UCLouvain

LOUVAIN
School of Management

Louvain School of Management

Portfolio Selection via Independent Component Analysis

Auteur : Emmanuel Masson
Promoteur : Frédéric Vrins
Lecteur: Nathan Lassance
Année académique 2018-2019

"A second reason why science cannot replace judgement is the behavior of financial markets."

Martin Feldstein

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

Abstract

Louvain School of Management

Master in Business Engineering

Portfolio Selection via Independent Component Analysis

by Emmanuel MASSON

Different authors have remarked that the first two moments of the return distribution do not adequately summarize the investor's preferences. However, higher-order moment portfolios perform poorly out-of-sample due to the estimation error and the curse of dimensionality. To circumvent this problem, we rely on minimum Modified Value-at-Risk portfolio based on independent components (ICs) which are the rotation of the principal components (PCs) that are independent as possible. On one hand, the ICs are useful for dimension reduction by retaining only the first high variance factors. On the other hand, we harness the near independence of the ICs to parsimoniously estimate the higher-order moments. Finally, out-of-sample studies reveal that minimizing the portfolio MVaR with this parsimonious estimation of the higher-order moments on the ICs yields solid financial performance in terms of higher-order risk. However, this strategy generally tends to generate higher turnover and less performing mean-variance trade-off.

Acknowledgements

Throughout the writing of this master's thesis I have received a great deal of support and assistance. First of all, I would like to thanks my thesis advisor, Pr. Frédéric Vrins for the valuable guidance and the support during my learning process. He gave me the opportunity to work on this exciting and complex field of portfolio selection.

I am particularly grateful for the assistance given by Mr. Nathan Lassance for his guidance that helped me in all the time of research and writing of this thesis. His constructive advice contributed a lot for the betterment of this thesis.

My acknowledgement would be incomplete without thanking my family and closest friends who contributed to making those university years in the best of my life.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Nowadays, asset management financial institutions play a key role in the modern economy. They help to finance new investment projects and put investors' savings at the service of the economy by investing their money in companies, governments and projects, with the aim of generating returns for millions of savers and investors. In 2017, total assets under management were $37.4 trillion (2007 - $24.1 trillion) in North America and $22.2 trillion (2007 - $14.3 trillion) in Europe[1]. In light of these figures, we cannot afford not to fully understand how to allocate these assets in an optimal way and for a corresponding certain level of risk. It is in this context that our work is focused by developing efficient asset allocation models for practitioners and academics.

In 1952, Harry Markotwitz paved the foundation of the modern portfolio theory in his article *Portfolio Selection* (Markowitz, 1952). He defined the optimal portfolio as the one that maximizes returns for a given risk level. Particularly thanks to its ease of implementation, the mean-variance optimization has quickly become the theoretical and industrial standard at that time. Traditionally, the sample mean and the sample covariance matrix have been used to implement these portfolio in practice. However, because of estimation error, portfolio selected using these estimators lead to concentration, high sensitivity of extreme positions and unstable portfolio, see for instance Michaud, 1989; Chopra and Ziemba, 1993; Broadie, 1993. For these reasons, the popularity of MV portfolios has been significantly reduced among portfolio managers, who were cautious about implementing such a strategy that requires significant change in the portfolio's composition. For the above reasons, the minimum-variance portfolio has demonstrated superior performance because it relies solely on estimates of the covariance matrix and perfoms usually better out-of-sample (Jagannathan & Ma, 2003). However, even this model is not perfect and remains vulnerable to the risk of estimation (Jagannathan & Ma, 2003).

[1] From Statista.com in *Global assets under management in selected years from 2007 to 2017, by region*

Moreover, researchers took a deeper look and remarked that the first two moments of the return distribution do not adequately summarize the investor's preferences (Martellini & Ziemann, 2010a). With this in mind, the question is how to take advantage of asset return time series to properly estimate the first fourth moments of the return distribution. In the literature, several avenues of research have been explored, including: shrinkage estimation (Ledoit & Wolf, 2001; James & Stein, 1961; Martellini & Ziemann, 2010b), robust estimation (Demiguel & Nogales, 2009), and factor model (Xu, Caramanis, & Mannor, 2016; Tan, 2012; Lassance, DeMiguel, & Vrins, 2018).

To circumvent this risk estimation issue, we study in this thesis a higher-moment minimum risk portfolio via Independent Component Analysis (ICA). This well-known machine learning technique takes a set of measured signal vector and extracts from them a new set of statistically vectors as independent as possible. In our context, we assume that the asset returns consist of a linear combination of independent factors that are provided by the ICA. Using ICA in portfolio selection offers two interesting features when estimating superior moments. On one hand, ICA performs dimensionality-reduction at the whitening step by selecting only K first independent factors. Selecting a limited number of factors is sufficient to correctly reflect the cross-section of asset returns (Xu et al., 2016). On the other hand, if we assume that the factors determined by the ICA are totally independent, all co-skewness and co-kurtosis don't need to be estimated because they vanish to zero.

Following the presentation of the context, academic and practical relevance of the subject, we identify a main research question in this thesis:

Does the strategy of reducing the number of factors selected and considering higher co-moments as zero reduce the risk of estimation and outperform out-of-sample a strategy where no dimension reduction has been applied and where all higher-moments are part of the estimation ?

More precisely, we compare the out-of-sample performance of five different policies:

1. Traditional minimum-variance portfolio (*MV*). This is a strategy that returns the lowest risk where the only estimation done is the sample covariance matrix as measure of risk.

2. Minimum MVaR portfolio (*MVaR*). This is a strategy that returns the lowest MVaR as measure of risk.

3. Minimum MVaR portfolio on principal components (*PCMVaR*). This is a strategy that returns the lowest MVaR on the K first principal components.

4. Minimum MVaR portfolio on independent components (*ICMVaR*). This is a strategy that returns the lowest MVaR on the K first independent components.

5. Minimum MVaR portfolio on independent components (*ICMVaRCM*). This is a strategy that returns the lowest MVaR on the K first independent components and where all third and fourth co-moments are considered to be equal to zero.

The purpose of this thesis is to clarify the theortical background, provide an answer to the research question and present empirical performance on both simulated and real data. This thesis is structured along the following chapters:

- Chapter 2 introduces the mean-variance paradigm, the minimum-variance portfolio and the higher-moment risk minimization portfolio with Modified Value-at-Risk as definition of risk.
- Chapter 3 presents the theoretical framework behind Principal Component Analysis and Independent Component Analysis.
- Chapter 4 details the risk minimization strategies based on the principal components and on the independent components.
- Chapter 5 presents both simulated and real data used as well as the methodology followed.
- Chapter 6 contains a presentation and a discussion of the results for the five different strategies considered in this thesis.
- Chapter 7 summarizes the main results and its implications, presents the research limitations and propose future avenues of research.

Chapter 2

Risk minimization portfolio selection

In this chapter, we present the main statistical concepts, the mean-variance paradigm ans its limitations, the minimum-variance portfolio and the minimum Modified Value-at-Risk portfolio.

2.1 Central moment, skewness & kurtosis

A central moment is a moment of a probability distribution of a random variable Z about the random variable's mean $\mu(Z)$. In other words, it is the expected value of a specified integer power of the deviation of the random variable from the mean. The kth central moment is equal to:

$$m_k(Z) = \mathbb{E}((Z - \mu(Z))^k) \tag{2.1}$$

The skewness $\gamma(Z)$ and kurtosis $\kappa(Z)$ are respectively the third and fourth standardized moments:

$$\gamma(Z) = \frac{m_3(Z)}{m_2(Z)^{3/2}} = \mathbb{E}\left[\left(\frac{Z - \mu(Z)}{\sigma(Z)}\right)^3\right] \tag{2.2}$$

$$K(Z) = \frac{m_4(Z)}{m_2(Z)^2} = \mathbb{E}\left[\left(\frac{Z - \mu(Z)}{\sigma(Z)}\right)^4\right] \tag{2.3}$$

where $\sigma(Z)$ is the standard deviation of the random variable Z. On the one hand, the skewness measures the degree of asymmetry of the distribution. Negative skew commonly indicates that the tail is on the left side of the distribution, and positive skew indicates that the tail is on the right. On the other hand, the kurtosis measures the tailedness of the distribution. For Gaussian distribution, we have $\gamma(Z) = 0$ and

$K(Z) = 3$. Generally, we rely on the excess kurtosis κ that equals to 0 for Gaussian distribution $\kappa(Z) = K(Z) - 3 = 0$.

Generally, asset returns have negative value for skewness i.e. a long tail to the left and so more downside risk than the Normal distribution and positive excess kurtosis i.e fatter tails than the Normal. These concepts are necessary in the next sections in order to select optimal portfolio based on higher-moment.

2.2 Portfolio moments

We consider a random asset-return vector $X = (X_1, ..., X_n) \in \mathbb{R}^N$ of dimension $n \times 1$ where N is the number of assets available for investment.

Let $P := w'X$ be the portfolio return where $w = (w_1, ..., w_n)'$ is the $n \times 1$ vector portfolio weights. The first fourth moments of the portfolio are determined by:

$$m_1(P) = w'\mu_X \tag{2.4}$$

$$m_2(P) = w' \sum\nolimits_X w \tag{2.5}$$

$$m_3(P) = w'\Phi_X(w \otimes w) \tag{2.6}$$

$$m_4(P) = w'\Psi_X(w \otimes w \otimes w) \tag{2.7}$$

where μ_X is the sample expected return of X, $\sum_X$ is the sample covariance matrix, $\Phi_X \in \mathbb{R}^{N \times N^2}$ sample co-skewness of X, $\Psi_X \in \mathbb{R}^{N \times N^3}$ sample co-kurtosis matrices of X and $\otimes$ is the Kronecker product.

2.3 The mean-variance paradigm

In 1952, Harry Markotwitz paved the foundation of the modern portfolio theory in his article *Portfolio Selection* (Markowitz, 1952). The mean-variance paradigm is founded on the assumption that investors are rational and select optimal portfolio exclusively based on expected portfolio return and risk measured by portfolio variance. Hence, the optimal portfolio is the one that maximizes returns for a given risk level. More precisely, the solution is a set of efficient portfolio depending on the risk level. On one hand, the portfolios above the efficient frontier are not feasible. On the hand, a rational investor would not choose a portfolio below the efficient frontier because higher return can be obtained for the same level of risk. The effective frontier is generally represented graphically as a hyperbola in the mean variance space called the "Markowitz bullet".

Despite its simplicity of application, Markowitz model has his drawbacks. First of all, the determination of expected returns and the covariance matrix are crucial choices. The expected returns are typically over-estimated while the risk associated is under-estimated (Michaud, 1989). In addition, this model tends to determine portfolios composed of combinations of extreme longs and extreme shorts positions that are very sensitive to a change in the inputs (Chopra & Ziemba, 1993). More particularly, the portfolio composition tend to be more sensitive to a change in the expected returns compared to a change in the covariance matrix (Best & Grauer, 1991). Kallberg and Ziemba, 1984 showed that estimation errors in means are about ten times more important than estimation errors in the covariance matrix. These results lead in a unstable portfolio composition with a significant fluctuation in the portfolio weights over time. Finally, these unstable portfolio performs very poorly when they are tested out-of-sample (Michaud, 1989; K. Chopra & Ziemba, 2016; Broadie, 1993).

Demiguel, Garlappi, and Uppal, 2009 compare the out-of-sample performance of the sample-based mean-variance model and 13 derived strategies relative to the naive 1/N portfolio. The author found that any optimal strategy is consistently better than the $1/N$ naive diversification in terms of Sharpe ratio, return and turnover. He also showed that the estimation window needed for the classical Markowitz optimization to outperform the $1/N$ strategy is around 3000 months for a portfolio with 25 assets. The gain from optimal diversification was more than balanced by an error in estimation in that case.

2.4 Variance minimization portfolio

For the above reasons, the minimum-variance portfolio has demonstrated superior performance because it only requires the estimation of the covariance matrix and is consequently less sensitive to the estimation risk (Jagannathan & Ma, 2003). Conversely, the minimum variance portfolio has also its drawback and remains quite subject to estimation risk due to the sensitivity of estimators to deviations in the distribution of asset returns from the assumed Gaussian distribution (Demiguel & Nogales, 2009).

The first optimization model that we consider in this thesis is the portfolio that minimize the sample covariance matrix.

Definition 2.4.1. The minimum variance portfolio (MV) is defined as the solution to the following optimization program:

$$w := \operatorname{argmin} m_2(P) \text{ subject to } \sum_{i=1}^{N} w_i = 1 \tag{2.8}$$

where w are the portfolio weights with least return portfolio variance out of all the portfolios.

In the presence of short-selling, there exists a closed form of the solution to this optimization problem (in contrary in absence of short-selling where optimal solutions are obtained via Quadratic Optimization). The optimal weights w of MV are given by (Kempf & Memmel, 2006):

$$w = \frac{\Sigma_X^{-1}\mathbb{1}}{\mathbb{1}\Sigma_X^{-1}\mathbb{1}} \tag{2.9}$$

2.5 Modified Value-at-Risk definition

Vaue-at-Risk (VaR) represents the maximum potential loss on a specific portfolio of financial assets given a specific time horizon and a confidence interval. While the traditional parametric VaR computation is limited at the second-order moment, the Modified Value-at-Risk (MVaR) incorporates higher moments. Introduced by Favre and Galeano, 2002, MVaR is based on a Cornish-Fisher expansion (Cornish & Fisher, 1938) that is a method to transform a standard Gaussian random variable into a non Gaussian random variable by taking into account superior moments in its computation.

Definition 2.5.1. The Modified VaR of the portfolio return with finite fourth central moments $P \in \mathcal{L}^4$ is defined from a second order Cornish-Fisher expansion of its quantile function and is given by

$$MVaR_P(\alpha) := -\mu(P) - \sqrt{m_2(P)}Z \tag{2.10}$$

with

$$Z := z_\alpha + \frac{1}{6}(z_\alpha^2 - 1)s(P) + \frac{1}{24}(z_\alpha^3 - 3z_\alpha)\kappa(P) - \frac{1}{36}(2z_\alpha^3 - 5z_\alpha)s(P)^2 \tag{2.11}$$

where $z_\alpha = \Phi^{-1}(\alpha)$ is the Gaussian quantile at the confidence level α.

Even if the Cornish-Fisher expansion is truncated at the second-order term, Baillie and Bollerslev, 1992 shows that the incorporation of additional terms does not improve the approximation much because of the increase of the number of parameters to estimate for large asset portfolios. This is why in Chapter 3, we follow the same approach as Xu et al., 2016 by relying on the principal components and on the independent components because it reduces considerably the number of parameters to be estimated.

Note that a value of alpha above 4,16% should never be used for the MVaR computation to correctly reflect investors preferences for kurtosis. Failure to comply with this restrictions on confidence levels results in an incorrect risk assessment and a potential overweighting of assets that have undesirable properties in terms of higher moments (Cavenaile & Lejeune, 2010).

2.6 MVaR minimization portfolio

The second optimization model that we consider here is the portfolio that minimize the Modified Value-at-Risk as definition of risk.

Definition 2.6.1. The MVaR minimum portfolio (*MVaR*) for a given α is defined as the solution to the following optimization program:

$$w^{\gamma} := \text{ argmin } MVaR(\alpha) \text{ subject to } \sum_{i=1}^{N} w_i^{\gamma} = 1 \tag{2.12}$$

where w^{γ} are the portfolio weights with least return MVaR out of all the portfolios and where the first fourth portfolio moments are respectively $m_1(P)$, $m_2(P)$, $m_3(P)$ and $m_4(P)$.

Chapter 3

PCA and ICA: from decorrelation to independence

In this chapter, we first discuss Principal Component Analysis (PCA), a well-known dimension-reduction technique that converts a set of observations of potentially correlated variables into a set of values of linearly non-correlated variables. We point out that Principal Components (PCs) do not set any constraint on dependence with respect to higher-order moment. For this reason we opt for Independent Component Analysis (ICA), a machine learning technique that rotates the PCs to make them as independent as possible under some conditions.

3.1 The *PC basis*

Principal Component Analysis (PCA) is a statistical method that is based on an orthogonal transformation to convert a set of observations of potentially correlated variables into a set of values of linearly non-correlated variables called Principal Components; see Section 6.4 of Campbell, Lo, and MacKinlay, 1997 or Section 6 of Aapo Hyvärinen, Karhunen, and Oja, 2001 for a good reference on the subject.

Before to detail Principal Component Analysis theory, we should define two important concepts in this chapter: rotation matrix and white basis.

Definition 3.1.1. A *rotation matrix* is a matrix that is used to perform a rotation in Euclidean space. The matrix $R \in \mathbb{R}^{N \times N}$ is a rotation matrix if and only if (Aapo Hyvärinen et al., 2001):

- The determinant of a rotation matrix is equal to 1 i.e. $det(R) = 1$
- The transpose of a rotation matrix is its inverse i.e. $RR' = \mathbb{1}_N$

where $\mathbb{1}_N$ is the identity matrix $\mathbb{1}_N \in \mathbb{R}^{N \times N}$

Definition 3.1.2. A basis $Y \in \mathbb{R}^{N\times N}$ is said to be *white* if its element are uncorrelated and have unit variances. In terms of the covariance matrix it means that it has a unit covariance matrix $\sum_Y = \mathbb{1}_N$ (Aapo Hyvärinen et al., 2001).

Let consider now a random asset-return vector $X = (X_1, ..., X_n) \in \mathbb{R}^N$ and his covariance matrix $\sum_X \in \mathbb{R}^{N\times N}$. Based on PCA, the covariance matrix can be diagonalized as:

$$\sum\nolimits_X = V\Lambda V' \tag{3.1}$$

where $\Lambda := diag(\lambda_1, ..., \lambda_N) \in \mathbb{R}^{N\times N}$ is the diagonal matrix containing the strictly positive eigenvalues sorted in decreasing order and where the columns of $V \in \mathbb{R}^{N\times N}$ are the eigenvectors of the covariance matrix.

Definition 3.1.3. The *PC basis* $Y^* \in \mathbb{R}^{N\times N}$ is the basis given by the orthogonal linear transformation of asset returns X (Lassance et al., 2018):

$$Y^* := \Lambda^{-1/2}V'X \tag{3.2}$$

The principal components forms the *PC Basis* and are standardized to have unit variance. Notice that a big advantage of relying on PC basis is that solutions of Equation 3.2 is a closed form.

The *PC Basis* is usefull for dimensionality reduction optimization (Xu et al., 2016) by only selecting the first principal components $K \leq N$ to constitute the *PC basis*. This results in a decrease of estimation error when dealing with large assets portfolio. From then on, we set the new notations to $\Lambda := diag(\lambda_1, ..., \lambda_K) \in \mathbb{R}^{K\times K}$, $Y^* \in \mathbb{R}^{N\times N}$ and $V \in \mathbb{R}^{N\times K}$ in that case.

An important point is the determination of the number of components K to retain to determine optimal portfolios. On one hand, the number of parameters to be estimated increases significantly when an additional dimension is added. On the other hand, having too few components may not adequately summarize the cross-section of asset returns. In order to circumvent this problem, we use the method of Velicer, 1976 where the author's approach is to determine the optimal number of components from the matrix of partial correlations. The advantage of this method is that it can be applied to any type of component analysis and therefore would be appropriate for principal components based on asset returns. Let consider the assset-return correlation matrix after the K PC has been partialed out as:

$$R(K) = R_X - VV' \tag{3.3}$$

where $R_X \in \mathbb{R}^{N \times N}$ is the asset-return correlation. Then, we should select $K \in [2, N-1]$ that minimizes the quantity $\sum_{i=1}^{N} \sum_{i \neq j}^{N} R_{ij}(K)^2$.

From the definition in 3.2, the PC basis Y^* is a white basis with covariance matrix $\sum_{Y^*} = I_K$. Let consider the special orthonormal group of order K $SO(K)$, that is, the set of $K \times K$ rotation matrices. For any rotation matrix $R \in SO(K)$, Lassance et al., 2018 showed that[1] :

$$Y = RY^* \tag{3.4}$$

remains uncorrelated. In addition, the *PC basis* set any constraint on dependence with respect to higher-order moment. To circumvent this problems, a solution is to rely on Independent Component Analysis in order to go further than decorrelation and look for the orthonormal basis that is as independent as possible.

3.2 The *IC basis*

3.2.1 What is ICA ?

Independent Component Analysis is a well-known machine learning technique that is being used more and more by the scientific community, see for instance Aapo Hyvärinen et al., 2001 for introduction to the topic.

The technique provides a solution to the problem of Blind source separation (BSS). This problem consists of the separation of a set of source signals from a set of linear mixed signals, without the aid of information about the source signals or the mixing process. For illustration, a classical application of this problem is the cocktail party where a group of people talking at the same time. Imagine you have multiple microphones picking up mixed signals, but you want to isolate the speech of a single person. ICA is able to separate the individual sources by using observed signal if some conditions are met.

Independent Component Analysis refers to the process of taking a set of measured signal vector, X, and extracting from them a new set of statistically independent vectors, Y, called the Independent Components (IC).

[1] $\sum_Y = R \sum_{Y^*} R' = R I_K R' = RR' = I_K$

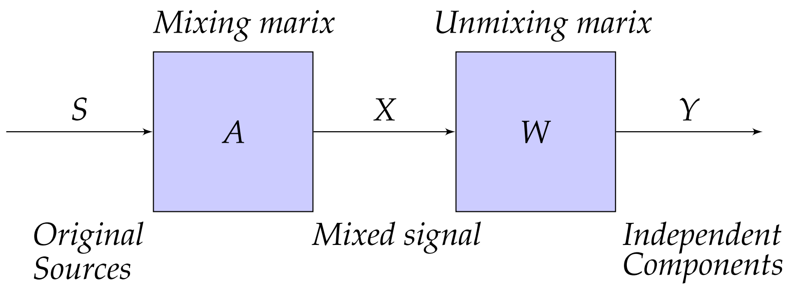

FIGURE 3.1: Shematic representation of ICA. The original independent sources S are mixed through a mixing matrix A. The IC can be retrieved by transforming the observed signal X thanks a demixing matrix W.

Figure 3.1 displays ICA process in a simplistic way. There are two mains process: mixing and demixing. The measured signal matrix X is the result of a mixing process such that

$$X = AS \tag{3.5}$$

where A is the unknown mixing matrix. The problem is to estimate a demixing matrix W such that

$$Y = WX \tag{3.6}$$

where Y are the IC. We will see later that, in general, is only possible to find a demixing matrix the IC are unique up to a change of sign and permutation.

As far as we know, only two authors have based their research on portfolio selection via ICA. Optimal *Portfolio Diversification via Independent Component Analysis* from Lassance et al., 2018 seems to be the most significant contribution on the subject. The authors mainly proposed a shrinkage portfolio that combines the minimum-variance portfolio and factor-risk-parity portfolio based on the independent components. They empirically showed that IC-based shrinkage portfolios outperform PC-based portfolios but also minimum risk portfolios obtained without shrinkage. The second article is *Portfolio selection with independent component analysis* from Hitaj, Mercuri, and Rroji, 2015. The authors' approach differs from that of the first paper as they estimate CARA utility function of an investor by assuming the independence of the IC factors.

3.2.2 *IC basis* definition

Independence goes beyond than decorrelation by asking independence for every transformation. In particular, independence does not require any dependence with regard to higher order moments. A measure of the independence level of a particular random vector is necessary in order to determine the white basis that is as independent as possible, that is, the *IC basis*. Such measure can be obtained thanks the concept of mutual information.

Definition 3.2.1. The *mutual information* $I(Y)$ of a random vector Y is the Kullback-Leibler divergence between its joint density and the corresponding product density :

$$I(Y) = \int f_Y(y) \frac{f_Y(y)}{\prod_{i=1}^{K} f_{Y_i}(y_i)} = \mathbb{E}\left(ln \frac{f_Y(Y)}{\prod_{i=1}^{K} f_{Y_i}(Y_i)} \right) \tag{3.7}$$

The Kullback-Leibler divergence can be considered as a kind of a distance between the two probability densities. Mutual information equals zero if and only if the components of Y are mutually independent and is always nonnegative $I(Y) \geq 0$ (Cover & Thomas, 2006). Moreover, the mutual information criterion can be standardized to be included between 0 and 1.

Definition 3.2.2. The *relative mutual information* $\bar{I}(Y)$ is the standardized mutual information to take values in [0,1] as follows:

$$\bar{I}(Y) = \frac{I(Y)}{\frac{1}{K}\sum_{i=1}^{K} h(Y_i)} \tag{3.8}$$

where

$$h(Y_i) := -\mathbb{E}(ln f_{Y_i}(Yi)) \tag{3.9}$$

is the differential entropy of Y_i.

Let X be a random asset-return vector with PC basis Y^*.

Definition 3.2.3. The *IC basis* $Y^\dagger \in \mathbb{R}^K$ is the rotation of the *PC basis* that minimizes mutual information (Lassance et al., 2018):

$$Y^\dagger := R^\dagger Y^*, R^\dagger \in \left\{ \underset{R \in SO(K)}{\text{argmin}} I(RY^*) \right\} \tag{3.10}$$

Unlike the *PC basis*, there is no closed form of the Equation 3.10. This basis is determined based on the Independent Component Analysis, introduced in 3.2.1.

Achieving perfect independence between IC factors is not common. In general, the mutual information of IC factors is strictly positive $I(Y^{\dagger}) \geq 0$. Indeed, *IC Basis* is computed based on the white basis that has the minimum dependence as measured from the mutual information. In other words, there are no specific reason to believe that the asset returns X can be decomposed as a linear combination of independent sources.

In addition, two important assumptions must be considered before opting for the *IC basis*:

1. As mutual information is invariant to permutation and to a change of sign of the independent components, the rotation matrix $R^{\dagger}$ is not unique. This means that the *IC basis* can be retrieved only up to a sign and a permutation indeterminacy.
2. At most one of the K principal components has a Normal distribution. This assumption is not a serious restriction when dealing with financial data given that asset returns are known to deviate from Normality. Because decorrelation and independence are equivalent concept for Normally random variable, ICA needs non-Normal distribution to discriminate the white basis that has the minimum mutual information among all the white basis.

3.2.3 *IC basis* computation

Independent Component Analysis is used to determine the *IC basis* i.e. the basis that minimizes mutual information among all the white bases. In contrast to *PC basis* that have a closed form via the PCA decomposition in 3.2, adaptive learning techniques are necessary to determine *IC basis*.

We will follow the same method as Lassance et al., 2018 in order to set up the framework for IC factors determination. More precisely, the Equation 3.10 simplifies greatly in our context:

$$Y^{\dagger} := R^{\dagger}Y^{*}, R^{\dagger} \in \left\{ \underset{R \in SO(K)}{\operatorname{argmin}} \sum_{i=1}^{K} h(R_i Y^{*}) \right\} \tag{3.11}$$

where R_i is the *ith* row of R and h(.) is the entropy defined in 3.9. In our context, the objective function defined in 3.10 collapses to the sum of individual entropies, which features only K one-dimensional distributions but not the K-dimensional joint density, see Vrins and Verleysen, 2005 for more information.

We use the approach of Hyvarinen, 1999 to implement Equation 3.11. A. Hyvärinen and Oja, 2000 have developed one of the most popular and efficient algorithm called *FactICA*[2]. This algorithm uses the Taylor expansion to estimate the marginal entropies $h(R_i Y^*)$:

$$h(Y) \approx h[\mathcal{N}(0,1)] - \frac{1}{2}\mathbb{E}(G(Y))^2 \tag{3.12}$$

where $h[\mathcal{N}(0,1)]$ is the entropy of a standard normal random variable and G is any smooth non-quadratic even function. *FactICA* then computes the matrix $R^\dagger$ using a gradient-descent scheme that shows at least quadratic local convergence. We follow the same approach as Lassance et al., 2018 by setting the function G to $G(x) = lncosh(x)$ as also suggested by Hyvarinen, 1999.

3.2.4 *IC basis* example

We will now illustrate the ICA with a hands-on example. Let consider two independent sources S_1 and S_2 simulated in a sinusoidal and uniform way respectively. These independent sources are mixed linearly with a random matrix A to obtain mixed signals $X = AY$. As any of these mixed signals follow a Gaussian distribution and have been obtained trough a linear mixture, ICA is able to retrieve the independent sources up to a change of sign and permutation i.e. mutual information of the independent sources is equal to zero $I(Y^\dagger) = 0$. Figure 3.2 displays the independent sources Y, the mixed signals X and the independent components estimated $Y^\dagger$. Note that in this case, the independent component have been permuted and that a change of sign have been applied on $Y_1^\dagger$.

[2] A **R** package implementing `FastICA` is available here: https://cran.r-project.org/web/packages/fastICA/fastICA.pdf

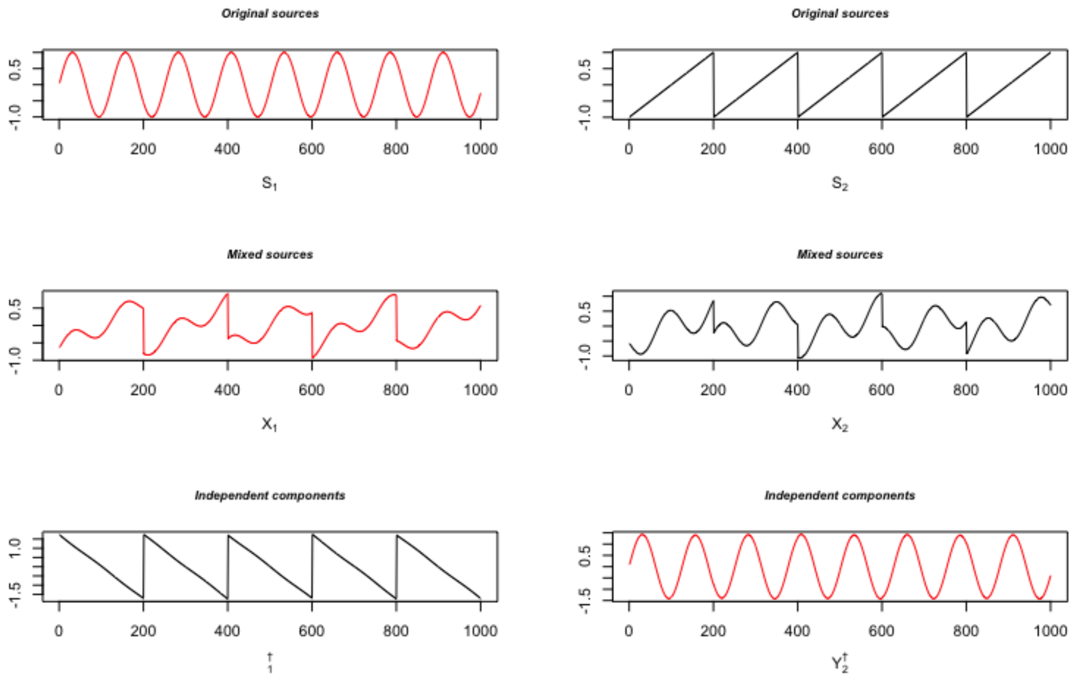

FIGURE 3.2: Theoretical example of ICA. The mixed signals X are obtained through a linear mixing of independent sources S. ICA estimated independent components $Y^{\dagger}$ up to a change of sign and permutation.

Chapter 4

Higher-moment risk mimimization portfolio via factor model

In this chapter, following a brief overview on factor model, we introduce the last three strategies of this thesis. (1) *PCMVaR* is the portfolio that returns the lowest MVaR on the K first principal components. (2) *ICMVaR* is the portfolio that returns the lowest MVaR on the K first independent components. (3) *ICMVaRCM* is the portfolio that returns the lowest MVaR on the K first independent components and where all third and fourth co-moments are considered to be equal to zero.

4.1 Introduction to factor model

A factor model is a financial model that uses multiple factors in its calculations to explain market phenomena and/or asset equilibrium prices. The model can be used to explain either an individual asset or a portfolio of asset.

In our context, factor models are used to select portfolio with certain features, such as risk, or to track indices. The number of factors to be used in the model is an important criterion in the construction of portfolio.

Factor models can be divided into three categories: macroeconomic models, fundamental models and statistical models.

- Under the macroeconomic factor model, factors are observable macro-financial variables. For example, it compares an asset's return to such factors as interest, inflation and employment.
- Under the fundamental factor model, factors are created from observable asset features. For example, it is used to analyse the relationship between a asset's return and its underlying financial, such as earnings. One of the most common

popular of fundamental model is the Fama and French Three-Factor Model (F. Fama Sr & French, 1996).

- Under the statistical factors model, factors are unobservable and extracted from the asset returns. In this paper, we will opt for this option by retrieving factors thanks Independent Component Analysis.

Connor, 1995 contrasts the explanatory power of the three types of factor models for securities market returns and concludes that statistical and fundamental factor models perform better than the macroeconomic model. However, the factor model is deficient in some way because the models are based on historical data, which may not accurately predict future values.

4.2 Higher-moment portfolio optimization via principal components

4.2.1 Estimation of higher-moment via PCA

Following Equation 3.2, we consider the *PC basis* given by the orthogonal linear transformation of asset returns X. We consider the *PC basis* of $K \leq N$ factors where the first fourth central moments of the projected portfolio return $P^* = w^{*\prime} Y^*$ are given by:

$$m_1(P^*) = w^{*\prime}\mu_{Y^*} \tag{4.1}$$

$$m_2(P^*) = w^{*\prime}\textstyle\sum_{Y^*} w^* \tag{4.2}$$

$$m_3(P^*) = w^{*\prime}\Phi_{Y^*}(w^* \otimes w^*) \tag{4.3}$$

$$m_4(P^*) = w^{*\prime}\Psi_{Y^*}(w^* \otimes w^* \otimes w^*) \tag{4.4}$$

where $\mu_{Y^*} \in \mathbb{R}^K$ is the sample expected return of the *PC basis*, $\sum_{Y^*} \in \mathbb{R}^{K \times K}$ the sample covariance matrix of the *PC basis*, $\Phi_{Y^*} \in \mathbb{R}^{K \times K^3}$ the sample co-skewness matrix of the *PC basis* and $\Psi_{Y^*} \in \mathbb{R}^{K \times K^3}$ the sample co-kurtosis matrix of the *PC basis*.

Kozak, Nagel, and Santosh, 2018 showed that limiting the number of high variance factor is sufficient to correctly summarize the cross-section of stock returns. In this sense, in order to limit the estimation risk associated with higher moments, we follow the same approach by selecting only $K \leq N$ factors.

4.2.2 MVaR minimization based on PC basis

The portfolio of assets $P = w'X$ can be considered as a portfolio of the principal components $P^* = w^{*\prime}Y^* = w^{*\prime}\Lambda^{-1/2}V'X$, so that the factor exposures w^* are determined as:

$$w^* := \Lambda^{1/2}V'w \tag{4.5}$$

So, the portfolio weight w can be retrieved from the factor exposure as: (Lassance et al., 2018):

$$w = (\Lambda^{1/2}V')^{+}w^* = V\Lambda^{-1/2}w^* \tag{4.6}$$

where A^+ is the pseudo-inverse of A.

Note that here we rely on the *PC basis* to determine optimal portfolio. We could have considered an arbitrary rotation of this basis since this basis is one specific choice among an infinite number of other white basis.

Definition 4.2.1. The minimum MVaR portfolio on *PC basis* (PCMVaR) for a given α is defined as the solution to the following optimization program:

$$w^* := \text{ argmin } MVaR(\alpha) \text{ subject to } \mathbb{1}\Lambda^{1/2}V'w^* = 1 \tag{4.7}$$

where w^* are the portfolio weights on principal components with least return MVaR out of all the portfolios and where the first fourth portfolio moments are respectively $m_1(P^*)$, $m_2(P^*)$, $m_3(P^*)$ and $m_4(P^*)$. It is noteworthy that the full investment constraint is not directly applied on exposure factors but on the portfolio weights directly[1].

In absence of short-selling, the following constraint is added to the optimization problem:

$$\Lambda^{1/2}V'w^* \geq \mathbb{O} \tag{4.8}$$

where $\mathbb{O} = (0, ..., 0)'$ is a $N \times 1$ vector of 0.

[1] In fact, exposure factors are also constrained since they are linked by Equation 4.6.

4.3 Higher-moment portfolio optimization via independent components

4.3.1 Estimation of higher-moment via ICA

As already explained, when estimating higher-moments portfolio, the estimation error are part of the optimization problem. Using the *PC basis* allow to to partially mitigate this estimation risk by selecting only K high-variance factors. This results in a reduction in the number of dimensions and therefore in a reduction of the parameters to be estimated.

More specifically, the problem is to estimate these co-moments matrices Φ and Ψ. The number of terms to be estimated grows exponentially in relation to the number of selected factors K. The number of parameters to be determined for the co-skewness matrix is equal to (Boudt, Lu, & Peeters, 2015):

$$\frac{K(K+1)(K+2)}{6}$$

and for the co-kurtosis matrix, the number of parameters to be determined is equal to:

$$\frac{K(K+1)(K+2)(K+3)}{24}$$

Besides retaining only K independent components, we propose to rely on the IC basis $Y^{\dagger} = R^{\dagger}Y^{*}$ in order to assume perfect independence between the IC factors. As a result, all co-moments vanish and only K parameters per matrix have to be estimated. We denote, respectively, $\Phi_{Y^{\dagger}}$ and $\Psi_{Y^{\dagger}}$ the co-skewness and co-kurtosis matrices if we assume that the IC factors are perfectly independent.

From this point, if we assume that IC factors are perfectly independent and have finite fourth central moments, $Y^{\dagger} \in \mathcal{L}^4$, then the expressions of the first fourth moment of the projected portfolio return, $P^{\dagger} = w^{\dagger\prime}Y^{\dagger}$, are greatly simplified in our context (Lassance et al., 2018)[2]:

$$m_1(P^\dagger) = w^{\dagger\prime}\mu_{Y^\dagger} \tag{4.9}$$
$$m_2(P^\dagger) = w^{\dagger\prime}\sum_{Y^\dagger} w^\dagger \tag{4.10}$$
$$m_3(P^\dagger) = \sum_{i=1}^{K}(w_i^\dagger)^3 m_3(Y_i^\dagger) \tag{4.11}$$
$$m_4(P^\dagger) = \sum_{i=1}^{K}(w_i^\dagger)^4 m_4(Y_i^\dagger) + 3\sum_{i=1}^{K}\sum_{i\neq j}(w_i^\dagger w_j^\dagger)^2 \tag{4.12}$$

where $\mu_{Y^\dagger} \in \mathbb{R}^K$ is the sample expected return of the *IC basis*, $\sum_{Y^\dagger} \in \mathbb{R}^{K\times K}$ the sample covariance matrix of the *IC basis* and where $m_3(Y_i^\dagger)$ and $m_4(Y_i^\dagger)$ are respectively the third and fourth central moments of the *ith* independent component.

4.3.2 MVaR minimization based on IC basis

Following the same approach as for the *PC basis*, the portfolio of assets $P = w'X$ can be considered as a portfolio of the independent components $P^\dagger = w^{\dagger\prime}Y^\dagger = w^{\dagger\prime}R^\dagger\Lambda^{-1/2}V'X$ where the factor exposures $w^\dagger$ are determined as:

$$w^\dagger := R^\dagger\Lambda^{1/2}V'w \tag{4.13}$$

where $R^\dagger$ is the rotation matrix on the *PC basis* that minimizes mutual information. So, the portfolio weight w can be retrieved from the factor exposure as (Lassance et al., 2018):

$$w = (R^\dagger\Lambda^{1/2}V')^{+}w^\dagger = V\Lambda^{-1/2}R^{\dagger\prime}w^\dagger \tag{4.14}$$

where A^+ is the pseudo-inverse of A.

As explained previously, the *PC basis* is one specific choice among an infinite number of other white basis. In order to highlight this point, we develop in this section two models based on the *IC basis*. On the one hand, the first model use the same estimation of higher-moments as for the *PC basis*. In this sense, since the IC factors are obtained through a linear rotation of the PC factors, we expect these two models to perform exactly the same way. On the other hand, we assume that the IC factors are completely independent and their superior moments can be estimated parsimoniously as detailed in Section 4.3.

[2]The authors use the fact that cumulants are additive for independent random variables, see Appendix A.5. of Lassance et al., 2018 for the entire proof.

We define the first optimization strategy on the *IC basis* as follow.
Definition 4.3.1. The minimum MVaR portfolio on *IC basis* (ICMVaR) for a given α is defined as the solution to the following optimization program:

$$w^{\dagger} := \text{ argmin } MVaR(\alpha) \text{ subject to } \mathbb{1}R^{\dagger}\Lambda^{1/2}V'w^{\dagger} = 1 \tag{4.15}$$

where $w^{\dagger}$ are the portfolio weights on independent components with least return MVaR out of all the portfolios and where the first fourth portfolio moments are estimated in the same way as for the *PC basis*.

Again, if we consider that short selling is not allowed, the following constraint is added to the optimization problem:

$$R^{\dagger}\Lambda^{1/2}V'w^{\dagger} \geq \mathbb{0} \tag{4.16}$$

The second strategy on the *IC basis* is defined as follows.
Definition 4.3.2. The minimum MVaR portfolio on *IC basis* (ICMVaRCM) for a given α is defined as the solution to the following optimization program:

$$w^{\dagger} := \text{ argmin } MVaR(\alpha) \text{ subject to } \mathbb{1}R^{\dagger}\Lambda^{1/2}V'w^{\dagger} = 1 \tag{4.17}$$

where $w^{\dagger}$ are the portfolio weights on independent components with least return MVaR out of all the portfolios and where the first fourth portfolio moments are respectively $m_1(P^{\dagger})$, $m_2(P^{\dagger})$, $m_3(P^{\dagger})$ and $m_4(P^{\dagger})$.

If short selling is not allowed, the same constraint is added to the optimization problem:

$$R^{\dagger}\Lambda^{1/2}V'w^{\dagger} \geq \mathbb{0} \tag{4.18}$$

Chapter 5

Data & Methodology

The purpose of this section is to detail the methodology for both simulated and real data. After a quick overview on the simulated data, we describe the time series of the arithmetic returns of the two selected datasets over the period considered. In particular, we want to determine whether it makes sense to rely on higher moment portfolio selection by assessing whether the data follow a Gaussian law. Next, we detail the methodology used to conduct our studies on the performance of the five models considered.

5.1 Simulated data

We first assess the theoretical performance of the different portfolio strategies on simulated data because relying on simulated data gives flexibility in the choice of the parameters we want to test. In our case, it is rather interesting to want to simulate dependent and independent data of different dimensions. For each strategy, we report the sample mean, sample volatility, sample skewness, sample kurtosis, the MVaR(1%), the Sharpe ratio and 1% Modified Sharpe ratio, see Section 5.2.3 for the definitions.

We evaluate the out-of-sample performance of the different strategies via 10-fold cross validation, a commonly used approach, see for instance DeMiguel, Garlappi, Nogales, and Uppal, 2009. The idea of K-fold cross validation is to randomly divide the data into K equal-sized parts. The k part is left out and the others $K-1$ combined parts are used to compute the optimal weights. Those portfolio weights are then used to compute out-of-sample portfolio returns on the kth part that has been previously removed. This process is repeated for each part $k = 1, 2, .., K$ and all results are combined.

5.1.1 First case: linear mixture of 2 independent sources

In the first case, we consider the case where $K = N = 2$, so that no dimension reduction is performed. Let $X = R(\theta)Z$ a linear mixture of two independent sources $Z_1 \sim T(\nu = 5) \perp Z_2 \sim T(\nu = 10)$ mixed with an angle $\theta = \pi/4$ and where the rotation matrix $R(\theta) \in SO(2)$ can be determined for a given angle θ as follow[1] :

$$R(\theta) = \begin{pmatrix} cos(\theta) & -sin(\theta) \\ sin(\theta) & cos(\theta) \end{pmatrix} \tag{5.1}$$

This matrix rotates points in the xy-plane counterclockwise through an angle θ about the origin of the two-dimensional Cartesian coordinate system.

We create pseudo daily returns by simulating $n = 2000$ observations of Z_1 and Z_2. In order to ease the results analysis, these simulated data are then scaled and centered to obtain plausible descriptive statistics for daily returns.

5.1.2 Second case: non-linear mixture of 2 independent sources

In the second case, we consider the same marginal distributions for the sources $Z_1 \sim T(\nu = 10) \perp Z_2 \sim T(\nu = 5)$ as in Section 5.1.1 but X is obtained by mixing them non-linearly with a Gaussian copula of parameter $\rho = 0.2$. Since X is not a linear mix of Z, we expect that complete independence cannot be achieved through a linear transformation of X i.e. we expect that estimating higher co-moments as null is therefore not a completely correct approximation of these moments for the *IC basis*.

5.2 Real data

5.2.1 Time series statistical description

In this paper, we work with two distinct equity datasets from K. French's library: five and thirty U.S. industry portfolios (*5Ind* and *30Ind* respectively). The two datasets correspond to a different number of assets present in the portfolios. We collect daily value-weighted arithmetic returns on the time period 1978-2017 (total length of the data series T = 10080). Daily return are used because it produces better estimation of risk compared to weekly or monthly return data (Jagannathan & Ma, 2003).

[1]Note that the *PC basis* can be obtained for $\theta^* = 0$ as $\mathbb{V}ar(Z_1) > \mathbb{V}ar(Z_2)$ and null correlation between Z_1 and Z_2

The two datasets are composed of *NYSE*, *AMEX* and *NASDAQ* stock where each stock has been assigned to an industrial portfolio at the end of June of each year based on its four-digit SIC code[2] at that time.

In Appendix A.1 and A.2, for each industry and for each dataset, we report the minimum, maximum, first quartile, third quartile, sample annual mean, sample annual volatility , sample daily skewness, sample daily excess kurtosis and p-value of Jarque-Bera test of Normality[3].

Based on these figures, preliminary observations can be drawn. We first notice that all industries have a positive mean. The minimum and maximum are very large given the impact of the October 1987 crash and the 2007-2009 financial crisis. With regard to the annual return, for the *5Ind*, the average annual return ranges from 12.76% (*HiTec*) to 14.49% (*Hlth*) while for the *30Ind*, the average annual return ranges from 9.90% (*Steel*) to 19.05% (*Smoke*). Moreover, some industries are more volatile than others. For the *5Ind*, the annual volatility ranges from 15.61% (*Cnsmr*) to 19.38% (*Other*) while for the *30Ind*, the annual volatility ranges from 14.59% (*Util*) to 39.66% (*Coal*). Then, all industries (except for *Mines*) have negative skewness and a positive excess kurtosis. To confirm non-Normality of the industry returns, we perform the test of Normality Jarque-Bera (Jarque & Bera, 1987) on each industry of each dataset. The observation is straightforward, normality can be rejected as all p-values are close to 0 (< 2.2e-16). From then on, as no industries follow a Normal distribution, this confirms the well-known feature of non-Normality of asset returns. This last observation is interesting in our research context because ICA requires non-Gaussian features to correctly discriminate between all the white bases, as detailed in Section 3.2.2.

Finally, Figure 5.1 illustrates the time evolution of the cumulative percentage variance explained by the first K PC factors for *5Ind* and *30Ind*. Based on this graph, we note that the K selection process detailed in the Section 3.1 works properly by selecting K factors that explain on average the same cumulative percentage variance over time (with an average of around 0.9 cumulative percentage variance for both datasets).

[2]The Standard Industrial Classification (SIC) is a system for classifying industries by a four-digit code.

[3]The null hypothesis for the test is that the data follow Gaussian distribution while the alternate hypothesis is that the data does not come from a Gaussian distribution

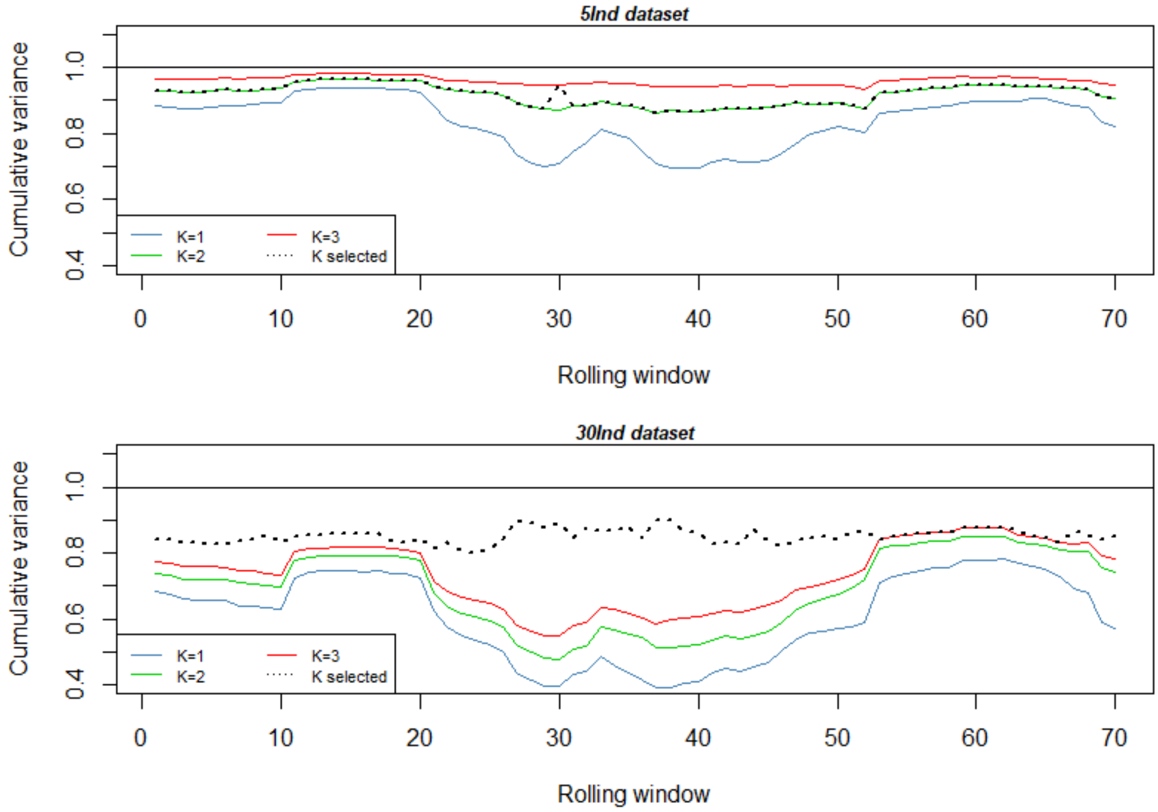

FIGURE 5.1: Time evolution of the cumulative percentage variance explained by the first K Factors for the two datasets. The cumulative percentage variances are computed every six months on a rolling window of five years.

5.2.2 Portfolio policies

We evaluate the performance of the five following strategies:

1. Traditional minimum-variance portfolio (*MV*). This is a strategy that returns the lowest risk where the only estimation done is the sample covariance matrix as measure of risk.
2. Minimum MVaR portfolio (*MVaR*). This is a strategy that returns the lowest MVaR as measure of risk.
3. Minimum MVaR portfolio on principal components (*PCMVaR*). This is a strategy that returns the lowest MVaR on the K first principal components.
4. Minimum MVaR portfolio on independent components (*ICMVaR*). This is a strategy that returns the lowest MVaR on the K first independent components.
5. Minimum MVaR portfolio on independent components (*ICMVaRCM*). This is a strategy that returns the lowest MVaR on the K first independent components

and where all third and fourth co-moments are considered to be equal to zero.

More precisely, we consider two different cases for each strategy. In the first case, short-selling is possible while in the second case it is not allowed. The first case is particularly useful in order to observe how strategies perform best when no constraints are imposed (except the full investment constraint). However, such strategies are not actually applicable because this results in high turnover, so that we will also discuss the second case.

We fix $\alpha = 0.01$, a common choice in the calculation of MVaR while complying with the 4.16% limit. Except for the *MV* in presence of short-selling where a closed form is available, optimal portfolio weights are obtained via Sequential Quadratic Programming[4]. All of our studies are performed with R Studio.

Regarding the portfolio rebalancing rules, we opt for a standard rebalancing approach to take into account time-varying distributions of return on assets. Optimal portfolio weights are determined from daily data from the last five years (an estimation window of $M = 1260$) and are only rebalanced every six months. As a result, we obtain 70 windows of six months of out-of-sample performance on which the different performance criteria detailed in Section 5.2.3 are evaluated.

Lastly, regarding the number of K components used in the *PCMVaR* and *ICMVaR* strategies, we use the Velicer, 1976 method as detailed in the Section 3.1. More specifically, the number of components retained is determined for each estimation window as shown in Figure 5.1.

5.2.3 Out-of-sample performance evaluation

Given the time series of daily out-of-sample returns generated by each strategy and in each dataset, several performance criteria are computed to evaluate the portfolio strategy. Regarding the mean-variance trade-off, we report the sample mean, volatility and Sharpe ratio, all expressed in daily terms. Developed by Sharpe, 1994, the Sharpe Ratio is used to help investors understand the return of an investment compared to its risk.

Definition 5.2.1. The *Sharpe ratio SR* is defined as the sample mean of portfolio returnsμ_k, divided by the sample standard deviation of strategy k, σ_k :

$$SR_k = \frac{\mu_k}{\sigma_k} \tag{5.2}$$

[4]A **R** package implementing this algorithm : `fmincon`, avaiable here: https://rdrr.io/cran/pracma/man/fmincon.html

Regarding the higher-order risk, we report the sample skewness, sample excess kurtosis, $MVaR(1\%)$ and $1\%MSR$, all expressed in daily terms.

Definition 5.2.2. The *Modified Sharpe ratio MSR* (Gregoriou & Gueyie, 2003) is defined as the sample mean of portfolio returns μ_k, divided by the Modified Value-at-Risk for a certain confidence level, $MVaR_k(\alpha)$:

$$MSR = \frac{\mu}{MVaR(\alpha)} \tag{5.3}$$

Concerning portfolio stability, we report the portfolio turnover, expressed in annual terms, to obtain an idea of the volume of transactions required to implement each portfolio strategy. This portfolio performance feature can be interpreted as the average percentage of wealth traded during each period (Demiguel et al., 2009).

Definition 5.2.3. The *Turnover* is defined as the average sum of the absolute value of the trades across the N available assets traded during each period:

$$Turnover = \frac{1}{T-M} \sum_{t=1}^{T-M} \sum_{j=1}^{N} \left(|\hat{w}_{j,t+1} - \hat{w}_{j,t}| \right) \tag{5.4}$$

where $\hat{w}_{j,t}$ is the portfolio weight in asset j at time t (before rebalancing) and $\hat{w}_{j,t+1}$ is the portfolio weight at time $t+1$ (after rebalancing).

Finally, an economic measure is also reported in order to indicate how proportional transactions costs, generated by this turnover, affect the returns from a particular strategy. The proportional transactions costs are set to be equal to 50 basis points per transaction as also assumed by Demiguel et al., 2009, based on the study of transaction cost by Balduzzi and Lynch, 1999.

Definition 5.2.4. Let R_p be the return on the portfolio of N assets before rebalancing:

$$R_p = \sum_{j=1}^{N} R_{j,t+1} \hat{w}_{j,t} \tag{5.5}$$

Designating c as the proportional transaction cost, the cost of such a transaction on all assets during $t+1$ and t is determined by:

$$TC = c \times \sum_{j=1}^{N} \left(|\hat{w}_{k,j,t+1} - \hat{w}_{k,j,t}| \right) \tag{5.6}$$

So, the evolution of wealth $W_{k,t}$ of stategy k at time t can be written as follows:

$$W_{t+1} = W_t(1 + R_p)\left(1 - c \times \sum_{j=1}^{N} |\hat{w}_{j,t+1} - \hat{w}_{j,t}|\right) \tag{5.7}$$

where the return net of transactions costs is given by $\frac{W_{t+1}}{W_t} - 1$.

Chapter 6

Minimum risk portfolios: Results & Further Analysis

In this section, we compare empirically the performance of the optimal risk minimization portfolios both on simulated and real data. In each of the below tables, the various strategies being examined are listed in column, while the rows refer to the different performance criteria.

6.1 Results on simulated data

6.1.1 Linear mixture of two independent sources

In this first example, we evaluate the out-of-sample performance (via 10-fold cross validation) of a linear mixture $X = R(\theta)Z$ of two independent sources $Z_1 \sim T(\nu = 5) \perp Z_2 \sim T(\nu = 10)$ mixed with an angle $\theta = \pi/4$. Table 6.1 displays the results. Based on these results, some observations of the five strategy's performance can be drawn:

- First, we note that the MVaR, *PCMVaR* and *ICMVaR* strategies return exactly the same results as expected for all criteria. Indeed, since no size reduction has been made, all the information present in the mixed data is contained in the PC and IC factors.
- In terms of mean-variance trade-off, the first 4 strategies perform equally well overall. For the ICMVaRCM strategy, we observe a slightly higher volatility compared to all others strategies.

- Regarding higher-order risk, the strategy ICMVaRCM returns the highest MVaR. However, we observe a better performance for this strategy in terms of skewness and kurtosis compared to other strategies. The MV strategy is the least performing.
- Finally, concerning portfolio stability, the MV strategy is the more stable while the ICMVaRCM strategy generates high turnover up to 10x higher than other strategies.

	MV	MVaR	PCMVaR	ICMVaR	ICMVaRCM
Mean	0.06%	0.08%	0.08%	0.08%	0.06%
Volatility	0.89%	0.92%	0.92%	0.92%	1.01%
SR	0.07	0.08	0.08	0.08	0.06
Skewness	-0.06	-0.04	-0.04	-0.04	0.01
Excess Kurtosis	2.04	1.65	1.65	1.65	1.23
1% MVaR	2.48%	2.45%	2.45%	2.45%	2.58%
1% MSR $x10^2$	2.51	3.07	3.07	3.07	2.34
Turnover	0.01	0.04	0.04	0.04	0.38

TABLE 6.1: Out-of-sample performance of a linear mixture $X = R(\theta)Z$ of two independent sources $Z_1 \sim T(\nu = 5) \perp Z_2 \sim T(\nu = 10)$ mixed with an angle $\theta = \pi/4$

6.1.2 Unlinear mixture of two independent sources

In this second example, we consider a Gaussian copula of parameter $\rho = 2$ to create an unlinear mixture of the same independent sources $Z_1 \sim T(\nu = 5) \perp Z_2 \sim T(\nu = 10)$. Clearly, total independence between IC factors can never be achieved in this case i.e. the mutual information is higher than zero between IC factors. Table 6.2 displays the results. Again, some points can be highlighted:

- The MVaR, PCMVaR and ICMVaR strategies return again exactly the same results for all criteria.
- Regarding mean-variance trade-off, all 5 strategies perform equally well overall.
- Concerning higher-order risk, the higher moment strategies performs as well as the MV strategy. ICMVaRCM strategy is even now less performant compared to other strategies.
- Finally, about portfolio stability, the MV strategy is again the more stable while the ICMVaRCM strategy's turnover has dramatically dropped.

	MV	MVaR	PCMVaR	ICMVaR	ICMVaRCM
Mean	0.06%	0.06%	0.06%	0.06%	0.06%
Volatility	1.03%	1.04%	1.04%	1.04%	1.03%
SR	0.06	0.05	0.05	0.05	0.05
Skewness	-0.20	-0.20	-0.20	-0.20	-0.22
Excess Kurtosis	1.19	1.18	1.18	1.18	1.22
1% MVaR	2.76%	2.80%	2.80%	2.80%	2.79%
1% MSR $x10^2$	2.10	2.01	2.01	2.01	2.03
Turnover	0.03	0.06	0.06	0.06	0.04

TABLE 6.2: Out-of-sample performance of a unlinear mixture of two independent sources $Z_1 \sim T(\nu = 5) \perp Z_2 \sim T(\nu = 10)$ mixed with an Gaussian copula with $\rho = 0.2$

In summary, the simulation results in this section indicates that the level of independence between IC factors is a crucial point. If this gain in independence is not significant enough, the hypothesis of zero higher co-moments cannot be taken and will probably lead to poor out-of-sample results. In addition, we evaluate the performance of a two-dimensional portfolio only, so the estimation risk is lower than that of a larger portfolio, as described in the Section 4.3. The advantage of the null co-moments hypothesis can be very small. These assumptions will now be tested on real data.

6.2 Results on real data

We now evaluate the out-of-sample performance of the five proposed strategies developed in this thesis for the *5Ind* and *30Ind*. This section is divided in four part: (1) Discussion of Portfolio Weight Stability, (2) Discussion of Mean-Variance trade-off, (3) Discussion of Higher-order risk.

In relation to the discussions above, the results in the presence and in the absence of short-selling are displayed respectively in Table 6.3 and 6.4. In general, it should be noted that the strategies *PCMVaR* and *ICMVaR* perform equally well in all performance criteria since the number of dimensions retained K is the same for both strategies and the higher moments are identically estimated[1].

[1]Slight differences between the two portfolios are due to optimization errors in R.

6.2.1 Discussion of Portfolio Weight Stability

In relation in Table 6.3 and 6.4, our first observation is that the minimum-variance *MV* strategy is the more stable one compared to the four others both in the presence and in the absence of short-selling constraint. The *ICMVaRCM* is the least stable strategy with, for example, a turnover (3.88%) up to 6 times the turnover of *MV's* strategy (0.56%) for *5Ind* dataset with no short-selling constraint.

In addition, Figures B.1 and B.2 in Appendix show a comparison of different turnover's time evolution of the 5 different portfolios for both datasets while Figures B.3 and B.4 show the same evolution in the absence of short-selling. From these figures, it is clear that the introduction of the no short-selling constraint substantially improves the stability of all strategies for both dataset.

Figures 6.1 and 6.2 show a comparison of wealth over time between the five strategies for the two datasets where transaction costs are proportional to the turnover generated by the given portfolio (the transaction cost is set to 50bps per transactions). We note that the turnover has a considerable impact on the portfolio's performance. In the absence of short selling, turnover is much lower, so that all portfolios outperform those without constraints in terms of this economic measure. There is no doubt that not limiting portfolio turnover makes the implementation of these strategies cost-intensive in practice.

We now turn to discuss the stability of the portfolios based on MVaR minimization. We observe that for the large dataset *30Ind*, the turnover of the factor models *PCMVaR* and *ICMVaR* are lower in comparison to the portfolio *MVaR*. Consequently, the results confirm the findings reported in much of the recent financial literature that factor model decreases estimation risk for large portfolio dimension.

Finally, these results are in line with those of the simulated data presented in Section 6.1. In each case, ICMVaRCM strategy is the most unstable portfolio with the higher turnover.

6.2.2 Discussion of Mean-Variance trade-off

In the presence of short selling, Policy *MV* performs the best by exhibiting the highest Sharpe Ratio among the four others policies. When short-selling is limited, *MVaR* strategy outperfoms slightly the others strategies in terms of Sharpe Ratio. In contrast, *ICMVaRCM* portfolios returns the lowest Sharpe Ratio in each scenario. In other words, portfolio *ICMVaRCM* tends to generate a relatively low annual return for higher volatility.

Unlike turnover, factor models *PCMVaR* and *ICMVaR* do not perform more efficiently than strategy *MVaR* in terms of mean-variance trade-off.

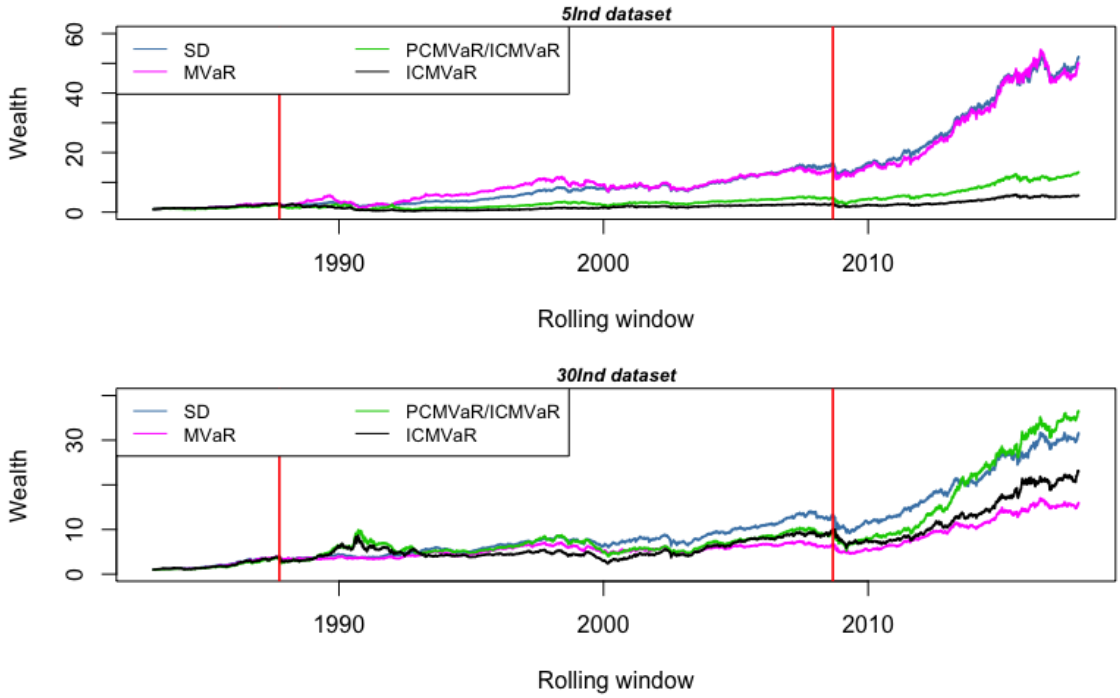

FIGURE 6.1: Comparison of the Wealth's Time evolution between the five strategies for *5Ind* and *30Ind* - Short-selling allowed.

6.2.3 Discussion of Higher-order risk

The first important point concerns the strategies *MV* and *MVaR*. In small dimension portfolio (*5Ind*), policy *MVaR* outperforms policy *MV* in terms of skewness and excess kurtosis both in presence and absence of short-selling. On the contrary, when considering a portfolio of higher dimensions (*30Ind*), we observe the opposite effect. This implies that for a high-dimensional portfolio, the estimation error becomes too important and nothing much is lost by ignoring the higher-order moment. These results are consistent with the findings of most financial literature about the curse of dimensionality and higher-order estimation risk.

In addition, for high dimension portfolio (*30Ind*), we observe that factor models *PCMVaR* and *ICMVaR* outperform MVaR and variance minimization strategies in terms of higher-order risk both in presence and absence of short-selling. In other

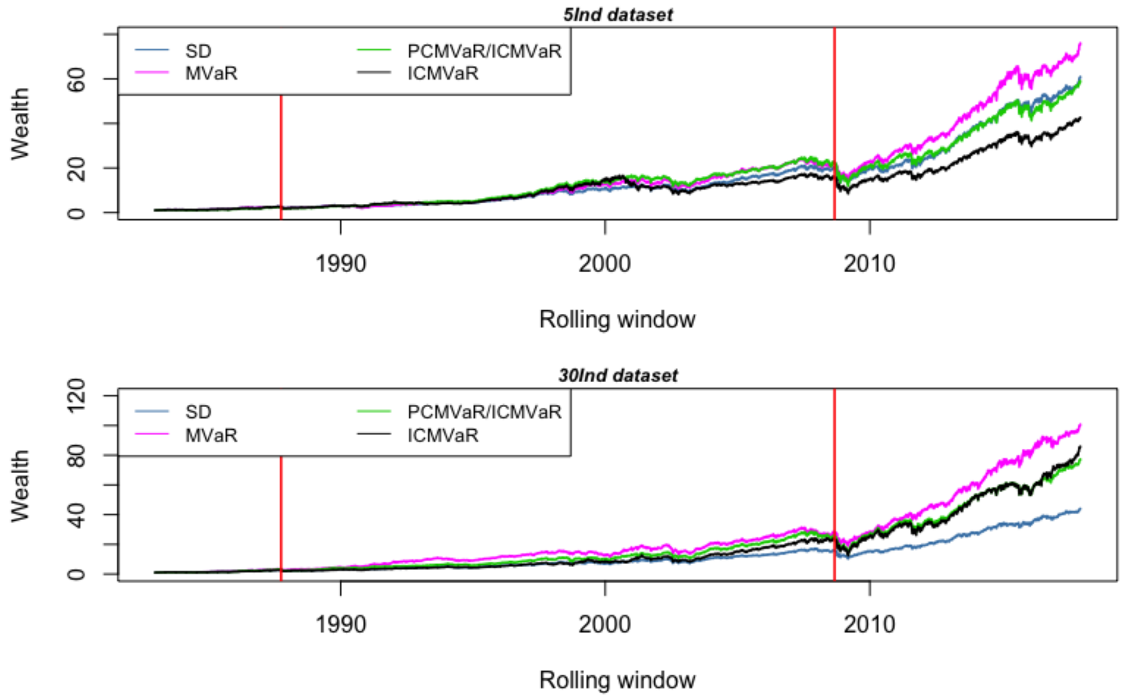

FIGURE 6.2: Comparison of the Wealth's Time evolution between the five strategies for *5Ind* and *30Ind* - No short selling.
The proportional transactions cost are set to be equal to 50bps per transaction.

words, the *PC basis* and the *IC basis* reduce the number of parameters to be estimated and consequently mitigates risk estimation of superior moments. These results are again in line with most of the financial literature's findings about the higher moment's estimation via Principal Component Analysis.

Moreover, *ICMVaRCM* strategy systematically outperforms *PCMVaR* and *ICMVaR* strategies in terms of higher-order risk (except for *30Ind* in presence of short-selling where *ICMVaRCM* returns a slightly lower skewness). In the absence of short selling, *ICMVaRCM* portfolio depicts the higher skewness and the lower excess kurtosis among all the strategies considered in this thesis. To put it another way, for factor models on large portfolios, the strategy of considering higher co-moments as null for the *IC basis* reduces the estimation risk of superior moments and outperforms in terms of higher-order risk out-of-sample a strategy where all co-moments are part of the estimation (*PCMVaR/ICMVaR)*.

Finally, by examining the MVaR and MSR, we observe that portfolio *ICMVaRCM* do not achieve overall a better balance between return and risk in comparison with the other four strategies even if this strategy provides the best performance in terms of higher-order risk. This result makes this strategy not appealing in practice.

5Ind	MV	MVaR	PCMVaR	ICMVaR	ICMVaRCM
Return (annual)	12.56%	**13.58%**	11.78%	11.83%	9.34%
Volatility (annual)	**13.95%**	17.48%	22.50%	22.48%	22.06%
SR (annual)	**0.90**	0.78	0.52	0.52	0.42
Skewness(daily)	-0.63	**-0.20**	-0.57	-0.58	-0.36
Excess Kurtosis (daily)	19.44	**7.86**	10.81	10.67	8.24
1% MVaR (daily)	6.26%	**4.68%**	7.25%	7.24%	6.17%
1% MSR $x10^2$ (daily)	0.80	**1.15**	0.65	0.65	0.60
Turnover (annual)	**0.56%**	1.71%	3.53%	3.49%	3.88%
30Ind	MV	MVaR	PCMVaR	ICMVaR	ICMVaRCM
Return (annual)	11.16%	10.19%	**12.74%**	**12.74%**	12.31%
Volatility (annual)	**11.71%**	13.29%	16.78%	16.78%	17.70%
SR (annual)	**0.95**	0.77	0.76	0.76	0.70
Skewness(daily)	-0.73	-0.96	**-0.39**	**-0.39**	-0.45
Excess Kurtosis (daily)	23.96	21.34	9.61	9.61	**8.65**
1% MVaR (daily)	6.05%	6.38%	**5.03%**	**5.03%**	5.08%
1% MSR $x10^2$ (daily)	0.73	0.63	**1.01**	**1.01**	0.96
Turnover (annual)	**1.21**	2.75	2.09	2.09	3.52

TABLE 6.3: Out-of-sample performance of the 5 different strategies for *5Ind* and *30Ind* - Short selling allowed.

5Ind	MV	MVaR	PCMVaR	ICMVaR	ICMVaRCM
Return (annual)	12.97%	**13.79%**	13.28%	13.40%	12.73%
Volatility (annual)	**14.61%**	15.14%	16.93%	17.09%	17.95%
SR (annual)	0.89	**0.91**	0.78	0.78	0.71
Skewness(daily)	-0.87	-0.67	-0.69	-0.68	**-0.43**
Excess Kurtosis (daily)	21.76	15.20	16.85	16.34	**13.56**
1% MVaR (daily)	7.10%	**5.86%**	6.98%	6.91%	6.45%
1% MSR $x10^2$ (daily)	0.72	**0.93**	0.75	0.77	0.78
Turnover (annual)	**0.30**	0.52	0.37	0.43	0.77
30Ind	MV	MVaR	PCMVaR	ICMVaR	ICMVaRCM
Return (annual)	11.74%	14.69%	14.14%	14.41%	**14.76%**
Volatility (annual)	**12.40%**	15.12%	16.62%	16.59%	17.23%
SR (annual)	0.95	**0.97**	0.85	0.87	0.86
Skewness(daily)	-0.65	-0.82	-0.56	-0.56	**-0.47**
Excess Kurtosis (daily)	23.60	22.79	15.09	15.17	**12.63**
1% MVaR (daily)	6.33%	7.56%	6.38%	6.38%	**5.96%**
1% MSR $x10^2$ (daily)	0.74	0.77	0.88	0.90	**0.98**
Turnover (annual)	**0.31**	0.71	0.65	0.72	1.08

TABLE 6.4: Out-of-sample performance of the 5 different strategies for *5Ind* and *30Ind* - No short-selling.
The portfolio are only rebalanced every six months on a rolling window of five years.

Chapter 7

Conclusion

The conclusion of this thesis begins with a brief summary of the results obtained. Then, we present the three main limitations of our research context and we end up proposing two directions of research.

7.1 Results summary and implications

To implement in practice higher-moment portfolio, one has to traditionally estimate the sample co-skewness and sample co-kurtosis matrix. When dealing with large assets portfolio, the number of parameters to be estimated increases rapidly due to the curse of dimensionality. Because of the estimation error, strategies constructing on these estimators do not provide satisfactory performance. Factor model based on Principal Component Analysis provides a powerful way of mitigating this risk estimation. It has been shown that only selecting a limited number of factors is sufficient to adequately summarize the cross-section of asset returns. Compared to traditional higher-moment portfolio, it has the advantage to perform dimension-reduction, so that the number of parameters to be estimated decreases dramatically. However, the basis formed by the first eigenvectors of the covariance matrix only imposes decorrelation and do not restrict any dependence structure in terms of higher-moment. To circumvent this issue, we proposed to look for a basis that features the least higher-moment dependence. Independent Component Analysis provides this such basis by discriminate among all the orthonormal basis in order to determine the factors that are as independent as possible. This has the advantage to mitigate even more the estimation risk by simplifying the decomposition of higher-order risk measure as one can simply ignore the co-moments if we assume perfect independence between the independent components.

Empirically, the numerical results shown three main findings:

1. For large assets portfolio, strategy that estimates all superior co-moments in their optimization are usually outperformed by factor model where dimension-reduction has been applied.

2. Perfect independence between the independent components is never achieved. However, the gain in independence generated is sufficient to select this basis because it reduces estimation risk by assuming superior co-moments as null.

3. In the presence of short selling, the strategy based on independent component where superior co-moments are assumed to be null outperforms in terms of higher-order risk traditional strategy based on principal components. In the absence of short selling, this strategy outperforms also variance and Modified Value-at-Risk minimization strategies. Nevertheless, this strategy generally tends to generate higher turnover and less performing mean-variance trade-off.

7.2 Research limitations

A careful reader is aware that all research has its limitations and that the findings of our research should be taken with hindsight. This section aims to present the most important ones.

1. The first limitation of our research concerns one of the application hypotheses of the *fastICA* algorithm. In order to be able to use it in our context, we have assumed that asset returns are a linear combination of independent factors. However, as demonstrated in our results, we can only find factors that are as independent as possible but the gain of independence generated between the factors was sufficient to consider the higher co-moments as zero. Perhaps sometimes this assumption does not lead to superior results, especially for smaller portfolios where the number of parameters to be estimated is limited.

2. A second limitation of our research concerns the data used and the methodology related to it. We only considered 2 datasets. Additional tests and further analyses should be performed on other datasets of different dimensions. We only collected daily data for out-of-sample performance testing. The robustness of the different strategies should also be evaluated on weekly or monthly data. In addition, we also only used a calendar rebalancing approach every 6 months. It would be interesting to consider different timing of rebalancing and different types of rebalancing such as percentage-of-portfolio rebalancing.

3. To the best of our knowledge, very few authors have devoted their research to portfolio selection via Independent Component Analysis. This machine learning technique remains very poorly known in the financial world, particularly because of its novelty and complexity. The use of ICA in portfolio selection by asset management practitioners remains unlikely at this time. Nevertheless, this technique can be a promising option in finance if additional researches are carried out on it.

7.3 Suggestions for future research

We conclude this thesis by proposing two avenues of research that are worth investigating.

1. In this thesis, we have only considered *fastICA* as algorithm to determine independent factors. However, even if this algorithm provides an efficient and fast solution, several new algorithms (some of which are able to deal with non linear mixture forms) have been developed in recent years. Among others, we should mention the following: Second-Order Blind Identification (*SOBI*), logistic Infomax (*Infomax*), Joint Approximation Diagonalization of Eigenmatrices (*JADE*) and Iterative Maximization of the Kurtosis Contrast (*RobustICA*). A further study of the performance of the various existing algorithms and their impact on portfolio performance is a possible area of research, see for instance Sahonero and Calderon, 2017.

2. In the same vein, it would be relevant to extend the application of the ICA with different portfolios selection strategy. A first possible direction of future research would be to consider a shrinkage estimation of the covariance matrix for the eigen decomposition instead of the sample covariance matrix. Secondly, it is known that estimating return is tricky and that most of the time nothing much is lost by ignoring their estimation. Hence, it would be interesting to consider models without return estimation in order to give more importance to the higher moments. As a matter of fact, a possible strategy would be to consider the minimum Modified Value-at-Risk without the expected return in his computation.

Appendix A

Tables

A.1 Descriptive statistics of *5Ind* and *30Ind*

	Cnsmr	Manuf	HiTec	Hlth	Other
Observations	10080	10080	10080	10080	10080
NAs	0	0	0	0	0
Minimum	-17.28	-18.23	-18.89	-17.89	-15.25
Quartile 1	-0.43	-0.43	-0.55	-0.50	-0.46
Median	0.08	0.07	0.08	0.07	0.08
Quartile 3	0.56	0.57	0.67	0.65	0.60
Maximum	9.88	14.52	14.25	11.10	11.27
Mean (annual)	13.79	12.87	12.76	14.49	13.08
Volatility (annual)	15.61	16.75	21.02	17.75	19.38
Skewness	-0.62	-0.63	-0.10	-0.50	-0.29
Excess Kurtosis	16.24	21.88	11.13	11.63	13.98
P-Value J-B	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16

TABLE A.1: Descriptive statistics of *5Ind* - All values are expressed in daily terms except for the return and volatility that are expressed in annual terms

	Food	Beer	Smoke	Games	Books	Hshld	Clths	Hlth	Chems	Txtls	Cnstr	Steel	FabPr	ElcEq	Autos
Observations	10080	10080	10080	10080	10080	10080	10080	10080	10080	10080	10080	10080	10080	10080	10080
NAs	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Minimum	-16.52	-14.73	-13.99	-22.00	-19.01	-21.46	-18.51	-17.89	-17.66	-18.40	-17.30	-23.95	-17.82	-19.70	-19.72
Quartile 1	-0.42	-0.53	-0.62	-0.65	-0.52	-0.49	-0.58	-0.50	-0.56	-0.56	-0.56	-0.73	-0.57	-0.66	-0.71
Median	0.08	0.06	0.06	0.08	0.05	0.04	0.08	0.07	0.06	0.07	0.07	0.04	0.07	0.05	0.03
Quartile 3	0.55	0.65	0.78	0.81	0.62	0.59	0.72	0.65	0.70	0.68	0.69	0.83	0.69	0.79	0.82
Maximum	10.11	10.12	14.99	12.51	18.82	11.03	12.69	11.10	13.07	19.50	11.90	20.26	13.79	14.08	11.70
Mean (annual)	14.37	16.97	19.05	15.35	11.74	12.31	14.00	14.49	13.77	14.03	13.19	9.90	12.50	15.48	11.26
Volatility (annual)	15.16	18.56	22.95	24.10	19.66	17.10	20.88	17.75	20.48	22.77	21.22	27.49	21.60	22.64	24.18
Skewness	-0.51	-0.06	-0.07	-0.49	-0.09	-1.15	-0.31	-0.50	-0.46	0.22	-0.34	-0.33	-0.36	-0.27	-0.25
Excess Kurtosis	17.17	7.41	10.32	11.31	18.91	29.08	9.93	11.63	10.96	16.03	10.41	13.28	10.72	9.42	8.54
P-Value J-B	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16

	Carry	Mines	Coal	Oil	Util	Telcm	Servs	BusEq	Paper	Trans	Whlsl	Rtail	Meals	Fin	Other
Observations	10080	10080	10080	10080	10080	10080	10080	10080	10080	10080	10080	10080	10080	10080	10080
NAs	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Minimum	-16.38	-14.30	-19.34	-19.47	-12.86	-16.69	-16.81	-20.16	-20.91	-17.56	-13.20	-18.01	-15.48	-14.84	-16.54
Quartile 1	-0.61	-0.87	-1.05	-0.66	-0.34	-0.50	-0.55	-0.69	-0.48	-0.60	-0.46	-0.55	-0.58	-0.48	-0.49
Median	0.08	0.00	-0.00	0.05	0.07	0.05	0.10	0.08	0.06	0.07	0.09	0.07	0.07	0.07	0.07
Quartile 3	0.75	0.95	1.12	0.76	0.47	0.61	0.72	0.81	0.60	0.69	0.59	0.66	0.70	0.62	0.60
Maximum	13.07	16.39	21.36	19.27	14.43	14.47	12.64	17.52	8.96	9.33	9.74	11.75	11.49	14.65	13.59
Mean (annual)	15.78	10.40	11.21	13.54	11.67	12.09	14.98	13.14	12.76	13.53	12.56	14.71	14.42	14.10	9.91
Volatility (annual)	20.82	28.12	39.66	22.65	14.59	18.66	21.16	25.25	17.52	19.92	16.77	19.06	18.75	21.26	19.01
Skewness	-0.39	0.10	-0.00	-0.20	-0.07	-0.11	-0.24	0.06	-0.81	-0.47	-0.43	-0.27	-0.27	-0.00	-0.41
Excess Kurtosis	8.90	7.04	7.59	13.40	20.36	13.07	8.88	9.13	18.37	9.41	8.07	10.97	7.53	18.03	13.42
P-Value J-B	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16

TABLE A.2: Descriptive statistics of *30Ind* - All values are expressed in daily terms except for the return and volatility that are expressed in annual terms

Appendix B

Figures

B.1 Portfolio stability

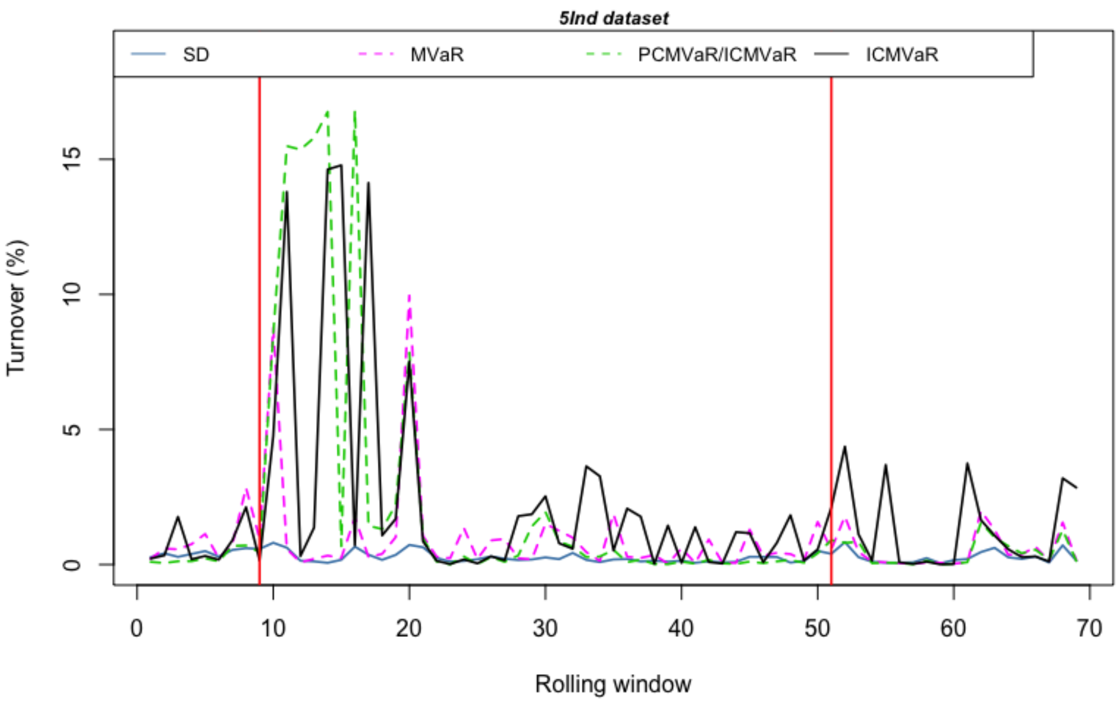

FIGURE B.1: Comparison of the turnover's time evolution for *5Ind* between the five strategies - Short selling allowed. The turnover is computed every six months on a rolling window of five years. The two red lines refer to the October 1987 crash and the 2007-2009 financial crisis.

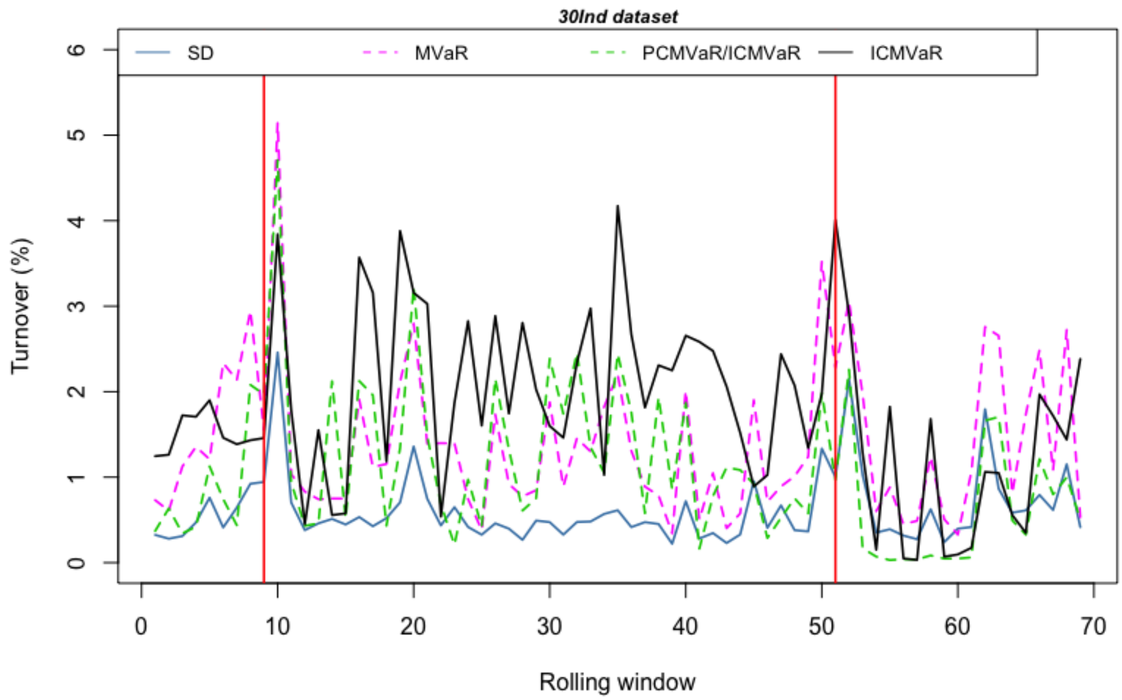

FIGURE B.2: Comparison of the turnover's time evolution for *30Ind* between the five strategies - Short selling allowed. The turnover is computed every six months on a rolling window of five years. The two red lines refer to the October 1987 crash and the 2007-2009 financial crisis.

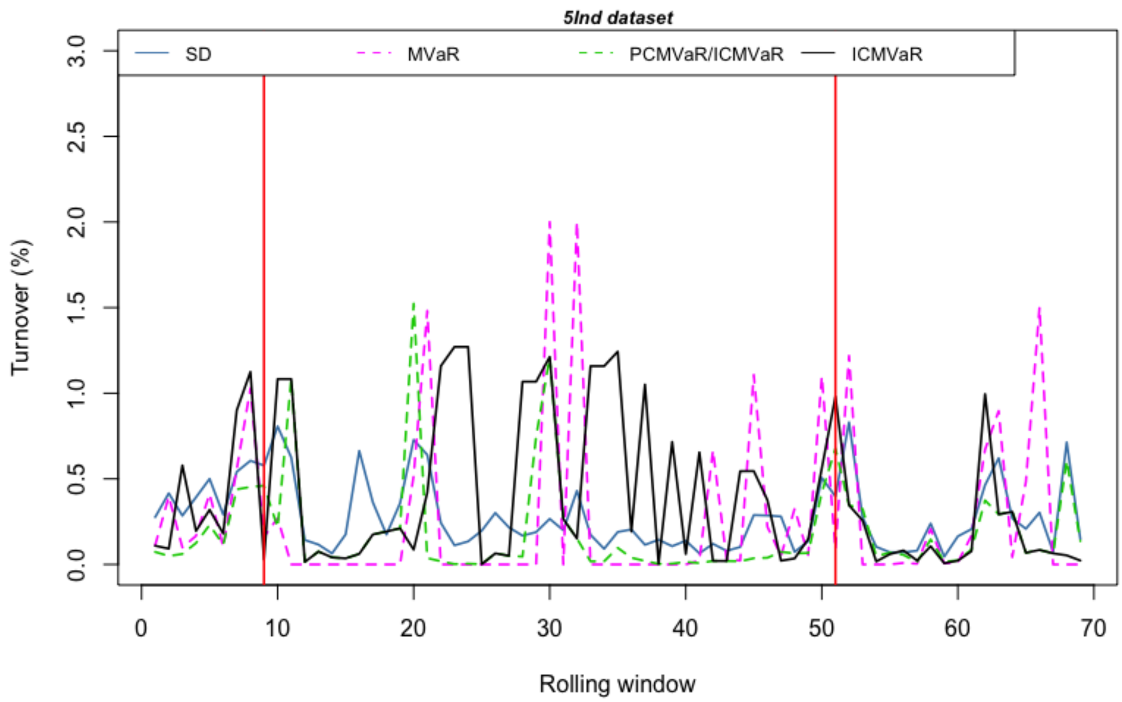

FIGURE B.3: Comparison of the turnover's time evolution for *5Ind* between the five strategies - No short selling. The turnover is computed every six months on a rolling window of five years. The two red lines refer to the October 1987 crash and the 2007-2009 financial crisis.

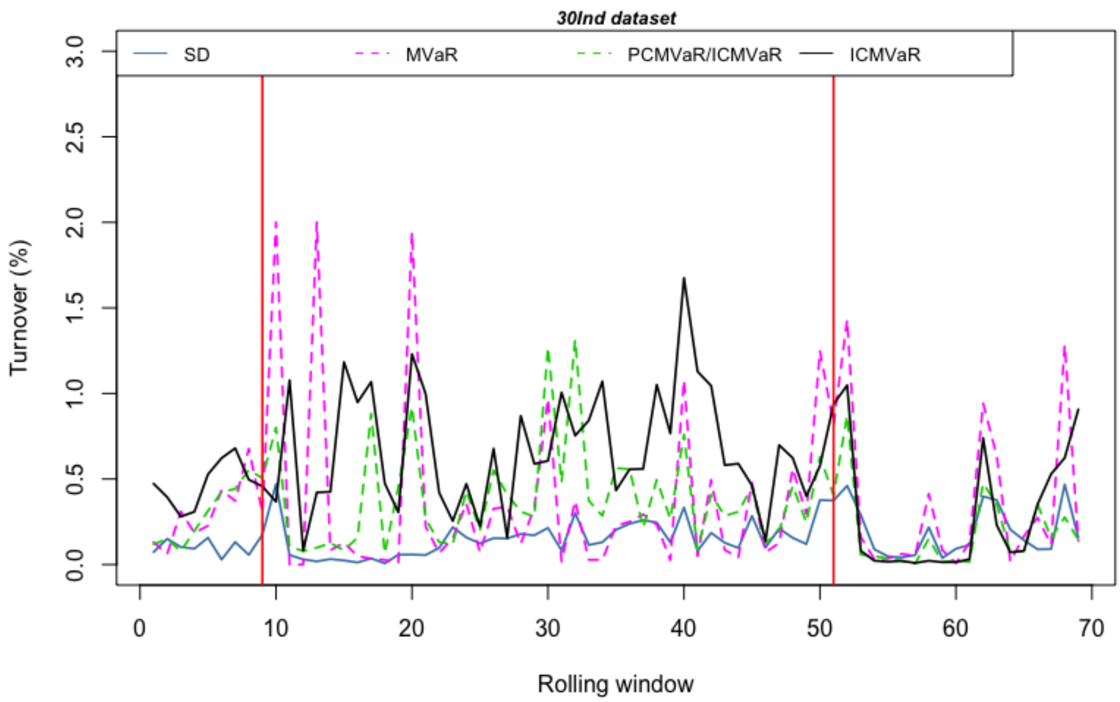

FIGURE B.4: Comparison of the turnover's time evolution for *30Ind* between the five strategies - No short selling. The turnover is computed every six months on a rolling window of five years. The two red lines refer to the October 1987 crash and the 2007-2009 financial crisis.

Bibliography

Baillie, R. T. & Bollerslev, T. (1992). Prediction in dynamic models with time-dependent conditional variances. *Journal of Econometrics*, *52*(1), 91–113. doi:https://doi.org/10.1016/0304-4076(92)90066-Z

Balduzzi, P. & Lynch, A. W. (1999). Transaction costs and predictability: Some utility cost calculations. *Journal of Financial Economics*, *52*(1), 47–78. Retrieved from https://EconPapers.repec.org/RePEc:eee:jfinec:v:52:y:1999:i:1:p:47-78

Best, M. & Grauer, R. (1991). On the sensitivity of mean-variance-efficient portfolios to changes in asset means: Some analytical and computational results. *Review of Financial Studies*, *4*, 315–42. doi:10.1093/rfs/4.2.315

Boudt, K., Lu, W., & Peeters, B. (2015). Higher order comoments of multifactor models and asset allocation. *Finance Research Letters*, *13*. doi:10.1016/j.frl.2014.12.008

Broadie, M. (1993). Computing efficient frontiers using estimated parameters. *Annals of Operations Research*, *45*, 21–58. doi:10.1007/BF02282040

Campbell, J. Y., Lo, A. W., & MacKinlay, A. (1997). *The econometrics of financial markets*. Princeton University Press. Retrieved from http://www.jstor.org/stable/j.ctt7skm5

Cavenaile, L. & Lejeune, T. (2010). A note on the use of the modified value-at-risk. *Journal of Alternative Investments*, *14*. doi:10.3905/jai.2012.14.4.079

Chopra, V. K. & Ziemba, W. T. [William T.]. (1993). The effect of errors in means, variances, and covariances on optimal portfolio choice. *The Journal of Portfolio Management*, *19*(2), 6–11. doi:10.3905/jpm.1993.409440. eprint: https://jpm.iijournals.com/content/19/2/6.full.pdf

Connor, G. (1995). The three types of factor models: A comparison of their explanatory power. *Financial Analysts Journal*, *51*(3), 42–46. Retrieved from http://www.jstor.org/stable/4479845

Cornish, E. A. & Fisher, R. A. (1938). Moments and cumulants in the specification of distributions. *Revue de l'Institut International de Statistique / Review of the International Statistical Institute*, *5*(4), 307–320. Retrieved from http://www.jstor.org/stable/1400905

Cover, T. M. & Thomas, J. A. (2006). *Elements of information theory (wiley series in telecommunications and signal processing)*. New York, NY, USA: Wiley-Interscience.

DeMiguel, V., Garlappi, L., Nogales, F. J., & Uppal, R. (2009). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science*, *55*(5), 798–812. Retrieved from http://www.jstor.org/stable/40539189

Demiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *Review of Financial Studies*, *22*. doi:10.1093/rfs/hhm075

Demiguel, V. & Nogales, F. (2009). Portfolio selection with robust estimation. *Operations Research*, *57*, 560–577. doi:10.2139/ssrn.911596

F. Fama Sr, E. & French, K. (1996). Multifactor explanation of asset pricing anomalies. *Journal of Finance*, *51*, 55–84. doi:10.1111/j.1540-6261.1996.tb05202.x

Favre, L. & Galeano, J.-A. (2002). Mean-modified value-at-risk optimization with hedge funds. *The Journal of Alternative Investments*, *5*(2), 21–25. doi:10.3905/jai.2002.319052. eprint: https://jai.iijournals.com/content/5/2/21.full.pdf

Gregoriou, G. & Gueyie, J.-P. (2003). Risk-adjusted performance of funds of hedge funds using a modified sharpe ratio. *The Journal of Wealth Management*, *6*, 77–83. doi:10.3905/jwm.2003.442378

Hitaj, A., Mercuri, L., & Rroji, E. (2015). Portfolio selection with independent component analysis. *Finance Research Letters*, *15*. doi:10.1016/j.frl.2015.09.005

Hyvarinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, *10*(3), 626–634. doi:10.1109/72.761722

Hyvärinen, A. [A.] & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks*, *13*(4), 411–430. doi:https://doi.org/10.1016/S0893-6080(00)00026-5

Hyvärinen, A. [Aapo], Karhunen, J., & Oja, E. (2001). *Independent component analysis*. doi:10.1002/0471221317

Jagannathan, R. & Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraint helps. *Journal of Finance*, *58*, 1651–1684. doi:10.1111/1540-6261.00580

James, W. & Stein, C. (1961). Estimation with quadratic loss. In *Proceedings of the fourth berkeley symposium on mathematical statistics and probability, volume 1: Contributions to the theory of statistics* (pp. 361–379). Berkeley, Calif.: University of California Press. Retrieved from https://projecteuclid.org/euclid.bsmsp/1200512173

Jarque, C. M. & Bera, A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review / Revue Internationale de Statistique*, *55*(2), 163–172. Retrieved from http://www.jstor.org/stable/1403192

K. Chopra, V. & Ziemba, W. (2016). The effect of errors in means, variances, and covariances on optimal portfolio choice. (Vol. 19, pp. 15–24). doi:10.1142/9789813144385_0002

Kallberg, J. G. & Ziemba, W. T. [W. T.]. (1984). Mis-specifications in portfolio selection problems. In G. Bamberg & K. Spremann (Eds.), *Risk and capital* (pp. 74–87). Berlin, Heidelberg: Springer Berlin Heidelberg.

Kempf, A. & Memmel, C. (2006). Estimating the global minimum variance portfolio. *Schmalenbach Business Review (sbr)*, *58*, 332–348. doi:10.1007/BF03396737

Kozak, S., Nagel, S., & Santosh, S. (2018). Interpreting factor models. *The Journal of Finance*. doi:10.1111/jofi.12612

Lassance, N., DeMiguel, V., & Vrins, F. D. (2018). Optimal portfolio diversification via independent component analysis.

Ledoit, O. & Wolf, M. (2001). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, *10*, 603–621. doi:10.1016/S0927-5398(03)00007-0

Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, *7*(1), 77–91. Retrieved from https://EconPapers.repec.org/RePEc:bla:jfinan:v:7:y:1952:i:1:p:77-91

Martellini, L. & Ziemann, V. (2010a). Improved estimates of higher-order comoments and implications for portfolio selection. *Review of Financial Studies*, *23*, 1467–1502. doi:10.1093/rfs/hhp099

Martellini, L. & Ziemann, V. (2010b). Improved estimates of higher-order comoments and implications for portfolio selection. *Review of Financial Studies*, *23*, 1467–1502. doi:10.1093/rfs/hhp099

Michaud, R. (1989). The markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal - FINANC ANAL J*, *45*, 31–42. doi:10.2469/faj.v45.n1.31

Sahonero, G. & Calderon, H. (2017). A comparison of sobi, fastica, jade and infomax algorithms.

Sharpe, W. F. (1994). The sharpe ratio. *The Journal of Portfolio Management*, *21*(1), 49–58. doi:10.3905/jpm.1994.409501. eprint: https://jpm.iijournals.com/content/21/1/49.full.pdf

Tan, j. (2012). Principal component analysis and portfolio optimization. *SSRN Electronic Journal*. doi:10.2139/ssrn.2213687

Velicer, W. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*, 321–327. doi:10.1007/BF02293557

Vrins, F. & Verleysen, M. (2005). On the entropy minimization of a linear mixture of variables for source separation. *Signal Process. 85*(5), 1029–1044. doi:10.1016/j.sigpro.2004.11.026

Xu, H., Caramanis, C., & Mannor, S. (2016). Statistical Optimization in High Dimensions. *Operations Research*, *64*(4), 958–979. Retrieved from https://ideas.repec.org/a/inm/oropre/v64y2016i4p958-979.html

LOUVAIN-LA-NEUVE | BRUXELLES | MONS | TOURNAI | CHARLEROI | NAMUR
Place des Doyens, 1 bte L2.01.01, 1348 Louvain-la-Neuve, Belgique | **www.uclouvain.be/lsm**