Impact of the Reference Covariance Matrix in Portfolio Optimization Techniques with Shrinkage

Antoine De Saint Moulin

August 2022

Outline

Abstract

This work focuses on efficient portfolio computation methods and more specifically on the estimation of the covariance matrix. First, theoretical reminders about portfolio optimization are presented. A special focus is made on shrinkage-based method (i.e. Ledoit & Wolf) and principal component analysis (PCA). Next, different strategies are compared and new shrinkage- and PCA-based portfolio optimization methods are presented. Then, the performance achieved with these methods and those from the literature are compared and discussed.

1 Acknowledgements

First of all, I would like to thank my supervisor, Mr. Frédéric Vrins which supported me during all the writing of this paper. He was always available and each of his feedbacks were both complete and interesting. Without it, this paper would have been absolutely different. He introduced me to this great topic that piqued my interest and allowed me to learn a lot.

I would also like to thank Mr. Nathan Lassance. His course on Big Data in Finance is extremely well done. The answers to my questions were very often inside and it allowed me to improve the quality of this document. I started to understand the workings of the different portfolio optimization strategies thanks to him.

I would like to thank my brother who took a lot of his time to help me. Thanks to his competent eye and his experience in writing articles, he has enabled me to write a scientific document that respects the standards and rules in force in this field. His love of mathematics and his skills of rigour and deep understanding of things were a real inspiration to me.

Finally I would also like to end by thanking my parents, who supported me unconditionally during my two bachelors and my master degrees. These eight years of study were marked by ups and downs but I could always count on them to help me overcome one obstacle after another.

2 Introduction

Finding the best way to allocate his wealth has always been a historical topic which is at the same time extremely interesting and difficult to solve because it is simply impossible to predict the future. Nevertheless, it exists a lot of methods that aim to find an optimal balance between risk and return using the information available at the moment investors are taking their decisions. And when the discussion is about portfolio strategies, it is totally impossible not to talk about Markowitz.

The Modern Portfolio Theory was born in 1952 with the publication of a paper by Harry Markowitz [1].It assumes that the risk of a portfolio can be correctly measured by the variance of its profitability, He formalized and made explicit the fundamental dilemma of modern finance: obtain a low but certain return, or accept to take a risk in the hope of increasing this return. That is how the **Mean Variance Portfolio** was born. It has become the most well-known strategy due to its intuitive theory and its easy implementation. The main difficulty in estimating the mean-variance portfolio resides in the estimation of the mean returns μ. This one is generally characterized by a much higher variance than the sample variances. As Fischer Black claimed in 1993 [2],*"Estimating expected return is hard. Daily data hardly help at all. Only a longer period of data can help. We need decades of data for accurate estimates of average expected return."* Moreover, the portfolio is very sensitive to this estimate.

In the following years, a lot of mathematicians managed to improve this strategy. One of the proposed solutions is to consider that all assets have the same sample mean. As explained by Jagannatha and Ma [3], the sample mean is such a poor estimator of the true mean that it is better to simply ignore the mean when no further information available. This leads to the **Minimum Variance Portfolio**. This one aims to minimize the risk and therefore suits to investors with infinite risk-aversion. Moreover, although this portfolio is more robust, it is still quite sensitive to estimation error in the covariance matrix [8].High levels of positive error result from the largest values and high levels of negative error from the smallest values in the covariance matrix.

In order to overcome this estimation problem, different shrinkage methods were developed. Shrinkage aims to pull these extreme values toward a central value. The method allows to compute a covariance matrix by finding the right balance between two matrices. This balance is defined by a coefficient shrinkage computed in different ways that will be discussed later.

The shrinkage method we are focusing in this work is the **Ledoit** & **Wolf strategy**. They propose to shrink an unbiased-high-variance sample estimate and a biased-low-variance target estimate. As a target estimate, they used the identity matrix in [4]. They also proposed a method to shrink the sample covariance matrix with the single-index covariance matrix in [5] and with the constant correlation covariance matrix in [6]. They explain that *"The shrinkage target should fulfill two requirements at the same time: it involves only a small number of free parameters (that is, a lot of structure) but it also reflects important characteristics of the unknown quantity being estimated."* However, for each method, the main challenge is to find the optimal shrinkage intensity to use.

In this work, the achieved performance with the mean variance portfolio, minimum variance portfolio and the Ledoit & Wolf strategies are presented. Then, another method is developed: as a target estimate, we use a covariance matrix computed after applying **Principal Component Analysis** [22], and the shrinkage intensity is found with a 10-fold cross-validation (this method is explained in [20]). Thanks to PCA, a low-dimensional representation of the matrix of the returns is computed, capturing most of the information. This is motivated by the performance achieved with the minimum variance strategy, which is less sensitive to estimation error of the covariance matrix when the matrix dimension is low [6]. Finally, the performance achieved with these four methods will be compared.

3 Mathematical Background

3.1 Mean and Minimum Variance Portfolios

Historically, investors have allocated their money to assets that exhibit the highest return, where the volatility is usually stronger, allowing them to gain bargains on expected returns [7]. However, it is really difficult to compute an exact forecast with any sort of asset because markets are unpredictable and volatility is sometimes huge. The most famous model that is used is the Modern Portfolio Theory introduced by Markowitz in 1952 [1]. As written in his paper, the economic sciences Nobel Prize winner explains that portfolios that seek to balance the volatility of their assets could perform better than standard portfolios. He details the mathematics of diversification and demonstrates how risk can be reduced through an efficient combination of assets.

Let us denote by R_i the return of the asset X_i in a portfolio. The **expected return** μ_p of the portfolio is defined as

$$\mu_p = \sum_{i=1}^{N} w_i \, \mu_i = \mathbf{w}'\boldsymbol{\mu} \tag{1}$$

where N is the number of assets, $\boldsymbol{\mu} = [\mu_1, ..., \mu_N]'$ with $\mu_i = \mathbb{E}[R_i]$ is a vector containing the mean returns, and $\mathbf{w} = [w_1, ..., w_N]'$ is a vector of weights the investor chose to put on each asset. The variance of the returns, corresponding in this field to the risk associated to the portfolio, is defined as

$$\sigma_p^2 = \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}, \tag{2}$$

where $\boldsymbol{\Sigma}$ is the covariance matrix of the assets returns, i.e. $[\boldsymbol{\Sigma}]_{ij} = \mathbb{E}[(R_i - \mu_i)(R_j - \mu_j)]$.

In order to maximize the expected return in one hand and to minimize the risk in the other hand, Markowitz proposes to compute the weights by solving the following quadratic problem:

$$\mathbf{w}^* = \underset{\mathbf{w}}{\operatorname{argmax}} \, \mu_p - \gamma\sigma_p^2, \tag{3}$$

with γ modelling the risk-aversion of the investor. This gives rise to the **efficient frontier**, i.e. the set of pairs (μ_p, σ_p^2) which are solutions of the maximisation problem. All the portfolios on this frontier are called **mean-variance portfolios**. With risk-free asset r_f, the efficient frontier becomes a line called capital **capital market line**, starting at $(r_f, 0)$ and tangent to the previous efficient frontier. The tangent point between the capital market line and the previous efficient frontier gives the portfolio which maximizes the **sharpe ratio** SR $= \frac{\mu_p - r_f}{\sigma_p}$, i.e. the ratio between the mean excess return and the volatility. This is illustrated in Figure 1.

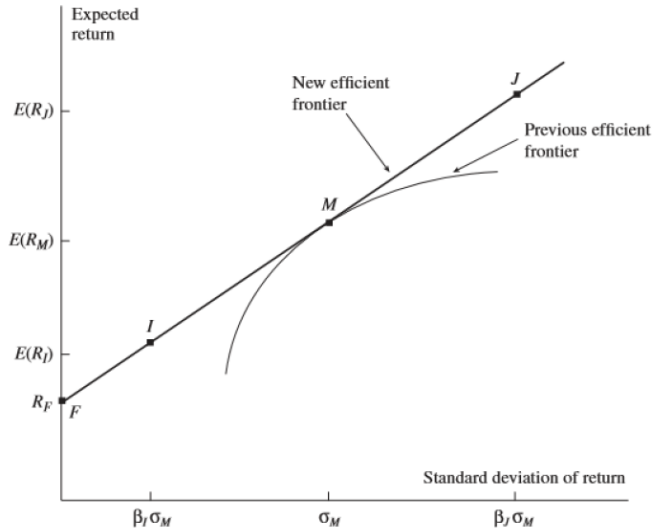


Figure 1: Efficient frontier and capital market line. The maximum sharpe-ratio portfolio is located at the point M [24].

This portfolio is called **maximum sharpe-ratio portfolio**. Therefore, to compute the maximum sharpe-ratio portfolio, Equation (3) is rewritten as

$$\mathbf{w}_{\text{MSR}} = \underset{\mathbf{w}}{\text{argmax}} \frac{\mu_p}{\sigma_p} = \underset{\mathbf{w}}{\text{argmax}} \frac{\mathbf{w}'\boldsymbol{\mu}}{\sqrt{\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}}}. \tag{4}$$

With the only constraint $\mathcal{W} = \left\{\mathbf{w} \in \mathbb{R}^N \mid \mathbf{w}'\mathbf{1} = 1\right\}$ where $\mathbf{1}$ is a all-ones vector of N elements, the solution is given by

$$\mathbf{w}_{\text{MSR}} = \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\mathbf{1}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}. \tag{5}$$

The classical estimate of the maximum sharpe-ratio portfolio is computed as

$$\hat{\mathbf{w}}_{\text{MSR}} = \frac{\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}}{\mathbf{1}'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}}, \tag{6}$$

where $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ are respectively the sample mean and sample covariance matrix obtained from a training set of T samples $\mathcal{T} = \{\mathbf{X}_t = [X_{1t}, ..., X_{Nt}]'\}_{t=1}^{T}$:

$$\hat{\boldsymbol{\mu}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t,$$

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{T-1}\sum_{t=1}^{T}\left(\mathbf{X}_t - \hat{\boldsymbol{\mu}}\right)\left(\mathbf{X}_t - \hat{\boldsymbol{\mu}}\right)'.$$

This is an unbiased but also high-variance estimator. Therefore, the main difficulty to obtain this mean-variance portfolio is the estimation of the mean returns $\hat{\boldsymbol{\mu}}$, owing to its

high variance (generally much higher than the variance of $\hat{\mathbf{\Sigma}}$), and the high sensitivity of the mean-variance portfolio to the sample mean $\boldsymbol{\mu}$.

One possibility to avoid facing the issue of the estimation of the sample mean is to assume that the sample mean of all the assets is the same. In that case, the mean-variance portfolio corresponds to the **minimum-variance portfolio**, and the weights are computed as

$$\mathbf{w}_{\text{MV}} = \frac{\mathbf{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}'\mathbf{\Sigma}^{-1}\mathbf{1}}. \tag{7}$$

Although the minimum-variance portfolio is more robust than the mean-variance portfolio, it is still sensitive to estimation error in the covariance matrix, especially when $N >> T$ since the cardinality of $\mathbf{\Sigma}$ increases quickly with N.

It is important to note that the sample covariance matrix is an efficient estimator (unbiased and minimum variance) when the asset returns are Gaussian. However, in practice, asset returns are not Gaussian, and there may be biased estimates which achieves better out-of-sample performance [8]. More specifically, the sample covariance matrix is the maximum likelihood estimator for normally distributed returns. However, even if this estimator has the smallest asymptotic variance for this distribution, it is highly sensitive to deviations of the asset-return distribution from the normal.

There exists well known strategies in order to overcome this problem like the **no-short-selling constraint** ("constraining portfolio weights to be nonnegative can reduce the risk in estimated optimal portfolios even when the constraints are wrong" [3]) or the **shrinkage method**. The shrinkage method is developed in details in Section 3.2.

3.2 Shrinkage-based Methods

As described in Section 3.1, the covariance matrix of the assets returns must be estimated accurately in order to compute the weights of the portfolio strategy. This topic and its related problems have been well documented in the literature [10–13]. In practice, when the number of stocks is large, especially relative to the available number of historical return observations (unfortunately, it is usually the case), numerous errors are committed in the estimation of the sample covariance matrix. This leads to large errors on the optimal weights, and therefore to a large variation of the weights over time. In order to avoid that, shrinkage-based methods have been developed. The **shrinkage** is a concept in statistics introduced by professor Charles Stein of Stanford university in 1955 [26]. The aim of this operation is to pull the most extreme coefficients of the sample covariance matrix towards more central values in order to reduce the impact of estimation errors when it matters the most. The challenge is to know what is the optimal shrinkage intensity.two shrinkage-based methods are presented in this Section: the Ledoit & Wolf strategy and the Stein's shrinkage estimator of eigenvalues.

3.2.1 Ledoit & Wolf Strategy

Ledoit & Wolf commented in [6] the attempts to use shrinkage in portfolio selection made in [14] and [15] but the proposed shrinkage techniques broke down when the number of stock exceeds the number of historical return observations, as often in practice. More

recently, [3] showed that mean-variance optimizers are already applying some form of shrinkage implicitely to the covariance matrix, and proved that it is generally beneficial to improve weights stability.

Based on that, Ledoit & Wolf also shown in [6] that the sample covariance matrix should not be used for portfolio optimization since the estimation errors perturb the mean-variance optimizer. Instead, they propose a shrinkage estimator in which the sample covariance matrix and another highly structured matrix are shrunk towards the center using a weighted average. Formally, the shrinkage estimator is computed as:

$$\hat{\mathbf{\Sigma}}_{\text{LW}} = \delta^* \mathbf{F} + (1 - \delta^*)\hat{\mathbf{\Sigma}}, \tag{8}$$

where $\delta \in [0, 1]$ is the shrinkage coefficient and $\mathbf{F}$ is a highly structured matrix. In [4–6], the matrix $\mathbf{F}$ is respectively chosen to be the identity matrix, the single-index covariance matrix and the constant correlation matrix. Then, each one of their three works previously mentioned propose a different way to estimate δ. The shrinkage coefficient is computed from the data itself and is not fixed exogenously. In order to compute the optimal shrinkage coefficient, Ledoit & Wolf first define a quadratic loss function in [5] as the Frobenius norm of the difference between the shrinkage estimator and the true covariance matrix:

$$L(\delta) = \|\hat{\mathbf{\Sigma}}_{\text{LW}} - \mathbf{\Sigma}\|_{\text{F}}^2, \tag{9}$$

where $\|\cdot\|_{\text{F}}$ is the Frobenius norm, defined for a symmetric matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$ with entries $[\mathbf{A}]_{ij} = a_{ij}$ and eigenvalues λ_i as

$$\|\mathbf{A}\|_{\text{F}}^2 = \text{tr}(\mathbf{A}^2) = \sum_{i=1}^{N} \sum_{i=1}^{N} a_{ij} = \sum_{i=1}^{N} \lambda_i^2. \tag{10}$$

Then, they define the risk function as $R(\delta) = \mathbb{E}\left[L(\delta)\right]$ and solve the following minimisation problem:

$$\delta^* = \arg\min_{\delta} R(\delta). \tag{11}$$

The exact solution of this problem can be expressed as

$$\delta^* = \frac{\kappa}{T} + \mathcal{O}\left(\frac{1}{T^2}\right), \tag{12}$$

where κ is a constant which depends on the unobservables. Note that the shrinkage coefficient decreases with the number of samples T. Denoting respectively by $\hat{\mu}_i = [\hat{\boldsymbol{\mu}}]_i$ and $\hat{\sigma}_{ij} = [\hat{\mathbf{\Sigma}}]_{ij}$ the elements of the sample mean and covariance matrix, and by $f_{ij} = [\mathbf{F}]_{ij}$ the elements of $\mathbf{F}$, they finally define a consistent estimator of κ as

$$\hat{\kappa} = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} (p_{ij} - r_{ij})}{\sum_{i=1}^{N} \sum_{j=1}^{N} c_{ij}}, \tag{13}$$

with

$$p_{ij} = \frac{1}{T}\sum_{t=1}^{T}\left[(X_{it}-\hat{\mu}_i)(X_{jt}-\hat{\mu}_j)-\hat{\sigma}_{ij}\right]^2, \tag{14}$$

$$r_{ij} = \frac{1}{T}\sum_{t=1}^{T}\frac{\hat{\sigma}_{j0}\hat{\sigma}_{00}(X_{it}-\hat{\mu}_i)+\hat{\sigma}_{i0}\hat{\sigma}_{00}(X_{jt}-\hat{\mu}_j)-\hat{\sigma}_{i0}\hat{\sigma}_{j0}(X_{0t}-\hat{\mu}_0)}{\hat{\sigma}_{00}^2} \cdot (X_{0t}-\hat{\mu}_0)(X_{it}-\hat{\mu}_i)(X_{jt}-\hat{\mu}_j) - f_{ij}\hat{\sigma}_{ij}, \tag{15}$$

$$c_{ij} = (f_{ij}-\hat{\sigma}_{ij})^2. \tag{16}$$

In summary, the approach of Ledoit & Wolf suggests to shrink the sample covariance matrix a given target matrix $\mathbf{F}$ involving only a small number of parameters (highly structured) and reflecting important characteristics about the quantity being estimated. Finally, this new matrix can be used for instance to compute the weights of the minimum variance portfolio:

$$\hat{\mathbf{w}}_{\text{MSR,LW}} = \frac{\hat{\mathbf{\Sigma}}_{\text{LW}}^{-1}\mathbf{1}}{\mathbf{1}'\hat{\mathbf{\Sigma}}_{\text{LW}}^{-1}\mathbf{1}}. \tag{17}$$

3.2.2 Stein's Shrinkage Estimator of Eigenvalues

Another method that will be used in this work the Stein's Shrinkage method. Stein's method is based on the observation that, unless $T >> N$, the eigenvalues of the sample covariance matrix $\hat{\mathbf{\Sigma}}$ are severly distorded compared to those of the true covariance matrix $\mathbf{\Sigma}$. The largest (resp. smallest) sample eigenvalues tend to significantly overestimate (resp. underestimate) their counterparts in the population [9].

The Stein's covariance matrix estimator is computed as

$$\hat{\mathbf{\Sigma}}_{\text{S}} = \mathbf{V}\tilde{\mathbf{\Lambda}}\mathbf{V}', \tag{18}$$

such that this matrix is as close as possible to the true covariance matrix $\hat{\mathbf{\Sigma}}$. In order to measure the distance between both matrices, Stein proposes to use the Kullback-Leibler divergence between two zero-mean Gaussian distribution with covariance matrices $\hat{\mathbf{\Sigma}}_{\text{S}}$ and $\hat{\mathbf{\Sigma}}$, defined in [8] for two multivariate random variables $\mathbf{X}$ and $\mathbf{Y}$ as

$$\text{KL}(\mathbf{X},\mathbf{Y}) = \int f_{\mathbf{X}}(\mathbf{z})\ln\frac{f_{\mathbf{X}}(\mathbf{z})}{f_{\mathbf{Y}}(\mathbf{z})}\,\text{d}\mathbf{z}. \tag{19}$$

This metric is always positive, and zero when $f_{\mathbf{X}} = f_{\mathbf{Y}}$. If $T \geq N$, the minimization of the Kullback-Leibler divergence leads to the following eigenvalues correction:

$$\tilde{\lambda}_i = \frac{T\lambda_i}{T-N+1+2\lambda_i\sum_{i\neq j}\frac{1}{\lambda_i-\lambda_j}}, \tag{20}$$

where $\tilde{\lambda}_i = [\tilde{\mathbf{\Lambda}}]_{ii}$ and $\lambda_i = [\mathbf{\Lambda}]_{ii}$. This solution leads to negative eigenvalues and their order is affected. Therefore, Stein also propose an isotonizing algorithm in order to solve these issues..

3.3 Principal Component Analysis

The **Principal Component Analysis** (PCA) is one of the most popular and powerful ways of tackling the curse of dimensionality and improve the out-of-sample performance of predictive models via **dimension reduction**. Reducing the dataset dimension is very useful for some methods (for instance [14,15]) which does not handle well large N values.

Let us consider a matrix $\mathbf{X} = [\mathbf{X}_1, ..., \mathbf{X}_N]' \in \mathbb{R}^{N \times T}$ containing T observations of N random variables. We assume that $\mathbf{X}$ is centered, i.e. each random variable is zero-mean (otherwise, the matrix is centered by subtracting the means). The goal is to find a low-dimensional representation of $\mathbf{X}$ that captures the most information. More specifically, the k^{th} principal component $\mathbf{Y}_k$ is defined as the standardized linear combination of the variables $\mathbf{X}_i$ that has maximum variance and is uncorrelated with the other principal components:

$$\mathbf{Y}_k = \sum_{i=1}^{N} v_{ki}\mathbf{X}_i \quad \text{s.t.} \quad \mathbf{v}_k'\mathbf{v}_k = 1 \quad \text{and} \quad \mathbf{v}_k'\mathbf{v}_l = 0 \,\forall\, l \neq k, \tag{21}$$

where $\mathbf{v}_k = [v_{k1}, ..., v_{kN}]'$ is the weight vector associated with the k^{th} principal component of $\mathbf{X}$. The iterative process is stopped after K iterations, giving the K principal components $\mathbf{Y} = [\mathbf{Y}_1, ..., \mathbf{Y}_K]' \in \mathbb{R}^{K \times T}$ computed as

$$\mathbf{Y} = \mathbf{V}'\mathbf{X}, \tag{22}$$

with $\mathbf{V} = [\mathbf{v}_1, ..., \mathbf{v}_K]$ and $\mathbf{V}'\mathbf{V} = \mathbf{I}_K$. These components represent the maximal-variance directions of the data, and thus are the most relevant to represent the sample matrix $\mathbf{X}$ with only K dimensions. The weight matrix $\mathbf{V}$ can be computed using an eigenvalue decomposition of the sample covariance matrix $\hat{\mathbf{\Sigma}}$ of $\mathbf{X}$: since the covariance matrix is symmetric and positive semi-definite, it can be written as

$$\hat{\mathbf{\Sigma}} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}', \tag{23}$$

where $\mathbf{Q} \in \mathbb{R}^{N \times N}$ is an orthogonal matrix ($\mathbf{Q}'\mathbf{Q} = \mathbf{Q}\mathbf{Q}' = \mathbf{I}_N$) and $\mathbf{\Lambda} \in \mathbb{R}^{N \times N}$ is a diagonal matrix. We are looking for a matrix $\mathbf{Z} = \mathbf{W}'\mathbf{X}$ composed of uncorrelated variables. If we choose $\mathbf{W}$ equal to $\mathbf{Q}$, the sample covariance matrix of $\mathbf{Z}$ is computed as

$$\hat{\mathbf{\Sigma}}_z = \mathbf{W}'\hat{\mathbf{\Sigma}}\mathbf{W} = \mathbf{Q}'\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}'\mathbf{Q} = \mathbf{\Lambda}, \tag{24}$$

which is diagonal. Therefore, K principal computed are obtained by keeping the K columns of $\mathbf{Q}$ associated with the largest eigenvalues to form the matrix $\mathbf{V}$, and finally $\mathbf{Y} = \mathbf{V}'\mathbf{X}$ is computed. This low-dimensional representation can be used to obtain more robust predictions, i.e. a more robust estimate of the covariance matrix $\hat{\mathbf{\Sigma}}$ of $\mathbf{X}$. Note that this method differs from the **independent component analysis**, which looks for components that are as independent as possible. It therefore accounts for non-linear dependence (beyond the covariance). Finally, if an high-dimensional representation of $\mathbf{Y}$ is needed, the matrix can be back-projected into the initial space by computing

$$\tilde{\mathbf{X}} = \mathbf{V}\mathbf{Y} = \mathbf{V}\mathbf{V}'\mathbf{X}. \tag{25}$$

Figure 2 illustrates the back-projection of 2D data which have been projected on the first principal component.

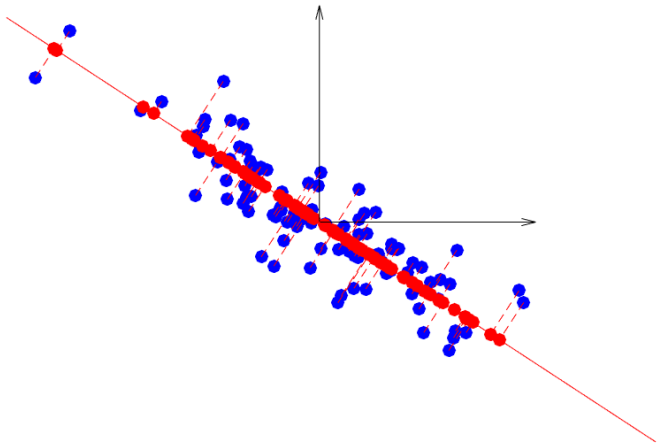


Figure 2: Projection and back-projection of 2D samples on the first principal component.

The main challenge of PCA is to choose the number of principal components wisely. In this work, we will focus on two methods from the literature: the **percentage variance** [3] and the **Marchenko-Pastur theorem** [5].

3.3.1 Percentage of Explained Variance

The total amount of variance in the correlation matrix is computed by adding the diagonal values of the sample covariance matrix $\hat{\mathbf{\Sigma}}$. Additionally, this sum is equal to the sum of the eigenvalues of $\hat{\mathbf{\Sigma}}$:

$$\sum_{i=1}^{N} \lambda_i = \sum_{i=1}^{N} \hat{\sigma}_i^2, \tag{26}$$

where $\lambda_i = [\mathbf{\Lambda}_{ii}]$ and $\hat{\sigma}_i^2 = [\hat{\mathbf{\Sigma}}]_{ii}$. Therefore, as detailed in [3], the percentage of the total variance of an eigenvalue is computed as the ratio between the eigenvalue and the total variance. The method of the percentage explained variance tries to find an optimal value of K by choosing the first value of K for which the sum of the percentage of the total variance of these K eigenvalues exceeds a given threshold $\alpha \in [0, 1]$:

$$\text{Select the smallest } K \text{ such that } \quad \frac{\sum_{i=1}^{K} \lambda_i}{\sum_{i=1}^{N} \lambda_i} > \alpha. \tag{27}$$

The threshold α should be close to one in order to select a sufficient number of components to represent $\mathbf{X}$. However, the issue is not really solved since the choice of K has been replaced by the choice of α.

3.3.2 Marchenko-Pastur Theorem

In order to choose K, Marchenko and Pastur described in [18] a method that is still very popular in order to choose K: K should be selected such that the K first principal

components represent relevant signals, whereas the $N-K$ remaining components represent noisy signals. Assuming that the entries of $\mathbf{X}$ are independent and identically distributed random variables with zero-mean and unit variance, the Marchenko-Pastur distribution describes the spreading of the eigenvalues of the covariance matrix of $\mathbf{X}$. When $N, T \to \infty$ with $1 < T/N < \infty$, each eigenvalue has the following probability density function:

$$f_\lambda(x) = \frac{T}{N} \frac{\sqrt{(\lambda_+ - x)(x - \lambda_-)}}{2\pi x} \mathbb{1}\left(x \in [\lambda_-, \lambda_+]\right), \tag{28}$$

with lower and upper bounds

$$\lambda_\pm = \left(1 \pm \sqrt{\frac{N}{T}}\right)^2. \tag{29}$$

This density is illustrated in Figure 3 for different N/T ratio.

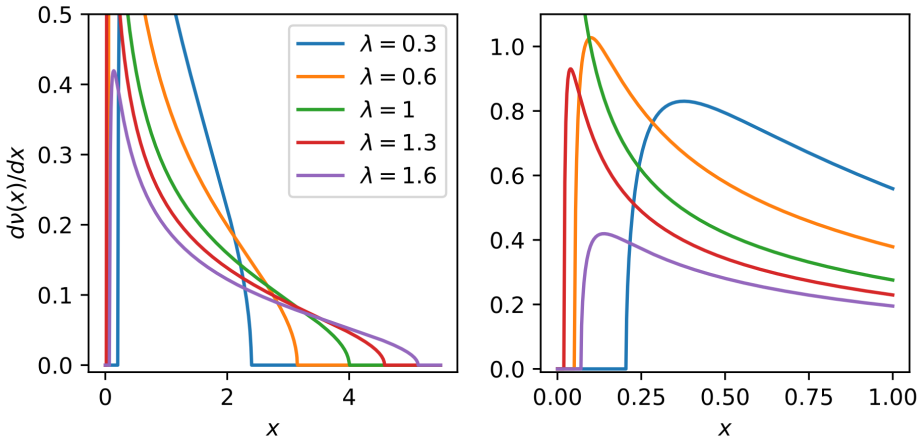


Figure 3: Marchenko-Pastur distribution for various ratio $\lambda = N/T$ [25].

Note that each eigenvalue has the same probability density function since it is assumed that each return is i.i.d with zero-mean and unit variance. This is indeed not the case in practice, and it explains why eigenvalues outside of the range $[\lambda_-, \lambda_+]$ are observed.

In practice, the eigenvalues which are smaller (resp. upper) than λ_- are considered to be associated with noisy (resp. relevant signals). Therefore, all the eigenvalues higher than λ_+ are kept, and $K = \#\{\lambda_i \mid \lambda_i > \lambda_+\}$ is therefore defined. [17]

3.3.3 Application to Portfolios

In [21], Chen & Yuan proposed to reduce the dimension of the dataset to work in another subspace of dimension K, typically the PCA subspace. The projection of the asset returns is computed following (22), leading to a sample mean and covariance matrix in low dimension given by

$$\hat{\boldsymbol{\mu}}_\mathbf{Y} = \mathbf{V}'\boldsymbol{\mu}, \quad \hat{\boldsymbol{\Sigma}}_\mathbf{Y} = \mathbf{V}'\hat{\boldsymbol{\Sigma}}\mathbf{V}. \tag{30}$$

Finally, the estimates of the maximum sharpe-ratio and minimum-variance weights computed through the low dimensional representation are respectively obtained as

$$\hat{\mathbf{w}}_{\text{MSR+PCA}} = \frac{\mathbf{V}\left(\mathbf{V}'\hat{\boldsymbol{\Sigma}}\mathbf{V}\right)^{-1}\mathbf{V}'\hat{\boldsymbol{\mu}}}{\mathbf{1}'\mathbf{V}\left(\mathbf{V}'\hat{\boldsymbol{\Sigma}}\mathbf{V}\right)^{-1}\mathbf{V}'\hat{\boldsymbol{\mu}}}, \quad \hat{\mathbf{w}}_{\text{MV+PCA}} = \frac{\mathbf{V}\left(\mathbf{V}'\hat{\boldsymbol{\Sigma}}\mathbf{V}\right)^{-1}\mathbf{V}'\mathbf{1}}{\mathbf{1}'\mathbf{V}\left(\mathbf{V}'\hat{\boldsymbol{\Sigma}}\mathbf{V}\right)^{-1}\mathbf{V}'\mathbf{1}}. \tag{31}$$

With the PCA subspace ,since $\hat{\boldsymbol{\Sigma}}_{\mathbf{Y}} = \boldsymbol{\Lambda}$, this is further simplified into

$$\hat{\mathbf{w}}_{\text{MSR+PCA}} = \frac{\mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{V}'\hat{\boldsymbol{\mu}}}{\mathbf{1}'\mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{V}'\hat{\boldsymbol{\mu}}}, \quad \hat{\mathbf{w}}_{\text{MV+PCA}} = \frac{\mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{V}'\mathbf{1}}{\mathbf{1}'\mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{V}'\mathbf{1}}. \tag{32}$$

4 Application of Portfolio Strategies on Datasets

4.1 Datasets

Through all the following analysis, the performance of the evaluated portfolio strategies are computed for 4 datasets:

1. The *48 industry portfolios* downloaded from Kenneth French's [1], regrouping the returns of 48 different industries, i.e. Agric, Food, Soda, Beer, Smoke, etc. Each NYSE, AMEX, and NASDAQ stock are assigned to an industry portfolio at the end of June of year t based on its four-digit Standard Industrial Classification code at that time.

2. An artificial dataset created with the same parameter as the *48 industry portfolios* and simulating shocks at a rate of one percent. Each shock will be represented by a fall of ten percent of the returns.

3. The *100 portfolios formed on size and book-to-market* dataset, downloaded from Kenneth French's. The portfolios are constructed at the end of each June. They are the intersections of 10 portfolios formed on size (market equity, ME) and 10 portfolios formed on the ratio of book equity to market equity (BE/ME).

4. An artificial dataset created with the same parameters as the *100 portfolios formed on size and book-to-market* and simulating shocks at a rate of one percent.

The time period is from January 1970 to December 2020, for a total of 612 months. We will use monthly frequencies. Values at -99.99% in the original datasets correspond to missing data and must be discarded. We decided to replace those values by 0. These variables could be suppressed but it would lead to the undesirable suppression of useful information. Other solutions are possible, i.e. replacing these values by the sample mean of the entire dataset or the sample mean of this variable. However, the number of missing values is very small and results are insensitive to the way they are replaced.

4.2 New Portfolio Strategies

In this work, we propose two new shrinkage-based strategies for the optimization of the porfolio. The shrinkage is performed as proposed by Ledoit & Wolf in (33):

$$\tilde{\mathbf{\Sigma}} = \delta^* \mathbf{F} + (1 - \delta^*)\hat{\mathbf{\Sigma}}, \tag{33}$$

However, instead of the scaled identity matrix, two other target matrices $\mathbf{F}$ are proposed:

1. The back-projected sample covariance matrix after performing PCA; Assuming that the number of principal components K has been determined, the following operations are performed:

   (a) Centering of $\mathbf{X}$.

   (b) Computation of the sample covariance matrix $\hat{\mathbf{\Sigma}}$.

(c) Eigenvalue decomposition of the sample covariance matrix $\hat{\mathbf{\Sigma}} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}'$.

(d) Computation of the matrix $\mathbf{V}$ which is composed of the K columns of $\mathbf{Q}$ associated with the highest eigenvalues in $\mathbf{\Lambda}$.

(e) Computation of the back-projection of the samples $\tilde{\mathbf{X}} = \mathbf{V}\mathbf{V}'\mathbf{X}$ (and compensation of the centering).

(f) Computation of the sample covariance matrix of the back-projected data $\hat{\mathbf{\Sigma}}_{\tilde{\mathbf{X}}}$.

This matrix has a rank of $K << N$. It is composed of the projection of the sample set on the principal directions obtained from its covariance matrix. Therefore, it is chosen as a candidate for the target matrix used in the shrinkage.

2. The diagonal matrix containing the eigenvalues of the sample covariance matrix; The sample covariance matrix of the centered data can be decomposed using eigenvalues decomposition: $\hat{\mathbf{\Sigma}} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}'$. Therefore, we choose $\mathbf{\Lambda}$ as a candidate for the target matrix used in the shrinkage. It is highly structured since it is a diagonal matrix, and it contains information about the sample set (intensity of the principal directions of the samples).

In order to estimate the shrinkage coefficient δ^*, 10-fold cross validation is used on the training set. More specifically, The training set is first splitted in ten folds. Then, each fold is considered one at a time as the testing set, and the other folds are merged to form the training set. Using that, the selected porfolio strategy is applied and the returns achieved for this fold are computed. At the end, the returns obtained on each fold are used to compute the expected return, the variance of the returns and the sharpe ratio. This is illustrated in Figure 4. This methodology is applied for different shrinkage coefficients (i.e. every multiple of 0.01 between 0 and 1), and the optimal shrinkage coefficient is finally chosen to be the one which achieves the greatest sharpe ratio.

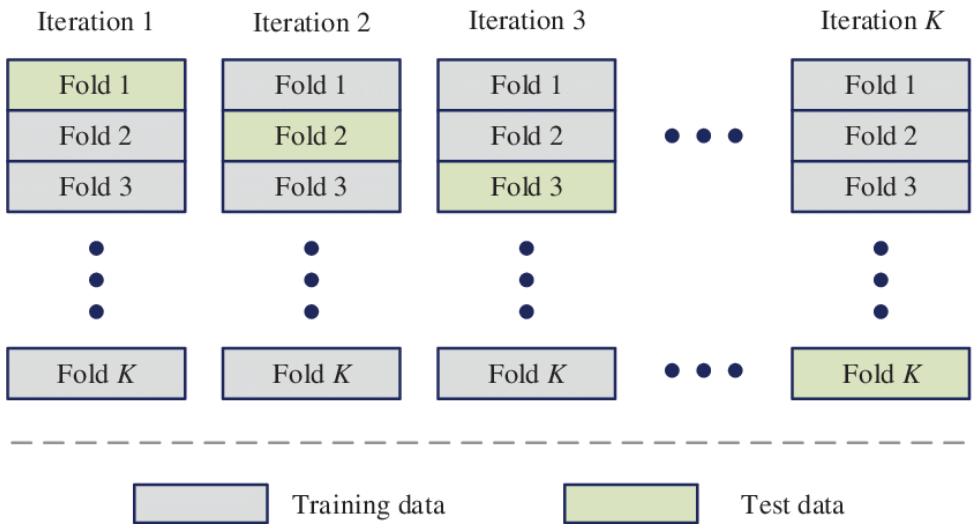


Figure 4: K-fold cross validation [23].

Finally, the portfolio weights are computed using the shrunk matrix following the minimum variance method:

$$\tilde{\mathbf{w}} = \frac{\tilde{\mathbf{\Sigma}}^{-1}\mathbf{1}}{\mathbf{1}'\tilde{\mathbf{\Sigma}}^{-1}\mathbf{1}}. \tag{34}$$

4.3 Methodology

The performance achieved with multiple portfolio strategies is compared. More specifically, the following strategies are implemented:

- The maximum sharpe-ratio portfolio presented in Section 3.1, where the weights are computed based on the sample mean and sample covariance matrix as shown in (6).
- The minimum variance portfolio presented in Section 3.1, where the weights are computed based only on the sample covariance matrix as shown in (7).
- The minimum variance portfolio with dimension reduction through PCA presented in Section 3.3.3. The weights are computed following (32) based on the sample covariance matrix.
- The Ledoit & Wolf strategy presented in Section 3.2.1, where the target matrix $\mathbf{F}$ is a scaled identity matrix and the shrinkage parameter is computed as in [4]. The weights are computed with the shrunk matrix as in (17).
- The new shrinkage-based portfolio strategies presented in Section 4.2. The weights are computed as shown in (34).

In order to compare the performance achieved with each strategy, the (annualized) mean sharpe ratio is computed, as developed by William Sharpe in 1966 [19]. It is computed as the ratio between the expected returns and the standard deviation of the returns, and describes how much excess return is received for the extra volatility endured for holding a riskier asset. The higher the ratio is, the more attractive the portfolio is for investors who are looking for well-balanced return-variance couple. The applied methodology is a rolling window analysis: starting from the first month, a window of 120 months is defined. Then, the weights of the different portfolio strategies are computed using these 120 months as the training set. Next, the returns achieved with the different strategies are computed in the 6 months following the window. These operations are repeated after shifting the window 6 month later one at a time, until the end of the dataset is reached. Since there is 612 months (51 years) available in the different datasets, the procedure generates 492 portfolio returns for each strategy. Using these returns achieved with each strategy over all the 6-month test windows, the performance metrics are finally evaluated. The rolling window analysis is illustrated in Figure 5.

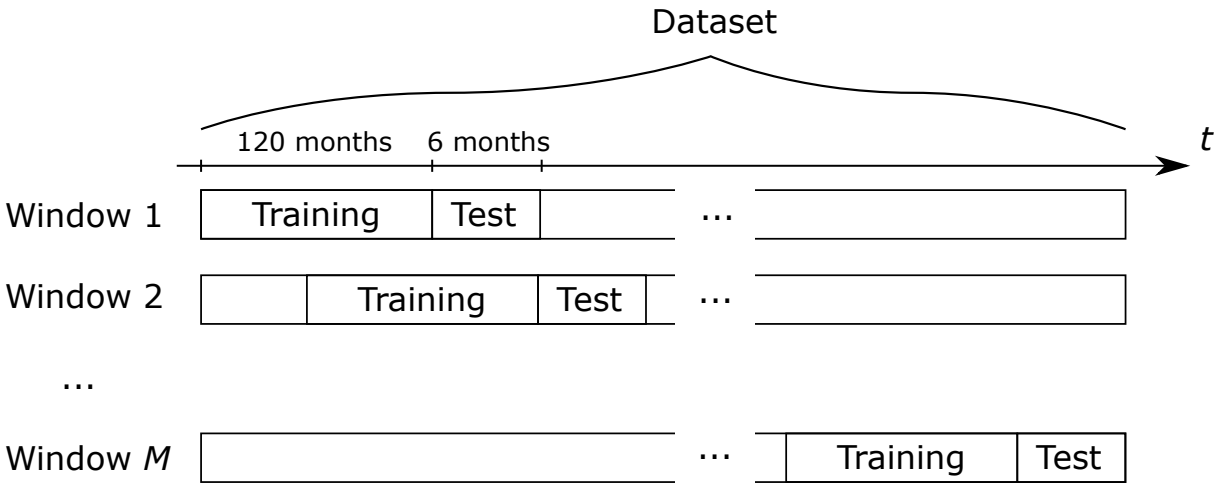


Figure 5: Rolling window.

In order to apply these methods, two meta-parameters should be optimized:

1. The number of principal components K for the dimension reduction through PCA. It is estimated using the percentage variance or the Marchenko-Pastur method, as respectively described in Section 3.3.1 and 3.3.2.

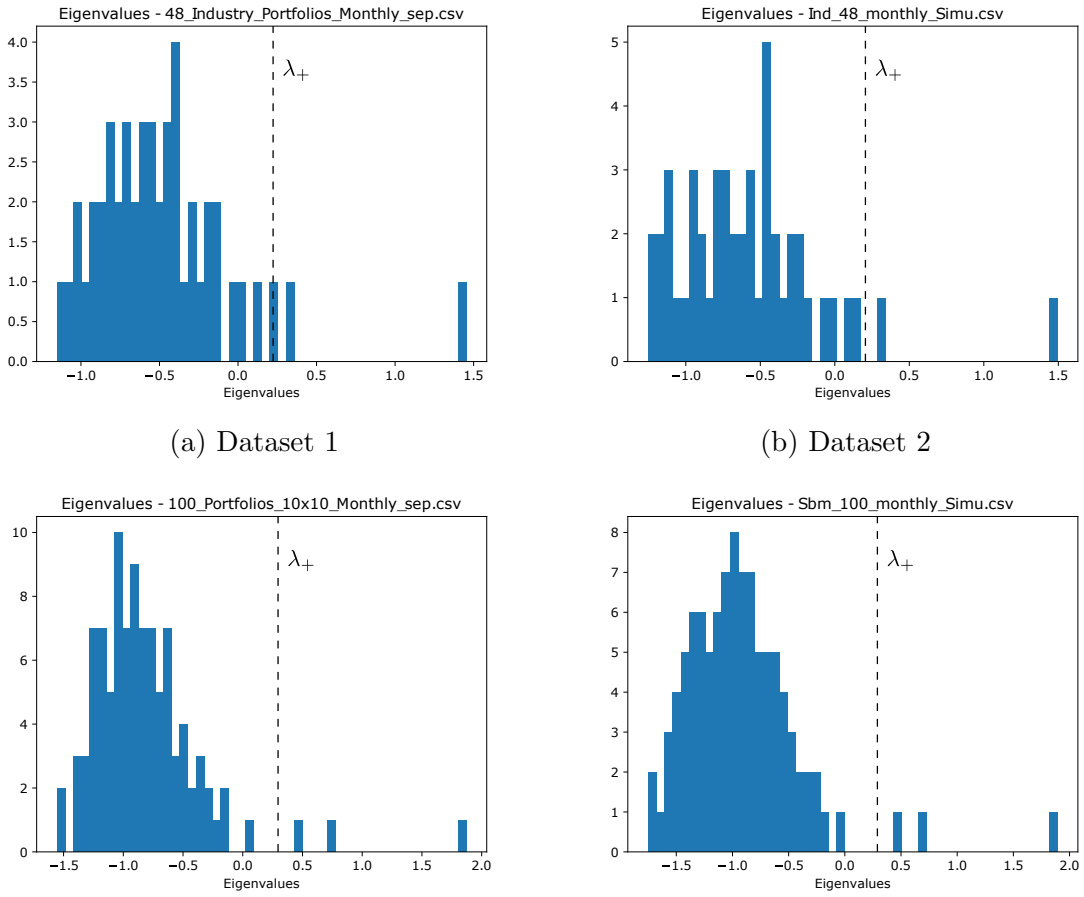
2. The shrinkage coefficients δ^* for the two new shrinkage-based portfolio strategies. It is estimated as described in Section 4.2, using a 10-fold cross validation method with the mean sharpe ratio as metric.

Two approaches have been followed in this work: either the meta-parameters are optimized on the whole dataset before the application of the rolling window, or the meta-parameters are optimized during the rolling window using only the training set. In the first approach, the meta-parameters are estimated at the start only once. Then they are applied on each window. The available number of sample is high, but it contains samples which are used in the test sets of the different windows. In the second approach, the meta-parameters are estimated for each window, and varies between different windows. It does not contain sample from the test sets, but the number of samples available for the estimation is limited.

4.4 Numerical Results

4.4.1 With meta-parameters Optimization on the Whole Datasets

In this section, the optimization of the meta-parameters on the whole datasets is detailed.



(a) Dataset 1

(b) Dataset 2

(c) Dataset 3

(d) Dataset 4

Figure 6: Eigenvalues distribution (in log scale) for the four datatsets.

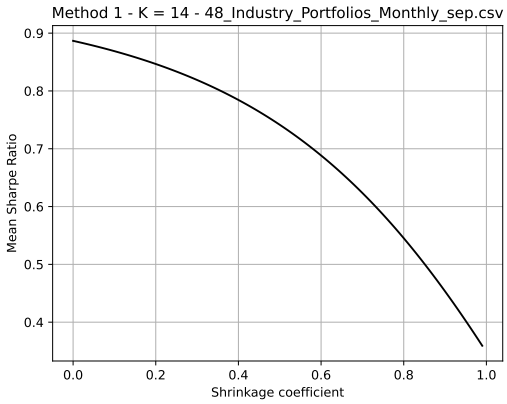
	Dataset 1	**Dataset 2**	**Dataset 3**	**Dataset 4**
N	48	100	48	100
K **(PV)**	14	6	11	3
K **(MP)**	2	3	2	3

Table 1: Estimation of the optimal K values using the percentage of explained variance (PV) and Marchenko-Pastur (MP) methods, on the four datasets.

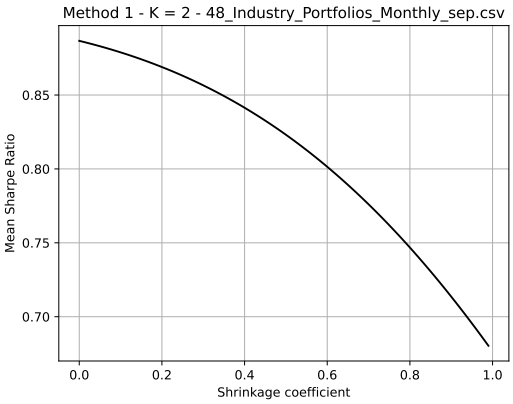
Table 1 summarizes the optimal K values computed on the whole four datasets with the percentage of explained variance (PV) and Marchenk-Pastur (MP) methods, and Figure 6 illustrates the eigenvalues distributions and the upper bound λ_+ of the MP distribution. The former compute K in order to preserve 85% of the total variance of the initial sample covariance matrix, while the latter keeps only the eigenvalues higher than the upper bound λ_+.

First, looking at the distribution of the eigenvalues in the four datasets, it can be observed than most of the eigenvalues are grouped around small values, while a limited

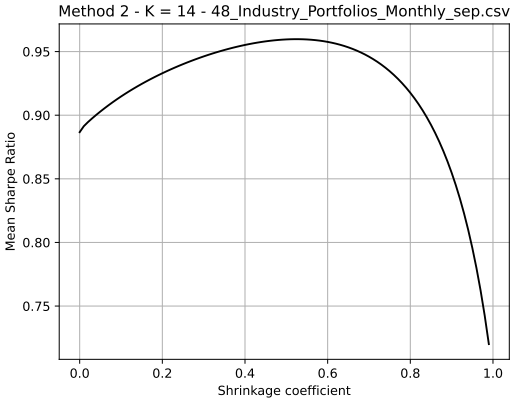
number of eigenvalues are present in large values. Therefore, low K values are obtained with the MP method. The PV method selects instead higher K values since multiple small eigenvalues should be summed in order to fulfill the total variance criterion. Note that the K value is not correlated to N: even if the number of assets increases, the number of principal component needed to represent the dataset does not increase if the new assets are correlated with the dataset.



(a) $K = 14$, method 1

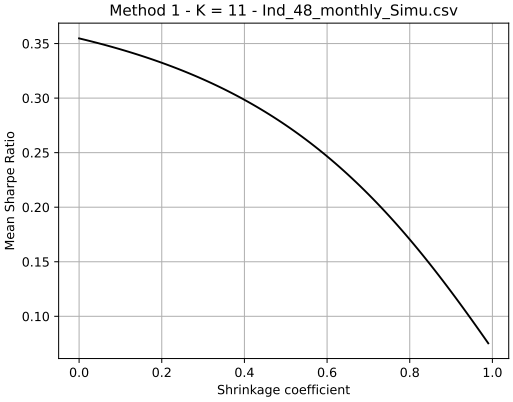


(b) $K = 2$, method 1

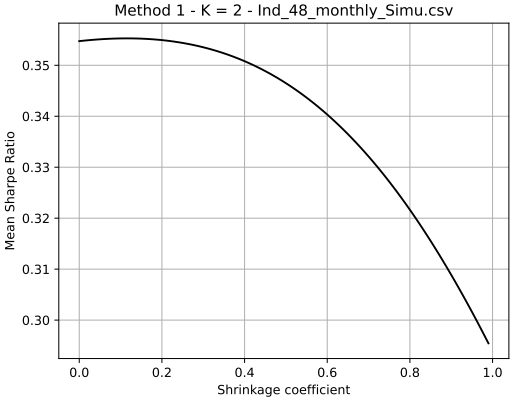


(c) $K = \{14, 2\}$, method 2

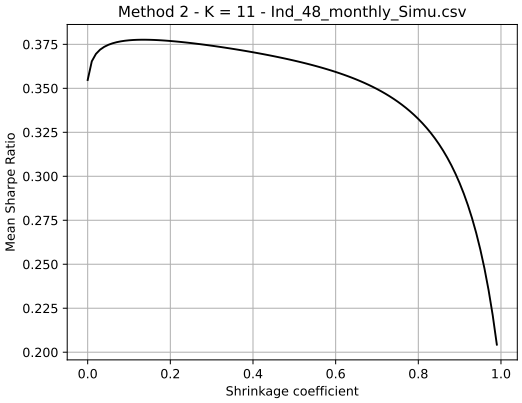
Figure 7: Mean Sharpe Ratio over the 10 folds for the two new methods with different K values on Dataset 1, in order to optimize the shrinkage value.



(a) $K = 11$, method 1

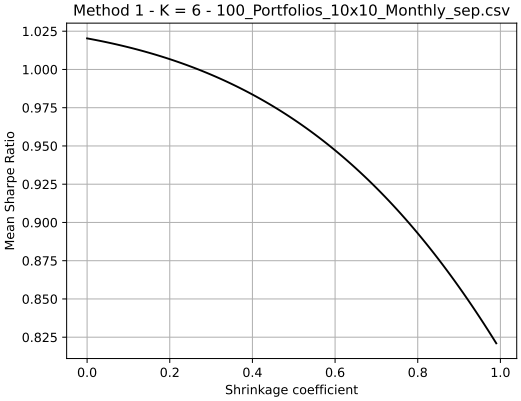


(b) $K = 2$, method 1

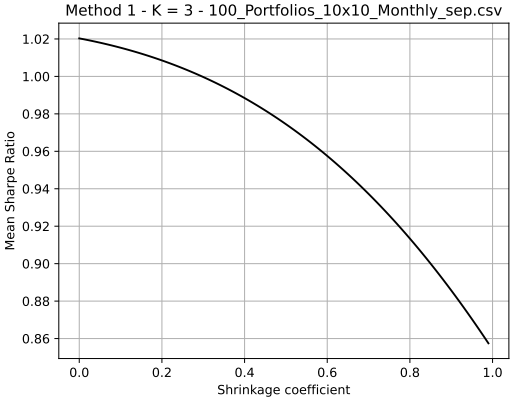


(c) $K = \{11, 2\}$, method 2

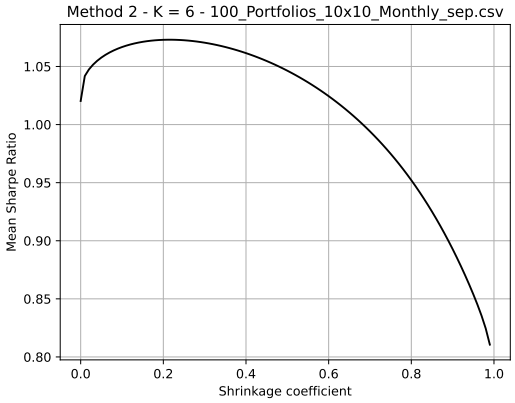
Figure 8: Mean Sharpe Ratio over the 10 folds for the two new methods with different K values on Dataset 2, in order to optimize the shrinkage value.



(a) $K = 6$, method 1

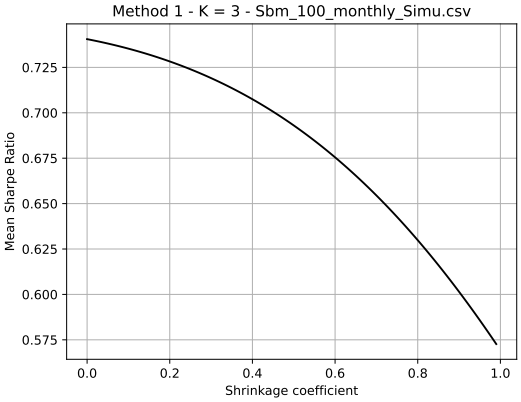


(b) $K = 3$, method 1

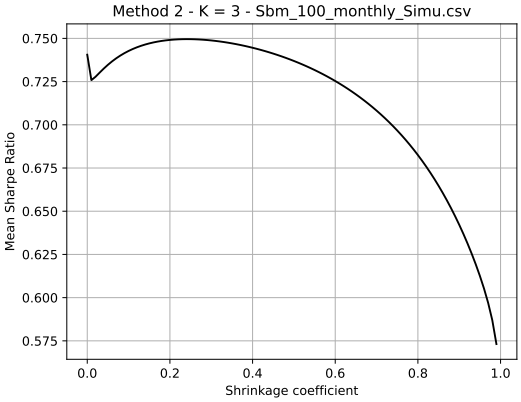


(c) $K = \{6, 3\}$, method 2

Figure 9: Mean Sharpe Ratio over the 10 folds for the two new methods with different K values on Dataset 3, in order to optimize the shrinkage value.



(a) $K = 3$, method 1



(b) $K = 3$, method 2

Figure 10: Mean Sharpe Ratio over the 10 folds for the two new methods with different K values on Dataset 4, in order to optimize the shrinkage value.

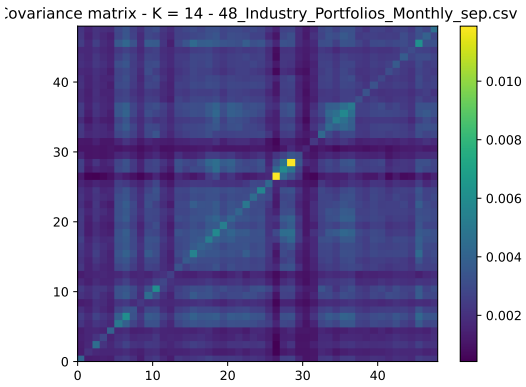
Dataset	1		2		3		4
	$K = 14$	$K = 2$	$K = 11$	$K = 2$	$K = 6$	$K = 3$	$K = 3$
Method 1	0	0	0	0.12	0	0	0
Method 2	0.52	0.52	0.13	0.13	0.22	0.22	0.24

Table 2: Estimations of the optimal shrinkage coefficient values using the percentage variance and Marchenko-Pastur methods, on the four Datasets.

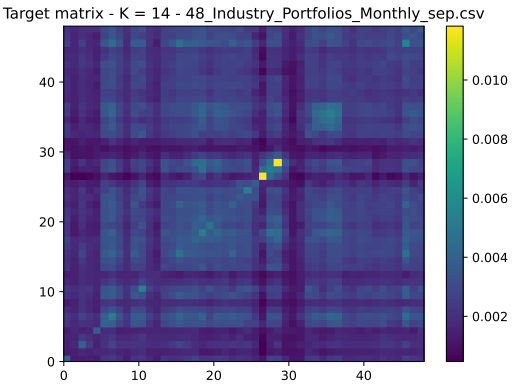
Table 2 illustrates the optimal shrinkage coefficients computed for the K values obtained with both the PV and MP methods on the four datasets. Figure 7 to 10 illustrates the shape of the mean sharpe ratio over the 10 folds evaluated for different shrinkage coefficients values. As a remainder, the first method uses the covariance matrix of the back-projected samples as a target matrix $\mathbf{F}$, while the second method uses the matrix of the eigenvalues of the sample covariance matrix.

First, it can be observed that the first method nearly always proposes a shrinkage coefficient of 0, meaning that no shrinkage is performed and the target matrix is ignored. Thus, it is already possible to anticipate that the achieved performance with this strategy will be similar to the one achieved with the minimum variance strategy. Even if the matrix $\mathbf{F}$ has a rank of $K << N$ since is generated from the back-projected samples of the initial dataset, as illustrated in Figure 11 and 12. it is still composed of a lot of non-zero coefficients with a structure close to the one of the sample covariance matrix. However, the diagonal terms weakens. Note that the low rank of the target matrix leads to numeric instabilities when the shrinkage coefficient is close to 1, since the shrunk matrix is inverted to compute the optimal weights.

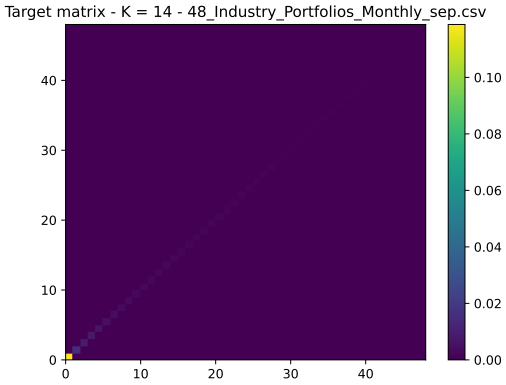
Contrariwise, the second method leads to various shrinkage coefficients depending on the studied dataset. There is no difference whatever the K value since there is no dimension reduction in this method. Figure 11 and 12 illustrates the shape of the of the target and shrunk matrix. The target matrix is highly structure since it only contains diagonal terms, i.e. the eigenvalues of the sample covariance matrix. After the shrinkage operation, new terms are integrated on the matrix, but the diagonal terms stays dominant.



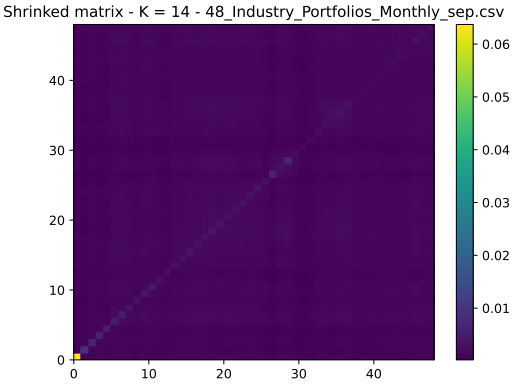
(a) Sample covariance matrix



(b) Target matrix, method 1

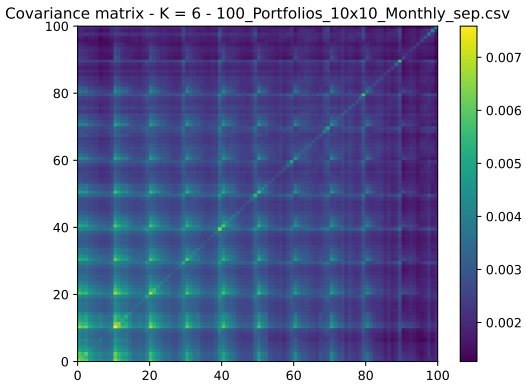


(c) Target matrix, method 2

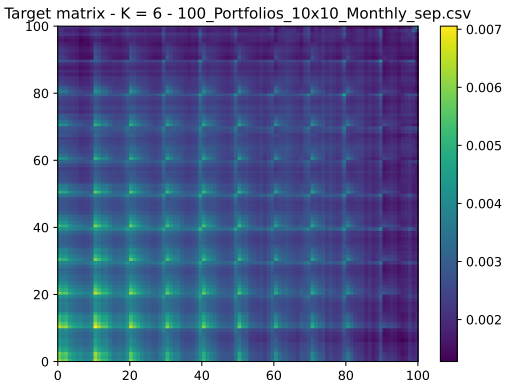


(d) Shrunk matrix, method 2

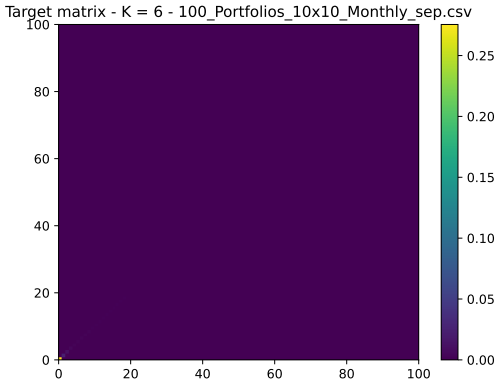
Figure 11: Covariance matrix, target matrix for both methods and Shrunk matrix for method 2 with Dataset 1 and $K = 14$.



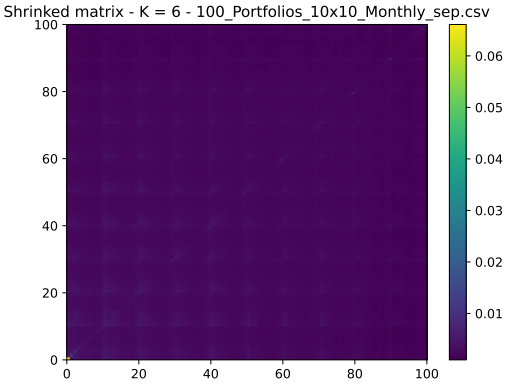
(a) Sample covariance matrix



(b) Target matrix, method 1



(c) Target matrix, method 2



(d) Shrunk matrix, method 2

Figure 12: Covariance matrix, target matrix for both methods and Shrunk matrix for method 2 with Dataset 3 and $K = 6$.

Sharpe Ratios	**1**		**2**		**3**		**4**
	$K = 14$	$K = 2$	$K = 11$	$K = 2$	$K = 6$	$K = 3$	$K = 3$
MSR In-Sample	1.5033		0.6292		2.4994		2.3170
MSR Out-Of-Sample	0.36		0.038		-0.163		-0.003
MinV In-Sample	1.0135		0.5381		1.3366		1.0091
MinV Out-Of-Sample	0.732		0.181		0.828		0.345
MinV with PCA	**1.104**	0.76	**0.456**	0.289	1.016	0.977	0.657
LW	0.961		0.335		**1.406**		**0.792**
Method 1	0.732	0.732	0.181	0.164	0.828	0.828	0.345
Method 2	0.98		0.297		1.11		0.699

Table 3: Mean sharpe ratios computed using all the windows and precomputed meta-parameters, with the percentage variance and Marchenko-Pastur methods for the selection of K, and on the four Datasets.

Table 3 summarizes the achieved mean sharpe ratio with the 6 presented methods

(i.e. the maximum sharpe-ratio (MSR), minimum variance (MV), minimum variance with PCA, Ledoit & Wolf (LW) and the two new methods), using the meta-parameters optimized previously in the four datasets.

A first observation on the performance achieved with the in- and out-of -sample sets of the MSR and MV strategies shows that the MSR (resp. MV) achieve better results than the MV (resp. MSR) for the in-sample (resp. out-of-sample) set. The first strategy looks for the efficient frontier and is therefore optimizing the sharpe ratio, leading to data overfitting. The second strategy aims to minimize the variance of the returns. This is therefore less sensitive to shocks and has better results out-of-sample. Note that this also leads to smaller difference between the achieved mean sharpe-ratio of the in- and out-of -sample sets.

Then, one can focus on the performance achieved when PCA are introduced in the minimum-variance strategy. In a first time, the introduction of PCA in the strategy increases in a significant manner the achieved performance. It confirms that the dimension reduction helps to limit the impairments introduced by the estimation of the covariance matrix. Additionally, the performance achieved with the K values estimated through the PV method are better than the ones achieved with the MP method. It seems that the MP method is too restrictive regarding the number of principal components associated with useful signals, while the PV method uses in most of the case more components.

Next, the shrinkage processed in the LW method gives always better performance than the MSR and MV methods. However, on these datasets, when the number of selected component is high (datasets 1 and 2), the LW method is outperformed by the MV with PCA method. In the other datasets, the number of principal components seems to limited, leading to information losses and decreasing the achieved performance.

Finally, the first proposed method achieves obviously similar performance than the minimum-variance portfolio since the shrinkage coefficients have been optimized to the zero value. With this value, the target matrix is not used at all and no shrinkage is performed. The only case where it is not the case shows that the performance of the method are worse than the performance achieved with the MV method. This translates probably an issue in the estimation of the shrinkage coefficient for this dataset. Regarding the second method, it performs a little bit less well than the LW method. This is either related to the choice of the shrinkage coefficient, done previously on the whole dataset, while the LW method computes it analytically. Despite that, the achieved results are better than the ones achieved with the MSR and MV methods. The eigenvalue matrix seems to be a decent choice as a target matrix, since it contains few coefficients translating information about the sample set.

4.4.2 Without meta-parameters Optimization on the Whole Datasets

In this section, the meta-parameters are not computed beforehand, but they are optimized for each window using the training samples before computing the weights.

Dataset	1		2		3		4	
	PV	**MP**	**PV**	**MP**	**PV**	**MP**	**PV**	**MP**
MSR	0.36		0.038		-0.163		-0.003	
MinV	0.732		0.181		0.828		0.345	
MinV with PCA	**1.106**	0.792	**0.437**	0.267	1.095	0.841	0.639	0.653
LW	0.961		0.335		**1.406**		**0.792**	
Method 1	0.731	0.732	0.181	0.129	0.827	0.828	0.34	0.341
Method 2	0.925		0.259		0.981		0.676	

Table 4: Mean sharpe ratios computed using all the windows and precomputed meta-parameters, with the percentage variance and Marchenko-Pastur methods for the selection of K, and on the four Datasets. Green (resp. red) numbers illustrates increases (resp. decreases) compared to Table 3.

Table 4 shows the mean sharpe ratio achieved without optimization of the meta-parameters beforehand, for the four datasets. Therefore, the meta-parameters are now varying from window to window. In most of the cases, the achieved performance are degraded compared to 3. This can be explained by two reasons: first, when optimization of the meta-parameters is done beforehand, the whole dataset is used. Therefore, samples included in the test windows have been already used to train the models, leading to increased performance. Then, using 10-fold cross validation into the rolling window leads to only 12 samples (one year) to estimate the meta-parameters. However, in some case, it helps to improve the performance achieved using the MV with PCA method. In these case, a better choice of K is made, leading to the good trade-off between dimension reduction and information loss. Regarding the first new method, the variations are negligible and are provoked by non-zero shrinkage coefficients, giving worse performance than when the shrinkage coefficient is set to zero.

5 Conclusion

In order to compute portfolio strategies, the estimation of the returns data statistics (mean and covariance) is crucial. With standard estimators, the portfolio strategies breaks down when the number of assets is too large.

In this work, multiple portfolio strategies have been presented to overcome this problem. More specifically, a particular focus has been made on shrinkage-based methods, and especially the Ledoit & Wolf strategy. In another hand, the introduction of PCA for dimension reduction of the sample set has shown excellent results to get around the problem. Therefore, two new methods involving shrinkage and PCA have been proposed. In the former, the covariance of the back-projection of the sample set after dimension reduction is used as a target matrix. In the latter, the matrix of the eigenvalues of the sample set covariance matrix is used as a target matrix.

Finally, the performance of all these strategies have been compared and discussed, after being applied on four datasets, whose two where generated artificially. In order to assess the performance, the mean sharpe ratio has been used as the main metric, and a rolling window has been performed. The meta-parameters have been optimized using 10-fold cross validation. Either the meta-parameters have been computed beforehand using the whole dataset, or they are computed in each window using the training set of the window. Unfortunately, the results have shown that the first method does not work, since it always suggests to only use the sample covariance matrix and avoid any shrinkage. However, the second method has shown good performance, similar to the ones achieved with the initial Ledoit & Wolf strategy.

In future works, the first method should be further studied. First, the choice of the target matrix could be improved in order to achieve an efficient shrinkage. Then, the choice of the optimal shrinkage coefficient could be improved, either through another optimization method or analytically.

6 Bibliography

References

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

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

[3] Jagannathan, R., & Ma, T. (2003). Risk Reduction in Large Portfolios : Why Imposing the Wrong Constraints Helps. *The Journal of Finance, 58*(4), 1651-1683. https://doi.org/10.1111/1540-6261.00580

[4] Ledoit, O., & Wolf, M. (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

[5] Ledoit, O., & Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance, 10*(5), 603-621. https://doi.org/10.1016/s0927-5398(03)00007-0

[6] Ledoit, O., & Wolf, M. (2004b). Honey, I Shrunk the Sample Covariance Matrix. *The Journal of Portfolio Management, 30*(4), 110-119. https://doi.org/10.3905/jpm.2004.110

[7] Michaud, R. O. (1989). The Markowitz Optimization Enigma : Is 'Optimized' Optimal ? *Financial Analysts Journal, 45*(1), 31-42. https://doi.org/10.2469/faj.v45.n1.31

[8] DeMiguel, V., & Nogales, F. J. (2009). Portfolio Selection with Robust Estimation. *Operations Research, 57*(3), 560-577. https://doi.org/10.1287/opre.1080.0566

[9] Rajaratnam, B., & Vincenzi, D. (2016). A theoretical study of Stein's covariance estimator. *Biometrika, 103*(3), 653-666. https://doi.org/10.1093/biomet/asw030

[10] Merton, R. C. (1980). On estimating the expected return on the market. *Journal of Financial Economics, 8*(4), 323-361. https://doi.org/10.1016/0304-405x(80)90007-0

[11] Jobson, J. D., & Korkie, B. (1980). Estimation for Markowitz Efficient Portfolios. *Journal of the American Statistical Association, 75*(371), 544-554. https://doi.org/10.1080/01621459.1980.10477507

[12] Devlin, S. J., Gnanadesikan, R., & Kettenring, J. R. (1981). Robust Estimation of Dispersion Matrices and Principal Components. *Journal of the American Statistical Association, 76*(374), 354-362. https://doi.org/10.1080/01621459.1981.10477654

[13] Huber, P. J., & Ronchetti, E. M. (2009). Robust Statistics. *Wiley Series in Probability and Statistics.* https://doi.org/10.1002/9780470434697

[14] Frost, P. A., & Savarino, J. E. (1986). An Empirical Bayes Approach to Efficient Portfolio Selection. *The Journal of Financial and Quantitative Analysis, 21*(3), 293. https://doi.org/10.2307/2331043

[15] Jorion, P. (1986). Bayes-Stein Estimation for Portfolio Analysis. *The Journal of Financial and Quantitative Analysis*, 21(3), 279. https://doi.org/10.2307/2331042

[16] Lorenzo-Seva, U. (2013). *How to report the percentage of explained common variance in exploratory factor analysis.* Technical Report. Department of Psychology, Universitat Rovira i Virgili, Tarragona.

[17] Yaskov, P. (2016). A short proof of the Marchenko–Pastur theorem. *Comptes Rendus Mathematique, 354*(3), 319-322. https://doi.org/10.1016/j.crma.2015.12.008

[18] Marčenko, V. A., & Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik, 1*(4), 457-483. https://doi.org/10.1070/sm1967v001n04abeh001994

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

[20] Hastie, T., Tibshirani, R., & Friedman, J. (2022). *The Elements of Statistical Learning : Data Mining, Inference, and Prediction, Second Edition (Springer Series in Statistics)* (2nd éd.). Springer.

[21] Chen, J. and Yuan, M. (2016), Efficient portfolio selection in a large market, *Journal of Financial Econometrics, 14*(3), 496–524.

[22] Law, J., & Jolliffe, I. T. (1987). Principal Component Analysis. *The Statistician, 36*(4), 432. https://doi.org/10.2307/2348864

[23] Ren, Q., Li, M., & Han, S. (2019). Tectonic discrimination of olivine in basalt using data mining techniques based on major elements : a comparative study from multiple perspectives. *Big Earth Data, 3*(1), 8-25. https://doi.org/10.1080/20964471.2019.1572452

[24] Hull, J. C. (2018). *Risk Management and Financial Institutions (Wiley Finance)* (5e éd.). Wiley.

[25] Wikipedia contributors. (2022, 20 mai). *Marchenko–Pastur distribution.* Wikipedia. https://en.wikipedia.org/wiki/Marchenko

[26] Stein, C. (1956), Inadmissibility of the usual estimator for the mean of a multivariate distribution, *Proc. Third Berkeley Symp. Math. Statist. Prob.*, vol. 1, pp. 197–206