolios. This allows to test whether shrinking towards a different matrix is more efficient than shrinking towards the identity matrix when the average correlation among the different assets is already known to be high.

Furthermore, data on the risk-free rate is needed to calculate excess returns. As the portfolio return data is obtained on a daily basis and the portfolio weight adjustments are assumed to occur on a monthly basis, the interest rate chosen to resemble the risk-free rate is the effective federal funds rate. The effective federal funds rate is the rate which is charged in the overnight market of depositary institutions in the financial American market. It is a short-term rate and thus matches the investment horizon of the different strategies. The effective federal funds rate can be retrieved from the Federal Reserve Economic Database. The rate is available from July 1954 onwards. To match the availability of both datasets, this thesis makes use of the portfolio data and interest rate data from July 1954 until and including March 2021. All datasets have been accessed and downloaded on the 8th of May 2021.

5.2 Data Preparation

The dataset of the interest rates includes data for every day, including weekends and holidays. Hence, the first step in preparing the dataset is to delete all days on which no trading took place in the American stock market. The number of daily returns per year in the cleaned-up dataset averages to 252 per year. Next, the dataset was adjusted to show daily interest rates as opposed to annualised data. This was done through geometric compounding. At the level of the return datasets not much data preparation was needed. Some portfolios have missing data especially towards the beginning of the dataset. The procedure on how to treat missing values is described by Acock (2005) as listwise deletion. The day on which any portfolio has a missing value is deleted and not taken into consideration for the calculations. Additionally, all datasets were divided by 100 to be shown as decimals instead of percentages.

The analysis, testing, and production of the graphs is conducted in Python 3.7.6. The testing consists of calculating different weights of the portfolios to optimise the overall asset allocation. The weights are calculated with the classical *MVO*, Pedersen et al.'s (2021) *EPO*, and the adjusted *EPO*, with the correlation matrix shrunk towards the average correlations. These weights are adjusted monthly. An additional portfolio that the *EPO* is compared to is the equally weighted portfolio, where each instrument receives the same weight.

5.3 Overview of the Different Strategies Applied

The equally weighted portfolio consists of all available assets with equal proportions. The equally weighted portfolio is a naïve view of an investor into the future without any expected returns or expected risk. Even though the strategy is technically not very complicated, its performance is difficult to beat historically. Ormos (2012) finds positive abnormal returns for equally weighted portfolios in the American stock market. Also, Pedersen et al. (2021) report the equally weighted portfolios to be a tough benchmark to beat. Furthermore, Bruder (2013) classifies the equally weighted portfolio to minimise the impact of estimation errors on the optimized portfolio.

The *MVO* and its advantages and disadvantages as well as the formulae used to calculate the weights have been described above. Nevertheless, as Pedersen et al. (2021) derive *EPO* from *MVO*, *MVO* weights can alternatively be obtained the same way as the *EPO* weights, just by using the original correlation and covariance matrix. As each portfolio is updated on a monthly basis, the weights for the *MVO* need to be adjusted accordingly. This requires the covariance matrix to be calculated each month. The *MVO* strategy returns weights for a portfolio that consists of a mix of the same set of assets which together yield the portfolio that minimises the total variance of returns. These weights are used to calculate the desired statistics on expected and realised returns and risk.

The *EPO* method is further split into two different strategies. Both *EPO* strategies calculate the weights based on the covariance matrix which is composed of the shrunk correlation matrix. However, the two strategies differ in the way the correlation matrices are shrunk. The first strategy (EPO 1) uses the same approach as described by Pedersen et al. (2021) and shrinks the correlation matrix towards the identity matrix:

$$\widetilde{\mathbf{\Omega}}_{EPO1} = (1-\theta)\mathbf{\Omega} + \theta\boldsymbol{I} \qquad (17),$$

where $\mathbf{\Omega}$ represents the original correlation matrix, $\boldsymbol{I}$ the identity matrix, and θ the shrinkage parameter dictating how much of the original correlation matrix is kept. θ can take on values between 0 and 1, where 0 would imply not applying any shrinkage at all. The second strategy (EPO 2) instead shrinks the correlation matrix to a matrix which consists of ones on the diagonals and the average correlations of all portfolios on the off diagonals.

$$\widetilde{\mathbf{\Omega}}_{EPO2} = (1-\theta)\mathbf{\Omega} + \theta\mathbf{M} \qquad (18),$$

where $\mathbf{\Omega}$ and θ are the same as in the formula 17, $\mathbf{M}$ is the matrix consisting of ones on the diagonals and the average correlations of all portfolios off the diagonals. This is also done monthly, where the average of the correlations is calculated each month on a rolling basis.

When choosing the number of days to include in the covariance and correlation calculations, a trade-off is faced. On the one hand, when estimating covariance matrices, including more data is always better as, among others, Fabozzi et al. (2010) or Tsay (2005) describe. On the other hand, when including more datapoints from the past, one runs the risk of estimating the average covariance and correlation over the past instead of showing the current covariance and correlation (Fabozzi et al., 2010). Covariances are likely to change over time and it is essential to use accurate covariances and correlations to forecast co-movements of stocks and adjust the weights accordingly. Gupta et al. (1987) advise to use between 670 and 5243 data points for the calculation of covariances with multivariate datasets. As 5243 data points corresponds to over twenty years of data, this number seems too high. The data used for all correlation and covariance calculations therefore consist of the last 750 daily returns, incorporating the last three years of daily returns. Fan, Fan and Lv (2008) also use a sample size including three years of daily returns.

The newly obtained covariance matrices are calculated with the original volatilities and the shrunk correlation matrices:

$$\widetilde{\mathbf{\Sigma}}_{EPO1} = \boldsymbol{\sigma}\widetilde{\mathbf{\Omega}}_{EPO1}\boldsymbol{\sigma} \qquad (19)$$

$$\widetilde{\mathbf{\Sigma}}_{EPO2} = \boldsymbol{\sigma}\widetilde{\mathbf{\Omega}}_{EPO2}\boldsymbol{\sigma} \qquad (20).$$

Further, the vector of signals $\boldsymbol{s}$ describes the expected returns of each instrument. The signal has been computed along the lines of time series momentum used by Pedersen et al.:

$$\boldsymbol{s} \;=\; 0.1 * \boldsymbol{\sigma^i} * \boldsymbol{sign}\left(r^i_{t-(t-12)}\right) \qquad (21),$$

consisting of the volatilities of each asset $\boldsymbol{\sigma^i}$ and the sign of the return each asset realised within the last year, where *t* denotes the current month and *t-12* the month one year ago. The factor of 0.1 is used because it is necessary to translate the volatilities into returns. The Sharpe ratio puts return and volatilities into perspective and allows to estimate the return based on the volatility by returning an expected return per unit of risk. Hence, the Sharpe ratio is assumed to be at a constant level of 0.1 for the analysis of this thesis. This assumption is in accordance with Babu et al. (2020), Moskowitz et al. (2012), and López de Prado (2020). Moreover, the $\boldsymbol{sign}$ variable is a binary variable which can take on the value of 1 or -1. It takes on 1 if the return of the last year was positive of the specific asset.

Likewise, it takes on -1 if the return over the last year was negative. This formula can be translated into the expectation of positive returns for any individual instrument if the return of that asset was positive for the last year. This can be derived from the fact that volatilities only take on positive values and the sign of $\boldsymbol{s}$ is solely determined by the variable $\boldsymbol{sign}$. This formulation of expectations is a typical time series momentum strategy, which is calculated at the end of each month using daily return data.

The adjusted covariance matrices, signals, and degree of risk aversion are used to establish weights of each instrument for the overall portfolio. The weights using the two different *EPO* strategies are calculated as:

$$EPO_1 = \frac{1}{\gamma} \widetilde{\boldsymbol{\Sigma}}^{-1}_{EPO1} \boldsymbol{s} \quad (22)$$

$$EPO_2 = \frac{1}{\gamma} \widetilde{\boldsymbol{\Sigma}}^{-1}_{EPO2} \boldsymbol{s} \quad (23),$$

consisting of the degree of risk aversion γ, the inverted newly configured covariance matrices, and the vector of signals $\boldsymbol{s}$. It is important to state that each variable is calculated for every date on which the portfolio weights are adjusted.

5.4 Comparing the Performance of the Strategies

As shrinking the correlation matrices reduces noise in both the covariance matrix and the expected returns, the best approach to compare the different strategies is to compare the Sharpe ratios of all strategies:

$$\frac{\left(r_{portfolio} - r_f\right)}{\sigma_{portfolio}} \quad (24),$$

with the return of each portfolio $r_{portfolio}$, the risk-free interest rate r_f, and the volatility of the portfolio $\sigma_{portfolio}$. To come up with an even more precise evaluation of the different strategies, the expected Sharpe ratios are compared to the realised ones. For the realised Sharpe ratios, the realised return is calculated as:

$$r_{portfolio} = w_i * r_{i,t+1} \quad (25),$$

multiplying the weight at the beginning of the month with the return that has been achieved throughout the entire month. The realised volatility is calculated based on the weights at the beginning of the month and the covariance matrix of each strategy at the end of the month. The formula can therefore be applied to all strategies, with a differing covariance matrix and the individual weights of each strategy:

$$\sigma_{portfolio} = \sqrt{\boldsymbol{w}^T \boldsymbol{\Sigma}_{\mathbf{t+1}} \boldsymbol{w}} \qquad (26).$$

Two variables influencing the calculations above still need to be discussed: θ and γ. The level of risk aversion γ is, for the purposes of this thesis, set to a level of one. Pedersen et al. (2021) suggest choosing a level of risk aversion between 1 and 10. However, it is not essential to decide on a specific level of risk aversion. What matters is to be consistent and use the same level throughout the different strategies to make the outcomes comparable. The shrinkage parameter θ instead is chosen empirically by running each strategy with different levels of θ. Based on the realised Sharpe ratios with each level of shrinkage, the ultimate shrinkage parameter is decided upon, where a higher realised Sharpe ratio is always preferred over a lower Sharpe ratio.

The performance of each method is judged based on the difference of the expected Sharpe ratio, volatility, and return, and the realised Sharpe ratio, volatility, and return. The method with a lower discrepancy between estimated and realised results is deemed to be the superior method. The reason for this choice of testing methods can be found in the research question. To show if machine learning technology, which is in this case represented by the two *EPO* approaches, produces more reliable weight allocations and portfolios than the classical *MVO* is at the core of this thesis. Consequently, the results of the testing method must enable the reader to choose between those alternatives. Under the assumption that investors only care about risk and return, the proposed testing method suffices to show which of the methods delivers a smaller gap between expected and realised results. The research question can also be answered by looking at the results. Pedersen et al. (2021) uses machine learning technology to counter the noise in data. If the *EPO* outperforms the *MVO*, machine learning can effectively address the problem of noise in data. Further, the adjusted *EPO* can be compared to the simple *EPO* based on the performance on the different datasets.

6. Results

This section first shows the results of the Principal Component Analysis, which is used to demonstrate how to differentiate between noise and signal. Second, the optimal shrinkage parameters for the two Enhanced Portfolio Optimisation strategies are determined. Last, the returns, volatilities, and Sharpe ratios for the different strategies are analysed. This is done for both datasets simultaneously.

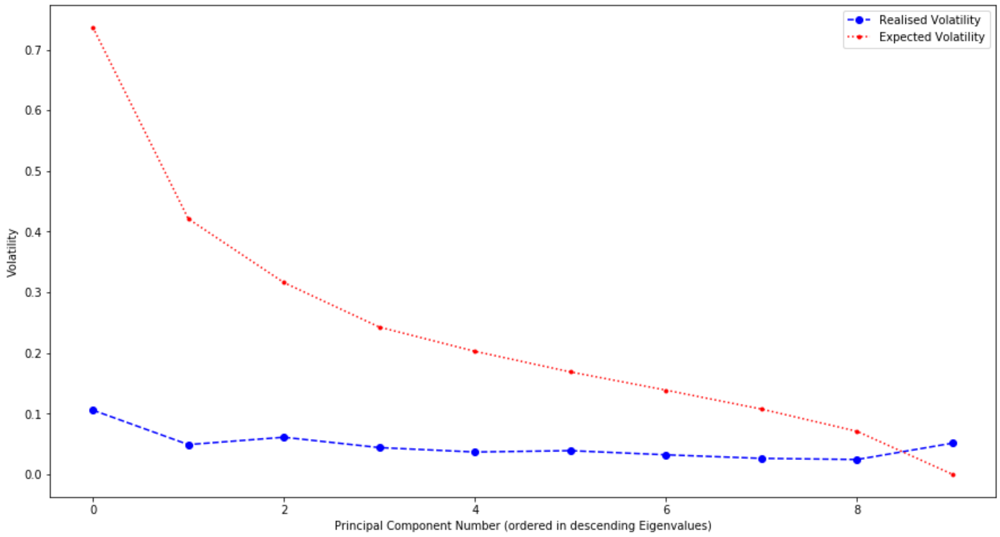

Figure 6 Volatility by principal component portfolio for the 10-industry dataset (own illustration)

The *PCA* decomposed each monthly correlation matrix into its eigenvectors and sorted them in a descending order by their eigenvalues. The figures below show the different eigenvectors on the x-axis which were used to construct portfolios. The eigenvector furthest to the left is therefore the eigenvector with the highest eigenvalue. The single elements within a principal component were used as weights to form a portfolio which has been adjusted monthly. Therefore, every month 10 new portfolios were calculated for the industry dataset, and 100 for the size dataset, as each month 10 and 100 eigenvectors are obtained.

The average expected and realised volatilities over time are depicted in Figure 6 for the 10 industry portfolios. The realised volatility is constant around 0.1, while the expected volatility is high for the first principal component and decreasing thereafter. The last principal component is the only principal component with a higher realised volatility than expected, which is in line with Pedersen et al.'s (2021) findings. Figure 7 shows the average expected and realised volatilities over time for the 100 size portfolios. Pedersen et al.'s (2021) findings are further supported here, showing that the principal components with the highest eigenvalues are the ones with the highest expected

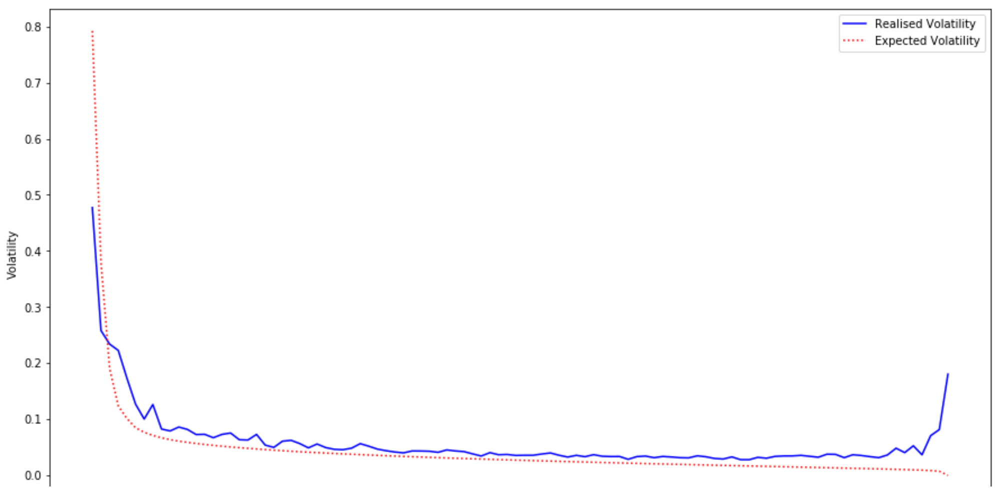

Figure 7 Volatility by principal component portfolio for the 100-size dataset (own illustration)

and realised volatilities. However, they are the only ones where the expected volatility is above the realised volatility. Most of the principal component portfolios realise a higher volatility than expected. Especially, those with the lowest eigenvalues have a large gap between realised and expected. This proves the random nature of the last principal components.

Similarly, the average expected and realised returns are examined for each principal component portfolio over time, shown in Figures 8 and 9. Especially the principal component portfolios from the 10 industry dataset show that the portfolios formed for the eigenvectors with low eigenvalues tend to realise lower returns than expected. While the first two eigenvectors realise the returns as expected, or slightly better; the others diverge drastically from the expectations. Together with Figure 6, one can conclude that the portfolios formed through the eigenvectors with the lowest eigenvalues are the ones where the expectations on risk and return deviate the most. This highlights the portfolios with low eigenvalues are random and composed of noise, while the portfolios which are obtained through the eigenvectors with high eigenvalues are based on underlying structure. The underlying structure leads to a low difference between expectations and realisations. These differences between realised and expected risk and return are in both datasets even favourable for the investor, as the realised volatility is lower than expected and the realised returns are higher than expected. The opposite is the case for the portfolios built through the eigenvectors with low eigenvalues.

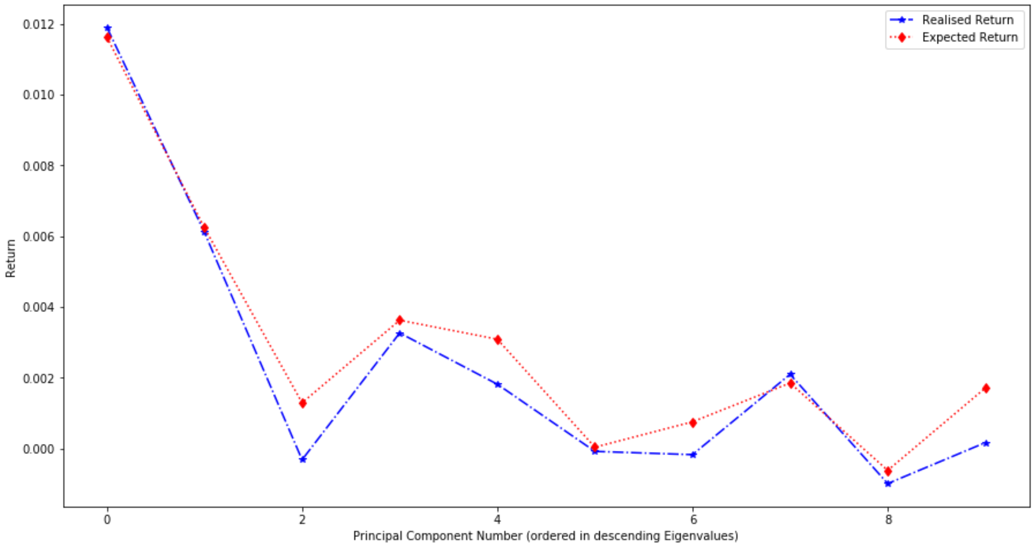

Figure 8 Return by principal component portfolio for the 10-industry dataset (own illustration)

This conclusion cannot be drawn as easily for the 100-size portfolio dataset. As seen in Figure 9, the expected and realised returns for each principal component are surprisingly close and do not differ much from each other. Overall, the expected and realised returns are the highest for the principal component portfolios with the highest eigenvalues and constantly low for the portfolios with lower eigenvalues. Nevertheless, the *PCA* also shows for this dataset a dangerous discrepancy of expected and realised volatilities. This is dangerous for investors who use *MVO* to allocate the weights of their portfolios because *MVO* would place high weights on the principal component portfolios with low eigenvalues, the ones to the right end on the Figures. This is for the reason that these portfolios expect high returns for each unit of expected risk. However, they realise low returns for each unit of risk, as the returns are less than expected and the risk higher than expected. The opposite is true for the portfolios which are to the left on the graphs and possess high eigenvalues. These portfolios are reasonably estimated based on the underlying structure found by *PCA*. Their expectations match the realised results. In case the expectations do not match, they tend to be favourable to the investor by being lower risk than expected and higher returns than expected.

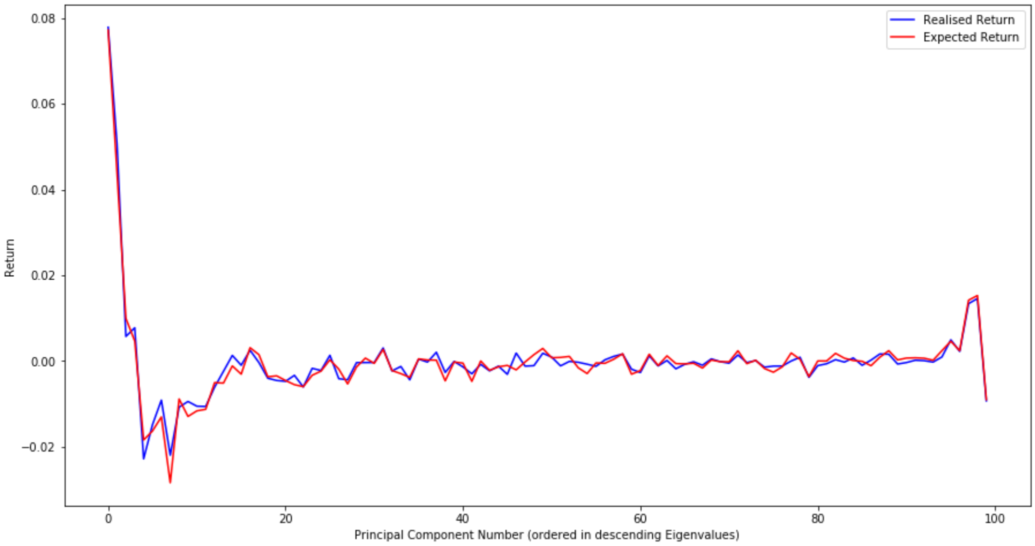

Figure 9 Return by principal component portfolio for the 100-size dataset (own illustration)

This section so far showed that it is possible to differentiate between noise and structure through *PCA*. The next step is to show if the Enhanced Portfolio Optimisation 1 as proposed by Pedersen et al. (2021) or the Enhanced Portfolio Optimisation 2 as developed in this thesis perform better than the *MVO* and the equally weighted portfolio. Before doing so, the optimal shrinkage parameters θ need to be decided upon.

EPO 1 shrinkage parameter simulation 10 portfolios

	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Realised Return	0.005641	0.005572	0.005509	0.005469	0.005444	0.005430	0.005425	0.005426	0.005432	0.005442	0.005455
Expected Return	0.005800	0.005703	0.005618	0.005557	0.005514	0.005484	0.005465	0.005454	0.005451	0.005454	0.005463
Absolute Difference Return	0.000159	0.000131	0.000109	0.000088	0.000069	0.000053	0.000040	0.000028	0.000019	0.000012	0.000008
Realised Risk	0.036492	0.035422	0.035018	0.034888	0.034973	0.035240	0.035666	0.036235	0.036936	0.037765	0.038719
Expected Risk	0.026811	0.027381	0.027283	0.026808	0.026026	0.024958	0.023594	0.021901	0.019810	0.017186	0.013732
Absolute Difference Risk	0.009682	0.008041	0.007735	0.008080	0.008947	0.010283	0.012072	0.014334	0.017127	0.020578	0.024987
Realised Sharpe Ratio	0.052352	0.053132	0.052257	0.051090	0.049743	0.048324	0.046924	0.045605	0.044393	0.043289	0.042273
Expected Sharpe Ratio	0.044478	0.043239	0.041822	0.041447	0.041992	0.043426	0.045825	0.049417	0.054687	0.062687	0.076055
Absolute Difference Sharpe Ratio	0.007874	0.009893	0.010435	0.009643	0.007750	0.004898	0.001099	0.003812	0.010293	0.019397	0.033782

Table 1 Shrinkage simulation EPO 1 10 industry portfolios (own illustration)

The optimal shrinkage parameter needs to be set four times, once for each strategy within each dataset. Tables 1 to 4 show simulations for each scenario. The strategies were simulated by using eleven different shrinkage parameters, ranging from 0 to 1 in steps of 0.1. The results of the average realised

EPO 2 shrinkage parameter simulation 10 portfolios

	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Realised Return	0.005641	0.005578	0.005533	0.005500	0.005477	0.005460	0.005448	0.005439	0.005431	0.005420	0.005401
Expected Return	0.005800	0.005750	0.005717	0.005697	0.005687	0.005683	0.005685	0.005690	0.005696	0.005700	0.005695
Absolute Difference Return	0.000159	0.000172	0.000184	0.000197	0.000210	0.000223	0.000237	0.000251	0.000265	0.000280	0.000295
Realised Risk	0.036492	0.036589	0.036808	0.037101	0.037455	0.037866	0.038343	0.038897	0.039548	0.040329	0.041301
Expected Risk	0.026811	0.026869	0.026847	0.026762	0.026621	0.026426	0.026176	0.025869	0.025499	0.025057	0.024526
Absolute Difference Risk	0.009682	0.009720	0.009961	0.010339	0.010834	0.011441	0.012167	0.013027	0.014048	0.015273	0.016774
Realised Sharpe Ratio	0.052352	0.050787	0.049551	0.048498	0.047527	0.046563	0.045537	0.044372	0.042974	0.041207	0.038859
Expected Sharpe Ratio	0.044478	0.042913	0.041981	0.041518	0.041416	0.041602	0.042022	0.042624	0.043343	0.044081	0.044658
Absolute Difference Sharpe Ratio	0.007874	0.007874	0.007571	0.006980	0.006111	0.004961	0.003515	0.001749	0.000369	0.002874	0.005799

Table 2 Shrinkage simulation EPO 2 10 industry portfolios (own illustration)

and expected risk, return, and Sharpe ratios are shown for each level of shrinkage. Furthermore, the absolute difference between the expectations and realisations are shown. The maximum values for each category are highlighted in red, while the minimum values are shown in green. Though Pedersen et al. (2021) choose the shrinkage parameter based on the realised Sharpe ratios, this thesis chooses the shrinkage parameter based on the smallest absolute difference between the expected and the realised Sharpe ratio. The reason to look at the Sharpe ratio is that shrinking the correlation matrices reduces the noise in returns as well as the noise in the risk assessment, as discussed in section 3.2 “Enhanced Portfolio Optimisation”. However, the fundamental idea of reducing noise is to achieve a level of predictability of returns and risk and thereby exclude the influence of the randomness of noise on the weight allocations. The key performance indicator (KPI) for the shrinkage parameter is therefore chosen as the absolute difference between the average realised and expected Sharpe ratios.

The EPO 1 strategy achieves the lowest difference between the average realised and expected Sharpe ratios in the 10-industry dataset with a level of shrinkage of 0.6 towards the identity matrix. The EPO 2 strategy achieves an even lower difference by setting the shrinkage parameter to 0.8 and shrinking towards the average correlations of all assets. These results are shown in Table 1 and Table 2. The shrinkage parameters used to show the overall performance against the *MVO* and the equally weighted portfolio are thus set accordingly.

EPO 1 shrinkage parameter simulation 100 portfolios

	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Realised Return	0.009565	0.009978	0.010189	0.010327	0.010429	0.010517	0.010608	0.010722	0.010892	0.011181	0.011747
Expected Return	0.011782	0.012233	0.012470	0.012623	0.012736	0.012834	0.012935	0.013061	0.013248	0.013561	0.014171
Absolute Difference Return	0.002217	0.002255	0.002281	0.002296	0.002307	0.002316	0.002327	0.002339	0.002356	0.002380	0.002424
Realised Risk	0.099552	0.099684	0.100155	0.100686	0.101263	0.101901	0.102636	0.103532	0.104709	0.106427	0.109289
Expected Risk	0.016843	0.018622	0.019303	0.019499	0.019342	0.018862	0.018035	0.016774	0.014874	0.011800	0.004726
Absolute Difference Risk	0.082709	0.081062	0.080852	0.081188	0.081921	0.083039	0.084601	0.086758	0.089835	0.094626	0.104563
Realised Sharpe Ratio	0.029407	0.020806	0.014910	0.010301	0.006277	0.002383	-0.001768	-0.006611	-0.012768	-0.021321	-0.034624
Expected Sharpe Ratio	1.474172	1.507009	1.535699	1.566955	1.606012	1.658328	1.732449	1.845200	2.037833	2.456406	4.990019
Absolute Difference Sharpe Ratio	1.444765	1.486203	1.520790	1.556655	1.599736	1.655945	1.734217	1.851811	2.050601	2.477727	5.024642

Table 4 Shrinkage simulation EPO 1 100 size portfolios (own illustration)

EPO 2 shrinkage parameter simulation 100 portfolios

	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Realised Return	0.009565	0.009569	0.009555	0.009539	0.009518	0.009486	0.009437	0.009361	0.009238	0.009025	0.008651
Expected Return	0.011782	0.011842	0.011913	0.011978	0.012031	0.012067	0.012082	0.012064	0.011995	0.011836	0.011540
Absolute Difference Return	0.002217	0.002273	0.002359	0.002439	0.002513	0.002581	0.002644	0.002703	0.002757	0.002811	0.002889
Realised Risk	0.099552	0.099426	0.099448	0.099563	0.099766	0.100066	0.100493	0.101108	0.102051	0.103719	0.108311
Expected Risk	0.016843	0.017241	0.017422	0.017453	0.017354	0.017132	0.016779	0.016271	0.015555	0.014504	0.012628
Absolute Difference Risk	0.082709	0.082186	0.082025	0.082111	0.082412	0.082934	0.083714	0.084837	0.086496	0.089214	0.095683
Realised Sharpe Ratio	0.029407	0.029113	0.029272	0.029149	0.028859	0.028468	0.028019	0.027525	0.026911	0.025632	0.018353
Expected Sharpe Ratio	1.474172	1.504564	1.543188	1.587003	1.636383	1.692509	1.757218	1.833329	1.925454	2.041819	2.193420
Absolute Difference Sharpe Ratio	1.444765	1.475451	1.513916	1.557854	1.607524	1.664040	1.729199	1.805804	1.898543	2.016187	2.175067

Table 3 Shrinkage simulation EPO 2 100 size portfolios (own illustration)

When looking at the realised results for each shrinkage parameter, the EPO 1 seems to work best without applying any shrinkage as the average realised return is highest here. The realised risk is lowest for a shrinkage parameter of 0.3. EPO 2 achieves the highest average of realised return and lowest average of realised risk for a shrinkage parameter of 0, indicating that not applying any shrinkage is also optimal in this scenario. In this case, EPO 2 also attains the highest realised Sharpe ratio by not applying any shrinkage. Nevertheless, the KPI is not simply the realised result, but the absolute difference between expectations and realisations, which is why 0.6 and 0.8 are chosen for both strategies as the best performing shrinkage parameters.

For the second dataset the optimal shrinkage parameter differs. As Table 3 and Table 4 show, the difference between the average realised and average expected Sharpe ratios is the lowest when not shrinking at all for both strategies. In that case, both strategies would converge towards the *MVO* strategy. As shown in both tables, both strategies would achieve the same results if they were executed without any shrinkage. For the purpose of showing the performance of the different strategies, the shrinkage parameters will be set to the values which maximised the performance on the smaller dataset. However, this adjustment will be taken into consideration when drawing conclusions on the effectiveness of both strategies.

When comparing both strategies to each other, EPO 2 seems to produce more reliable expectations than EPO 1. The absolute difference of realised Sharpe ratios and expected Sharpe ratios is less for a fully shrunk EPO 2 than for a fully shrunk EPO 1. Nevertheless, the fully shrunk EPO 1 realised higher returns than the fully shrunk EPO 2. EPO 2 lead to a lower realised risk instead.

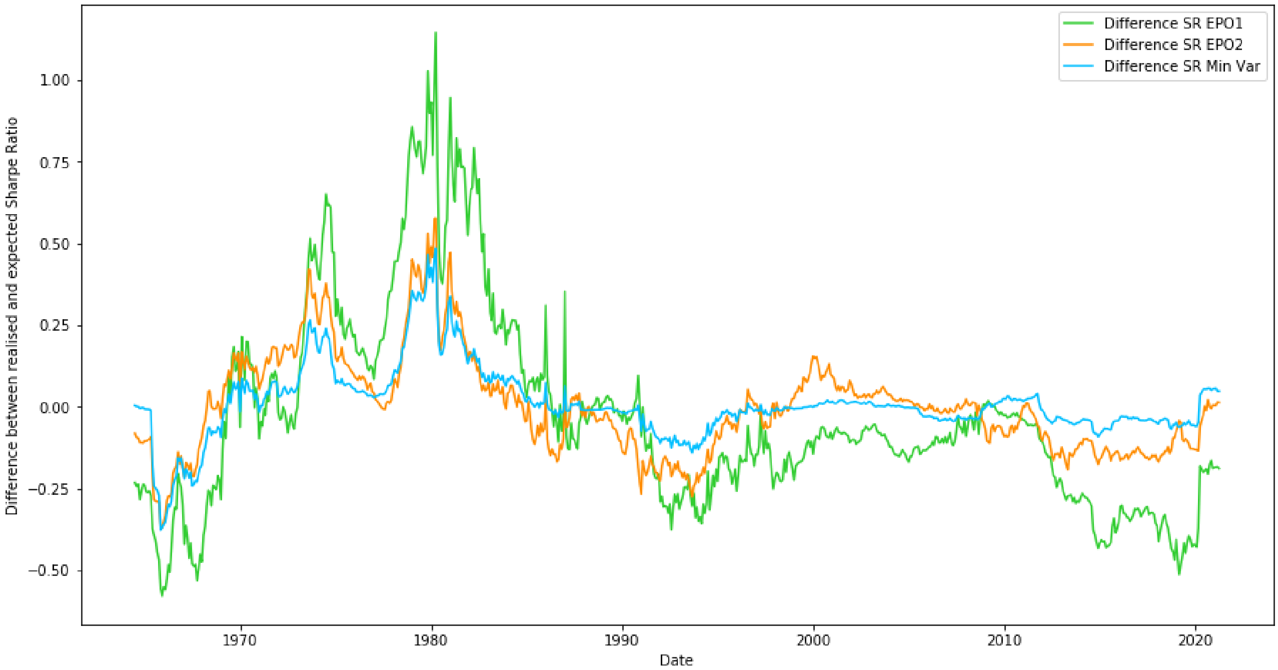

Figure 10 Prediction accuracy Sharpe ratios 10 industry dataset (own illustration)

Investigating the Sharpe ratio within the 10-industry portfolio dataset further, Figure 10 shows the differences over time between expected Sharpe ratio and realised Sharpe ratio for the *MVO*, EPO 1 (with $\theta = 0.6$), and EPO 2 (with $\theta = 0.8$) strategies. Generally, differences of zero are preferred as this relates to accurate predictions by the strategies. If differences occur, positive differences are in this case preferred by investors as this would mean that the realised Sharpe ratios exceeded the predicted ones. EPO 1 seems to produce the most extreme differences between expectations and realised Sharpe ratios. In the first ten years of the dataset, the realised Sharpe ratios were lower than the expected Sharpe ratios. The next ten years were the opposite, with higher realised Sharpe ratios than expected. However, since then the difference was rather constant in the negative space, implying lower realised Sharpe ratios than expected. The predictions of EPO 1 are the least accurate among the three strategies. EPO 2 and the *MVO* are very similar; however, the *MVO* seems to be more stable.

Including the equally weighted portfolio into the set of strategies that are evaluated, the realised Sharpe ratios of each strategy are compared in Figure 11 using the 100-size-portfolios dataset. The equally weighted portfolio seems to produce the least volatile Sharpe ratios over time, providing more evidence for Bruder (2013). For the first ten years, the equally weighted portfolio was the best performing strategy, even though it delivered negative Sharpe ratios. The economic context in the 1970s and 1980s serves as an explanation for the overall poorly performing strategies within that time. The 1970s marked the end of the post-World-War II boom and was paired with high unemployment as well as high inflation rates in the United States. EPO 2 yielded the highest Sharpe ratios from the late 1970s until the end of the 1980s. In times of expansionary monetary policies EPO 2 was the only strategy leading to positive Sharpe ratios. From the beginning of the 1990s until the Financial Crisis in 2008, all strategies achieved positive Sharpe ratios. The best performing strategy in that time was the *MVO*. From the time of the Financial Crisis until approx. 2015, the equally weighted portfolio produced highest Sharpe ratios. It is overall worth mentioning that the equally

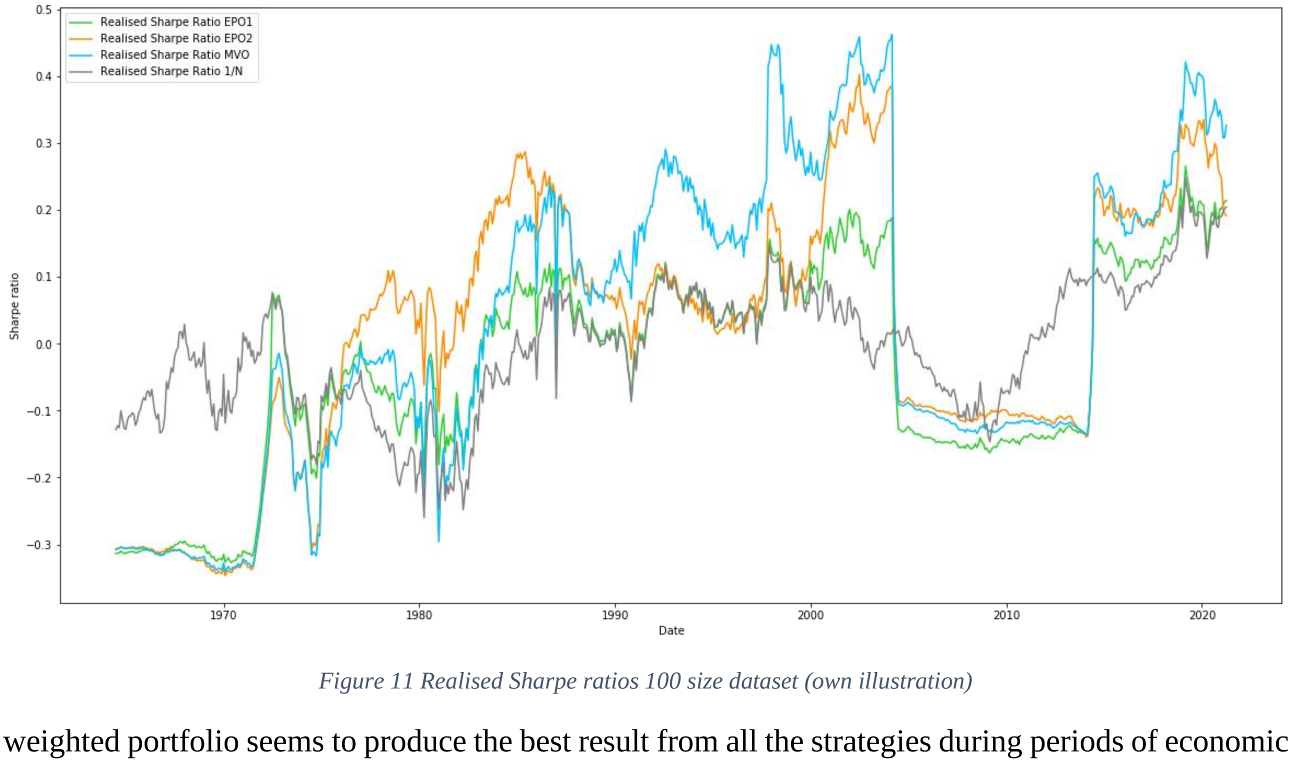

Figure 11 Realised Sharpe ratios 100 size dataset (own illustration)

weighted portfolio seems to produce the best result from all the strategies during periods of economic recessions or financial crises. EPO 1 performed the worst, yielding a lower Sharpe ratio than EPO 2 and *MVO* and only marginally better than the equally weighted portfolio.

7. Discussion

This part discusses the main findings from section 6. The differently performing strategies are discussed as well as interpretations on why the different strategies might have performed better than the others are outlined. Furthermore, the discussion will analyse the components leading towards a well-functioning strategy in setting expectations of risk, return, and Sharpe ratios. Additionally, the practical usefulness of Enhanced Portfolio Optimisation with high-dimensional data is analysed.

Two performance indicators are discussed: the (1) forecasting accuracy regarding Sharpe ratio and the (2) realised Sharpe ratio. In terms of forecasting accuracy, the best performing strategy is the Mean-Variance Optimisation. Its difference between the realised and expected Sharpe ratios is constantly the closest to zero out of all strategies (Figure 10). The fact that *MVO* outperforms EPO 1 and EPO 2 is surprising, given that *EPO* is built to reduce the impact of noise on the expected returns and expected risk. A lower impact of noise can only be translated into a more precise prediction. However, the opposite seems to be the case. The reason for the poorer prediction accuracy of EPO 1 could potentially be explained by a changing optimal shrinkage parameter over time.

It is possible to differentiate between the performance of the active portfolio optimising strategies (EPO 1, EPO 2, *MVO*) and the equally weighted portfolio based on the economic cycle. It is obvious from Figure 11 that the equally weighted portfolio performs better in financial downturns. Procacci and Aste (2019) find that the correlation between different assets changes between bull and bear phases. This would imply on the one hand, that the correlation matrices estimated on past data during a specific economic cycle and used to calculate the portfolio optimising weights are not valid anymore for all three active strategies for the next economic cycles. On the other hand, a changing correlation structure depending on the state of the economy also implies that the optimal shrinkage parameter for the EPO 1 strategy might change over time, too. This finding could explain the poor performance of the active strategies in the 1970s as well as during the financial crisis in the 2010s. The shrinkage parameter has been set to a level that minimizes the difference of the average realised and expected Sharpe ratios, and is not constantly adjusted. Nevertheless, Dzikevičius and Vetrov (2013) refer to realised returns and find that actively managed strategies outperform the equally weighted portfolio in a complete economic cycle. In periods of downturns, the equally weighted approach yields superior results. But the returns achieved with actively managed portfolios in periods of economic recoveries and expansions are higher than the difference to the returns achieved by the equally weighted portfolio in economic downturns and slowdowns. Even though this does not increase the prediction accuracy, it is still a point worth mentioning in favour of *EPO*.

The same reasoning cannot be applied to justify the worse performance of EPO 2 against *MVO* though, as it is shrunk towards the average prevailing correlation matrix each month. It should therefore caption changes of the underlying correlations. However, Andersen et al. (2006) present Constant Conditional Correlations (CCC) and Dynamic Conditional Correlations (DCC) as means to estimate the correlation matrix on a rolling basis. The key difference towards the computation of the correlation matrix applied by EPO 2 is the application of an autoregressive moving average (ARMA) model. ARMA models allow to estimate correlation matrices based on historical correlations and not just the historical return data displaying an advantage of the model. The forecasting accuracy has been shown to be superior (Andersen et al., 2006). Applying EPO 2 on a correlation matrix estimated with CCC or DCC could thus yield better results than the current EPO 2 due its potentially better estimation of the correlation matrix.

Referring to the second performance indicator of interest, namely the realised Sharpe ratio, EPO 1 is constantly outperformed by EPO 2 and *MVO*. In certain times, it is even outperformed by the equally weighted portfolio. Table 3 and 4 show that EPO 2 produces an overall higher realised Sharpe ratio than EPO 1. One reason for the difference between EPO 1 and EPO 2 is certainly the fact that shrinking towards the average correlations is a better approximation for the true underlying correlation than shrinking towards the identity matrix. The average correlations in the two datasets throughout the whole periods are 0.72 and 0.58 in the 10-industry portfolio and the 100-size portfolios, respectively. Shrinking towards zero might therefore not be the optimal solution and yield inferior Sharpe ratios compared to the EPO 2.

Despite EPO 2 outperforming EPO 1 in most characteristics, in comparison to the *MVO* it is also not constantly producing better results than the *MVO*. The prediction accuracy is better than EPO 1 but inferior to *MVO*. Its realised Sharpe ratio is only for a comparatively small period of ten years better than the realised Sharpe ratio of *MVO*. Throughout both datasets limited evidence is found in favour of either of the *EPO* strategies to outperform the other strategies. Depending on the performance indicator used, they were even outperformed by most strategies. The realised Sharpe ratio in the dataset with 100 portfolios was worse than the *MVO* strategy, as well as the equally weighted in financial crises. The realised Sharpe ratio of the EPO 1 in the 10-industry dataset was better than the other strategies, even though the difference to the predicted Sharpe ratio is quite large. Overall, the question arises of how useful the *EPO* is when working with high dimensional datasets. Pedersen et al. (2021) do not try it out on large datasets. This has been highlighted by Tables 3 and 4, showing that the optimal shrinkage parameters for EPO 1 and EPO 2 in the large dataset are 0 and can be interpreted as the *MVO* always outperforming the *EPO* strategies because not shrinking means the weights are the same as *MVO*.

This thesis provides evidence against the *EPO* in the space of highly dimensional datasets. The nonlinear common factor (NCF) approach taken on by Zhang et al. (2020) might be able to solve that problem. The NFC estimates correlation matrices of high-dimensional datasets over time. Their algorithm provides the advantage that it does not assume linear relationships but instead allows for non-linearity. The authors claim their algorithm to be more accurate at forecasting correlations than the classical shrinkage approach. The shrinkage approach taken on by Pedersen et al. (2021) is not the classical method. However, future research could entail an application of *EPO* with a correlation matrix estimated through the NCF.

One further point worth mentioning is made by Halbleib and Voev (2016) who compute the covariance matrix forecasts based on correlation matrices and volatilities in a similar way as Pedersen et al. (2021). However, they do not use the original volatilities but instead calculate the correlation matrix and the diagonal volatility matrix "independently" from each other. This approach yields the advantage that noise in the estimation of volatilities is also reduced. The *EPO* approach used in this thesis and developed by Pedersen et al. (2021) does not address noise in volatility estimation. The lack of doing so could explain the partial underperformance by EPO 1 and EPO 2 against the equally weighted portfolio and *MVO*.

The discussion so far highlighted the importance of the correlation matrix, as well as the dependence of the performance of *EPO* on it. Pedersen et al. (2021) use the original correlation matrix and shrink it towards the identity matrix. The reason for them to do so can be found by linking the analysis done on the level of the Principal Component portfolios with the weight allocations done at the level of the original assets. Comparing Figure 5.A (Pedersen et al., 2021) and Figures 6 and 7 in this thesis, the variances of the Principal Component portfolios seem to be estimated too high for the eigenvectors with high eigenvalues and too low for the eigenvectors with low eigenvalues. Pedersen et al. (2021) argue that shrinking the correlation matrix towards the identity matrix achieves the estimated variances of the principal components to decrease for the principal component portfolios with high eigenvalues and increase for those with low eigenvalues. This thesis showed that shrinking towards the average correlations is more precise at predicting Sharpe ratios and achieves higher Sharpe ratios. Future research could target different correlation estimates to shrink towards and apply *EPO*.

7.1 Limitations

This thesis is subject to limitations. First, EPO 2 has only been applied to two datasets in this thesis. The dataset of 10 industry portfolios does not contain high-dimensional data. It therefore does not reflect the highest impact estimation noise can have. Second, the way the optimal shrinkage parameter has been selected could be optimised. Instead of only looking at the prediction accuracy of Sharpe ratios, one could take other components into account. Prediction accuracy of volatility and returns could possibly be used to evaluate which shrinkage parameter to use. Third, to place a higher focus on the *PCA* and the machine learning component of this thesis, comparing Pedersen et al.'s (2021) approach to another approach which used machine learning for portfolio optimisation would be beneficial. Due to short time frame available to write this thesis, this has not been undertaken. Last, the results presented in the tables have not been statistically tested and might therefore not be statistically significant. The interpretation of the results is based entirely on the empirical findings which have not been tested for significance.

8. Conclusion

Portfolio optimisation can benefit substantially from the usage of machine learning algorithms such as Principal Component Analysis. The detection and extraction of signals in the variance of estimated correlation matrices is used by Enhanced Portfolio Optimisation. The *PCA* applied to the two datasets of 10-industry portfolios and 100-size portfolios revealed that it is possible to differentiate eigenvectors which are composed of noise and eigenvectors which are based on some underlying structure. Generally, it is not known what the variable underlying those eigenvectors is, yet, as Figures 8,9,10, and 11 show, it is possible to detect those eigenvectors with high eigenvalues and the ones with low eigenvalues. The ones with high eigenvalues are able to explain a larger part of the total variance. This is the contribution of machine learning in addressing the problem of noise in correlation and covariance matrices. Machine learning enables researchers, academics, and practitioners to exactly pinpoint those eigenvectors which cause problems in portfolio optimisation because they are entirely random, due to noise.

By shrinking the original correlation matrix, one can reduce the impact of the noisy eigenvectors on the weight allocation in optimised portfolios. Pedersen et al. (2021) shrink towards the identity matrix, while this thesis applies shrinkage of the correlation matrix towards the average correlations of the underlying assets. The resulting covariance matrices are used to perform portfolio optimisation. Their performance is judged based on the prediction accuracy against the classical *MVO* as well as on the overall achieved Sharpe ratios against the *MVO* and the equally weighted portfolio. It proved difficult to beat the *MVO* portfolio in terms of prediction accuracy as well as realised Sharpe ratio for the *EPO* by Pedersen et al. (2021). The modified *EPO* introduced in this thesis managed to beat *MVO* for a time frame of about ten years in realising a higher Sharpe ratio. The prediction accuracy performed better than the strategy proposed by Pedersen et al. (2021), however, not quite as strong as the *MVO*.

Reasons for the poor performance can be assumed to be found in the correlation matrix. Correlations are found to change depending on the economic cycle, which has implications on the optimal shrinkage parameter used by *EPO* (Procacci & Aste, 2019). Further, there are more possibilities to estimate the correlation matrix than solely basing it on historical returns. The reason for Pedersen et al. (2021) to use this simple approach is its simplicity, however, it has been proven to be a rather poor predictor of future correlations (Andersen et al., 2016). Overall, the outstanding performance of *EPO* reported by Pedersen et al. (2021) cannot be confirmed by the findings of this thesis. The modified version yielded superior results but struggled to beat the classical *MVO*. In spite of the performance, it has been shown that Machine Learning can play a major role in the detection and circumvention of noise in correlation and covariance matrices. Several other approaches and algorithms have been shown to calculate correlation matrices. Combining those with the Enhanced Portfolio Optimisation could be a promising opportunity on making Markowitz' Mean Variance Optimisation finally work.

References

Abdi, H., & Williams, L. J. (2010). Principal Component Analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2-47.

Acock, A. C. (2005). Working With Missing Values. *Journal of Marriage and Family*, 1012-1028.

Andersen, T. G., Bollerslev, T., Christoffersen, P. F., & Diebold, F. X. (2006). Volatility and Correlation Forecasting. In G. Elliott, C. Granger, & A. Timmermann, *Handbook of Economic Forecasting* (pp. 778-878). Amsterdam: North-Holland.

Ang, A. (2012). Mean-Variance Investing. *Columbia Business School Research Paper*.

Babu, A., Levine, A., Ooi, Y. H., Pedersen, L. H., & Stamelos, E. (2020). Trends Everywhere. *Journal of Investment Management*, 52-68.

Ball, C. A., & Torous, W. N. (2000). Stochastic correlation across international stock markets. *Journal of Empirical Finance*, 373-388.

Becker, F., Gürtler, M., & Hibbeln, M. (2009). Markowitz versus Michaud: Portfolio optimization strategies reconsidered. *Working Paper Series*.

Berk, J., & DeMarzo, P. (2014). *Corporate Finance.* Harlow: Pearson.

Black, F., & Litterman, R. (1992). Global Portfolio Optimization. *Financial Analysts Journal*, 28-44.

Brodie, J., Daubechies, I., De Mol, C., Giannone, D., & Loris, I. (2009). Sparse and stable Markowitz portfolios. *PNAS*, 12267-12272.

Bruder, B., Gaussel, N., Richard, J.-C., & Roncalli, T. (2013). Regularization of Portfolio Allocation. *SSRN Electronic Journal*.

DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy? *The review of Financial Studies*, 1915-1953.

Dzikevičius, A., & Vetrov, J. (2013). Investment Portfolio Management Using the Business Cycle Approach. *Business: Theory and Practice*, 57-63.

Elman, M. R., Minnier, J., Chang, X., & Choi, D. (2020). Noise Accumulation in High Dimensional Classification and Total Signal Index. *Journal of Machine Learning Research*, 1-23.

Fabozzi, F. J., Focardi, S. M., & Kolm, P. N. (2010). *Quantitative Equity Investing.* Hoboken: John Wiley & Sons, Inc.

Fan, J., Fan, Y., & Lv, J. (2008). Jigh dimensional covariance matrix estimation using a factor model. *Journal of Econometrics*, 186-197.

Gupta, P. L., & Gupta, R. D. (1987). Sample size determination in estimating a covariance matrix. *Computational Statistics & Data Analysis*, 185-192.

Halbleib, R., & Voev, V. (2016). Forecasting Covariance Matrices: A Mixed Approach. *Journal of Financial Econometrics*, 383-417.

Harris, C., Millman, K., & van der Walt, S. (2020). Array programming with NumPy. *Nature* , 357-362.

Heyde, C. (2014). Central Limit Theorem. *Encyclopedia of Actuarial Science*.

Hunter, J. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 90-95.

Jagannathan, R., & Ma, T. (2003). Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps. *The Journal of Finance*, 1651-1683.

Jorion, P. (1992). Portfolio Optimization in Practice. *Financial Analysts Journal*, 68-74.

Kirby, C., & Ostdiek, B. (2012). Optimizing the Performance of Sample Mean-Variance Efficient Portfolios.

Laloux, L., Cizeau, P., Bouchaud, J.-P., & Potters, M. (1999). Noise Dressing of Financial Correlation Matrices. *Physical review letters*, 109-111.

Laloux, L., Cizeau, P., Potters, M., & Bouchaud, J.-P. (2000). Random Matrix Theory and Financial Correlations. *International Journal of Theoretical and Applied Finance*, 391-397.

Ledoit, O., & Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of empirical finance*, 603-621.

Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 365-411.

Ledoit, O., & Wolf, M. (2020). The power of (non-)linear shrinking: A review and guide to covariance matrix estimation. *Working Paper*.

Liu, H., Aue, A., & Paul, D. (2015). On the Marcenko-Pastur Law for Linear Time Series. *The Annuals of Statistics*, 675-712.

Longin, F., & Solnik, B. (1995). Is the correlation in international equity returns constant: 1960-1990? *Journal of International Money and Finance*, 3-26.

López de Prado, M. M. (2020). *Machine Learning for Asset Managers.* New York: Cambridge University Press.

Marčenko, V. A., & Pastur, L. A. (1967). Distribution of Eigenvalues for some sets of Random Matrices. *Mathematics of the USSR-Sbornik*, 457-483.

Markowitz, H. (1952). Portfolio Selection. *The Journal of finance*, 77-91.

McKinney, W. (2010). *Data Structures for Statistical Computing in Python.*

Michaud, R. O. (1989). The Markowitz Optimization Enigma: Is "Optimized" Optimal? *Financial Analysts Journal*, 31-42.

Michaud, R., & Michaud, R. (1998). *Efficient Asset Management: A Practical Guide to Stock Portfolio Management and Asset Allocation.* New York: Oxford University Press.

Moskowitz, T. J., Ooi, Y. H., & Pedersen, L. H. (2012). Time series momentum. *Journal of Financial Economics*, 228-250.

Negahban, S., & Wainwright, M. (2011). Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *The Annuals of Statistics*, 1069-1097.

Ormos, M. (2012). Performance Analysis of Equally weighted Portfolios: USA and Hungary. *Acta Polytechnica Hungarica*, 155-168.

Pafka, S., & Kondor. (2001). Noisy covariance matrices and portfolio optimization. *Physics of Condensed Matter*.

Pedersen, L. H., Babu, A., & Levine, A. (2021). Enhanced Portfolio Optimization. *Financial Analysts Journal*, 124-151.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Duchesnay, E. (2011). Scikit-learn: Machine Learning in {P}ython. *Journal of Machine Learning Research,* 2825-2830.

Plerou, V., Gopikrishnan, P., Rosenow, B., Nunes Amaral, L. A., & Stanley, H. E. (1999). Universal and non-universal properties of cross-correlations in financial time series. *Physical Rview Letters*.

Procacci, P. F., & Aste, T. (2019). Forecasting market states. *Quantitative Finance*, 1491-1498.

Riksbank, T. S. (1990, October 16). *The Nobel Prize*. Retrieved from The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 1990: https://www.nobelprize.org/prizes/economic-sciences/1990/press-release/

Ringnér, M. (2008). What is principal component analysis. *nature Biotechnology*, 303-304.

Roncalli, T. (2010). Understanding the Impact of Weights Constraints in Portfolio Theory.

Saunders, M. N., Lewis, P., & Thornhill, A. (2019). *Research Methods for Business Students.* New York: Pearson.

Sharpe, W. F. (1963). A Simplified Model for Portfolio Analysis. *Management Science*, 277-293.

Sharpe, W. F. (1964). Capital Asset Prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 425-442.

Shaw, D. X., Liu, S., & Kopman, L. (2008). Lagrangian relaxation procedure for cardinality-constrained portfolio optimization. *Optimization Methods and Software*, 411-420.

Shlens, J. (2014). A Tutorial on Principal Component Analysis. *Google Research*, 1-12.

Tempelaar, D., & Kerckhoffs, C. (2015). *Sharpe, Business Statistics, and extra texts.* Harlow: Pearson.

Tsay, R. S. (2005). *Analysis of Financial Time Series.* Hoboken: John Wiley & Sons, Inc.

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 37-52.

Zakamulin, V. (2014). A Test of Covariance Matrix Forecasting Methods. *Journal of Portfolio Management*, 97-108.

Zhang, Y., Rolling, C., & Yang, Y. (2020). Estimating and forecasting dynamic correlation matrices: A nonlinear common factor approach. *Journal of Multivariate Analysis*, 104710.

Zhu, X., & Wu, X. (2004). Class Noise vs. Attribute Noise: A Quantitative Study of Their Impacts. *Artificial Intelligence Review*, 177-210.