UCLouvain
FUCaM MONS

LOUVAIN
School of Management

Louvain School of Management

Portfolio Selection via Principal Component Analysis

Author: Clément Piedboeuf
Supervisor: Nathan Lassance
Academic Year: 2022-2023
Business Engineering – Financial Engineering

Summary

Portfolio Selection is a critical process for investors and portfolio managers seeking to optimise returns while managing their risks.

We provide an in-depth analysis of four different portfolio optimisation strategies that uses principal components. The strategies are: the Bounded-Noise portfolio, the PC-Mean-Variance and PC-Minimum-Variance portfolios, the portfolios of Severini and finally the PC-Variance-Parity portfolio.

In first instance, we compare the performance of the different strategies to each other. We have found that the Bounded-Noise portfolio performs significantly better than the other portfolios. But closely followed by the PC-Mean-Variance portfolio. We also saw that the second portfolio of Severini yield very poor performance.

In second instance, we compare the PC-based strategies to the classical strategies. The Mean-variance portfolio, the Minimum-Variance portfolio, the equally weighted portfolio and the Asset-Variance-Parity portfolio. We saw that most of PC-based strategies outperformed the classical strategies at the exception of the second portfolio of Severini that performs significantly worse.

Finally, we highlight the positive impact of an optimal selection of principal components on the performance of the PC-based strategies.

Overall, the thesis contributes to the understanding of portfolio optimization strategies and provides recommendations for investors based on empirical analysis.

Acknowledgements

I want to start by acknowledging my supervisor, Pr. Nathan Lassance. Your expert guidance and valuable feedback have been crucial in shaping this Master's thesis. The intellectual rigor you brought to our discussions and your keen eye for detail have undoubtedly elevated the quality of my work. I am deeply grateful for your mentorship.

Next, my heartfelt appreciation goes to my parents. Thank you for your limitless love and your unwavering belief in my capabilities through the years. You've taught me the value of hard work and determination, and your encouragements have been essential in reaching this milestone. You have always been there to lift me when I stumbled and to celebrate each small success along the way. For this, and everything else, I am profoundly grateful.

I would also like to extend a special thanks to Ms. Hila Peer, my English teacher. Your attention to detail and valuable advice significantly enhanced the quality of my thesis. I am sincerely thankful for your expertise.

Finally, to my friends, my university years have been a remarkable journey, thanks to you. The friendship and the shared experience have made this journey not only possible but genuinely enriching. I am thankful to have shared this experience with you, these past five years would not have been the same without you.

Contents

List of Figures

List of Tables

Chapter I

Introduction

Portfolio selection is a critical process that allows investors and portfolio managers to distribute their investments across a range of assets to optimize returns while effectively managing risk. Despite extensive literature in this field, gaps remain, particularly concerning the role of Principal Component Analysis (PCA)-based strategies and how they compare to more traditional methods. This research aims to fill these gaps by providing a comprehensive examination of different optimization strategies, their theoretical foundations, advantages, and limitations.

The main research question that this thesis addresses is: "What is the empirical performance of different portfolio optimization techniques based on principal component analysis (PCA) compared to simpler strategies?" . The significance of this question lies in its potential to shape future investment strategies and portfolio management practices. It serves to improve the decision-making process of portfolio managers and investors and contributes to the academic discourse on portfolio optimization strategies.

This master thesis starts with a theoretical review of different PCA-based optimization strategies presented by Meucci (2010), Chen and Yuan (2016), Zhao et al. (2019) and Severini (2022). The study will then compare the performance of the discussed strategies on real data. The statistical significance of the different Sharpe ratios resulting from the strategies will be assessed with a specific hypothesis test.

Key findings from this study reveal that the bounded-noise portfolio of Zhao et al. (2019) consistently outperforms other strategies, with most of the Sharpe ratios of the strategies being significantly different from the bounded-noise portfolio at a 95% confidence level. The study also underscores the crucial role of selecting the appropriate number of principal components and shows that the use of a biased covariance matrix estimator enhances the performance of the mean-variance and minimum-variance strategy.

Chapter II

Literature Review & Research Question

This chapter provides the theoretical background needed to analyse the portfolio optimization methods. It provides a general overview of portfolio theory, and it then discusses the Principal Component Analysis (PCA) and the techniques based on it. The literature on this topic has been developed over the last seven decades and forms the basis for the knowledge that we have today.

1 Portfolio Theory

1.1 Foundational Concepts

For this work, let denote:

- x_t: The vector of the returns of N assets for observation t out of T
- μ: The mean of the returns
- Σ: Covariance matrix of the returns
- w: The vector of the weights of the portfolio such that $1^T w = 1$

The portfolio return is then, $R^P = w^T x_t$ with average expected return μ and variance $w^T \Sigma w$.

Markowitz (1952) explains that often two assumptions are considered. The first assumption assumes that investors tend to maximise their expected returns. The second assumption is based on the fact that investors do consider expected return a desirable thing and variance of return an undesirable thing. This creates a trade-off between risk and expected return. Investors are well aware of this balance, the higher the risks taken, the greater the potential for return (Hull, 2018).

Markowitz (1952) was among the first to understand this. A further development of this analysis has been made by Sharpe (1964), who introduced the capital asset pricing model. This model describes the relationship between the expected return of an asset, the risk-free rate of return, and the expected return of the market portfolio.

When using the expected return and standard deviation to describe each investment option, a graphical representation of all the investments can be created (Hull, 2019). This can be seen in figure II.1 below. In addition, a linear relationship can be witnessed between risk and return. Indeed, an increase in the risk often leads towards an increase in mean return.

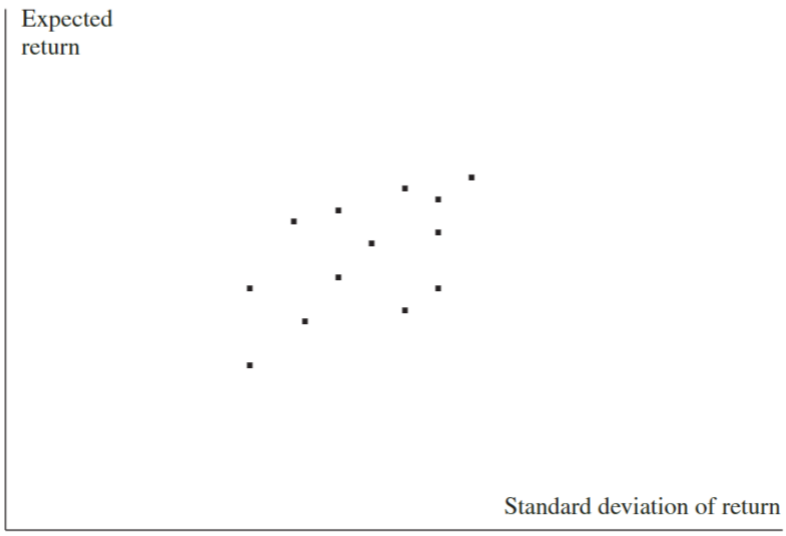


Figure II.1: Linear relationship between Risk and Return (Hull, 2018)

1.2 Mean-Variance portfolio

A well-known portfolio-selection framework is the mean-variance portfolio. It was presented by Markowitz in the Journal of Finance in 1952. He saw that often the portfolio-selection framework did not imply diversification. And this, no matter how the expected returns were formed, or which different discount rates were used. The hypothesis was that the investor always placed his funds in the security yielding the highest discounted expected return. Diversified portfolios were, under any circumstances, never preferred over all non-diversified portfolios.

To overcome this obstacle Markowitz (1952) assumed that there is an efficient portfolio leading to efficient asset-selection which are almost all diversified.

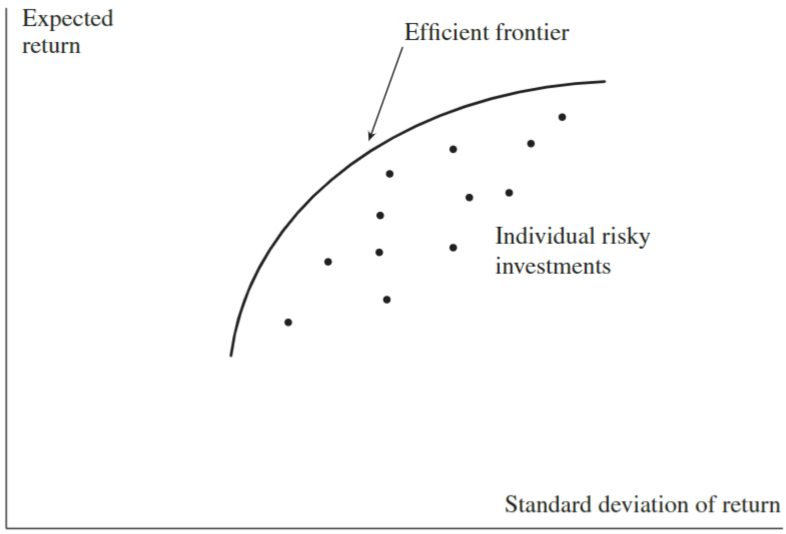


Figure II.2: Efficient Frontier of Markowtiz (Hull, 2018)

The efficient frontier can be witnessed on figure II.2 above and is a combination of every possible portfolio of the available risky investments. There is no investment that dominates on the efficient frontier but the area under the curve is not attractive since for any of these points, there is a point on the efficient frontier that has higher expected return and lower variance (Hull, 2018).

In addition, Markowitz (1952) find that the Mean-Variance Portfolio also implies diversification between industries. This is due to the fact that generally firms within the same industry tend to do poorly at the same time while firms in different industry less.

The so-called mean-variance portfolio maximizes the expected returns while minimising the risk (variance), in other words, it finds a trade-off between expected return and risk.

$$\begin{aligned} \max \quad & \frac{w^T\mu}{\sqrt{w^T\Sigma w}} \\ \text{subject to} \quad & 1^T w = 1 \end{aligned} \tag{II.1}$$

with λ being the parameter determining the weight of the risk in the optimisation.

The solution to this optimisation problem is given by the equation that follows.

$$w_{\text{MSR}} = \frac{\Sigma^{-1}\mu}{1^T\Sigma^{-1}\mu} \tag{II.2}$$

Or when estimated on a dataset:

$$\hat{w}_{\text{MSR}} = \frac{\hat{\Sigma}^{-1}\hat{\mu}}{1^T\hat{\Sigma}^{-1}\hat{\mu}} \tag{II.3}$$

The efficient portfolios produce a combination of (σ_P, μ_P) that leads to the creation of the efficient frontier.

But the mean-variance portfolio has several problems. A first drawback is that it is very difficult to estimate $\hat{\mu}$. Black (1993), Michaud (1989) and Broadie (1993) argue that historical returns are not reliable predictors to estimate expected returns. There are too many factors that can impact returns beyond the historical record which can result in significant errors. And even if returns are independent and identically distributed over time it is hard to estimate them. These issues can lead to instability in the portfolio optimization. Of course, alternative methods using robust estimators of $\hat{\mu}$ can be more dependable in specific situations. But generally, expected returns are by nature uncertain and investors should be careful when relying on estimation of expected return.

A second drawback is that the mean-variance portfolio has a high sensitivity to $\hat{\mu}$. This sensitivity arises because the mean-variance approach searches for the portfolio with the highest expected return for a given level (Best and Grauer, 1991).

These two drawbacks result in poor out-of-sample performance of the mean-variance approach (DeMiguel, 2009; Chen and Yuan, 2016).

1.3 Minimum-variance portfolio

To get around the problem of estimating $\hat{\mu}$ and therefore, enhance the out-of-sample performance, a modification can be applied to the mean-variance portfolio. By assuming that the sample mean of all assets is equal, an alternative strategy known as the "Minimum-Variance Portfolio" can be implemented.

The minimum-variance portfolio is the strategy which minimises the portfolio variance, where the solution is given by the following.

$$\hat{w}_{\text{MV}} = \frac{\hat{\Sigma}^{-1}1}{1^T\hat{\Sigma}^{-1}1} \tag{II.4}$$

It is interesting to note that the minimum-variance portfolio represents the most left point on the efficient frontier.

By removing the expected returns out of the optimisation constraints, the strategy is inevitably less subject to estimation error. Nothing is lost when omitting $\hat{\mu}$ since estimation errors in the sample mean are so large (Ma and Jagannathan , 2002).

The minimum-variance portfolio has generally better out-of-sample performance than the mean-variance portfolio (DeMiguel and Nogales, 2007; Clarke et al., 2006). Indeed, next to the problem the strategy solves, Haugen and Heins (1972) have shown that stock portfolios with lower variance achieved higher average returns than their riskier counterparts over the long term.

The fact that low-risk stocks provide greater returns than high-risk stocks is called the low-risk anomaly. In addition, the portfolio can also be an attractive strategy for risk averse people. Overall, the minimum-variance portfolio is in many cases better a alternative to the mean-variance optimization (Clarke et al. 2006).

However, there are some challenges associated with the minimum-variance portfolio. The main challenge is that the minimum-variance portfolio is still sensitive to estimation error in the covariance matrix

1.4 Equally weighted portfolio

Another well-known portfolio is the equally weighted portfolio. DeMiguel et al. (2009) explain that this portfolio is a naïve diversification strategy which involves investing equal weights in all assets. The solution to this strategy is given by:

$$w_{\text{EW}} = \frac{1}{N} \tag{II.5}$$

DeMiguel et al. (2009) find that the mean-variance strategy has a tendency to overfitting and instability when applied to real-world data. In addition, the strategy also has high transaction costs and estimation error. The naïve diversification does not have those problems due to its simplicity. This makes the strategy more robust and sometimes a better choice for investors.

The equally weighted portfolio, thus, often outperforms the mean-variance portfolio out-of-sample. But the drawback is that if individual risks are significantly different the strategy can lead to a very limited diversified portfolio (Maillard et al., 2008).

1.5 Asset-Variance-Parity portfolio

The asset-variance-parity portfolio (AVP) constitutes a middle-ground between minimum-variance and equally weighted portfolio. The goal of this strategy is to create a portfolio where each asset contributes equally to the portfolio-return volatility (Maillard et al. 2008; Bai et al., 2015).

The volatility of the portfolio $\sigma(w) = \sqrt{w^T \Sigma w}$ can be decomposed as follows using Euler's theorem:

$$\sigma(w) = \sum_{i=1}^{N} \frac{w_i (\Sigma w)_i}{\sqrt{w^T \Sigma w}} \tag{II.6}$$

The risk contribution (RC) from the ith asset to the total risk $\sigma(w)$ can then be calculated as:

$$RC_i = \frac{w_i (\Sigma w)_i}{\sqrt{w^T \Sigma w}} \tag{II.7}$$

The asset-variance-parity portfolio (AVP) equalises the risk contributions:

$$RC_i = \frac{1}{N} \sigma(w) \tag{II.8}$$

When the correlation between the assets is 0, Maillard et al. (2010) find that the solution to this strategy can be given by the following formula.

$$w = \frac{\sigma^{-1}}{1^T \sigma^{-1}} \tag{II.9}$$

If the correlation is non-zero, no closed-form solution can be found. Indeed, an optimisation has to be made to find the right weights.

$$\min \quad w^T \Sigma w - \sum_{i=1}^{N} ln(w_i) \tag{II.10}$$

The variance of this portfolio is smaller than that of equally weighted. Ji and Lejeune (2018) suggest that this strategy can provide a useful benchmark for portfolio optimization. However, assets are often correlated which can distort the results. Therefore, it is often not the best strategy to create a portfolio with optimal risk diversification and returns. This is where the PCA-Risk-Parity portfolio comes into play, by using decorrelated principal components. This strategy will be discussed more in detail in section 4.2.

2 Robust estimation for covariance matrix

In this subsection we will review how to create a robust estimation of the covariance matrix to obtain better out-of-sample results. We will first focus on the shrinkage estimator of Ledoit-Wolf and then we will briefly discuss the estimation of the covariance matrix with principal components.

2.1 Definition robustness

The definition of robustness is given by Markowitz (1952) and goes as follows.

Definition 1 (Robustness). Robustness in portfolio optimization refers to the portfolio's ability to maintain performance and stability despite changes in market conditions and uncertainties.

The robustness of a portfolio strategy can be assessed with in-sample and out-of-sample tests using rolling-windows strategies (See Section III.2).

2.2 Covariance Estimation with Ledoit-Wolf

One way to reduce error in the covariance matrix is to use shrinkage methods. Indeed, these methods can significantly improve the out-of-sample performance of portfolio strategies. For example, the mean-variance portfolio can be more robust to estimation errors and tail risks (Best and Grauer, 1991).

A common method consists of shrinking the sample covariance matrix towards a target covariance matrix. The shrinkage of the covariance Matrix with the estimator of Ledoit and Wolf (2004) is well-known. This method is considered to be more stable and robust and is given by the following formula.

$$\hat{\Sigma}^{LW} = (1-\delta)\hat{\Sigma} + \delta\hat{T} \tag{II.11}$$

Σ^{LW} being the linear combination of the unbiased-high-variance sample estimate $\hat{\Sigma}$ and a biased-low-variance target estimate T. For the target estimate one can use the average variance of the returns, $T = \sigma^2 I_n$.

δ being the shrinkage intensity which minimises the distance between Σ^{LW} and Σ and where the distance is calculated as the sum of all the distances a_{ij} between each other or as: $\sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}^2}$

2.3 Covariance Estimation with PC's

It can be challenging to estimate the covariance matrix in a high-dimensional database. Therefore, the principal component analysis can be used to reduce the dimensions and create a more robust estimation of the covariance matrix (See section 4.1).

3 Principal Component Analysis

This section will introduce the concept of PCA that will be usefull in section 4. We will first define what a PCA and the principal components are and we will then analyse how they are computed. Second, we will analyse the main criticisms and limitations of this strategy in detail. And third, we will present methods to define the optimal number of principal components to select.

3.1 Conceptual framework and mathematical formulation

It becomes increasingly difficult to interpret large datasets. Thus, it is important to use methods to reduce their dimensionality in a way that most of the information of the data is preserved. The most widely used technique is called the principal component analysis (PCA) and was conceived by Pearson in 1901 and later developed by Hotelling (1933). Its goal is to find a low dimensional representation of the dataset that preserves most of the information or "variability" (Jolliffe and Cadima, 2016; Mishra et al. , 2017).

PCA is widely used, and its main idea is that a dataset is defined in a p-dimensional space, but these dimensions are often not all useful. The principal component analysis will only look at a small number of dimensions that are useful. Indeed, it creates new variables (or dimensions) that are linear combinations of the prior variables in the datasets, that maximises variance and which are not correlated. Those can then be used for other purposes (Jolliffe and Cadima, 2016).

The new variables are called principal components (PCs) and they can be defined, here by Lassance et al. (2021), as follows:

Definition 2 (PCA). The principal components are the standardized projection of the asset returns X onto the first K eigenvectors:

$$Y := \Lambda^{-\frac{1}{2}} V' X \tag{II.12}$$

Note that the covariance matrix of X can be written as $\Sigma = Var(X) = V\Lambda V$. With the diagonal matrix Λ representing the variance of Y, containing all the K eigenvalues (λ_n) sorted in a decreasing order and with matrix V containing all the K associated orthonormal eigenvectors (v_n) (Dray, 2008; Jolliffe and Cadima, 2016; Jolliffe, 2002).

Mishra et al. (2017) explains that PCA was in previous days not very popular due to its computational complexity. But, with the advent of electronic computers, reducing the dimensionality of a dataset has become costless and very quick (Xu et al., 2016). Therefore, PCA has become widely used in data analysis and more specifically in portfolio optimisation.

Another advantage of PCA is that it aims to extract the most important information from the data while reducing the size of the data. This makes the interpretation and the analysis of the structure of the observations and variables simpler and faster (Mishra et al., 2017).

Abdi and Williams (2010) and Jolliffe and Cadima (2016) explain that the first eigenvector v_1 used to form the first principal component Y_1 can be found by solving:

Definition 3 (Eigenvector 1) . Define V1 as the maximisation of the variance of the combination:

$$v_1 = \underset{v'v=1}{\text{argmax}} \ \text{Var}(v'X) = \underset{v'v=1}{\text{argmax}} \ v'\Sigma v \tag{II.13}$$

An additional restriction has been imposed to have a well-defined solution for this problem. Using unit-norm vectors is one of the most common restrictions (Makamo, 2020).

Note that a second restriction is added to find the second eigenvector v_2. Indeed, eigenvector 1 and 2 have to be uncorrelated (Abdi and Williams, 2010). The conditions required for the second eigenvector can be written as:

$$v_2^T v_2 = 1, \quad v_1^T v_2 = 0$$

Repeating this process K times yields the matrix of eigenvectors V which can be plugged in formula II.12.

Going in the other direction is also possible. This is called reconstruction of variables. The matrix X is reconstructed using the principal components.

$$\hat{X} = VY = VV^T X$$

3.2 Criticisms and limitations

Despite the utility of PCs to reduce the dimensions of a data set, the method suffers from several issues.

To perform a PCA, a complete dataset is needed. If any data is missing, special methods, such as deletion, imputation or modelling techniques, will have to be used to handle this. This can heavily impact the accuracy of the covariance and correlation matrices, resulting in bias or distortion (Tabachnick and Fidell, 2013).

In addition, the PCA is limited by the fact that it produces only uncorrelated, rather than truly independent outputs, which presents certain drawbacks. Sometimes, correlation may not be enough. For instance, when assuming a linear relationship and a multivariate Gaussian distribution, the risk associated with joint extreme events may be underestimated. In such cases, the use of PCs can prove to be suboptimal (Lassance et al. 2022).

3.3 Optimal number of Principal components

Since the number of principal components have a huge impact on the analysis, an important task is to estimate the right number of components to retain. This choice can be seen as a bias-variance trade-off. It is crucial because it could lead to a loss of information (underestimation) or to the introduction of random noise (overestimation). This can be problematic for the interpretation of the results or for other analyses (Dray, 2008). Thus, the number of PCs (K) should be chosen such that the K first PCs are relevant, whereas all the ones after K are irrelevant.

To compute the quality of a K-dimensional approximation, the variability that the selected PCs retain can be measured. A simple method is the use of the percentage of explained variance. This is simple and intuitive but reliant on the choice of α (Jolliffe and Cadima, 2016).

$$\frac{\sum_{i=1}^{K} \lambda_i}{\sum_{i=1}^{N} \lambda_i} > \alpha, \tag{II.14}$$

where $\alpha \in [0, 1]$ and normally quite close to 1.

A common practice is to use 95% as predefined percentage of total variance explained to compute the number of selected PCs. Often the first few PCs retain most of the variance but there are circumstances where the last ones may be interesting, for example outlier detection (Jolliffe and Cadima, 2016).

Another method consists of plotting the eigenvalues based on their size. Then, it becomes easy to see if there is a point such that the slope of the graph goes from "steep" to "flat". The PCs before this change are then kept. This is called the "elbow test" (Abdi and Williams, 2010).

A last method developed by Bai and Ng (2002) permits to select K that is consistent in high dimension (N, $T \to \infty$) by using the formula:

$$\hat{K} = \arg \min_{K \in [1, K_{\max}]} \left\{ \ln \left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \varepsilon_{it}^2 \right) + K \left(\frac{N+T}{NT} \right) \ln \left(\frac{NT}{N+T} \right) \right\} \tag{II.15}$$

where $\varepsilon = X - \hat{X} = X - VV^T X$ and $K_{\max} \leq N$.

4 Portfolio optimization techniques based on PCs

4.1 PC-subspace portfolio

In high-dimensional databases, the sample estimator of the covariance matrix can be a poor estimator of the true estimator. Chen and Yuan (2016) proposed to reduce the dimensions of the dataset. By using only the first K Principal components we obtain a more robust estimation of the covariance matrix than can alleviate estimation error and improve out-of-sample performance. Tan (2012) explains that a robust estimation of Σ can be estimated by the covariance matrix of the reconstructed asset returns X. Remember the following formula:

$$\hat{\Sigma_K} = \mathrm{Var}(X^*) = \mathrm{Var}(VY) = V\mathrm{Var}(Y)V' = V\Lambda V'$$

where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_K)$ is a diagonal matrix of the variances of the K PCs $Y_1, \ldots, Y_K$. Based on how many principal components are retained, the covariance matrix will account for more or less variance. This robust estimation of the covariance matrix can then be used in a portfolio optimization technique.

For example, this robust covariance matrix can be plugged into the mean-variance and the minimum-variance portfolio. The formula to obtain the PC-mean-variance portfolio is then:

$$\hat{w}^{\mathrm{K}}_{\mathrm{MSR}} = \frac{\hat{\Sigma}_K^{-1}\hat{\mu}}{1'\hat{\Sigma}_K^{-1}\hat{\mu}} = \frac{V\Lambda^{-1}V'\hat{\mu}}{1'(V\Lambda^{-1}V'\hat{\mu}} \tag{II.16}$$

And for the minimum-variance portfolio, it is:

$$\hat{w}^{\mathrm{K}}_{\mathrm{MV}} = \frac{\hat{\Sigma}_K^{-1}1}{1'\hat{\Sigma}_K^{-1}1} = \frac{V\Lambda^{-1}V'1}{1'V\Lambda^{-1}V'1} \tag{II.17}$$

It is important to note that reducing the dimensions does not help in estimating $\hat{\mu}$. Nevertheless, by removing the last small eigenvalues, it should reduce the sensitivity of the portfolio optimization strategies to changes in $\hat{\mu}$. Therefore, the PC-mean-variance portfolio should work well in cases where T is low or N is large and thus the optimal Sharpe ratio can be achieved by restricting to a number of leading eigenvectors (Chen and Yuan, 2016).

4.2 PC-Variance-Parity portfolio

While the concept of risk parity based on assets can be attractive, it is often not the best approach to optimise diversification. Indeed, the issue is that asset returns are often correlated, making diversification across assets suboptimal. Alternative methods that consider the underlying correlation of the assets can be used.

Meucci (2009) introduces in his article “Managing diversification” the PC-variance-parity (PCVP) portfolio, which is the portfolio for which each of the N PCs contributes equally to the portfolio-return volatility. In other words, the PCVP portfolio diversifies across the underlying risk factors instead of the individual assets, creating a more diversified portfolio.

Starting from the PCA representation of asset returns: $X = V_N \Lambda_N^{1/2} Y$. We know we can represent the portfolio returns as: $R^P = w^T X = \tilde{w}^T Y$. With the weights of the portfolio being equal to $\tilde{w} = \Lambda_N^{1/2} V_N' w$.

The portfolio-return variance is given by

$$\hat{\Sigma}_w = \tilde{w}^T \Sigma_Y \tilde{w}$$

Note that $\Sigma_Y = I$ because of the use of the PCA.

Now, the PCVP portfolio is such that $\tilde{w}_i^2 = \tilde{w}_j^2$ for all i and j. It is given by:

$$W_{PCVP} = \left\{ w \in \mathbb{R}^N \| \; w = \frac{V_N \Lambda_N^{-1/2} 1^{\pm}}{1^T V_N \Lambda_N^{-1} 1^{\pm}} \right\} \tag{II.18}$$

Note that $1^{\pm} = \text{sign}(\Lambda^{-1/2} V^T 1)$ is chosen to minimise the variance of the portfolio.

4.3 Bounded Noise Portfolio

When a minimum-variance portfolio is constructed using the sample estimates, estimation errors in the covariance matrix are often blamed for the poor portfolio performance. But instead of trying to improve the estimation or using norm-constraints, Zhao et al. (2019) found that by separating the well-estimated aspects from the poorly estimated aspects of the covariance matrix excellent performance could be achieved. This is done by using a single parameter which is constant through the datasets and time periods.

This approach is called the bounded noise (BN) portfolio and consists of 4 steps. First, the BN portfolio looks at the covariance matrix and it turns out that some eigenvectors are easier to estimate than others. Thus, a split is needed between the well estimated (signal) and the wrongly estimated (noise) eigenvectors. The split at the k^{th} eigenvector is dictated by the impact of the estimation errors on the portfolio objective. In the following graph (figure II.3) the distribution of the true and the estimated eigenvalues can be witnessed.

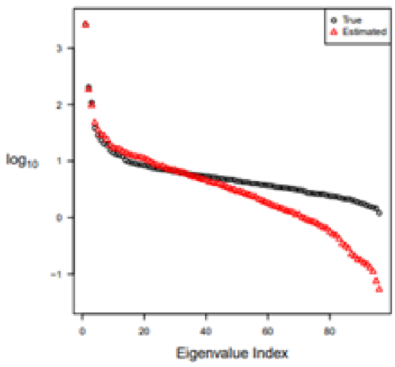
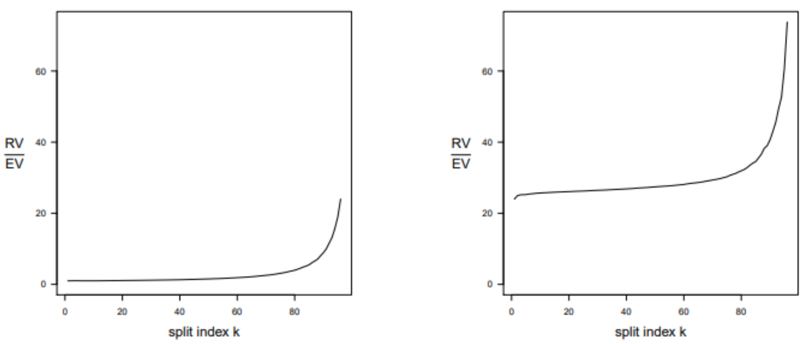


Figure II.3: Difference between True & Estimated Eigenvalues Zhao et al. (2019)

It is obvious that the largest estimated eigenvalues, and thus their related eigenvectors, are close to the true values. However, when examining the smallest eigenvalues, a significant estimation error can be observed, which also affects their related eigenvectors.

It is important to note the realised variance $RV(w) = w^T \Sigma w$ and the estimated variance $EV(w) = w^T \hat{\Sigma} w$. When the ratio between both is taken in function of the the split k, a huge difference between the top-k eigenvectors and the others can be witnessed. The difference is clear in figure II.4 below.



(a) The top-k eigenvector portfolio $\hat{\boldsymbol{w}}_S^*$ (b) The bottom $(p-k)$-eigenvector portfolio $\hat{\boldsymbol{w}}_N^*$

Figure II.4: $\frac{RV}{EV}$ for Signal-Only and Noise-Only portfolios Zhao et al.(2019)

We see that the ratio is rising when more eigenvector are added to the portfolio. This indicates that the portfolio variance is poorly estimated. This is particularly true for the noise-only portfolio (right). It is also interesting to see that the two graphs are giving a good explanation of why the bounded-noise portfolio is a good idea. We see that the first k eigenvectors are doing a great job. In contrast, the noise-only portfolio has a lower estimated variance. Nevertheless, it is stable until a certain point, so this stable part is useful for computation.

Second, a signal-only portfolio is constructed from the well-estimated eigenvectors from index 1 to k. Mathematically speaking, this portfolio is equivalent to a PCA-based portfolio that ignores a certain number of the low eigenvalues and their corresponding eigenvectors.

Third, a conservative noise-only portfolio is constructed from the upper bound of the poorly estimated eigenvectors. Zhao et al. (2019) argue that the space spanned by the poorly estimated eigenvectors taken together is well-estimated. This can be explained by the fact that this space is orthogonal to the space spanned by the signal eigenvectors. Thus, a portfolio from index k+1 to p from the noise space can improve the performance when combined with the signal-only portfolio. Indeed, the estimated variance of a portfolio from the noise space might be low, its realized variance might be much higher.

The conservative noise-only portfolio (w^N_{BN}) is found by minimising the upper-bound of the realised variance.

$$\begin{aligned} \min_{w} \quad & \mathrm{BRV}(w) \\ \text{subject to} \quad & w^T\mathbf{1} = 1, \\ & w \in \hat{\mathcal{N}} \end{aligned} \tag{II.19}$$

With $\mathrm{BRV}(w) = \mathrm{EV}(w) + m\|w\|_2^2$, where m is the largest eigenvalue of the matrix $\hat{N}^T(\Sigma - \hat{\Sigma})\hat{N}$ and where $\hat{N}$ is the matrix containing the noise eigenvectors of $\hat{\Sigma}$ in the space generated by the same eigenvectors $\hat{\mathcal{N}}$. Not that m can be estimated via cross-validation or bootstrapping.

Finally, in the fourth step the signal-only and the conservative noise-only portfolio are combined to form the bounded-noise portfolio. The BN portfolio is given by the following optimisation:

$$\begin{aligned} \min_{w} \quad & w^T(\hat{\Sigma} + \mathbf{M})w \\ \text{subject to} \quad & w^T\mathbf{1} = 1, \\ & \text{where } \mathbf{M} = m\hat{\mathbf{N}}\hat{\mathbf{N}}^T \end{aligned} \tag{II.20}$$

The solution w^{BN} is the following formula

$$\hat{w}_{\mathrm{BN}} = \frac{\sum_{i=1}^{K} \frac{\hat{v}_i^T 1}{\hat{\lambda}_i}\hat{v}_i + \sum_{i=K+1}^{N} \frac{\hat{v}_i^T 1}{\hat{\lambda}_i + m}\hat{v}_i}{\sum_{i=1}^{K} \left(\frac{\hat{v}_i^T 1}{\hat{\lambda}_i}\right)^2 + \sum_{i=K+1}^{N} \left(\frac{\hat{v}_i^T 1}{\hat{\lambda}_i + m}\right)^2} \tag{II.21}$$

And estimating K and m is done with bootstrapping or cross-validation.

It is important to note that the portfolio is created by using the same split defining scalar threshold parameter for all datasets. This ensures that the performance of the portfolio does not rely on the fine-tune of a very sensitive parameter.

4.4 PC-Portfolio of Severini

$V \equiv (v_1, ..., v_N)$ are the respective eigenvectors. The eigenvectors define a set of N uncorrelated portfolios, the principal portfolios, whose returns are decreasingly responsible for the randomness in the market. Indeed, the eigenvalues Λ correspond to the variances of these uncorrelated portfolios, and are all sorted in a decreasing order.

If the elements of v_i have a non zero sum, we can define one principal component portfolio per eigenvector (Severini, 2022). The weights of the i^{th} portfolio can be found as follows.

$$w_i = \frac{v_i}{1'v_i} \tag{II.22}$$

Note that these portfolio do not consider μ thus this portfolios could be undesirable as potential investments.

4.5 Research question

A limited number of papers are effectively comparing the different portfolio strategies different selected in this paper. Consequently, no consensus on the optimal method for portfolio optimisation has been found in the financial literature. As mentioned in the previous sections each strategy has its benefits but also its drawbacks. This master thesis aims at filling this gap by providing a systematic comparison of the different strategies and lining up the different strengths and weaknesses of those strategies.

For all those reasons the following research question will be analysed and answered in this paper.

Research question: *What is the empirical performance of different portfolio optimization techniques based on principal component analysis (PCA) compared to simpler strategies?*

This research question can be divided into two sub-questions.
Sub-questions 1: How does the empirical performance differ across different PCA-based portfolios?

Sub-question 2: How does the performance of PCA-based portfolios compare to that of strategies without dimension reduction but with similar objectives?

Portfolio selection is often seen as an essential task for investors. This research question provides practical relevance for investors seeking guidance in their investments.

Chapter III

Data & Methodology

1 Description of the financial data

To perform the analysis, datasets from the Kenneth French's library are considered. Four datasets are used, the 25 portfolios formed on size and book-to-market, the 100 portfolios formed on size and book-to-market, the 10 industry portfolios and the 48 industry portfolios. The industry portfolios are build on the returns of different sectors that fairly well represent the US economy. The size and book-to-market portfolios are constructed based on the size of the company yielding the return. The returns are value-weighted monthly returns from January 1970 to January 2023.

In the following graphs (figure III.1) the distribution of the returns of the different datasets can be seen. Note that each histogram represents all the returns of one particular dataset.

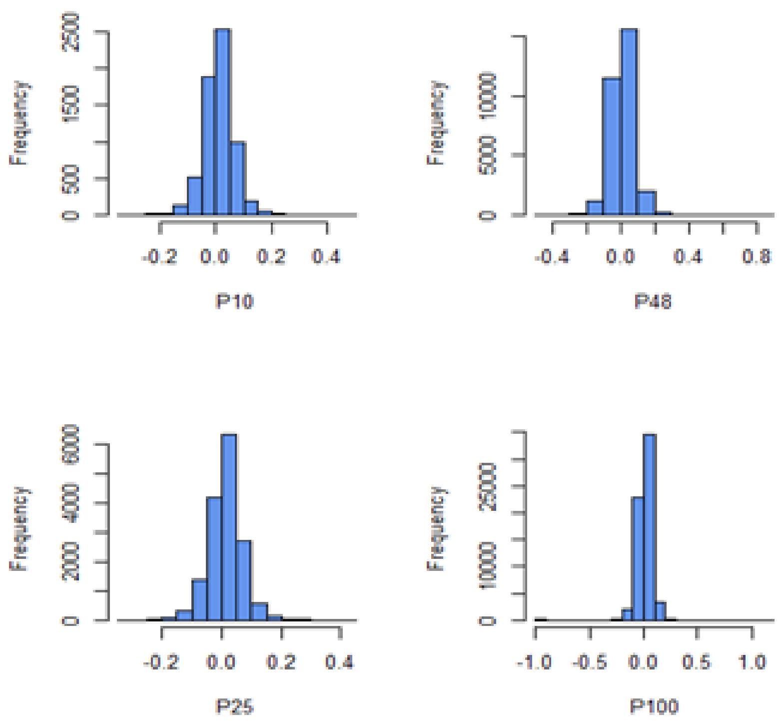


Figure III.1: Distribution of the returns of each dataset

Upon examining the descriptive statistics, it is easy to see that the average monthly return is close to 1% for the four datasets. However, in the datasets with higher dimensions extreme returns can be witnessed during crisis periods. The standard deviation also points in that direction with a value slightly above 5% for the datasets with 10 and 25 variables but respectively 7 and 8% for the P48 dataset and the P100. Important outliers can also be seen in those datasets.

Here the kurtosis and the skewness for each asset in each data set can be seen (See figure III.2).

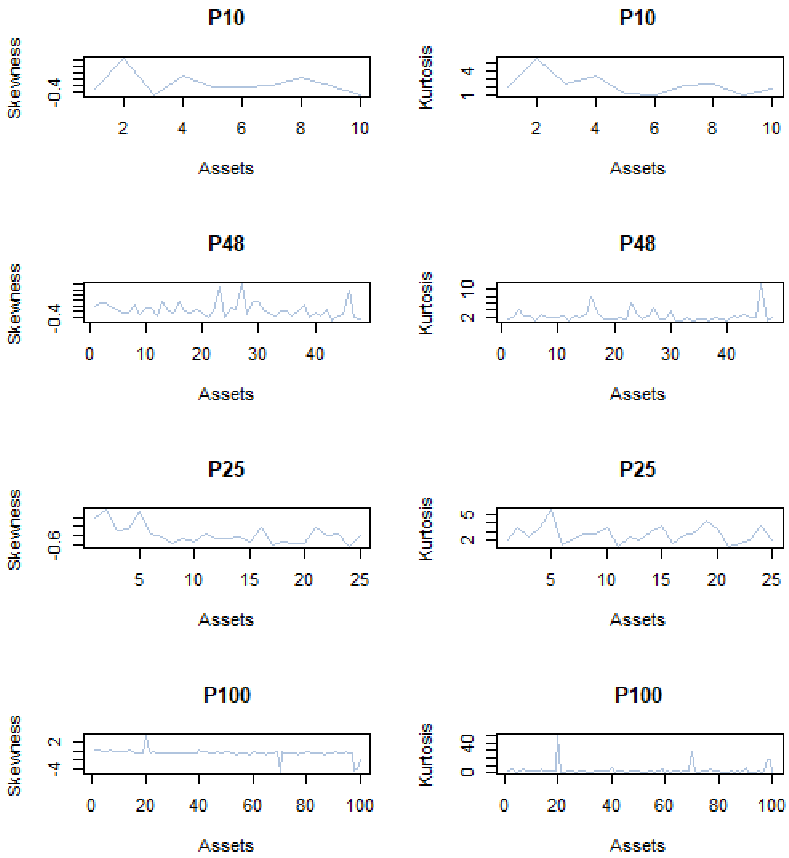


Figure III.2: Skewness (left) and Kurtosis (right) of each dataset

When analysing the skewness, negatively skewed values for the four datasets can be witnessed. In other words, this indicates that the extreme returns are more often negative than positive.

Note that on figure III.2 the Kurtosis has already been adjusted with regards to the Normal distribution by substracting 3. The Kurtosis values are for all the datasets (P10, P25, P48 and P100) larger than 3 which indicates it is leptokurtic, this means that the dataset will have more extreme returns than the normal.

2 Implementation of the portfolios

Eight types of portfolios will be considered during this master thesis. Four well-known portfolios are presented and four other portfolios which use principal components to optimise their weights. All the eight portfolios are presented in the first part of this work. The first four portfolio strategies are: mean-variance, minimum-variance, equally weighted and asset-risk-parity portfolio. The other four portfolios are pc-variance-parity, pc-subspace, bounded-noise and Severini's portfolios.

In the first instance we will implement four types of pc-portfolios, and their performance will be compared to each other. After that, we will implement the four methods without dimension reduction, and their performance will be compared to the pc-based portfolios.

In a second instance, we will assess the impact of the use of a robust estimation of the covariance matrix on the performance of the optimisation strategy. In addition, we will assess the choice of principal components on the performance of the strategies.

The performance of the different strategies will be backtested by generating out-of-sample returns. Those will be generated with a rolling window of 6 months. This procedure goes as follows.

First, a selection of the length of the estimation window $\mathcal{T}$ out of the T observations is made. Here 10 years or 120 months is used. Second, by using the first $\mathcal{T}$ returns in the sample, a computation of the optimal portfolio weights is done. Then those weights are used to the out-of-sample portfolio return on the next 6 months after the estimation window. Third, the latest 6 months are removed from the estimation window $\mathcal{T}$, and the next 6 months are added. By repeating this until the end of the dataset, $T - \mathcal{T}$ weight vectors and out-of-sample returns are generated. In the fourth and last step it becomes possible to compute out-of-sample performance metrics.

Three main performance metrics will be used in this paper. The average annual return, obtained by multiplying the monthly portfolio returns R_i^P by 12, will be used to compare the strategies. $R_{\text{annual}}^P = mean(R_i^P) * 12$

The annualised volatility, obtained by multiplying the standard deviation of the portfolio returns R_i^P by the square root of 12. The annualized volatility of the portoflio is given by: $\sigma_{\text{annual}}^P = \sigma^P\sqrt{12}$

And finally, the Sharpe ratio (SR) will be used. The Sharpe ratio is a metric that takes into account the risk a return entails, the measure is thus risk-adjusted. It was developed by Sharpe in 1966 to help investors understand the trade-off between return and risk. The SR can be calculated by dividing the portfolio average return by the associated standard deviation of the portfolio. If some portfolio strategy has a higher Sharpe ratio than another, this indicates that the portfolio yields a higher return for a given level of risk or that the portfolio yields a lower risk for a given return. This measure is thus very useful to compare the performance of the different strategies selected in this paper. The

Sharpe ratio is given by:

$$\text{Sharpe Ratio} = \frac{R^P_{\text{annual}}}{\sigma^P_{\text{annual}}}$$

To be able to determine if differences in the ratios are statistically significant a statistical test is used specifically, designed to work with Sharpe ratios. This test was developed by Ledoit and Wolf (2008) to overcome the limitations of classic tests which do not perform well in the case of non-normal distributed returns.

Chapter IV

Results & Analysis

1 Presentation and interpretation of the results

In this section the different results of the portfolio strategies will be presented. The metrics are all computed on out-of-sample data.

1.1 Performance comparison of PC-based optimisation techniques

The K used for each dataset was select via explained variance at a level of 90% as explained in section II.3.3. For the P10, P25, P48 and P100 dataset we used the rolling window method described in section III.2 and we obtained on average respectively K equal to 5, 17, 3 and 12.

PC-Variance-Parity portfolio.

This strategy performs at reasonably well out-of-sample. While it achieves a stable average return which is always above 11%, we see that the volatility is also not very high. Indeed, we see that the volatility is kept under 20% in most of the cases.

It is important to note that even though, the PCVP was not designed to maximise the Sharpe ratio or the returns it does not achieve bad results. Instead, the strategy was developed to minimise the variance and this can be seen in the results in table IV.1. Note, however, that there is a high estimation risk linked to the covariance matrix when estimating in a high dimensional dataset. This can explain the poor performance in the P100 dataset.

In the following table (table IV.1) the exact results of the strategy in function of the different datasets (P10, P48, P25 and P100) can be found.

Table IV.1: Performance Metrics of PCVP

PCVP	**Ann Return**	**Ann Volatility**	**SR**
P10	0.111	0.147	0.751
P48	0.156	0.169	0.926
P25	0.133	0.207	0.642
P100	0.1148	0.183	0.627

We see that the best risk-adjusted return is achieved in the P48 dataset, mainly due to its higher return.

When looking at the boxplot (figure IV.1) of the weights from the first 10 assets for the different datasets we see that they are not very different from a dataset to another. In addition, we see that the weighs are stable in most of the cases.

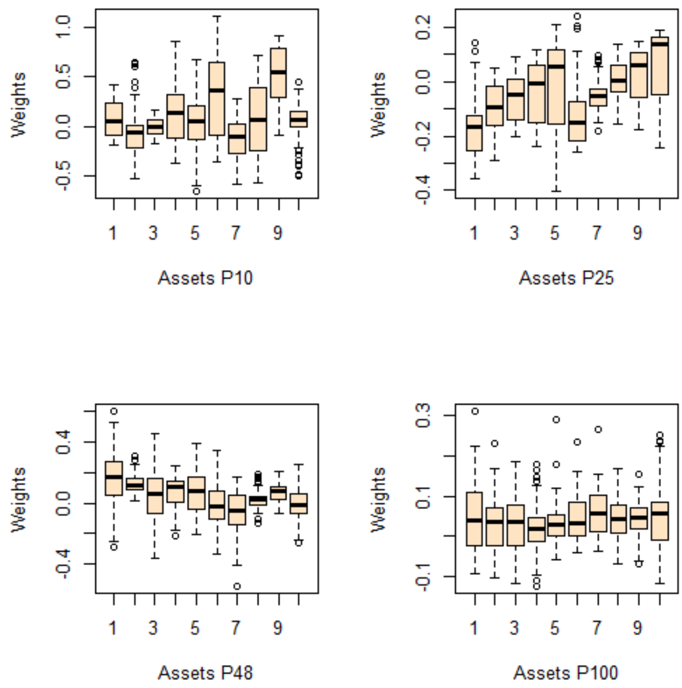


Figure IV.1: Boxplot of the weights obtained by PCVP for each dataset

PC-Mean-Variance portfolio.

We see in table IV.2 that results of this strategy are heavily dependent on the dataset used. Indeed, we see a large difference between the Sharpe ratios of the P100 and the P48 dataset. This can also be witnessed when considering the returns of the strategies on the datasets which are more than four times higher for the P100 than for the P48. Of course this higher return has a price, a higher volatility is generated by those returns. This balances the Sharpe ratios. The strategy thus delivers very volatile returns. A similar conclusion can be made for the annualised volatility, extreme results are achieved.

Since the PC-Mean-Variance portfolio strategy reduces the dimensionality of the datasets, it can simplify the optimisation process and give good results. However, this can also lead to extreme results in the annual volatility and the average annual return. The strategy is thus heavily dependent on the choice of K which can incorporate more or less extreme returns.

We see in the table (Table IV.2) below that the best risk-adjusted return is achieved by using the P100 dataset. Mostly due to its very high return.

Table IV.2: Performance Metrics of PC-MSR

PC-MSR	**Ann Return**	**Ann Volatility**	**SR**
P10	0.108	0.163	0.660
P48	0.110	0.247	0.443
P25	0.174	0.252	0.690
P100	0.442	0.403	1.097

When looking at the boxplot of the first 10 assets of the datasets we see that the weights tend to be very stable. This is particularly true for the P48 and the P100.

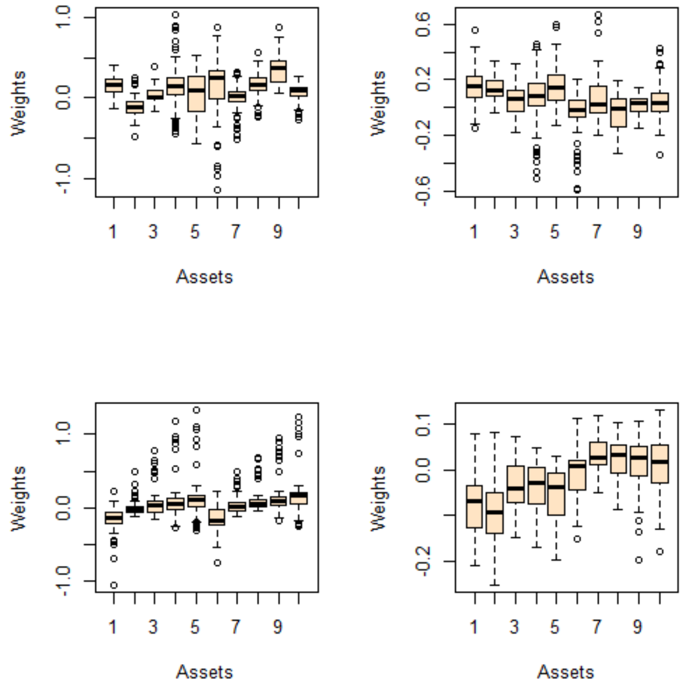


Figure IV.2: Boxplot of the weights obtained by PCMSR for each dataset (P10 and P48 are in the top of the firgure).

PC-Minimum-Variance portfolio.

We see in table IV.3 that this strategy performs very well out-of-sample. The PC-minimum-variance provides high Sharpe ratios. This is mainly due to its very low annual volatility.

Note that the extremely low annual volatility makes totally sense since the strategy was designed to minimise the variance. Interestingly, even though the strategy does not take care of optimising the returns, significantly high annual returns are achieved. It is also important to note that those average annual returns are very stable.

Table IV.3: Performance Metrics of PC-MV

PC-MV	Ann Return	Ann Volatility	SR
P10	0.123	0.125	0.984
P48	0.142	0.130	1.089
P25	0.126	0.173	0.728
P100	0.153	0.152	1.008

In the table above, most of the Sharpe ratios are high. The best results are obtained by using the P48 dataset, with an SR of 1.068.

The boxplot in figure IV.2 is interesting since it shows for each dataset how stable the portfolio is over the first 10 assets.

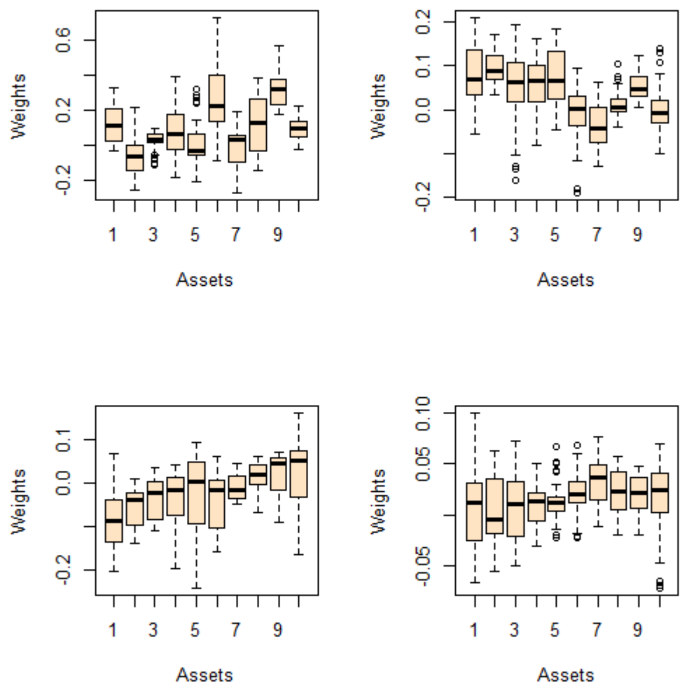


Figure IV.3: Boxplot of the weights obtained by PCMV for each dataset

The weights do not change drastically.

Severini's portfolios.

When using the first eigenvector as input for the portfolio strategy, it achieves average results in general. Indeed, the strategy has a low annual volatility with average returns which are consistent in most of the cases except when its used on the P100 dataset.

In table IV.4 you can see the different results obtained by using the first eigenvector.

Table IV.4: Performance Metrics of Severini E1

Severini E1	**Ann Return**	**Ann Volatility**	**SR**
P10	0.129	0.162	0.797
P48	0.129	0.176	0.730
P25	0.133	0.183	0.726
P100	0.003	0.222	0.013

In the table above we can see that the strategy achieves the best risk-weighted results when used on the P10 dataset.

When considering the portfolio based on the second eigenvector the results do not particularly stand out. Indeed, the strategy gives returns that are very volatile, sometimes reaching more than 100% return but sometimes yielding negative returns. A very high volatility can also be witnessed in table IV.5. The Sharpe ratios are very poorly performing and are each time worse compared to the ones generated by the portfolio on the first eigenvector.

These results can be seen in the table IV.5 below. Note that the strategy achieves the highest Sharpe ratio when used on the P25 dataset, with a SR of 0.165.

Table IV.5: Performance Metrics of Severini E2

Severini E2	**Ann Return**	**Ann Volatility**	**SR**
P10	−0.343	1.466	−0.234
P48	0.693	5.875	0.118
P25	1.445	8.755	0.165
P100	−1.657	9.584	−0.173

The impact of the eigenvector considered to build the portfolio on, can be witnessed in the four graphs below (figure IV.4).

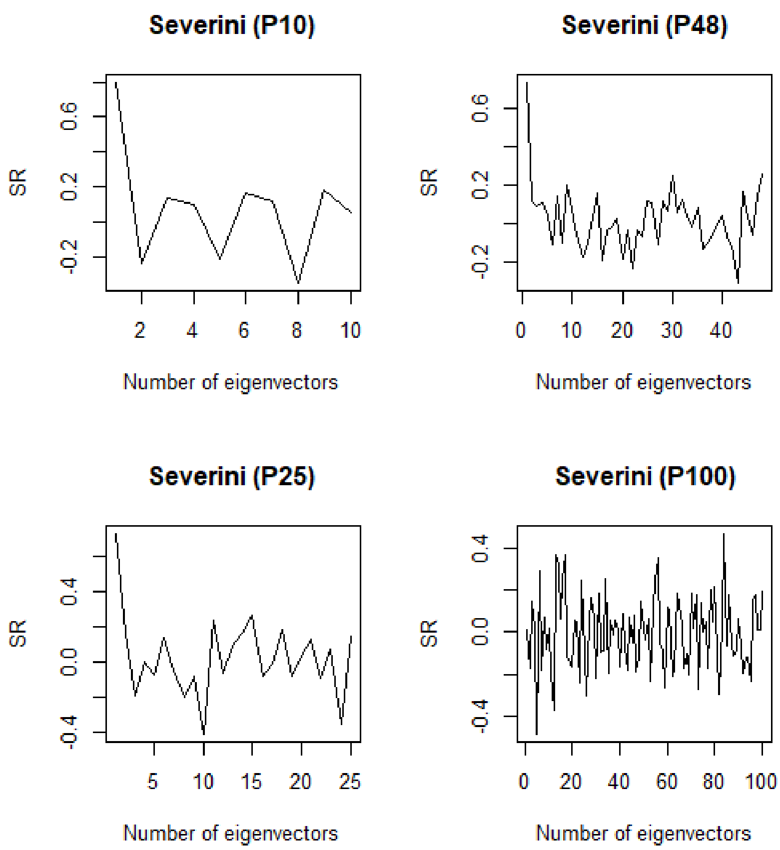


Figure IV.4: Impact of number of eigenvectors on the SR when using Severini

We see that in the first three data sets dataset the highest Sharpe ratio is achieved when using the first eigenvector. This is not surprising since most of the information is carried by those vector. After the first eigenvector the Sharpe ratios stay relatively stable. Interestingly the optimal eigenvector for the P100 dataset is the 84^{th}.

Bounded-Noise portfolio.

The bounded-noise portfolio has very good results. It achieves extremely low variance in combination with high returns. This creates Sharpe ratios which are very impressive and all close or above one. Note that this is the only PC-based strategy that achieves lower and stabler volatility than the PC-minimum-variance.

Table IV.6: Performance Metrics of BN

BN	Ann Return	Ann Volatility	SR
P10	0.122	0.122	0.998
P48	0.131	0.122	1.078
P25	0.168	0.132	1.275
P100	0.185	0.141	1.319

In table IV.6 we can see that the highest Sharpe ratio is achieved when applying the strategy to the biggest dataset. This makes sense since a larger dataset will provide a broader and more precise choice in where to separate the signal portfolio from the noise portfolio. In addition, the upper bound of the noise portfolio will also be easier to determine thanks to a broader choice.

Performance comparison of the PC-based strategies.

Now that the results of each strategy have been reviewed we will analyse how they compare to each other.

The following table (Table IV.7) summarises the out of sample (OoS) Sharpe ratios of the strategies. Each result being coloured in function of its out-of-sample ranking, green represents the best Sharpe ratio and red represents the worst Sharpe ratio when considering a specific dataset.

Table IV.7: SR of PC-based Strategies

SR OoS	**P10**	**P48**	**P25**	**P100**
PCVP	0.751	0.926	0.642	0.627
PCMV	0.984	1.089	0.728	1.008
PCMSR	0.660	0.443	0.690	1.097
Severini 1	0.797	0.730	0.726	0.013
Severini 2	−0.234	0.118	0.165	−0.173
BN	0.998	1.078	1.275	1.319

The table shows that the Bounded-Noise performs better than any other strategy. Indeed, the BN portfolio consistently outperforms other portfolios regardless of the datasets chosen. The combination of a signal only and a conservative noise only portfolio shows great results. This superiority is even more evident when considering the dataset with the largest amount of assets (P100).

The PC-minimum-variance portfolio also performs very well out-of-sample. In most of the datasets, the strategy yields better results than the other strategies. Also, when comparing the portfolio with the PC-mean-variance, it is interesting to see that it achieves a higher Sharpe ratio. This is not surprising since the PC-MSR suffers from estimation errors in the mean. By removing the mean from the strategy we obtain the PC-MV portfolio which alleviates this estimation error, yielding better results.

When comparing both portfolio of Severini we see a large difference. Indeed, while the first portfolio performs fairly well the second portfolio yield very poor results and by far.

More generally, it is not surprising that the portfolios of Severini does not perform very well compared to the other PC-based strategy since they were not designed to maximise the Sharpe ratio. Indeed, the strategies only invest in the assets that compose the first (or second) eigenvector that explains most of the variance. This yields weights that are higher for assets that contribute a lot to the firsts eigenvectors and low for those that do not contribute significantly to the first eigenvector. Making it subject to extreme results sometimes.

However, when considering only the PC-MSR, the PCVP and the first portfolio of Severini, much less disparities between the strategies are to find. We see that Severini's first portfolio and the PCVP portfolio perform quiet well, followed by the PC-MSR. The first portfolio of Severini and the PC-variance-parity portfolio show, in most of the cases, good and stable results. However, they are often outperformed by the PCVP. This is interesting since the portfolio does not optimise the Sharpe ratio.

We also see that the first portfolio of Severini encounters some problems when applied to high dimensional datasets. This bad Sharpe ratio is obtained due to particularly low returns generated by the strategy.

The hypothesis test of Ledoit and Wolf (2008) is now applied on the Sharpe ratios shown above to determine if the results are significantly different from each other. The obtained results can be seen in the tables 8 and 9 below, where we use 5% as significance level. Each column represents a strategy which is then compared to another. If it is significantly different from a strategy in a row, the cell will be coloured in green if the portfolio in the column performs better than the portfolio in the row. Conversely, if the portfolio performs worse than the strategy in the row it will be coloured red.

Table IV.8: P-values P10

P-values P10	**PCVP**	**PCMV**	**PCMSR**	**Severini 1**	**Severini 2**	**BN**
PCVP		0.015	0.92	0.603	0.001	0.016
PCMV	0.015		0.027	0.14	0.001	0.586
PCMSR	0.92	0.027		0.525	0.001	0.02
Severini 1	0.603	0.14	0.525		0.001	0.13
Severini 2	0.001	0.001	0.001	0.001		0.001
BN	0.016	0.586	0.02	0.13	0.001	

Table IV.9: P-values P48

P-values P48	**PCVP**	**PCMV**	**PCMSR**	**Severini 1**	**Severini 2**	**BN**
PCVP		0.232	0.005	0.24	0.001	0.283
PCMV	0.232		0.001	0.007	0.001	0.891
PCMSR	0.005	0.001		0.108	0.226	0.001
Severini 1	0.24	0.007	0.108		0.024	0.008
Severini 2	0.001	0.001	0.226	0.024		0.001
BN	0.283	0.891	0.001	0.008	0.001	

The p-values for the P25 and P100 are similar to the one showed above.

In the P10 and the P48 dataset we see that the bounded-noise portfolio has several p-values smaller than 5% this means that the Sharpe ratio obtained by the strategy is significantly different from that of the corresponding strategy.

The PC-minimum-variance portfolio is also significantly different from most other portfolios. This is in line with the Sharpe ratios we have seen previously. It is interesting to see that the null hypothesis between the PC-MSR and the PC-MV is easily rejected.

In contrast, we see that all of the portfolios yield significantly different results from the second portfolio of Severini.

We also see when comparing the PC-MSR, the PCVP and the first portfolio of Severini that they are on average not statistaclly significantly different from each other at a level of 5%.

In conclusion, the Bounded-Noise portfolio strategy yields, by far, the best results out-of-sample. It yields the best Sharpe ratios based on the highest returns and the lowest variance. The PC-Minimum-Variance portfolios are generating very attractive results too. They achieve low variance with high and consistent returns. The PC-MSR, the first strategy of Severini and the PCVP yield good results. Nevertheless, those results are sometimes less stable compared to the top perfromers, the BN nad the PC-MV portfolios. Ultimately, the last strategy, the Severini's second portfolio do not show good returns at all. This leads to poor performing Sharpe ratios. In addition, the portfolio is not stable over time in function of the dataset used.

1.2 Performance comparison of PC-based and classical optimisation techniques

Classical strategies.

In table IV.10 we clearly see that the Mean-Variance portfolio systematically outperforms the other strategies in most of the datasets. This was to expect, and this is in line with what has been seen in chapter 2. It is interesting to note that in many cases the Mean-Variance portfolio underperforms the other strategies except in the P25 dataset where it outperforms the rest. This also shows the inconsistency of the MSR portfolio.

Table IV.10: SR OoS of Classical Strategies

SR OoS	P10	P48	P25	P100
EW	0.866	0.778	0.760	0.609
MV	1.004	0.950	0.889	0.763
MSR	0.670	0.372	1.183	0.098
AVP	0.915	0.825	0.782	0.667

The Equally Weighted portfolio performs quiet well in the four datasets, but we see that its performance decrease in function of the number of assets considered. This could be explained by the fact that not all assets are equal. A larger dataset will include more less profitable investments, and we have seen that this proportion is higher for the P100 dataset. Since the strategy will give the same weight to any asset, they will not be filtered out the portfolio.

The same argument can be used for the Asset-Variance-Parity portfolio. The strategy performs reasonably well but its performance decreases when higher dimensional datasets are considered. Since the AVP portfolio tries to invest the same risk in each asset it will have to incorporate the stable bad performing asset in the same measure as the well performing stable assets.

Comparison with the PC-strategies.

Here a comparison of the performance of the different PC-based strategies with the simpler strategies will be done. The statistical significance of those results will also be analysed.

When we compare the Sharpe ratios obtained from the classical strategies and their principal component counterpart, we see that in most of the cases the principal component strategies outperform the classical strategies. Indeed, in table IV.11 we see the actual Sharpe ratios generated by each strategy.

Table IV.11: SR OoS - PC/Classic

SR OoS	**P10**	**P48**	**P25**	**P100**
MV	1.004	0.950	0.889	0.763
PCMV	0.984	1.089	0.728	1.008
MSR	0.670	0.372	1.183	0.098
PCMSR	0.660	0.443	0.690	1.097
AVP	0.915	0.825	0.782	0.667
PCVP	0.751	0.926	0.642	0.627

The use of principal components generally enhances the overall results of the strategies. This is particularly true for the PC-minimum-variance portfolio. This portfolio performs in most of the cases better than the its classical counterpart. It is also important to note that this better performance is best reflected in the P100 dataset. This is totally predictable since the use of principal components in portfolio optimisation was developed to reduce the dimensionality of datasets with a high number of variables.

When a look is given at the significance of the results, we see that most of the results of the PC portfolios are significantly lower. Table 12 summarises this observation and colours the cells with a significant p-value in red if the PC-strategy underperforms and in green when the PC-strategy outperforms the classical strategy.

Table IV.12: P-Value/ PC

P-Value/ PC	**P10**	**P48**	**P25**	**P100**
MV - PCMV	0.423	0.179	0.933	0.141
MSR - PCMSR	0.761	0.708	0.009	0.001
AVP - PCVP	0.179	0.496	0.720	0.782

For the PC-MV and the PCVP strategies it is difficult to reject the null hypothesis with the MV and the AVP portfolios. Indeed, even if most of the Sharpe ratios are higher no p-value falls under the 5% even for the large datasets. The null hypotheses concerning the PCMSR portfolio can be reject in the two last datasets.

The first portfolio of Severini generates considerable results. Nevertheless, compared to the classical portfolio strategies, Severini's first portfolio do not significantly outperform. Actually, Severini portfolio is often outperformed by the classical strategies. In table IV.13 we see the out-of-sample Sharpe ratios of the different strategies.

Table IV.13: SR OoS - Severini/Classic

SR OoS	P10	P48	P25	P100
EW	0.866	0.778	0.760	0.609
MV	1.004	0.950	0.889	0.763
MSR	0.670	0.372	1.183	0.098
AVP	0.915	0.825	0.782	0.667
Severini 1	0.797	0.730	0.726	0.013
Severini 2	-0.234	0.118	0.165	-0.173

We see that in most of the cases the minimum-variance portfolio yields the best results. The AVP and the equally weighted portfolios are yielding good results too. They are closely followed by the mean-variance portfolio and the portfolios of Severini which are not yielding very good results compared too the other strategies. Then comes the second portfolio of Severini which yield very bad results. It is interesting to know that the portfolio strategy of Severini, particularly the one base on the first eigenvector, is in fact very close to the equally weighted portfolio. Indeed, most of the first eigenvector in a dataset tend to incorporate the same amount of explained variance through the estimation windows. This creates a stable nominator when computing the weights with the portfolio strategy of Severini. This resemblance can be witnessed when looking at the Sharpe ratios of both strategies which are often not fare away from each other.

To see if this results are significant we performed a hypothesis test that can be find in the table 14.

Table IV.14: P-Value/ SEV1

P-Value/ SEV1	P10	P48	P25	P100
MSR	0.475	0.108	0.036	0.798
MV	0.044	0.033	0.011	0.013
EW	0.001	0.002	0.015	0.001
AVP	0.001	0.002	0.012	0.475

We see that most null hypothesis are easily rejected in favor of the classical portfolios. In red, every significant results in favor of the classical strategies can be witnessed.

When looking at Severini's second portfolio, the conclusions is even more extreme. The majority of the null hypotheses can be rejected. This can be seen in table IV.15 below.

Table IV.15: P-Value/ SEV2

P-Value/ SEV2	**P10**	**P48**	**P25**	**P100**
EW	0.001	0.015	0.002	0.001
MV	0.001	0.002	0.001	0.001
MSR	0.001	0.329	0.001	0.293
AVP	0.001	0.009	0.001	0.001

Both strategies of Severini are on average performing worse than other strategies.

In table IV.16 we see that the BN-portfolio keeps performing systematically better than the simpler strategies. The MV portfolio manages to surpass the BN on one occasion, when used on the P10 portfolio. This is possible becasue the BN portfolio needs to perform the right split between the signal and the noise principal components, this is more difficult to do on a smaller dataset.

Table IV.16: SR OoS - BN/Classic

SR OoS	**P10**	**P48**	**P25**	**P100**
EW	0.866	0.778	0.760	0.609
MV	1.004	0.950	0.889	0.763
MSR	0.670	0.372	1.183	0.098
AVP	0.915	0.825	0.782	0.667
BN	0.998	1.078	1.275	1.319

When we look at the significance of the different Sharpe ratios in table IV.17 we see that in most of the cases the BN portfolio generated out-of-sample Sharpe ratios that are significantly higher than the ones obtained from the classic strategies. This is particularly true for the results obtained from the P100 dataset.

Table IV.17: P-Value/ BN

P-Value/ BN	**P10**	**P48**	**P25**	**P100**
EW	0.283	0.016	0.001	0.001
MV	0.919	0.098	0.001	0.001
MSR	0.008	0.001	0.615	0.001
AVP	0.459	0.03	0.001	0.001

The Bounded-Noise portfolio is thus performing significantly better than any other strategy and in multiple datasets.

2 Impact of the number of principal components

We computed the optimal number of principal components with the explained variance method. This method yielded a K for each estimation window $\mathcal{T}$ with a mean of 5, 3, 17 and 12 for the datasets P10, P48, P25 and P100. When comparing these results with the optimal results given by the method of Bai and Ng (2002) when applied on the whole dataset some differences appear. Indeed, the Bai and Ng method yielded the optimal K 9, 4, 20 and 6. Note that the optimal K obtained by using Bai and Ng (2002) cannot be find since it uses the whole dataset and is thus looking ahead of the windows.

The figures IV.5 and IV.6 below represent the explained variance for each of the principal components of the P10 and P25 datasets, we see that most of the variance is explained by the first component and even more by the first couple components.

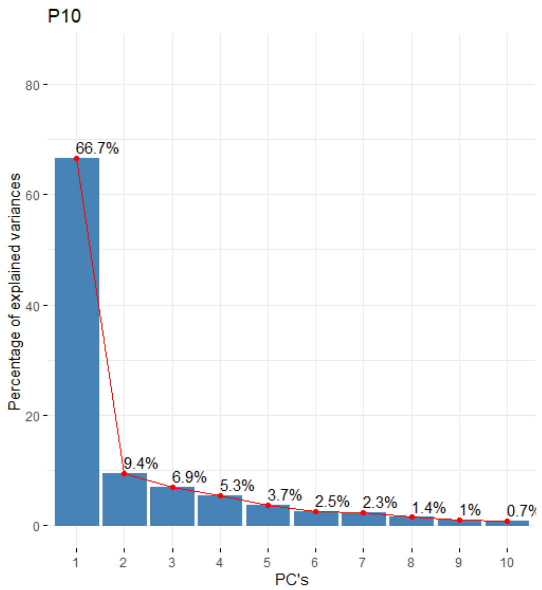


Figure IV.5: Percentage of variance explained in function of number of PC for P10

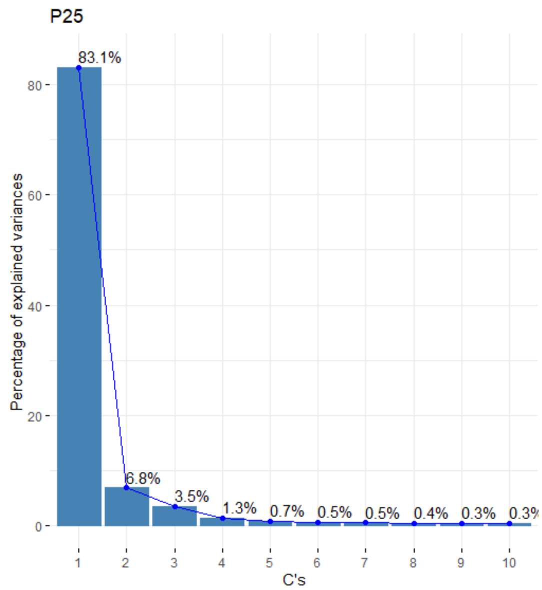


Figure IV.6: Percentage of variance explained in function of number of PC for P25

The two other figures below (figure IV.6 and IV.7) represents the explained variance of the P48 and the P100 datasets.

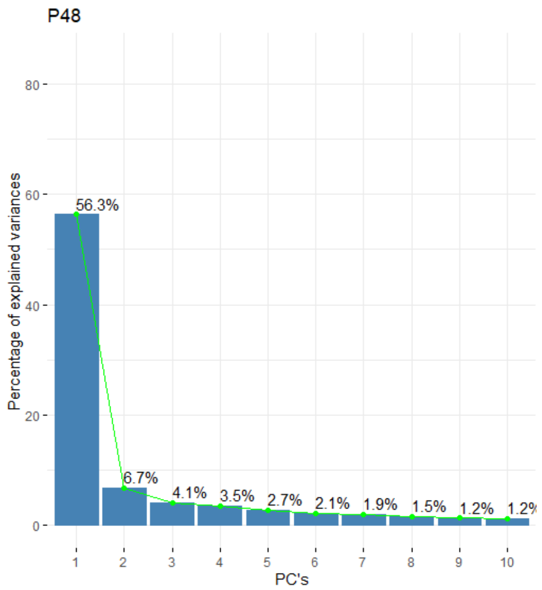


Figure IV.7: Percentage of variance explained in function of number of PC for P48

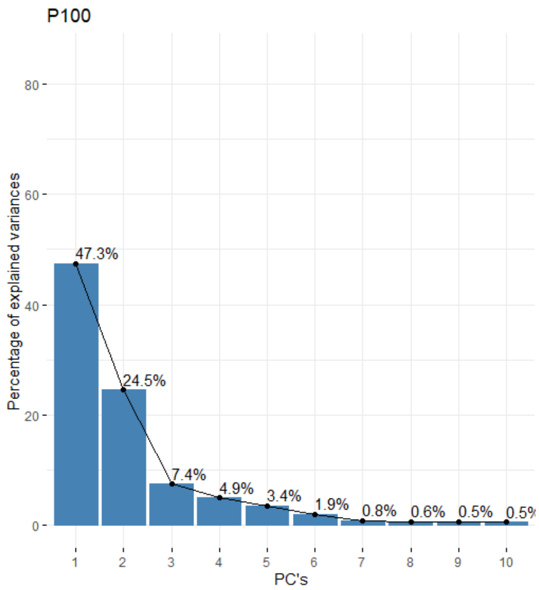


Figure IV.8: Percentage of variance explained in function of number of PC for P100

Like in the previous figure, the two larger datasets concentrate the variance in the first principal components. However, an important difference can be seen, the smaller datasets concentrate much more explained variance in the first PC than then the larger datasets do. Indeed, more than 80% of the variance of the P25 dataset can be explained by the first dimension. On the other hand, we see that the first principal component of the P100 dataset can only explain 40% of the variance.

To see the impact of the choice of the number of principal components, it is interesting to see the evolution of the out-of-sample Sharpe ratio in function of the number of principal components selected. For example, the mean-variance strategy can be used to illustrated this. In figure IV.9 we see that by adding more principal components the strategy's SR increases rapidly but past a certain, optimal, point it will start to decrease. It is compelling to see that change in SR is more important in high dimensional datasets. This is a logical outcome since the use of PC's in portfolio optimisation was design to enhance the performance in high dimensions.

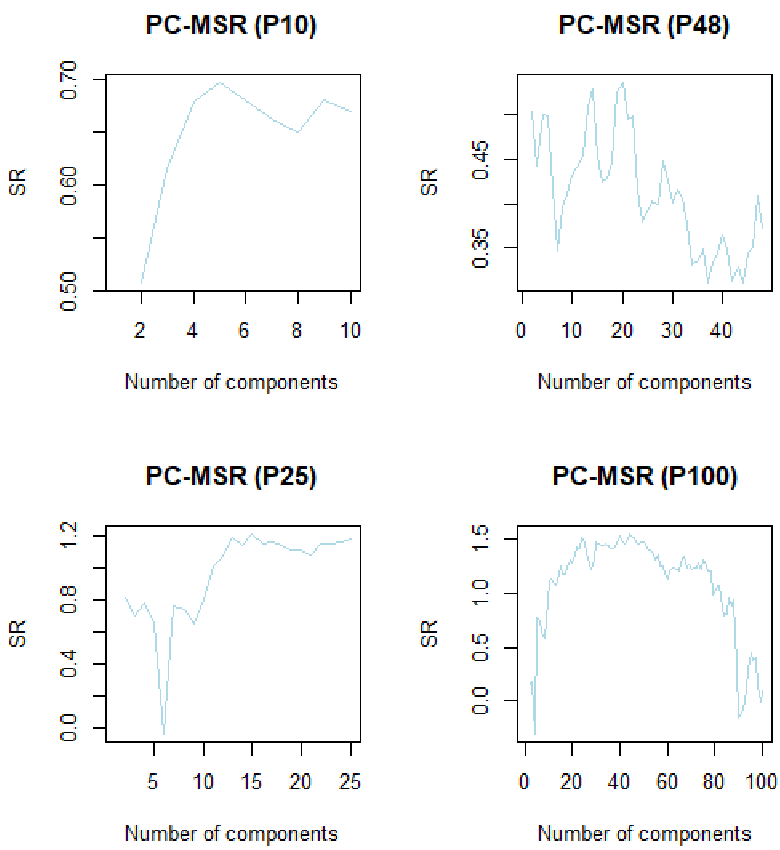


Figure IV.9: Evolution of SR in function of number of principal components for each dataset

It is important to note that this method can not be used to estimate the number of principal components to select since this uses out-of-sample data that an investor will not have when investing. These figure just give additional information regarding the impact of the number of principal components on the Sharpe ratio.

3 Impact of the biased covariance matrix estimator of Ledoit-wolf

When the covariance matrix is shrunk by using the method of Ledoit-Wolf we obtain a biased covariance matrix. After plugging this in the Mean-variance and the minimum-variance strategy we obtain the results in table 18.

Table IV.18: SR OoS

SR OoS	**P10**	**P48**	**P25**	**P100**
MV	1.004	0.950	0.889	0.763
LWMV	1.011	0.959	0.887	0.774
MSR	0.670	0.372	1.183	0.098
LWMSR	0.716	0.447	1.273	1.720

It is interesting to see that a strategy using the covariance matrix of Ledoit-Wolf outperforms their standard counterpart each time. But this performance is often only a bit higher than the classical strategy. To see if these results are significant or not, we performed a hypothesis test, which you can see in table 19.

Table IV.19: P-Value/ LW

P-Value/ LW	**P10**	**P48**	**P25**	**P100**
MSR - LWMSR	0.089	0.326	0.270	0,001
MV - LWMV	0.442	0.565	0.897	0.406

We see that the LW-MSR portfolio performs better than its classical counterpart when used on the P100 dataset. For the Minimum-variance it is nearly the same, at 5% certainty we see that we are unable to reject the null hypotheses.

4 Interpretation and significance of the results

Most of the PC-strategies outperform their classical counterpart. The PC-MSR, the PC-MV and the PCVP portfolios show results that are often slightly above the average of the MSR, the MV and the AVP portfolios. Yet not all of these results are significantly different at a level of 5%. Due to the small differences between the Sharpe ratios we have difficulties to reject the null hypothesis.

The portfolios designed by Severini are not yielding very interesting results. The first strategy fall somewhere in the middle in terms of performance, not ranking as the best choice overall but also not consistently the worst. When we look at the significance of the results for the first portfolio of Severini, we can argue that the strategy performs worse than the classical strategies at a level of 5%. On the other hand, the second portfolio of Severini shows very poor results, this makes it even easier to determine if the portfolio performs better or worse than the classical strategies at a level of 5% certainty.

The Bounded-Noise portfolio is the only portfolio that out-performs any other strategy. Most of these strong results are also significant at a level of 5%.

Chapter V

Conclusion

This paper has reviewed the state-of-the-art strategies based on the use of principal components. The goal was to analyse and understand how the PC-based strategies compare to each other. In a second analysis this master thesis had as purpose to compare those strategies with the more classical strategies which do not use PCs.

The portfolios based on the principal components have been compared to each other and hypothesis tests have been performed to see if the results were significant or not. The bounded-noise portfolio yields by far the best results, followed by the PC-minimum-variance portfolio. They are then followed by the PC-mean-variance portfolio, Severini's first portfolio and the PCVP portfolio. Finally, the second portfolio of Severini comes as last. Most of the Sharpe ratios of the strategies where significantly different from each other at a level of 5%.

We then compared the PC-strategies with the so-called classical strategies, mean-variance, minimum-variance, asset-variance-parity and equally weighted portfolio. The MV, MSR and AVP performed slightly worse than their PC-counterparts in terms of Sharpe ratios. But it was hard to reject the null hypotheses at a level of 5% for most of them. For the portfolios of Severini, the conclusion was the total opposite. Indeed, the two strategy are not top performer compared to the different classical portfolios. This is confirmed by most of the null hypothesis on the Sharpe ratios which could easily be rejected at a level of 5%. We could thus demonstrate that the strategies yield significantly different results. Finally, we compared the bounded-noise portfolio with the classical strategies. Here the results were significantly higher that in any other strategy. It was thus easy to reject the null hypothesis against any strategy and in any dataset.

We have seen that the number of selected principal components has an important impact on the performance of the strategy. By selecting principal components achieving a certain level of explained variance the Sharpe ratio can attain its highest level, but if the selection was poorly done it could yield results way under the optimal Sharpe ratio. Finally, we have seen that the performance of the mean-variance and the minimum-variance strategy could be enhanced by using the biased covariance matrix estimator of Ledoit-wolf. Nevertheless, most of the results could not be demonstrated to be significantly different from each other at a level of 5%. This is due to the small performance difference of the strategies.

Several recommendations can be given to investors. First, we encourage to use the BN-portfolio based on his great and robust performances. Second, the portfolios of Severini are not an interesting alternative to the classical portfolios. Third, we encourage to use the PC-MV, the PC-MSR and the PCVP instead of their classical counterpart even if their results are only slightly higher. Fourth, the number of principal components has a significant impact of the performance of the portfolio. It is important to carefully select them. Fifth, the AVP portfolio performed very well while allocating the same risk to each asset. This can be a good balance between risk and return. The equally weighted portfolio can also be interesting for investors seeking to diversify without making to complex calculations, thanks to its simplicity and effectiveness.

One of the most important limitations of the study is that is uses datasets with monthly returns. These datasets do not incorporate a lot of information compared to datasets build on daily observations.

In future research it could be interesting to improve the performance PCA-based strategies by combining them with other techniques that are less sensitive to extreme returns. This could for example explore machine learning approaches or robust optimisation techniques. It could also be interesting to consider strategies that incorporate transaction costs to penalise strategies that heavily re-balance their portfolio. A model considering shocks or black swan events could also be interesting. The goal would be to construct a portfolio specially designed to withstand those extreme events. Taxes and legislation constraints could also be added to a future study. Finally, it could be interesting to include the multiple eigenvectors in the strategy of Severini. Indeed, by summing up multiple eigenvectors the strategy could include more explained variance and thus probably perform better.

Bibliography

[1] Abdi, H., & Williams, L. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews : Computational Statistics, 2*(4), 433-459. https://doi.org/10.1002/wics.101

[2] Akaike, H. (2011). Akaike's Information Criterion. *Dans Springer eBooks* (p. 25). https://doi.org/10.1007/978-3-642-04898-2_110

[3] Bai, J., & Ng, S. (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica, 70*(1), 191-221. http://www.jstor.org/stable/2692167

[4] Bai, X., Scheinberg, K., & Tütüncü, R. (2015). Least-Squares Approach to Risk Parity in Portfolio Selection. *Social Science Research Network.* https://doi.org/10.2139/ssrn.2343406

[5] Best, M. (1991). On the Sensitivity of Mean-Variance-Efficient Portfolios to Changes in Asset Means : Some Analytical and Computational Results. https://www.semanticscholar.org/paper/On-the-Sensitivity-of-Mean-Variance-Efficient-to-in-Best-Grauer./10bf02a77f8d99d987bb80e27ef462c19fc1e0c6

[6] Best, M. D., & Grauer, R. T. (1992). Positively Weighted Minimum-Variance Portfolios and the Structure of Asset Expected Returns. *Journal of Financial and Quantitative Analysis, 27*(4), 513. https://doi.org/10.2307/2331138

[7] Black, F. (1993). Estimating Expected Return. *Financial Analysts Journal, 49*(5), 36-38. https://doi.org/10.2469/faj.v49.n5.36

[8] Broadie, M. (1993). Computing efficient frontiers using estimated parameters. *Annals of Operations Research, 45*(1), 21-58. https://doi.org/10.1007/bf02282040

[9] Campa, D., & Camacho-Miñano, M. (2015). The impact of SME's pre-bankruptcy financial distress on earnings management tools. *International Review of Financial Analysis, 42*, 222-234. https://doi.org/10.1016/j.irfa.2015.07.004

[10] Chen, J., & Yuan, M. (2016). Efficient Portfolio Selection in a Large Market. *Journal of Financial Econometrics, 14*(3), 496-524. https://doi.org/10.1093/jjfinec/nbw003

[11] Chopra, V. K., & Ziemba, W. T. (1993). The Effect of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice. *The Journal of Portfolio Management, 19*(2), 6-11. https://doi.org/10.3905/jpm.1993.409440

[12] Clarke, R., De Silva, H., & Thorley, S. (2006). Minimum-Variance Portfolios in the U.S. Equity Market. *The Journal of Portfolio Management, 33*(1), 10-24. https://doi.org/10.3905/jpm.2006.661366

[13] DeMiguel, V. (2009, 1 mai). Optimal Versus Naive Diversification : How Inefficient is the 1/N Portfolio Strategy ? https://ssrn.com/abstract=1376199

[14] DeMiguel, V., Garlappi, L., Nogales, F. F., & Uppal, R. (2009). A Generalized Approach to Portfolio Optimization : Improving Performance by Constraining Portfolio Norms. *Management Science, 55*(5), 798-812. https://doi.org/10.1287/mnsc.1080.0986

[15] DeMiguel, V., & Nogales, F. F. (2007). Portfolio Selection With Robust Estimation. *Social Science Research Network*. https://doi.org/10.2139/ssrn.911596

[16] Dray, S. (2008). On the number of principal components : A test of dimensionality based on measurements of similarity between matrices. *Computational Statistics & Data Analysis, 52*(4), 2228-2237. https://doi.org/10.1016/j.csda.2007.07.015

[17] FRS, K. P. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh and Dublin philosophical magazine and journal of science, 2*(11), 559-572. https://doi.org/10.1080/14786440109462720

[18] Haugen, R. A., & Heins, A. J. (1972). On the Evidence Supporting the Existence of Risk Premiums in the Capital Market. *The Journal of Finance, 27*(5), 881-887. https://doi.org/10.2139/ssrn.1783797

[19] Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology, 24*(6), 417-441. https://doi.org/10.1037/h0071325

[20] Hull (2018).*Risk Management and Financial Institutions*, 5th Edition – Instructor Companion Site. (s. d.). https://bcs.wiley.com/he-bcs/Books?action=index&itemId=1119448115&bcsId=11147

[21] Ji, R., & Lejeune, M. A. (2018). Risk-budgeting multi-portfolio optimization with portfolio and marginal risk constraints. *Annals of Operations Research, 262*(2), 547-578. https://doi.org/10.1007/s10479-015-2044-9

[22] Jolliffe, I. (2002). *Principal Component Analysis*. Springer Science & Business Media.

[23] Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis : a review and recent developments. *Philosophical Transactions of the Royal Society A, 374*(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202

[24] Lassance, N., DeMiguel, V., & Vrins, F. (2021). Optimal Portfolio Diversification via Independent Component Analysis. *Operations Research, 70*(1), 55-72. https://doi.org/10.1287/opre.2021.2140

[25] Ledoit, O., & Wolf, M. S. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis, 88*(2), 365-411. https://doi.org/10.1016/s0047-259x(03)00096-4

[26] Ledoit, O., & Wolf, M. S. (2008). Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance, 15*(5), 850-859. https://doi.org/10.1016/j.jempfin.2008.03.002

[27] Ma, T., & Jagannathan, R. (2002). *Risk Reduction in Large Portfolios : Why Imposing the Wrong Constraints Helps.* https://doi.org/10.3386/w8922

[28] Maillard, S., Roncalli, T., & Teiletche, J. (2008). On the Properties of Equally-Weighted Risk Contributions Portfolios. *Social Science Research Network.* https://doi.org/10.2139/ssrn.1271972

[29] Makamo, E. (2020). Understanding Principal Component Analysis (PCA) and its Application in Factor-Based Risk Attribution Model. *ResearchGate.* https://doi.org/10.13140/RG.2.2.24845.08168

[30] Marčenko, V. (1967). *DISTRIBUTION OF EIGENVALUES FOR SOME SETS OF RANDOM MATRICES.* https://www.semanticscholar.org/paper/DISTRIBUTION-OF-EIGENVALUES-FOR-SOME-SETS-OF-RANDOM-Mar%C4%8Denko-Pastur/88950c60307cc8f2a365fe0e566744a2f3427802

[31] Markowitz, H. M. (1952). *Portfolio Selection. Journal of Finance, 7*(1), 77. https://doi.org/10.2307/2975974

[32] Memmel, C. (s. d.). *Estimating the Global Minimum Variance Portfolio.* https://ssrn.com/abstract=940367

[33] Meucci, A. (2010, 1 avril). *Managing Diversification.* https://ssrn.com/abstract=1358533

[34] Meuci, A. (2005). *Risk and Asset Allocation.* Springer.

[35] Michaud, R. O. (1989). The Markowitz optimization enigma : Is “optimized” optimal ? *Financial Analysts Journal*, 45(1), 31-42.

[36] Mishra, S. P., Sarkar, U. K., Taraphder, S., Datta, S. K., Swain, D. P., Saikhom, R., Panda, S., & Laishram, M. (2017). Principal Component Analysis. *International Journal of Livestock Research*, 1. https://doi.org/10.5455/ijlr.20170415115235

[37] Severini, T. A. (2022). Some properties of portfolios constructed from principal components of asset returns. *Annals of Finance, 18*(4), 457-483. https://doi.org/10.1007/s10436-022-00412-z

[38] Sharpe, W. F. (1964). CAPITAL ASSET PRICES : A THEORY OF MARKET EQUILIBRIUM UNDER CONDITIONS OF RISK*. *Journal of Finance, 19*(3), 425-442. https://doi.org/10.1111/j.1540-6261.1964.tb02865.x

[39] Sharpe, W. F. (1994). The Sharpe Ratio. *The Journal of Portfolio Management, 21*(1), 49-58. https://doi.org/10.3905/jpm.1994.409501

[40] Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (5e éd.). Pearson Education. https://psycnet.apa.org/record/2006-03883-000

[41] Tan, J. (2012). Principal Component Analysis and Portfolio Optimization. *Social Science Research Network*. https://doi.org/10.2139/ssrn.2213687

[42] The Markowitz Optimization Enigma : is "Optimized" Optimal ? (2014). *Social Science Research Network*. https://doi.org/10.2139/ssrn.2387669

[43] Tsukuma, H., & Kubokawa, T. (2020). Estimation of the Covariance Matrix.*SpringerBriefs in statistics*. Springer International Publishing. https://doi.org/10.1007/978-981-15-1596-5_7

[44] Xu, H., Caramanis, C., & Mannor, S. (2016). Statistical Optimization in High Dimensions. *Operations Research, 64*(4), 958-979. https://doi.org/10.1287/opre.2016.1504

[45] Zhao, L., Chakrabarti, D., & Muthuraman, K. (2019). Portfolio Construction by Mitigating Error Amplification : The Bounded-Noise Portfolio. *Operations Research*. https://doi.org/10.1287/opre.2019.1858

[46] Zhao, L., & Gao, R. (2021). An Efficient Dimension Reduction for Portfolio Optimization via the Most Robust Solution. *Social Science Research Network*. https://doi.org/10.2139/ssrn.3991393

[47] Zou, H., Hastie, T., & Tibshirani, R. (2006). Sparse Principal Component Analysis. *Journal of Computational and Graphical Statistics, 15*(2), 265-286. https://doi.org/10.1198/106186006x113430

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Louvain School of Management
Chaussée de Binche 151, 7000 Mons, Belgique | **www.uclouvain.be/lsm**