UNIVERSITA' DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE ECONOMICHE ED AZIENDALI "M. FANNO"

DIPARTIMENTO DI SCIENZE STATISTICHE

Corso di Laurea Magistrale in Economics and Finance

"Portfolio allocation with penalized regression for sparse index tracking"

Relatore

Chiar.mo Prof. Massimiliano Caporin

Laureando

Serra Luca

Matricola n°

1149676

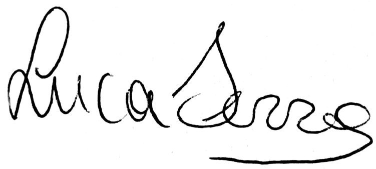
A.A. 2017/2018

Alla mia famiglia...

Declaration of Authorship

Il candidato dichiara che il presente lavoro è originale e non è già stato sottoposto, in tutto o in parte, per il conseguimento di un titolo accademico in altre Università italiane o straniere. Il candidato dichiara altresì che tutti i materiali utilizzati durante la preparazione dell'elaborato sono stati indicati nel testo e nella sezione "Riferimenti bibliografici" e che le eventuali citazioni testuali sono individuabili attraverso l'esplicito richiamo alla pubblicazione originale.

FIRMA DEL CANDIDATO



Abstract

Index tracking is a well-known passive portfolio management strategy that aims to replicate the performance, the holdings and the behavior of a designated index. In this dissertation, we show how, by applying a ℓq-norm regularization to the index tracking optimization problem, it is possible to closely replicate the performances of the benchmark and concurrently promote portfolio sparsity. This permits to reduce transaction costs and avoid illiquid positions using only a small fraction of the index constituents. The empirical analysis on real-world financial data, performed considering the Standard and Poor's 100 Global Index, allows to highlight the validity of the approach both in terms of low transaction costs, but also and above all in terms of tracking accuracy.

KEY WORDS: Sparse Index Tracking, ℓq-norm regularization, Passive portfolio management

Contents

Introduction **1**

1 The Modern Portfolio Theory **5**

1.1 Mean-Variance analysis . . . 5

1.2 Mean-Variance approach instability . . . 8

1.3 GMV Portfolio as a solution for the estimation errors problem . . . 11

1.4 Other solutions for the estimation errors problem . . . 12

2 Portfolio Optimization with Penalized Regression **15**

2.1 The benefits of using penalized regressions . . . 15

2.2 Linear regression model . . . 17

2.3 The Bias-Variance trade-off . . . 18

2.4 Portfolio optimization as a linear regression model . . . 20

2.5 Shrinking weights using penalized regressions . . . 22

2.5.1 The Ridge . . . 22

2.5.2 The Lasso . . . 25

2.5.3 The Elastic Net . . . 27

2.5.4 The q-norm constraint and the non-convex optimization problem . . . 28

3 Index Tracking: a Passive Portfolio Management Strategy **33**

3.1 Active vs passive investing . . . 33

3.1.1 Advantages of passive investing . . . 34

3.1.2 Advantages of active investing . . . 37

3.2 Index tracking . . . 38

3.3 Index tracking using penalized regression . . . 40

4 Experimental Set-up **43**

4.1 Data description . . . 43

4.2 Working environment . . . 45

4.3 Methodology . . . 45

4.4 Performance indicators 50

5 Empirical Evidence **55**

5.1 Results for the static approach 55

5.2 Results for the dynamic approaches 61

5.2.1 Portfolio reallocation at index constituents variation 61

5.2.2 Portfolio reallocation on a monthly basis 66

Conclusions and future works **73**

A Markovitz Optimization **75**

B Figures and Tabels **77**

B.1 Summary table for regularization methods 77

B.2 Data descriptive analysis 77

B.3 Summary statistics for all stocks 78

B.3.1 Grouped by super-sector 78

B.3.2 Grouped by geographical area 79

B.4 Buy and Hold portfolios composition 80

C Matlab Codes **81**

C.1 Example 1 81

C.2 GlobalSearch with static approach 83

C.3 GlobalSearch with dynamic approach: index variation. 86

C.4 GlobalSearch with dynamic approach: monthly reallocation. 92

Bibliography **99**

List of Figures

Figure 1.1: The Markoviitz efficient frontier 7
Figure 1.2: Sensibility of portfolio weights to variation in inputs estimation. 10
Figure 1.3: The Markovitz efficient frontier with GMV portfolio 11

Figure 2.1: Number of inputs needed in the MVO model 16
Figure 2.2: A graphical illustration of bias and variance 18
Figure 2.3: The Bias and Variance contribution to total error 19
Figure 2.4: The graphical solution of Ridge regression 23
Figure 2.5: The graphical solution of LASSO regression 26
Figure 2.6: Two-dimensional contour plots for Ridge, LASSO and Elastic net penalty 28
Figure 2.7: The contours of $\|w\|_q$ for given values of q 29
Figure 2.8: Graphical representation of convex and non-convex sets. 30
Figure 2.9: Graphical representation of convex and non-convex optimization problem 31

Figure 3.1: Ten-year annualized returns for funds in the lowest-cost and highest-cost quartiles. 35
Figure 3.2: Annual percentage change of the S&P 500 index from 1988 to 2017. . 36
Figure 3.3: Cash flow exchanges between passive and active managed funds . . . 37

Figure 4.1: Correlation between S&P 100 Global Index and its constituents 44
Figure 4.2: Number of assets in the tracking portfolios across q and k. 47
Figure 4.3: Distribution of the portfolio IRs across q and k 49

Figure 5.1: Number of Assets and IR distributions with No-Short selling constraint. 56
Figure 5.2: Number of assets and IR distributions with 0.2 Upper Bound. 56
Figure 5.3: Number of assets and IR distributions with 0.1 upper bound. 57
Figure 5.4: Cumulated returns for different Buy and Hold strategies. 59
Figure 5.5: Annualized volatility for different Buy and Hold strategies. 60
Figure 5.6: Legends of the area plots containing the portfolio rotations. 61
Figure 5.7: Rotation of portfolio weights with No-Short selling constraint 62
Figure 5.8: Rotation of portfolio weights with 0.2 upper bound 62
Figure 5.9: Rotation of portfolio weights with 0.1 upper bound 62

Figure 5.10: Cumulated returns obtained reallocating at index variation. 63
Figure 5.11: Annualized volatility for different strategies when reallocating at index variation. 64
Figure 5.12: Approximated turnover considering reallocation at index variation . . . 65
Figure 5.13: Legends of the area plots containing the monthly portfolio Rotations. . 66
Figure 5.14: Monthly rotation of portfolio weights with No-Short selling 67
Figure 5.15: Monthly rotation of portfolio weights with 0.2 upper bound 67
Figure 5.16: Monthly rotation of portfolio weights with 0.1 upper bound 67
Figure 5.17: Cumulated returns obtained with monthly reallocation. 68
Figure 5.18: Annualized volatility for different strategies with monthly reallocation. 69
Figure 5.19: Approximated turnover for strategies with monthly reallocation. 70

List of Tables

Table 1.1: Sensitivity of portfolio weights to perturbation in asset expected returns 9
Table 1.2: Sensitivity of portfolio weights to perturbation in asset variances. 10

Table 3.1: Average expense ratio of different funds until December 2013. 35

Table 4.1: Summary statistics for S&P 100 Global. 44

Table 5.1: Portfolio weights for Buy and Hold strategies. 58
Table 5.2: Summary statistics for Buy and Hold strategies. 60
Table 5.3: Summary statistics for trackers adjusted at index variation gross of transaction costs 64
Table 5.4: Summary statistics for trackers adjusted at index variation net of transaction costs. 65
Table 5.5: Summary statistics in the case of monthly reallocation gross of transaction costs. 69
Table 5.6: Summary statistics in the case of monthly reallocation net of transaction costs. 71

Introduction

Since the classical mean-variance model was introduced by Hanry Markowitz in 1952, the asset allocation problem receives considerable interest both from practitioners and academics. This topic has been widely studied in the financial literature not only in its classical formulation, but more recently it has been extended and adapted in order to reach other alternative goals as, for example, establishing an optimal portfolio allocation in order to replicate the performances of a benchmark index, so-called *"Index tracking"*.

Index tracking can be defined as a type of passive management strategy which consists in designing a portfolio (tracking portfolio or index fund) that replicates the performance and the behavior of a broad market index. The popularity of index funds, as mentioned in Chavez-Bedoya and Bridge (2014), relies on both theoretical and empirical aspects. If the market is supposed to follow the Efficient Market Hypothesis (EMH), stock prices, at any time, reflect all the available information about securities. In this way, theoretically, neither technical nor fundamental analysis can allow active portfolios to produce excess returns with respect to the market. For instance, since the market portfolio captures the efficiency of the market through diversification, it is a reasonable strategy to invest in an index fund. Moreover, many empirical studies have shown how, on average, active portfolio managers do not outperform the major indexes: active management generally incurs in costly research activities and compensation to the fund managers that might be avoided by relying on an index tracking strategy.

In recent years, two main ways to achieve this objective have been developed: full replication and partial replication (Strub & Baumann, 2018). The full replication strategy offers the most intuitive and straightforward solution to the weights selection problem. Indeed, by using this method the initial wealth is invested in all index constituents in the exact proportions as they appear in the index composition. In this way, the constructed portfolio perfectly reflects the trend of the index, having the exact same returns over time. Even if this is the strategy that more than any other minimize the deviation (tracking error) with respect to the benchmark; however, when an index with a large number of securities is taken into consideration, transaction costs dramatically increase and liquidity problems arise if some of the assets to be included cannot be easily bought in the market (Beasley & Meade, 2003).

On the other hand, in the partial replication strategy, also called *"sparse index tracking"*, instead of creating a tracking portfolio containing the same number of securities included in the benchmark, a portfolio with a lower number of assets is constructed. The wealth is allocated only on the securities that provide the most representative sample of the index based on correlations, exposure and risks. In this way, the tracking error will increase on one side (imperfect tracking), but on the other transaction costs are reduced and illiquid positions are avoided.

Usually, the sparse index tracking problem is formulated simply imposing a cardinality constraint to restrict the number of assets (Fastrich, Paterlini, & Winker, 2014), so that it is possible to retrieve a tracking portfolio selecting an optimal subset of benchmark constituents. However, recently, statistical regularization methods have found strong application in the mean-variance portfolio settings (Brodie, Daubechies, De Mol, Giannone, & Loris, 2009; DeMiguel & Garlappi, 2009) in order to promote the construction of sparse portfolios with good out-of-sample properties and restricted turnover.

The aforementioned approaches rely on imposing bounds on the $\ell 2$-norm or on the $\ell 1$-norm of the vector of the portfolio weights as suggested by the Ridge regression (Hoerl and Kennard, 1970) and the LASSO (Tibshirani, 1996) approach, respectively. Empirical results in a mean-variance framework support the use of the LASSO method when short selling is allowed. However, the latter results to be unable in promoting sparsity when the budget and the no short-selling constraint are imposed (typical in index tracking), since the 1-norm of the asset weights will have a constant value of one.

One valid alternative could be to consider a constraint on the q-norm with $0 < q < 1$. The lower the upper bound on the ℓq-norm is, the sparser and less diversified (with larger weights) the portfolios will result. Indeed, in the implementation of the ℓq-norm constraint, the latter should be considered as a measure of diversity of the portfolio. When the no-short selling limitation is imposed, this measure has maximum value for the equally weighted portfolio and minimum value for a portfolio totally invested in a single asset (Fernholz, Garvy, & Hannon, 1998). Therefore, by imposing an upper bound on the q-norm, we might be able to identify the tracking portfolio with the desirable maximum number of assets, but some difficulties might be preliminarily addressed. Because of the presence of a non-convex constraint, the problem is very challenging from an optimization point of view (NP-hard).

In this dissertation, starting from the seminal contribution of Fan (2012) that recast the risk minimization problem as a regression model, we propose a sparse index tracking strategy that allows to minimize a given tracking error measure, namely the TEV, by imposing a ℓq-norm constraint on the portfolio weights. This strategy allows to determine in one single step the number of active positions and their optimal weights. Moreover, considering the *S&P 100 Global*, we highlight the validity of the approach in providing portfolio sparsity, reducing transaction costs and guaranteeing high tracking accuracy with respect to the main index.

The remainder of the paper is organized as follows. In Chapter 1, we present an overview of the Markowitz Portfolio Theory deepening in its functioning, analytical formulation and some issues arising from its implementation. In Chapter 2, we describe the use of penalized regression models (Ridge, Lasso, Elastic Net and ℓq-norm) as a remedy for the error estimation and variables selection problems. In Chapter 3, we introduce the general concepts of passive portfolio management and index tracking, focusing on the benefits that the implementation of a penalized regression in this particular context might produce. In Chapter 4, we present the experimental set-up for implementation of the empirical analysis. In Chapter 5, we present the empirical results for the *S&P 100 Global* considering different types of constraints.

Chapter 1

The Modern Portfolio Theory

In this Chapter, a brief introduction and overview of the Markowitz Portfolio Theory are going to be provided; underlying its functioning, analytical formulation, but also some problematics arising from its implementation.

1.1 Mean-Variance analysis

In 1952 Hanry Markowitz, professor of finance the University of California, published an article in "The Journal of Finance", titled "Portfolio Selection", that would have introduced probably the most influential theory in the practice of portfolio management. This theory, better known as *Modern Portfolio Theory* (MPT) or *Markowitz Mean-Variance Model* (MV), has been widely used, from its development, as a framework for optimal portfolio selection in active asset allocation.

This classical model has the aim to determine, analyze and evaluate the optimal investment decision of a rational agent[1], willing to allocate his wealth across N financial assets under a basic assumption: return distribution can be completely characterized by its first two moments. Therefore, the model seeks a portfolio weight vector which allows obtaining the highest expected return at a given level of portfolio risk, defined as the variance of the portfolio return. Similarly, for a given level of expected return, the investor would choose the portfolio with the lowest risk. In other words, the MV model characterizes a series of efficient portfolios (MV Efficient), that minimize the portfolio risk for each level of portfolio return, ruling out all the others with higher volatility.

Suppose an investor desires to allocate his wealth in a portfolio that contains N assets, having

- $r = (r_1, ..., r_N)$ the vector of asset returns where r_i is the return of ith asset

[1] In economics, game theory, decision theory, a rational agent is an agent that has clear preferences and always chooses to perform the action with the optimal expected outcome for itself from among all feasible actions.

- $w = (w_1, ..., w_N)$ the vector of weights of portfolio where w_i, is the weight of ith asset

- $\mu(w) = \mathbf{E}[r_w]$ the vector of the expected returns of assets

The portfolio expected return μ_p can be defined as the weighted average of its individual asset expected returns:

$$\mu_p = \sum_{i=1}^{N} \mu_i w_i = w'\mu \tag{1.1}$$

and the portfolio variance results to be:

$$\begin{aligned} \sigma^2(w) &= \mathbf{E}[(r(w) - \mu(w))(r(w) - \mu(w))'] \\ &= \mathbf{E}[(w'r - w'\mu)(w'r - w'\mu)'] \\ &= \mathbf{E}[(w'(r - \mu)(r - \mu)w'] \\ &= \mathbf{w}'E[(r - \mu)(r - \mu)]w' \\ &= w'\Sigma w\ . \end{aligned} \tag{1.2}$$

Σ represents a N-by-N positive semi-definite matrix containing variance and covariance of individual assets,

$$\Sigma = \begin{bmatrix} \sigma_{(1,1)} & \cdots & \sigma_{(1,N)} \\ \vdots & \ddots & \vdots \\ \sigma_{(N,1)} & \cdots & \sigma_{(N,N)} \end{bmatrix}$$

with σ_{ii} being the variance on the return of asset i, and σij being the covariance between the returns of asset i and j. Equation (1.2) just represents the objective function that has to be minimized in order to obtain the set of portfolios with the minimum level of risk for given levels of expected returns.

In minimizing this function two important constraints must hold. First, as already mentioned, the expected return has to be fixed and given. On the other hand, the admissibility of the portfolio must be guaranteed; meaning that it is not possible to invest more or less than the available total wealth (*Budget Constraint*). These two constraints could be easily expressed mathematically as:

$$w'\mu = \mu_p \qquad \text{and} \qquad w'1_N = 1$$

and the MV optimization problem, can at this point be defined as follow[2]:

$$\begin{aligned} \min_{w} \quad & w'\Sigma w \\ s.t. \quad & w'\mu = \mu_p \\ & w'1_N = 1 \ . \end{aligned} \tag{1.3}$$

After some mathematical steps, shown in Appendix A, Equation 1.3 is solved and the expression for the so-called *"Efficient frontier"* is obtained:

$$\sigma^2 = \frac{1}{ac - b^2}(c\mu_p^2 - 2b\mu_p + a) \tag{1.4}$$

The Efficient frontier, introduced again for the first time by Markovitz, is defined as the set of all efficient portfolios. Any rational investor, using mean-variance analysis, would choose a portfolio on the efficient frontier that suits its risk preference (Markowitz, 1952). For any level of risk/return, the corresponding point on the efficient frontier denotes the portfolio which has the maximum return at specified risk, or minimum volatility at given expected return. The Efficient frontier is shaped in the plane (μ_p, σ_p^2) as a parabola, but following standard practices, it is usually represented the space (σ_p^2, μ_p) by the portion of the right branch of the hyperbola lying above the vertex, as shown in Figure 1.1.

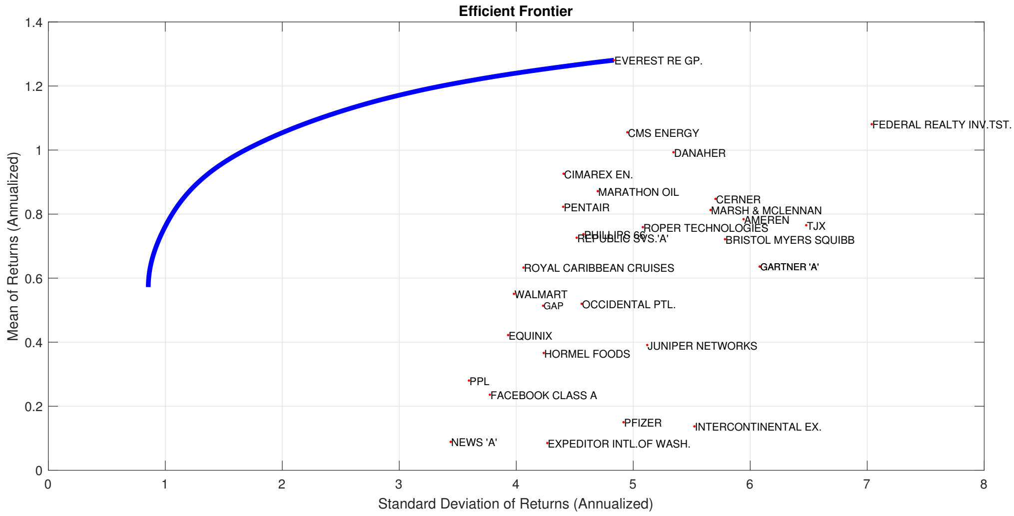


Figure 1.1: The Markovitz efficient frontier[3].

The area below and to the right of efficient frontier represents the feasible area including

[2] 1_N is an N-dimensional vector of ones.

[3] The Figure represents the no short-selling MV efficient frontier constructed taking, as investment universe, 30 random assets of the S&P 500 picked.

all portfolios that can be constructed yet not efficient, while portfolios above and to the left are impossible by construction.

Even if the MV model acts as a powerful tool to transform the views of the portfolio manager into investment decisions, it has suffered during the years a lot of criticism from both academics and practitioners.

1.2 Mean-Variance approach instability

The first bundle of criticism moved against the MV approach refers to the strong assumptions on which it relies, but that do not properly represent the reality. In particular, asset returns are supposed to follow a Normal distribution. In reality, instead, empirical evidence proves that asset returns follow a leptokurtic distribution or heavy-tailed distribution (Sheikh & Qiao, 2009)[4]. In this way, in order to describe the distribution of assets returns, is not enough to know their means and the variances.

Furthermore, Markowitz's theory assumes that investors are rational and avoid risk when possible, that there are not large enough investors to influence market prices, and that investors have unlimited access in borrowing and lending money at the risk-free interest rate. In reality, though, many professional investors and academics have observed that:

- investors habitually fail to consider and correctly interpret all the relevant information that might drive their investment decisions;
- institutional barriers prevent them from acting on certain information;
- even with all the relevant information at disposal, they persist in making irrational choices.

Another clue point that has been largely criticized by academics (Michaud, 1989; Jorion, 1992; Broadie, 1993; Ledoit & Wolf, 2003) or more recently (DeMiguel, Martin-Utrera, & Nogales, 2013), is that the estimation procedure for the input parameters of the optimization program necessarily introduces sources of estimation error and instability in the optimal solution. On this topic Dickinson (1974), Jobson and Korkie (1980), and Frost and Savarino (1988) already provided strong empirical evidence that testifies the unreliability of the estimates of expected returns and variances.

For this reasons, all investors involved into the Mean-Variance approach need to carefully evaluate not only the market risk caused by fluctuations in macroeconomic factors but also the estimation risk depending both on the allocation method used and on the number of parameters taken into consideration. On the whole, it has been proved that the largest source of estimation

[4]More information available on: *https://am.jpmorgan.com/blobcontent/*1383169198442/83456/11_438.*pdf*

error comes from the estimation of expected returns, while the estimation of covariance is more stable and causes fewer concerns in term of accuracy (Merton, 1980).

In order to clarify which is the impact of a small perturbation of input parameters on the mean-variance portfolio, the following example is proposed.

Example 1

Consider, for simplicity, an investment universe of four assets. The expected returns, randomly chosen, are $\mu_1 = 4\%$, $\mu_2 = 5\%$, $\mu_3 = 6\%$, $\mu_4 = 7\%$, the variances of the assets are equal to $\sigma_1^2 = 13\%$, $\sigma_2^2 = 15\%$, $\sigma_3^2 = 16\%$, $\sigma_4^2 = 17\%$, and suppose that the correlations between them are $\rho_{i,j} = 60\%$ (constant for all the assets). Using these input parameters, the unconstrained Mean-Variance problem is solved (Appendix C.1) and the following optimal portfolio weights are retrieved[5]:

$$w_1 = -12.86\%; \quad w_2 = 12.83\%; \quad w_3 = 38.61\%; \quad w_4 = 62.12\%$$

Now, in order to show the sensitivity of the optimal solution with respect to a small perturbation in the input parameters, the same problem is solved by including, one at the time, variations (increase) of 1% in the values of expected returns. The variation of weights, with respect to the starting Max Sharpe portfolio, is at this point calculated.

Table 1.1 clearly shows how, even a small change in the estimation of asset expected returns, induces a dramatic change in the portfolio weights, that, however, is not accompanied by an equal variation of portfolio returns. In fact, despite the changes in portfolio weights, the expected performance of the portfolio remains almost stable.

	Δw_1	Δw_2	Δw_3	Δw_4	$\Delta wTotal$	$\Delta Port.Ret$
Case 1	-231%	-68%	-26%	-19%	343%	7%
Case 2	54%	207%	-22%	-15%	298%	2%
Case 3	54%	-60%	59%	-14%	187%	11%
Case 4	53%	-58%	20%	35%	165%	19%

Table 1.1: Sensitivity of portfolio Weights to perturbation in asset expected returns[6].

[5]The values are the ones referred to the Max Sharpe portfolio defined as the portfolio, on the efficient frontier, settled at the point in which th line drawn from the point [0, risk-free rate] is tangent to the efficient frontier.

[6]The first column represents the different iteration in which time by time the return of one assets is modified by 1%. For example in "Case 3" the return of the third asset is increased by 1% and so on.

At this point, in order to investigate which is the sensibility of the same portfolio to a perturbation of the variance of assets, a similar analysis is conducted, considering this time variations (increase) of 1% in the assets volatility (Table 1.2).

	Δw_1	Δw_2	Δw_3	Δw_4	$\Delta wTotal$
Case 1	31%	10%	3%	2%	47%
Case 2	-9%	-37%	4%	3%	52%
Case 3	-9%	12%	-13%	4%	38%
Case 4	-9%	14%	6%	-8%	37%

Table 1.2: Sensitivity of portfolio weights to perturbation in asset variances.

The results, reported in Table 1.1 and Table 1.2, confirm on one hand what has been already obtained by Merton: the covariance estimation is more stable and causes fewer problems in term of portfolio weights variation. On the other side, they prove how the stability of the allocation is actually a real problem, justifying in some sense the hypothesis, proposed by Michaud, that Mean-Variance maximization is, indeed, an "error maximization" model:

> *"MV optimization significantly overweights (underweights) those securities that have large (small) estimated returns, negative (positive) correlations and small (large) variances. These securities are, of course, the ones most likely to have large estimation errors"* (Michaud, 1989, pp. 33).

Exploiting the former evidence, that the estimation of input returns has a much greater impact on the variation of portfolio weights (Figure 1.2), expertises came up with a simple but clever methodology to reduce the estimation risk, consisting in completely avoid the estimation of expected returns.

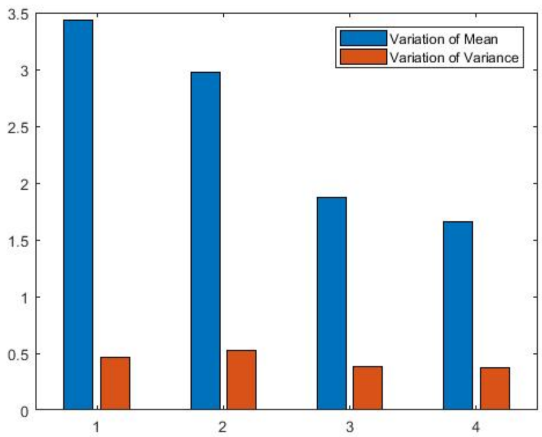


Figure 1.2: Sensibility of portfolio weights to variation in inputs estimation.

1.3 GMV Portfolio as a solution for the estimation errors problem

Expected stock returns are hard to estimate and lead to estimation errors that might result in the selection of non-optimal portfolios. In fact, after estimating the input parameters, the optimization is performed as if these quantities were perfectly certain, without considering that estimation errors are introduced into the allocation process.

A quite large group of researchers have suggested to neglect the estimation of expected returns, by relying only on the covariance structure and assuming that all stocks have equal expected returns (Chan, Karceski, & Lakonishok, 1999; Jagannathan & Ma, 2003; Ledoit & Wolf, 2003; DeMiguel & Nogales, 2009; Fan et al., 2012; Fernandes, Rocha, & Souza, 2012; Behr, Guettler, & Truebenbach, 2012). Under this assumption, all the assets differ only in term of their risk and, therefore, the only possible efficient portfolio is the one having the smallest risk: the so-called global minimum variance portfolio. The global minimum variance portfolio (GMV) is the only portfolio, located on the efficient frontier, that does not depend on any assumptions about expected returns, but its composition just depends on the covariance matrix of the assets. Graphically the GMV corresponds to the portfolio on the efficient frontier with the lowest volatility and, thus, is located at the vertex of the hyperbola as shown in Figure 1.3:

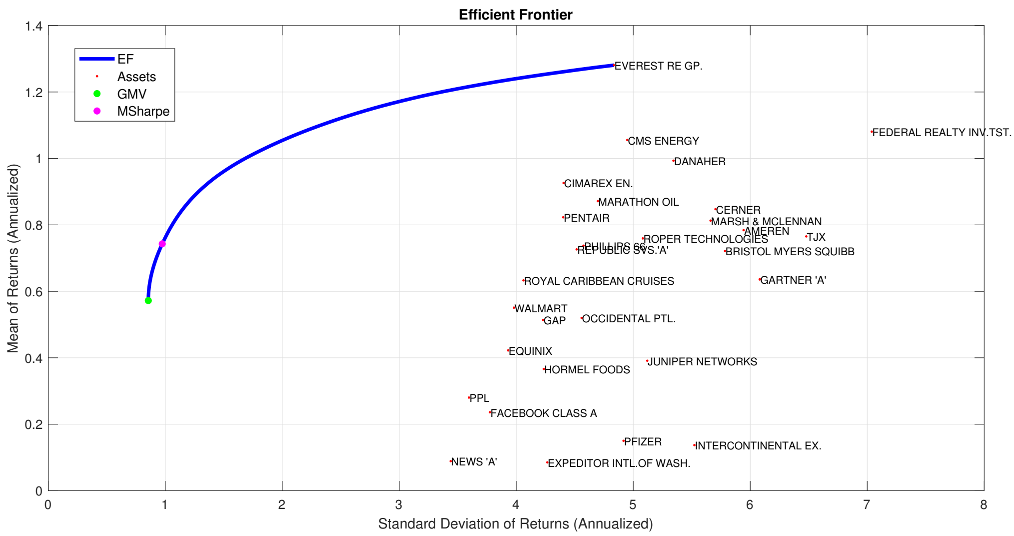


Figure 1.3: Markovitz efficient frontier with GMV portfolio.

Since the covariance matrix can be estimated much more precisely, compared to the expected returns, the estimation risk of the investor is expected to be strongly reduced. The GMV portfolio can be calculated by minimizing the portfolio variance, this time considering only the

"Budget Constraint". The optimization problem in this case becomes:

$$\begin{aligned} \min_{w} \quad & w'\Sigma w \\ s.t. \quad & w'1_N = 1 \end{aligned} \tag{1.5}$$

and its solution end to be

$$w_{GMV} = \frac{\Sigma^{-1}1_N}{1_N'\Sigma^{-1}1_N} \tag{1.6}$$

It is important to notice that the choice of using the GMV portfolio is not a limiting one. Past literature has already shown how this portfolio is characterized by an out-of-sample Sharpe ratio[7] which is as good as that of other efficient portfolios (Ingersoll, 1987; Jorion, 1985).

Despite the use of the GMV portfolio is widely spread as a remedy for the estimation error problem, in reality, it is still not able to sufficiently reduce the former issue (Chan et al., 1999; Jagannathan & Ma, 2003; Ledoit & Wolf, 2003).

1.4 Other solutions for the estimation errors problem

Although the introduction of the GMV portfolio had represented a first step in solving the problem of estimation errors, however, many academics and practitioners continued to look for other more sophisticated and more effective solutions.

Some of them consist in reducing the estimation errors of the input parameters by using econometric methods. For instance, Michaud (1989) propose the uses of the so-called *"Resampling approach"*, that consists in averaging many realizations of optimized MV solutions, with the aim to improve out-of-sample performance thanks to statistical diversification. Unfortunately, this procedure has no economic justification on its behind and the Resampled efficient portfolio is not mean-variance efficient anymore, by definition. Alongside, Black and Litterman propose to overcome the problem of unintuitive, highly-concentrated and input-sensitive portfolios, using the *Black-Litterman asset allocation model* (Black & Litterman, 1992). This equilibrium model[8] uses a Bayesian approach to combine the subjective views of investors, regarding the expected returns, with the market equilibrium vector of expected return; forming a new, mixed estimate of expected returns.

[7] It represent a key portfolio performance measure. It is defined as the ratio between the portfolio return and its standard deviation.

[8] Equilibrium ,according to Litterman ("Modern Investment Management- An Equilibrium Approach", Litterman et.al.2003), is an idealized state in which supply equals demand. The market is not assumed to be in equilibrium, but equilibrium is viewed as a "center of gravity": the market deviates from this state, but there are forces pushing it towards the equilibrium.

More recently, instead of using econometric methods, some other approaches aimed to directly shrink the portfolio weights using weight bounds, penalization of the objective function or regularization of input parameters. Jagannathan and Ma (2003) have shown that imposing constraints on the mean-variance optimization can be interpreted as a modification of the covariance matrix. In particular, lower bounds (upper bounds) decrease (increase) asset return volatilities. Constraints on weights reduce the degree of freedom of the optimization and the allocation is forced to remain in certain intervals.

Despite all these solutions have been elaborated during the years, the correction of estimation errors is such a difficult task that several studies still promote the use of heuristic[9] allocation strategy, since they perform better than the MV allocation in terms of Sharpe Ratio. In this respect, investing in the *"Equally Weighted portfolio"* (EW) might represent a much more appealing alternative (DeMiguel & Nogales, 2009).

[9]It defined heuristic, any approach to problem solving that employs a practical method, not guaranteed to be optimal, perfect, logical, or rational, but instead sufficient for reaching an immediate goal.

Chapter 2

Portfolio Optimization with Penalized Regression

In this Chapter, the use of penalized regressions as a remedy for error estimation and variable selection problem is presented. Starting from the seminal contribution of Fan, the reformulation of the MV optimization problem in the form of linear regression is described, proposing, in addition, a deepening on the most used penalized regression models.

2.1 The benefits of using penalized regressions

As already shown in the previous Chapter, financial literature has largely proved how the use of sample estimates can hardly provide reliable out-of-sample performances in asset allocation for practical implementation. In addition to that, in the actual financial environment, portfolio managers are very interested in having the opportunity to select the portfolio constituents among a large bulk of alternatives. This, in fact, would allow them to not preliminarily restrict their investment opportunities and, moreover, it would guarantee the possibility to diversify the risk exposure among different assets.

Unfortunately, exploiting this potential is very difficult, since historical data provide a noisy estimation of the future. As already observed by Shrerer (2002), this noise dramatically rises when the number of securities included in the investment universe N increases, relative to the number of observations T, causing a worsening in the out of the sample performance of the portfolio. Thus, if, as it often happens, the number of assets N is large compared to T, only an insufficient amount of data is available to precisely estimate the parameters needed for the implementation of the MV optimization. In this sense, Ledoit and Wolf (2003) remark that relevant problems in covariance matrix estimation, such as severe estimation errors and numerical instability, may occur whenever T is not at least ten times larger than N[1].

Just to give an idea, if the universe of assets considered is composed of n risky assets, then n expected returns, n volatilities, and $\frac{1}{2}(n-1)$ correlations are needed to be known. For example, with 100 assets in the investment universe, 5150 parameters have to be estimated, as shown in Figure 2.1, and so an enormous number of observations is requested.

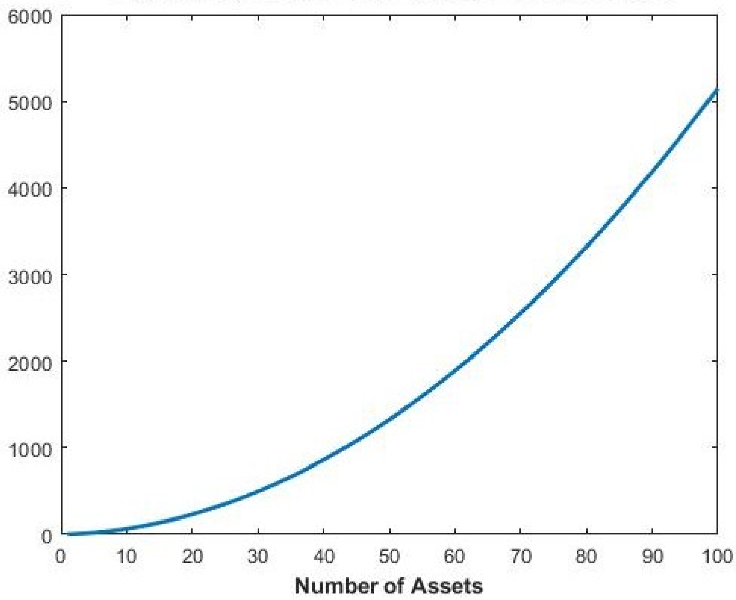


Figure 2.1: Number of inputs needed in the MVO model

In order to simultaneously solve the problem of estimation errors, guarantee a better out-of-sample performances and allow managers to choose from a large universe of assets and obtain, in this way, the benefits of diversification, one research stream has recently focused on shrinking optimal portfolio weights by using regularization methods[2]. These methods have already been largely used in statistics when, dealing with regression models, standard linear models perform poorly in the presence of large multivariate datasets containing a number of variables superior to the number of observation. In these contexts, *"penalized regressions"*, might represent advantageous alternatives: they give the opportunity, through the use of specific constraints, to create linear regression problems that are penalized for having too many explanatory variables (James, Witten, Hastie, & Tibshirani, 2013; Bruce & Bruce, 2017).

The consequence of imposing this penalty is to reduce (i.e. shrink) the coefficient values towards zero, pushing the less contributive variables to have coefficients close or equal to zero. This constraint or penalty on the size of the regression coefficients induces, on one side, an increase in the bias of the model, but on the other side it also reduces the overall prediction

[1]Note that sample covariance matrix estimates are singular when $T < N$.

[2]In mathematics, statistics, and computer science, particularly in the fields of machine learning, regularization is a process of introducing additional information in order to solve an ill-posed problem or to prevent over-fitting.

error and the variance of the estimated coefficients. This property, called Bias-Variance trade-off, result to be crucial in understanding the sense of using penalized regressions.

Suppose to have a statistical modeling problem with many possible predictors and your goal is to find a simple model that has also good predictive performance. In this case, a model with fewer predictor variables is desirable: it is easy to interpret and gives the opportunity to better understand the underlying process that generates the data.

2.2 Linear regression model

A linear regression model is a linear approach allowing to shape the relationship between a response (dependent) variable Y and one or more explanatory (independent) variables $X_1, X_2, ..., X_p$, plus a random noise.

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p + \varepsilon$$

where $\beta_0, \beta_1, ..., \beta_p$ are the regression parameters and ε is the error term.

Usually, linear models are preferred to more complicated statistical models, since it is possible to fit them relatively easily and, in many practical situations, they provide simpler models with good predictive performance (Friedman, Hastie, & Tibshirani, 2001). Moreover, linearity with respect to fixed functions of the predictors is often an adequate first approximation to more complex behaviors. Other methods, such as nonlinear or nonparametric models, often fail as the number of variables increase, becoming very complicated in term of computation and interpretability.

The most common method used to estimate regression coefficients for a linear model is the Least Squares estimation. It finds the coefficients β that minimize the Residual Sum of Squares (RSS):

$$RSS = \mathbf{E}[Y - \beta_0 - \beta_1 X_1 - ... - \beta_p X_p]^2$$

According to the Gauss-Markov theorem, the least squares estimate has the smallest variance among all linear unbiased estimates under certain assumptions[3]. However, if there are correlated predictor variables and their number is quite large, some of these assumptions are violated. As result, least squares estimates become highly variable (unstable) and the resulting model exhibits poor predictive performance.

Therefore, the least squares estimation does not necessarily lead to a simple model that identifies the parameter of interest when there are many predictors. In these cases, the use of penalized regressions might be preferable.

[3] The Gauss-Markov theorem states that in a linear regression model in which the errors have expectation equals to zero, are uncorrelated and have equal variances (Homoskedasticity), the best linear unbiased estimator (BLUE) of the coefficients is given by the ordinary least squares (OLS) estimator, provided it exists.

2.3 The Bias-Variance trade-off

The bias-variance trade-off is the property of a set of predictive models whereby models with lower bias in parameter estimation, have higher variance in the estimated parameter across samples, and vice versa. Thus, the bias-variance dilemma constitutes a real problem for who is trying to simultaneously minimize these two sources of error, that differently affect the total estimation error of the regression problem.

The *error due to bias* is taken as the difference between, the expected (or average) prediction obtained by the model and the correct value which we are trying to predict. If the model building process is repeated more than once, a range of predictor will be obtained. The bias measures how far these predictions are from the correct value. On the other side, the *error due to variance* is taken as the variability of a model prediction for a given data point. Again imagine it is possible to repeat the entire model building process multiple times. The variance indicates how much the predictions for each given point vary between different realizations of the model.

In order to give a more clear idea of what this two sources of error are, Figure 2.2 proposes a graphical visualization of bias and variance using bulls-eye diagrams. Imagine that the center of the target represents the model that perfectly predicts the correct values and that, moving away from the bulls-eye, the predictions get worse and worse. Sometimes it is possible to have a good distribution of the data so that the prediction power is high and the points are located close to the center, while sometimes the data might be full of outliers or non-standard values, resulting in poorer predictive performances with high variance. In this case, different realizations are scattered among the target (High Bias and High Variance).

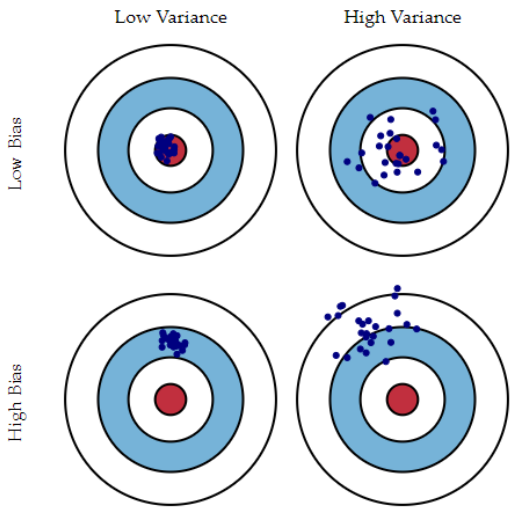


Figure 2.2: Graphical illustration of bias and variance[4].

[4]Source: Understanding the Bias-Variance trade-off, http://scott.fortmann-roe.com/docs/BiasVariance.html

The bias-variance trade-off can be also described through the use of a mathematical formulation, since, as aforementioned, summing these two quantities the total estimation error of the model, usually expressed by the mean square error (MSE)[5], is obtained:

$$MSE(x) = \mathbf{E}\left[(Y - \hat{f}(x))^2\right]$$
$$MSE(x) = \left(\mathbf{E}[\hat{f}(x)] - f(x)\right)^2 + \mathbf{E}\left[\hat{f}(x) - \mathbf{E}[\hat{f}(x)]\right]^2 + \sigma_e^2$$
$$MSE(x) = Bias^2 + Variance + Irreducible\ Error.$$

In this case, $\hat{f}(x)$ is the estimated model of $f(x)$ and the third term σ_e^2 represents the noise term in the true relationship, that fundamentally cannot be reduced by any model.

In the case in which unbiased estimators are used, the MSE and variance are equivalent ($Bias = 0$). Instead, if penalized regression methods are considered, the bias grows on one side, since distortion in coefficient estimation is introduced by continuously shrinking the regression coefficients, and on the other side the variance is reduced in relation to model complexity. As more and more parameters are shrunk toward zero, the complexity and the variance of the model fall and so the bias becomes the primary concern (Figure 2.3).

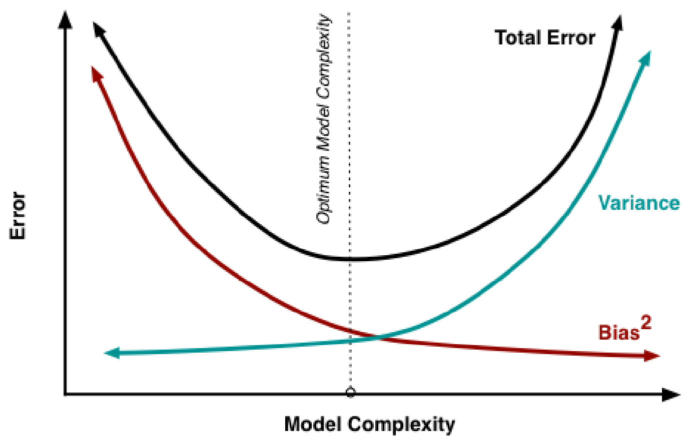


Figure 2.3: Bias and Variance contribution to total error[6].

Often, when using penalized regressions, the increase in bias is lower than the decrease in variance; hence the resulting model would have a smaller MSE than the unbiased Least Squared estimator. Therefore, penalized regression methods can produce models that have stronger predictive performances than standard linear regression models (Gunes, 2015).

[5]The MSE or mean squared deviation of an estimator measures the average squared difference between the estimated values and observed value. The MSE is a measure of the quality of an estimator, it is always non-negative, and the smaller its value is, the more precise its estimation will be.

[6]Source: Understanding the Bias-Variance trade-off, http://scott.fortmann-roe.com/docs/BiasVariance.html

2.4 Portfolio optimization as a linear regression model

The introduction of the concepts of linear regressions and penalized regressions might, at first, appear curious and inappropriate in an elaborated that is trying to deal with the problematics arising from the Modern Portfolio Theory. However, in reality, past literature has already shown how a link between these two topics exists. In fact, the risk minimization problem can be easily recast as a regression problem in which the estimated coefficients correspond to the optimal portfolio weights (Efron, Hastie, Johnstone, Tibshirani, et al., 2004).

Suppose to have the opportunity to invest in n assets with returns $R_1,...,R_n$. Let R be the row vector of asset returns, Σ be their covariance matrix and w the portfolio allocation vector satisfying the budget constraint $w'1 = 1$. As shown by Fan, Zhang and Yu in the article *"Vast Portfolio Selection With Gross-Exposure Constraints"* [7] the risk minimization problem

$$\begin{aligned} \min Var(R_p) = \min Var(w'R) = \min_{w} \quad & w'\Sigma w \\ s.t. \quad & w'1 = 1 \end{aligned} \tag{2.1}$$

can be assimilated to a regression problem that implicitly includes the budget constraint.

Starting from the definition of portfolio return

$$R_p = w'R = w_1R_1 + w_2R_2 + ... + w_{n-1}R_{n-1} + w_nR_n \tag{2.2}$$

if the quantity $t = w_1R_n + w_2R_n + ... + w_{n-1}R_n$ is subtracted and added, and the portfolio weights w are collected, it is possible to obtain that

$$R_p = w_1(R_1 - R_n) + w_2(R_2 - R_n) + ... + w_{n-1}(R_{n-1} - R_n) + R_n(w_1 + w_2 + ... + w_{n-1} + w_n).$$

Considering that portfolio weights must sum to one, in order to respect the budget constraint, a companion representation is derived

$$\begin{aligned} R_p &= w_1(R_1 - R_n) + w_2(R_2 - R_n) + ... + w_{n-1}(R_{n-1} - R_n) + R_n \\ R_p &= R_n - w_1(R_n - R_1) + w_2(R_n - R_2) + ... + w_{n-1}(R_n - R_{n-1}) \end{aligned} \tag{2.3}$$

and, by replacing $R^* = R_n - R_1$ [8], it is possible to show that the portfolio return can be reformu-

[7] Source: Fan, J., Zhang, J. & Yu, K.(2012). Vast portfolio selection with gross-exposure constraints.*Journal of the American Statistical Association*, 107(498), 592-606.

[8] The vector R^* represents the deviation in returns for each asset with respect to the numeraire security.

lated as

$$R_p = w_1 R_1^* + w_2 R_2^* + ... + w_{n-1} R_n - 1^* + R_n$$
$$R_p = R_n - w_1 R_1^* + w_2 R_2^* + ... + w_{n-1} R_{n-1}^* \tag{2.4}$$
$$R_p = R_n - w' R^*$$

where R_n represents a randomly picked numeraire asset and R^* indicates the vector of the $n-1$ asset returns deviations.

At this point, coming back to the formulation of the risk minimization problem, it is possible to substitute the portfolio return (2.4) in (2.1), obtaining

$$\min_{w \in R^n} w' \Sigma w = \min_{w_{-n} \in R_n} Var(R_n - w' R^*) = \min_{w_{-n} \in R_n} \mathbf{E}[R_n - w' R^*]^2 \tag{2.5}$$

where w_{-n} denotes the vector of weights excluding w_n. The weight of the last asset, taken as numeraire, could be easily determined by imposing that the budget constraint is satisfied

$$w_n = 1 - \sum_{i=1}^{n-1} w_i. \tag{2.6}$$

In Equation (2.5), $\mathbf{E}[R_n - w' R^*]^2$ corresponds to the variance of the errors for the linear regression of the return of asset n (R_n) with respect to R^*, without taking into account the effect due to the intercept that, from a risk minimization point of view, have to be managed by using centered returns[9]. Furthermore, estimating the coefficients $w_1, ..., w_n$, is equivalent of finding the GMV portfolio weights (Fan et al., 2012) and even if the numeraire asset (response variable of the model) is randomly chosen, this does not affect the resulting estimated coefficients (Bonaccolto, Caporin, & Paterlini, 2018).

Therefore, to conclude, it is possible to minimize the quantity $w' \Sigma w$ by minimizing the residual sum of squared of a linear regression model, with response variable R_n and covariates $R_i^*, ..., R_{n-1}^*$. However, since the MV optimization framework could present a large number of assets (regressors), that strongly correlates one to each other, the benefits of using one of the different type of penalized regressions, already introduced by past literature, might be exploited.

[9]In fact, the proper equation for the variance of errors should be $\mathbf{E}[R_n - w' R^* - b]^2$. However, since it is preferable to have a model being independent of the mean of the Y vector, all the variables are centered with respect to their mean.

2.5 Shrinking weights using penalized regressions

2.5.1 The Ridge

The Ridge regression, introduced by Hoerl and Kennard (1988) as a more stable alternative to the standard least squares estimator, is a remedial measure widely used in literature to alleviate the multicollinearity problem amongst regression predictor variables in a model. When multicollinearity occurs, even if least squares estimates are considered to be still unbiased, their variances result to be very large and the estimated parameters might be far from the true values. In fact, since the regression variables used in a model might be sometimes highly correlated, the regression coefficient of any variable strongly depends on which other predictor variables are included into the model, and which ones are left out. In this way, the prediction variables are not able to reflect any inherent effect of that particular predictor on the response variable, but only a partial or marginal effect. For example, a wildly large positive coefficient on one variable can be canceled out by a similarly large negative coefficient on its correlated cousin (Trevor, Robert, & JH, 2009). This creates a model with high variance, that becomes an increasingly unrealistic model as the correlation increases.

Ridge regression produces models with lower variance and shrinks the regression coefficients by adding a positive constant (penalty), proportional to the squared values of w, to the usual linear regression problem[10]. The Ridge coefficients minimize the penalized residual sum of squares,

$$\hat{w}^{Ridge} = \min_{w} \quad \sum_{i=1}^{N}(y_i - w_0 - \sum_{j=1}^{p} w_j x_{ij})^2 + \lambda \sum_{j=}^{p} w_j^2. \tag{2.7}$$

In equation (2.7), $\lambda \geq 0$ represents the complexity parameter that controls the amount of shrinkage: the larger the value of λ, the greater is the amount of shrinkage imposed on the regression parameters.

While in the latter equation the Ridge regression is expressed in its Lagrangian form, with λ as the tuning parameter, another equivalent and interesting way to formulate the problem, is by explicating the size constraint:

$$\begin{aligned} \hat{w}^{Ridge} = \min_{w} \quad & \sum_{i=1}^{N}(y_i - w_0 - \sum_{j=1}^{p} w_j x_{ij})^2. \\ s.t. \quad & \|w\|_2^2 \leq t^2 \end{aligned} \tag{2.8}$$

where $\|w\|_2 = (\sum_{j=1}^{p} w_j^2)^{\frac{1}{2}}$ is the Euclidean norm[11] of the vector of weights. Notice that there is a one-to-one correspondence between the parameters λ in (2.7) and t in (2.8).

[10] This is the reason why the ridge regression is also referred to as ℓ-2 regularization.

[11] It is a function that assigns to each vector in a vector space, excluding the zero vector, a strictly positive length.

In both cases, the coefficients are shrunk toward zero, but they never reach the null value (no variable selection is performed). In order to clarify why this happens, a geometrical representation and an example might be used. Let $f(w)$ be the objective function in (2.8) and consider the counter plot of this function in the bi-dimensional space (w_1, w_2)[12], as shown in red in Figure 2.4.

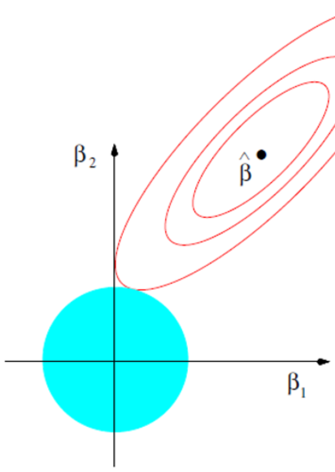


Figure 2.4: Estimation picture for the Ridge regression. Shown are contours of the error and constraint functions. The solid blue area is the constraint region $w_1^2 + w_2^2 \leq t$ .While the red ellipse is the contour of the least squares error function. [13]

The center of the contour represents the minimum of the objective function, given by the non-penalized solution of $f(x)$ obtained imposing no constraint on the parameters ($\lambda = 0$). Suppose now to add a different objective $g(w) = \lambda(w_1^2 + w_2^2)$, representing the ℓ-2 constraint imposed, which contour plot is given in blue. This contour plot shows all the admissible combinations of w_1 and w_2, with a distance from the center that is no more than the parameter t chosen in (2.5). So the larger t is, the greater the admissible area will be or, considering the Lagrangian form, the larger λ is, the faster $g(x)$ will grow and narrower the contour plot will result. At this point, the whole problem in (2.8) can be recast as the minimization of the sum of this two objectives: $f(w) + g(w)$. The solution to this problem is achieved, graphically, when the two contour plots meet each other. In fact, even if the final objective is to minimize the error function, only the values of w included in the blue area are allowed to be accepted. In some sense, by not allowing the w of getting too big, it is possible to keep the variance under control.

[12]In order to have a simple graphical representation an example with only two explanatory variable is considered.

[13]Source: Hastie T, Tibshirani R., Friedman J. (2009). "The Elements of Statistical Learning Data Mining, Inference, and Prediction", New York, NY: Springer, https://web.stanford.edu/ hastie/ElemStatLearn/printings/ESLII_print12.pdf, pp. 71.

The larger the penalty, the narrower the blue contour is, and so the plots meet each other at a point closer to zero. Vice-versa the smaller the penalty, the more contour expands, and the intersection of blue and red plots comes closer to the center of the red circle (non-penalized solution). In the specific case of Ridge regression, the two contour will difficulty meet where the corner of regularizers will be $w_1 = 0$ or $w_2 = 0$, due to the shape of the function $g(x)$ and this is the reason why no parameters are reduced exactly to zero.

In solving equation (2.7) or (2.8), it is important to specify that the quantity w_0 is not penalized. In fact, it is preferable to have a model being independent from the mean of the Y vector. Moreover, since the solutions are not equivariant under scaling of the inputs, usually a standardization of the parameters is needed. So the data used in the regression are usually centered by replacing each x_{ij} by $x_{ij} - \bar{x}_j$ and finally, the quantity w_0 is estimated by $\bar{y} = \frac{1}{N}\sum_{j=1}^{N} y_i$. The remaining coefficients are retrieved by the implementation of the Ridge regression without intercept, using the centered x_{ij}.

Henceforth it is assumed that this centering has been done, in such a way that it is possible to write the equation in (2.7) in matrix form,

$$RSS(\lambda) = (y - w'X)'(y - w'X) + \lambda w'w \tag{2.9}$$

and its solution can be retrieved as follow:

$$\hat{w}^{Ridge} = (X'X + \lambda I)^{-1} X'y \tag{2.10}$$

where I is the *pxp* identity matrix. Notice that choosing the quadratic penalty $w'w$, the Ridge regression solution is again a linear function of y.

The main motivation for the Ridge regression to be used, when it was first introduced in statistics, was that the solution adds a positive constant to the diagonal of $X'X$ before the inversion take place. This makes the problem nonsingular[14], even if $X'X$ is not of full rank. In addition, this form of regularization has some other intuitive advantages (Boyd & Vandenberghe, 2004):

- in terms of prediction, avoiding large values of w, allows eliminating large variations in the estimated parameters and also to achieve higher predictive performance;
- in terms of optimization, it gives a compromise between solving the risk minimization problem and have small values for w.

[14] A matrix is not invertible (also singular) iff its determinant is 0. In this case does not exist an n-by-n square matrix B such that $\mathbf{AB} = \mathbf{BA} = \mathbf{I}_n$ where $\mathbf{I}_n$ denotes the n-by-n identity matrix and the multiplication used is ordinary matrix multiplication.

2.5.2 The Lasso

Even if the Ridge regression brings important benefits and improvements to the usual least squares regression, it is not able to address the issue of variable selection, affecting the Markowitz optimization problem when a great number of assets are taken into consideration. For this reason recent studies (Brodie et al., 2009, DeMiguel et al., 2009b, Yen, 2010, Yen and Yen, 2011, Carrasco and Noumon, 2011, Fan et al., 2012, Fernandes et al., 2012) give considerable attention to another penalization approach in the portfolio optimization context: the so-called Least Absolute Shrinkage and Selection Operator (LASSO). In fact, while the ℓ-2 regularization is an effective means of achieving numerical stability and increasing predictive performances, it does not address another two problems affecting Least Squares estimates: parsimony[15] of the model and interpretability of the coefficient values[16]. While the size of the coefficient values is bounded, minimizing the variance of a regression with a penalty on the ℓ-2 norm does not encourage sparsity[17], inducing the resulting models to typically have all non-zero values associated with the regression coefficients.

With the purpose of resolving these issues, Robert Tibshirani in 1996 in the paper *"Regression Shrinkage and Selection via the LASSO"* (Tibshirani, 1996), proposed to replace the ℓ-2 norm with an ℓ-1 norm. The latter preserves many of the beneficial properties of Ridge regularization, but it generates sparse models that are more easy to interpret and often outperform those produced with an ℓ-2 penalty (Trevor et al., 2009). The LASSO estimate can be simply defined by

$$\begin{aligned} \hat{w}^{LASSO} = \min_{w} \quad & \sum_{i=1}^{N}(y_i - w_0 - \sum_{j=1}^{p} w_j x_{ij})^2. \\ s.t. \quad & \sum_{j=1}^{p} |w_j| \leq t. \end{aligned} \tag{2.11}$$

As done for the Ridge regression in Section 2.5.1, it is possible to re-parameterize the constant w_0 by standardizing the predictors. In this way, again, the solution for $\hat{w}_0$ is $\bar{y}$, and thereafter it is possible to fit a model without an intercept and rewrite the LASSO problem in the equivalent Lagrangian form

$$\hat{w}^{LASSO} = \min_{w} \quad \sum_{i=1}^{N}(y_i - \sum_{j=1}^{p} w_j x_{ij})^2 + \lambda \sum_{j=}^{p} |w_j|. \tag{2.12}$$

This time the ℓ-2 penalty $\sum_{j=1}^{p} w_j^2$, is replaced by the ℓ-1 penalty $\sum_{j=}^{p} |w_j|$ and differently from

[15]Parsimonious models are simple models with great explanatory predictive power. They explain data with a minimum number of parameters, or predictor variables.

[16]Interpretability is closely connected with the ability of users to understand the model.

[17]A sparse statistical model is one having only a small number of nonzero parameters or weights. It represents a classic case of “less is more”: a sparse model can be much easier to estimate and interpret than a dense model (Hastie, Tibshirani, & Wainwright, 2015).

the previous case, where the linearity of the function is preserved, the latter constraint makes the solutions nonlinear in y_i, so that computing the LASSO solution could be considered a quadratic programming problem[18].

The fundamental characteristic that convinced lots of researchers and portfolio managers to use the LASSO, is that making t sufficiently small (λ sufficiently large) will cause some of the coefficients to be exactly zero. This ability to generate a sparse model is due to the particular shape that ℓ-1 constraint have. In fact, if the constraint region for the Ridge regression is represented by a disk (Figure 2.4), that for LASSO is represented by a diamond given by the constraint $|w_1| + |w_2| \leq t$ (Figure 2.5).

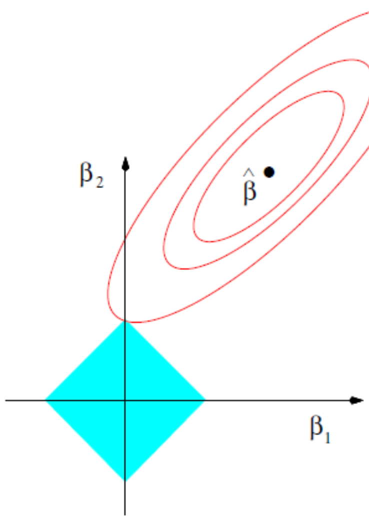


Figure 2.5: Estimation picture for the LASSO regression. Shown are contours of the error and constraint functions. The solid blue area is the constraint region $|w_1| + |w_2 \leq t$. While the red ellipse is the contour of the least squares error function[19].

As explained in the previous section, also in the case of the LASSO regression the solution is found in the first point where the elliptical contour hits the constraint region. Unlike the disk, the diamond has corners, so if the solution occurs there, then it is possible to have one parameter w_j set equal to zero. When the number of predictors is more than $p > 2$, the diamond becomes a rhomboid. So in this case, there are many more opportunities for the estimated parameters to be set to zero.

Empirical results in a mean-variance framework (DeMiguel et al., (2009); Brodie et al., (2009)) strongly support the use of the LASSO method when short selling is allowed. However,

[18] Quadratic programming (QP) is a particular type of nonlinear programming, that allows to optimize (minimize or maximize) a quadratic function of several variables subject to linear constraints.

[19] Source: Hastie T, Tibshirani R., Friedman J. (2009). "The Elements of Statistical Learning Data Mining, Inference, and Prediction", New York, NY: Springer, https://web.stanford.edu/ hastie/ElemStatLearn/printings/ESLII_print12.pdf, pp. 71.

in presence of no-short selling constraints, it shows some difficulties related to potentially biased estimates of large absolute coefficients and ineffectiveness in promoting sparsity (Fan & Li, 2001). In fact, the ℓ-1 norm of the asset weights will have a constant value of one and so, usually, a large number of very small position on assets are going to be included in the optimal portfolio. This could have a positive impact in term of portfolio diversification, but could be detrimental if the high transaction and administrative costs of keeping such small and sometimes illiquid positions are taken into consideration.

2.5.3 The Elastic Net

For what has been mentioned until now, the LASSO regression might seem to be a better alternative, compared to Ridge, in order to perform the variable selection. Although, in some specific situations, it exhibits some important limitations:

1. in the case in which the number of variables is greater than the number of observation ($p > n$), the LASSO is able to select at most n variables before it saturates[20];
2. if there is a group of variables among which the pairwise correlations are very high, then the LASSO tends to randomly select only one variable from the group, without considering all the others (Zou & Hastie, 2005);
3. in the common $n > p$ scenarios, if there are high correlations between predictors, the prediction performance of the LASSO is dominated by Ridge regression (Tibshirani, 1996).

Among the solution proposed, in order to overcome these limitations of LASSO, great attention was given to the one elaborated by Zou and Hastie in (2005). They, in fact, proposed a new regularization method that simultaneously performs an automatic variable selection and a continuous shrinkage, including also groups of correlated variables between the selected ones (*"grouped selection"*)[21].

This method takes the name of *"Elastic net"*, and it is expressed in the form

$$\begin{aligned} \hat{w} = \min_{w} \quad & \sum_{i=1}^{N}(y_i - \sum_{j=1}^{p} x_{ij}w_j)^2. \\ s.t. \quad & \sum_{j=1}^{p}\left(\alpha|w_j| + (1-\alpha)w_j^2\right) \leq t. \end{aligned} \tag{2.13}$$

[20] A model is saturated when there are as many estimated parameters as data points. By definition, this will often lead to a perfect fit, but also to extremely high-variance predictors.

[21] A regression method exhibits the grouping effect if the regression coefficients of a group of highly correlated variables tend to be equal, apart from the case in which they are negatively correlated. Identical coefficients must be assign in the extreme case in which the variable are identical too.

As highlighted above, the penalization term is a convex combination of the LASSO and Ridge penalties. The ℓ-1 part of the penalty term encourages a sparse solution in the coefficients. The quadratic part, instead, induce highly correlated features to be averaged, stabilizing the ℓ-1 regularization path.

The parameter α determines the mix of the penalties: when $\alpha = 1$, the elastic net can be simply considered as Ridge regression, if its value is $0 < \alpha < 1$, the elastic net penalty function is singular[22] at the origin and strictly convex, gathering in this way the characteristics of both the LASSO and Ridge (Figure 2.6).

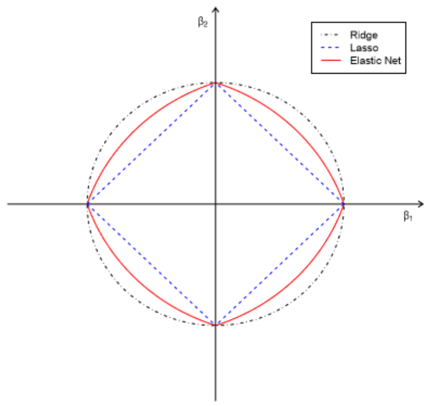


Figure 2.6: Two-dimensional contour plots for Ridge, LASSO and Elastic net penalty with $\alpha = 0.5$[23].

2.5.4 The q-norm constraint and the non-convex optimization problem

Another method, introduced in order to overcome the problematics arising from the LASSO, was the one proposed by Gasso et al. in 2009. In this respect, they suggested to add some penalties that are singular at the origin, just like the ℓ-1 penalty, in order to promote sparsity, but non-convex, in order to countervail biases (Gasso, Rakotomamonjy, & Canu, 2009).

In fact, the Ridge and the LASSO regressions that have been analyzed until now can be considered as specific versions of the more general (B)ridge regression approach introduced by Frank and Friendman (1993), where an upper bound is imposed on the q-norm considering $0 < q < \infty$[24]. Thus the previous equations with ℓ-1 and ℓ-2 constraints, might be generalize in

[22] A singularity is a point at which a function does not possess a derivative. In other words, a singularity function is discontinuous at its singular points.

[23] Source: Hui Zou and Trevor Hastie, Department of Statistics Stanford University. "Regularization and Variable Selection via the Elastic Net" https://web.stanford.edu/~hastie/TALKS/enet_talk.pdf.

[24] As in all other works in this field, in this paper is maintained the common practice to refer to a norm despite the fact that for cases with $0 < q < 1$ it does not define a norm but a quasi-norm, since the triangle inequality is not respected.

the form

$$\hat{w} = \min_{w} \quad \sum_{i=1}^{N} (y_i - \sum_{j=1}^{p} w_j x_{ij})^2. \qquad s.t. \quad \|w_j\|_q \leq t. \tag{2.14}$$

or in Lagrangian form

$$\hat{w} = \min_{w} \quad \sum_{i=1}^{N} (y_i - \sum_{j=1}^{p} w_j x_{ij})^2 + \lambda \|w\|_q \tag{2.15}$$

where $\|w\|_q = (\sum_{j=1}^{p} w_j^q)^{\frac{1}{q}}$ represents the q-norm of the vector of assets weight[25].

As has been done for all the other types of penalized regressions, the contours of constant value of $\|w\|_q$ for some values of q, are shown in Figure 2.7, considering the case of two inputs w_1 and w_2.

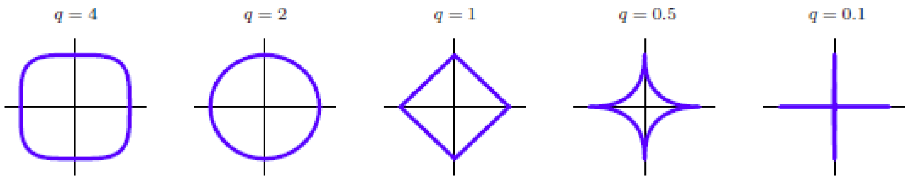


Figure 2.7: Contours of $\|w\|_q$ for given values of q [26].

Notice that $q = 1$ corresponds to the LASSO, while $q = 2$ to the Ridge regression. In addition, it is interesting to underline the fact that for $q \leq 1$, the contours are not uniform in direction, but concentrates more mass on the coordinates, allowing to obtain a higher number of inputs to be set to zero. The lower the upper bound on the q-norm is (e.i $q = 0.1$), the more sparse and less diversified (with larger weights) the portfolios are. In fact, in the implementation of the q-norm constraint, the latter should be considered as a measure of diversity of the portfolio. When the no-short selling limitation is imposed, this measure has maximum value for the equally weighted portfolio and minimum value for a portfolio totally invested in a single asset (Fernholz et al., 1998).

Gasso et al. (2009) have already shown the goodness of considering norms with $q < 1$ when the number of predictors in the model and their levels of correlation are high. The basic idea behind this class is to heavily penalize gains in small (absolute) portfolio weights on one side, and on the other to set a weak penalization in large (absolute) weights. This causes the invested wealth to be concentrated in more extreme positions than in the LASSO penalty, allowing a

[25] Notice that, in this case $q < 1$, there is no qth root on the right, which is the correct form of the ℓq.

[26] Source: Hastie T., Tibshirani R., Wainwright M.(2015). "Statistical learning with sparsity: the lasso and generalizations", pp. 22.

reduction of the estimation risk without increasing too much the value of the penalty term. Thus, the q-penalty provides a particularly strong incentive to avoid small and presumably dispensable positions in favor of a small subset of presumably indispensable assets (Fastrich, Paterlini, & Winker, 2015)[27], allowing the construction of sparse and less diversified portfolios that, as already said, are preferable in term of transaction costs.

Even though the ℓ-q penalization provide all the benefits just cited, it poses some difficulties and complications in term of optimization. In fact, if Equations (2.8) and (2.12) represent convex optimization problems and, thus, are easy to solve by using linear programming in the first case and quadratic programming in the second; Equation (2.14) represents a non-convex optimization problem, that is very challenging from an optimization viewpoint (called for this reason NP-Hard problem)[28]. Basically, convex optimization is a subfield of optimization that studies the problem of minimizing convex functions over convex sets. A function is said to be convex if the line segment between any two points lies above or on the graph of the function, in a space of at least two dimensions; and if, for a twice differentiable function of a single variable, the second derivative is always greater than or equal to zero in its entire domain. On the other side a set is said convex if, for any couple of points considered, it is possible to draw a line segment in between, that also entirely lies within the set (Figure 2.8).

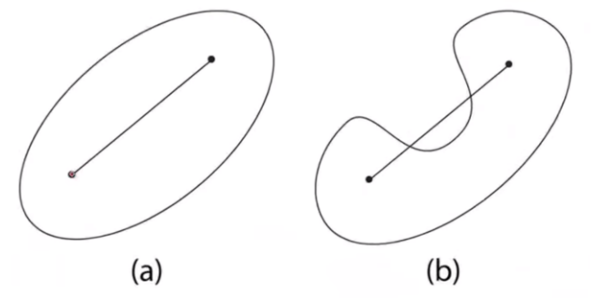


Figure 2.8: Representation of convex (a) and non-convex (b) sets in an optimization problem.

The convexity makes optimization much easier compared to the general case, since local minimum must be for sure a global minimum, and first-order conditions are sufficient for optimality. Instead, non-convex optimization[29] might have multiple locally optimal points and so might be difficult, or sometimes impossible, to obtain an identification. The difference in the two types of optimization problems can be graphically seen in Figure 2.9. The point Z is a local minimum but clearly, it does not represent a global minimum. This creates some complexities in the optimization procedure.

[27]Source: https://link.springer.com/content/pdf/10.1007%2Fs10287-014-0227-5.pdf

[28]NP-hard problems are problems for which there is no known polynomial algorithm, so that the time to find a solution grows exponentially with problem size. An algorithm is said to be solvable in polynomial time if the number of steps required to complete the algorithm for a given input is known.

[29]A function is defined to be non-convex when taking its Hessian, at least one of its Eigenvalues is negative.

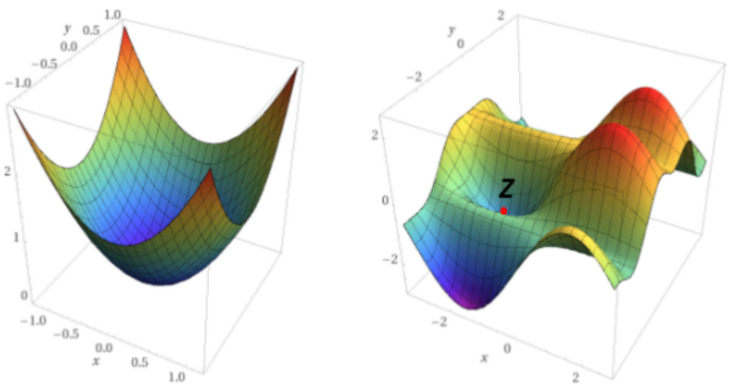


Figure 2.9: Representation of convex and non-convex optimization problem[30].

[30]Source: Reza Zadeh, November 16 2016, https://www.oreilly.com/ideas/the-hard-thing-about-deep-learning

Chapter 3

Index Tracking: a Passive Portfolio Management Strategy

The statistical regularization methods just explained (Appendix B.1), have recently found numerous applications in the mean-variance portfolio setting, due to their ability to promote the identification of sparse (with a small number of constituents) portfolios with good out-of-sample properties and low turnover. Specifically, this chapter will introduce the general concepts of passive management and index tracking strategy, showing how the advantages guaranteed by regularization, could be crucial also in this particular context.

3.1 Active vs passive investing

Since the existence of passive investment strategies, investors and managers have tried to understand, whether is better to allocate their wealth by using an active or a passive management strategy. Passive fund management was first introduced in 1970 as an academic concept when Burton Malkiel published his book, *"A Random Walk Down Wall Street"*[1], but it did not really gain traction until Vanguard entered into the market scene with low-cost indexed funds. In the 21^{st} century, however, probably due to poor returns of active management in the previous years and to the recommendation of some influential financiers, passive funds had become very popular, so that a huge amount of resources has flooded into them.

The dilemma of choosing between an active or passive strategy, can shortly be reduced to one fateful question: does the manager/investor believe in the Efficient Market Hypothesis or not? The Efficient Market Hypothesis (EMH), elaborated by Eugene Fama, professor of economics at the University of Chicago in 1960s, is an investment theory whereby stock prices,

[1] In this book was argued that prices typically exhibit signs of random walk and that it is not possible to consistently outperform market averages, challenging active managers' profitability over time.

at any time, reflect all available information about securities[2]. In this way theoretically, neither technical nor fundamental analysis can allow active portfolios to produce excess returns with respect to the market. According to this theory, stocks always trade at their fair value on stock exchanges, making impossible for investors to systematically identify and trade stocks that are mispriced. Stock price movements are largely random and are primarily driven by unpredictable events. In the light of these hypotheses, it seems clear that no active investor will consistently beat the market over long periods of time, meaning that active management strategies cannot add enough value to outperform passive management strategies.

On the other side instead, investors who do not follow the EMH, but are convinced that markets are not able to immediately reflect all available past and future information about the stocks, believe that it is possible to profit from the stock market through the use of strategies that aim to identify mispriced securities. The investors belonging to this category, are more prone to active investing. They will rely on the human element to actively manage a portfolio, on analytical research, on forecasts, and on their own judgment in making investment decisions.

Even nowadays, the endless debate between these two sides appears to be still bitter and broad. On one side, some expertises sustain the merits of passive versus active management, underlining that active investment managers are not able to pick enough winners to justify their high fees, in particular during bull market periods. On the other, the supporters of active investing enact the flexibility and the possibility of obtaining extra profit, given by an accurate active management strategy. In this elaborate empirical evidence and characteristics of both strategies are going to be described, without feeding in any way the debate and without taking any of the two sides.

3.1.1 Advantages of passive investing

As explained above, the underlying assumption of passive investment strategy is that it is not possible to beat the market in the long run (EMH). Passive investing, given that markets usually have positive returns and grow over time, has the goal to replicate market performances by constructing well-diversified portfolios of single stocks and, in this way, build wealth gradually (long-term investment horizon), seeking to avoid the fees and limited performance that may occur with frequent trading. In this view, perhaps, the most solid advantage provided by passive investing is the reduction of costs. As noted in some empirical researchers as *"Vanguard's Principles for Investing Success"*, lower cost funds tend to outperform their higher cost peers (Figure 3.1).

[2]The EMH exists in various degree: weak, semi-strong and strong, depending on the inclusion or not of non-public information in market prices
Source: Morningstar http://www.morningstar.com/InvGlossary/efficient_market_hypothesis_definition_what_is.aspx

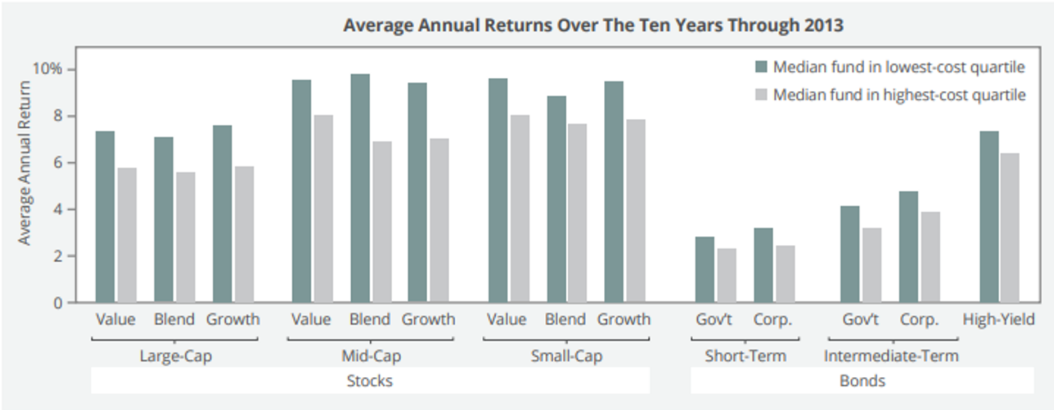


Figure 3.1: Ten-year annualized returns for funds in the lowest-cost and highest-cost quartiles. Returns are net of expenses, excluding loads and taxes. Both actively managed and indexed funds are included[3]

Since most often passive funds have had the lowest Average Expense Ration between all other types of funds (Table 3.1), it is possible to conclude that consequently, they are indeed those who guaranteed the best Average Annual return and outperformed all other peers, at least until 2013.

	Investment Type	Actively Managed Funds	Index Funds	ETFs
U.S.Stocks	Large-Cap	0.80%	0.11%	0.14%
	Mid-Cap	0.97	0.18	0.25
	Small-Cap	1.04	0.19	0.23
U.S. Sectors	Industry Sector	0.94	0.44	0.37
	Real Estate	0.92	0.13	0.20
International Stocks	Developed Market	0.91	0.17	0.29
	Emerging Market	1.16	0.21	0.42

Table 3.1: Average expense ratio of different funds until December 2013 [4].

This evidence is in general supported by Kent Smetters, professor of business economics at Wharton University, who assures that: *"On an after-tax basis, managers of stock funds for large- and mid-sized companies produced lower returns than their index-style competitors 97%*

[3] Source: Morningstar and Vanguard,12/31/13. Vanguard calculations, using data from Morningstar. All Mutual funds in each Morningstar category were ranked by their expense ratios as of 12/31/13. They were then divided into four equal groups. http://clients0.brinkercapital.com/wp-content/uploads/2015/03/WP_ACT_PASS-Active-Vs.-Passive-FINAL.pdf, pp.2.

[4] Source: Data taken from Morningstar and Vanguard, 12/31/13. Vanguard calculations, using data from Morningstar.

of the time, while managers of small-cap stocks trailed 77% of the time"[5]. The reason why this happens is very clear. It results to be difficult for an active manager to outperforms the market by such an amount that can compensate the fees charged for the research and stock picking operations. On the other side fees for passive managers are very low, 0.2% or 0.3%, guaranteeing a huge competitive advantage in term of costs.

In addition to cost advantages, passive investing bring benefits to investors by giving the opportunity to avoid the well-known phenomenon of *"cash drag"*. Indeed, in order to be able to quickly catch new investments opportunities and provide a boost during a down market environment; active managers tend to hold more cash within the investment when compared to their passive counterparts, that instead are mandated to stay fully invested. However, since empirical evidence shows that the S&P 500[6] has been positive for 23 of the last 30 years (Figure 3.2), cash holding often represents a performance drag on active managers and not an advantage[7]. So even if cash could have helped boosting active funds performances in the seven years of negative returns, that would have represented a load and a missed gain in the remaining years.

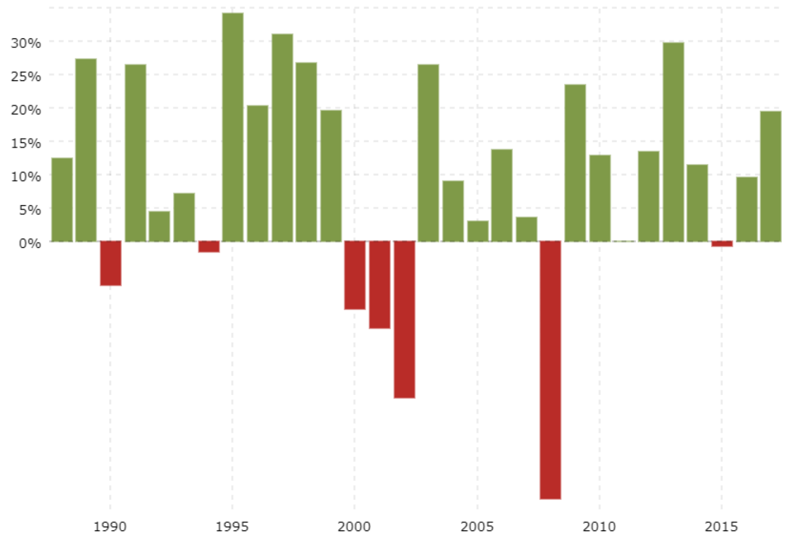


Figure 3.2: Annual percentage change of the S&P 500 index from 1988 to 2017, calculated from the last trading day of each year to the last trading day of the previous year[8]

[5]Source: https://executiveeducation.wharton.upenn.edu/thought-leadership/wharton-wealth-management-initiative/wmi-thought-leadership/active-vs-passive-investing-which-approach-offers-better-returns

[6]The S&P 500, has been used,in this case, as a benchmarks for the whole stock market.

[7]Source: http://clients0.brinkercapital.com/wp-content/uploads/2015/03/WP_ACT_PASS-Active-Vs.-Passive-FINAL.pdf

[8]Source: Macrotrends, https://www.macrotrends.net/2526/sp-500-historical-annual-returns

Other two important advantages, related to the use of a passive management strategy, are with no doubt transparency[9], because investors at all times know what stocks or bonds an indexed investment contains, and tax efficiency. In fact, since the passive funds are used as buy-and-hold strategies, the amount of executed trades will yield lower portfolio turnover and less concentrated realized gains, triggering a small annual capital taxation.

3.1.2 Advantages of active investing

At this point, the debate would seem to have no reason to exist and active management already seems to have lost the battle, given the enormous advantages of passive management. However, in reality, even the active one reserves very important advantages in some particular situations, that could induce some investors to prefer it to the former.

Indeed, active investing might be extremely advantageous, when the market is characterized by high sector dispersions. In other words, when sectors are not performing at about the same level, it is much easier for expert managers to add value by picking sectors that are meaningfully better than the average.

An additional circumstance where active managers could be considered preferable is during bear markets. In fact, just during bull markets, it is almost impossible to beat passive investing because market returns are extremely high and hedging usually hurts performance, as do transaction fees on switching stocks. However, during bear markets, active managers and hedgers nearly always beat passive investing. This tendency is clearly well known by both investors and managers, such that it is possible to identify a sort of cyclical path in the cash allocation between active and passive investing. As shown in Figure 3.3, large flows of cash into passive funds have been registered during periods of good market performances.

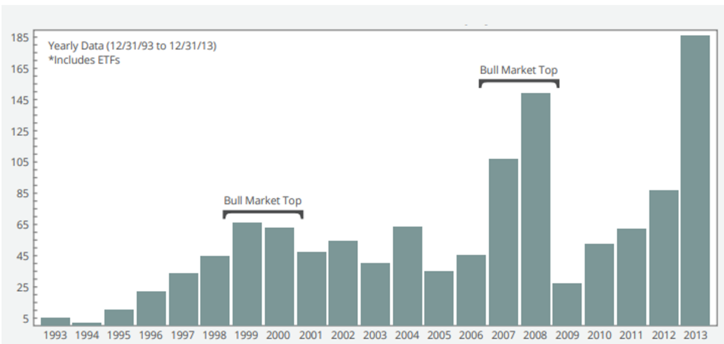


Figure 3.3: Cash flow exchanges between passive and active managed funds[10].

[9]It represents a huge advantage in order to avoid the principal agent problem". The problem arises when two parties have different interests and asymmetric information, such that the principal cannot directly ensure that the agent is always acting in their best interest

This cyclical shifts between active versus passive investing might be observed also from an informational point of view[11]. In fact, when a great number of highly skilled analysts and market operators compete to obtain relevant information on a stock before everyone else, it is difficult to beat the peers and obtain information before others. So, when a lot of resources are allocated through active investing, the competition leads to have no advantages to keep working and spending resources, to obtaining information that are already incorporated into the prices. In this environment, passive investing might appear to be a good alternative: cash increasingly flows towards index funds. However, if lots of participants become discouraged from the lack of return and retract from active funds, the gain from information edge start increasing again, due to the lack of analysts and information included into the market prices. At this point, wealth will move back to active funds, thanks to higher profit opportunities. In this way, the cycle continues over and over.

Other interesting characteristics of active investing are flexibility, managers are usually free in the stock selection process without being constrained to a specific index, and risk mitigation opportunities. In fact, active portfolio managers might use short selling and derivatives to protect portfolios.

After this quick description of the advantages of both passive and active investing, it is possible to conclude that there is not a clear right choice. Both of the strategies have their reasons to be applied and thus, when considering the choice between the two, an investor has to be open to both alternatives, not just one.

3.2 Index tracking

As already specified in Section 3.1.1 the aim of passive investing is to replicate a market index performance over time, but in practice, it is not possible to directly trade a financial index. In order to get access to it, investors must rely on different financial instruments such as options and futures, that, however, do not track the value of an index explicitly, reflecting only its value; or on exchange-traded funds (ETFs) that on the other side explicitly track the benchmark performances. Nonetheless, not all indexes or sectors have an associated ETF that could be purchased and so, in these cases, would be impossible to enact a passive strategy that explicitly reflects the benchmark. If this situation arises, the only possible alternative is to explicitly track the value of an index by constructing from scratch a portfolio of assets, so-called *"tracking portfolio"*, whose value follows the value of the given index (Benidis, Feng, Palomar, et al., 2018).

[10]Source: http://clients0.brinkercapital.com/wp-content/uploads/2015/03/WP_ACT_PASS-Active-Vs.-Passive-FINAL.pdf, pp.2.

[11]Source: https://www.forbes.com/sites/investor/2015/03/30/active-versus-passive-management-which-is-better/#1fe424327025

Index tracking (or benchmark replication) consists in designing a portfolio that replicates the behavior, the holdings and the performance of a designated index. The popularity of index funds relies on its ability to allow investors to be exposed to an entire index at a low cost[12].

Consider for example a general market index composed by a weighted collection of N assets, with a vector of weights b, representing the portion of wealth allocated in each security. The aim of index tracking is to retrieve the optimal tracking portfolio weights w, that minimizes a given function, expressing the distance between the index returns and those of the tracking portfolio (tracking error). Two main ways to achieve this objective has been developed in recent years: full replication and partial replication (Strub & Baumann, 2018).

The full replication strategy offers a straightforward solution to the weights selection problem. In fact, by using full replication the weights assigned to the tracking portfolio are perfectly equal to the weights of the benchmark index. In this way, the constructed portfolio perfectly reflects the trend of the index, having the exact same returns over time. Even if this is the strategy that more than any other minimize the deviation (tracking error) with respect to the benchmark; however, when an index with a large number of securities is taken into consideration, transaction costs dramatically increase and liquidity problems arise if some of the assets to be included cannot be easily bought in the market[13].

A possible solution to overcome these limitations is the implementation of a partial replication strategy, also called *"sparse index tracking"*. Instead of creating a tracking portfolio containing the same number of securities included in the benchmark, a portfolio with a lower number of assets is constructed, holding securities that provide the most representative sample of the index based on correlations, exposure and risks. In this way, the tracking error will increase on one side (imperfect tracking), but on the other transaction costs are reduced and illiquid positions are avoided.

This second approach, however, is not as simple as in the case of full replication, since two main challenges have to be addressed: primarily, it has to be chosen which of the N assets should be included into the tracking portfolio; secondly has to be determined which should be their relative weights. In order to solve these issues, usually, a two-step approach method is followed[14]. The first stage is the so-called "Stock selection", in which the securities to be included are chosen, selecting assets with higher market capitalization, or higher correlation to the index, or both. In the second stage, called "Capital allocation", the effective weights of the tracking portfolio are determined, or by using a "Naive allocation", meaning that the weights are proportional to the original weights in the benchmark portfolio; or by an "Optimized allocation". The latter consists in solving a convex minimization problem, having as objective function a

[12]This guarantees great advantages in term of portfolio diversification.
[13]Source: http://www.ece.ust.hk/ palomar/ELEC5470_lectures/14/slides_index_tracking.pdf
[14]Source: http://www.ece.ust.hk/ palomar/ELEC5470_lectures/14/slides_index_tracking.pdf

tracking error measure $f(w)$

$$w^{opt} = \min_{w} \quad f(w). \tag{3.1}$$

Some of tracking error measures most commonly used in the literature are: the average Tracking Error (3.2), Tracking Error Volatility (3.3), the Root Mean Squared Error (3.4) and the Empirical Tracking Error (3.5).

$$TE = \mathbf{E}\Big[Y - wX\Big] \tag{3.2}$$

$$TEV = \mathbf{V}\Big[Y - wX\Big] \tag{3.3}$$

$$RMSE = \sqrt{\frac{(Y - wX)'(Y - wX)}{T}} \tag{3.4}$$

$$ETE = \frac{1}{T}||Y - wX||_2^2 \tag{3.5}$$

In these equations: Y represents the vector of centered return of the benchmark index, w is the variable of interest representing the vector of weights of the tracking error portfolio and X represents the matrix of centered returns of the securities contained into the index.

3.3 Index tracking using penalized regression

Despite the standard two-steps approach for sparse index tracking, commonly used in practical implementations, offers the possibility to create acceptable portfolios with low tracking error,

Indeed, the first step of "Stock selection", almost arbitrarily eliminates possible interesting candidates for the portfolio construction, without including them into the optimization procedure. Furthermore, even if tracking error represents an optimal indicator to describe the deviation between index and tracking portfolio performances, however, it does not explain the whole story. In fact, past literature (Frino & Gallagher, 2001)[15], already underline the fact that tracking error measures do not take into account the transaction costs associated with trading securities, that strongly influence the ability of the tracking portfolio to provide identical returns with respect to the benchmark index. The index is calculated as a "paper"[16] portfolio that assumes transactions can occur instantaneously, in unlimited quantities and without cost (Frino & Gallagher, 2001). On the other side, an index fund, when dealing with portfolio construction and rebalancing, undoubtedly produce transaction costs that reduce the final return for investors. Additional transaction costs may be registered, when index composition change: usually com-

[15]Source: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.199.4563&rep=rep1&type=pdf
[16]Considered as an imaginary portfolio that could never be implemented in reality.

panies are added and deleted from the market index. In this way, passive funds are forced to trade in order to realign their portfolio.

For these reasons, in order to improve the wealth allocation among all the assets contained into the benchmark, to streamline the selection procedure in one single step and to minimize the tracking error, while having transaction costs and number of assets included under control; the tracking error problem in Equation (3.1) could be reformulated by including a penalty constraint (ℓq-norm constraint), as follow

$$w^{opt} = \min_{w} \quad f(w) + \lambda ||w||_q. \tag{3.6}$$

Again, λ represents the tuning parameter that regulates the intensity of the penalty, and $||w||_q = \sum_j^p |w_j^q|$ represents the q-norm of the weights vector w with $0 < q < 1$.

As already described in Chapter 2, the inclusion of a penalization term, abundantly used in other fields of statistics and portfolio management, guarantees enormous advantages for the assets selection procedure. In a single step, it allows to reduce toward zero the weights of assets that less correlates with the main index[17], and at the same time to optimally allocate the wealth among the selected assets. Moreover, it guarantees a reduction in portfolio turnover and so in transaction cost, giving the opportunity to increase the probability of obtaining higher final returns. The higher is the tuning parameter λ, the higher is the penalization imposed (more weights collapse to zero), the lower is the norm index q, the faster the portfolio increase its sparsity and reduce diversification. In some sense, the tracking error problem, reformulated in this way, coincide with the usual problem in Equation 3.1, but including a cardinality constraint that limits the number of admissible assets in the tracking portfolio (Fastrich et al., 2014).

[17] The less contributive securities are completely excluded from the tracking portfolio.

Chapter 4

Experimental Set-up

In this Chapter, the data and the methodology used for the implementation of the project are described. Starting from the dataset and the tools (software and hardware) used, an overview on how the project has been carried on is provided, focusing also on the performance measures considered in order to evaluate the goodness of the different strategies proposed.

4.1 Data description

The dataset used for the implementation of the sparse index tracking procedure, is composed by the time series of weekly prices for the past 10 years, starting from the 2^{nd} of January 2008 to the 20^{th} of December 2017, for a total of 521 observations. In particular, the weekly prices of the *Standard and Poors' 100 Global* and of its constituents in each period have been taken into account: since the composition of the index changes over time, it has been worked on the weekly prices for all 134 assets, that at least for once have been included into it[1].

The choice of considering weekly returns has been actually driven by two important consideration: on one side, since we are working with a large investment universe, we need a frequency of data providing us with enough amount of observation, at least equals to the number of assets; on the other side practitioners and researchers argue that the use of daily prices might introduce in the optimization framework the "noise" and volatility typical of the day-to-day fluctuations, creating huge distortions if we want to focus on the predominant longer-term trend.

All the data have been retrieved from *DataStream - Thomson Reuters* and each stock price corresponds exactly to the dividend adjusted, market closing price.

[1]The list of the components of the index over time has been obtained in a two-step procedure: first the reference codes for the stocks included monthly within the index has been retrieved. Once that has been done, each code is isolated and so the stock price is at this point downloaded.

Starting from these prices and using the standard formula in 4.1

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{P_t}{P_{t-1}} - 1, \tag{4.1}$$

where P_t represents the price of a given stock at time t, the 520 weekly historical percentage returns have been calculated, both for the benchmark index and for its constituents. They are going to be the inputs for our penalized regression model.

Table 4.1 reports some summary statistics for the weekly returns of the benchmark index.

	Mean	Median	StDev	Min	Max	Skew	Kurt
S&P 100 Global	0.07	0.16	2.67	-14.54	13.90	-0.22	7.70

Table 4.1: Summary statistics for S&P 100 Global.

As shown, the time series of the index returns exhibits the typical patterns of financial time series: mean values around zeros, light asymmetry and fat tails.

Moreover, in order to give an idea of which is the relationship between the index return and the return of each constituent, the correlation between them has been considered.

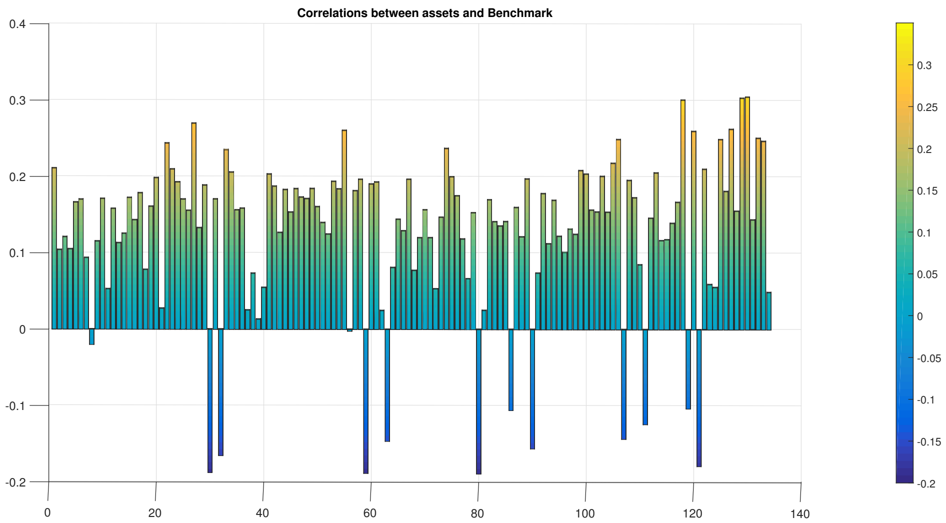


Figure 4.1: Correlation between S&P 100 Global Index and its constituents.

As explained in Chapter 2, it is likely to believe that the assets which are going to be included into the tracking portfolios, would be the ones that exhibit the highest correlations with respect to the benchmark.

In Appendix B.3 some other descriptive figures about price evolution, cumulated returns and summary statistics for the index and for all the stocks grouped for Sector and Geography are reported.

4.2 Working environment

The entire project implementation has been run in a Lenovo ideapad 320-15ABR with AMD A12-9720P RADEON R7 2.70GHz, and 8 GB RAM. All the codes used are written in Matlab R2017a and, in particular, the solution of the non-convex minimization problem, allowing us to retrieve the optimal portfolio weights, has been performed using a function included in the Matlab Global Optimization Toolbox called *"GlobalSearch"*

It consists in an algorithm that permits to find the global minimum of a given objective function, by repeatedly running a local solver (fmincon in this case) and generating a "GlobalOptimSolution" object. When run, the solver attempts to locate the solution that exhibits the lowest objective function value, starting from a given initial point x_0. The function permits the inclusion of linear equality constraints, linear inequality constraints and also of a set of lower and upper bounds on the design variables x, in such a way that the obtained solution is always included in the range $lb \leq x \leq ub$. More information about the GlobalSearch algorithm are available on the MathWork Website[2].

4.3 Methodology

After the dataset has been downloaded and managed[3], excluding all the stocks for which the time series of returns contain only few observations, it has been supposed to play the part of a passive portfolio manager who has the objective to build a mid-long term sparse index tracker being able to track the behavior of a benchmark index, in this case the *S&P 100 Global*, and to closely replicate its returns, by keeping the transaction costs under control and by avoiding illiquid positions.

In order to implement such a strategy, the first step consists in finding, among the all feasible portfolios that might be constructed using the index constituents, the set having minimum Tracking Error Volatility (Eq. 3.3) with respect to the benchmark. In fact, let's suppose we want to verify whether a portfolio closely replicates the benchmark to which it refers. To do so, it is enough to measure how much the tracker time series deviates with respect to the index one: the

[2]GlobalSearch algorithm description: https://it.mathworks.com/help/gads/how-globalsearch-and-multistart-work.html

[3]The constituents prices were expressed in different currencies. Considering the weekly time series of exchange rates, all values has been converted into US dollars.

logic is to subtract the benchmark return to the return of the aforementioned portfolio for each period of the time series and then calculate the variance of the sum of these deviations. The lowest the variance results, the better the index is tracked.

Thus, the project implementation starts by modifying the approach proposed by Fan (2012) in order to perform, instead of the minimization of the portfolio risk (variance), as shown in Section 2.4, the minimization of the TEV. Indeed, if we consider the minimization problem,

$$w^{opt} = \min_{w} \quad TEV = \min_{w} \quad Var\Big[R_p - R_B\Big] \tag{4.2}$$

where R_p is the tracker return and R_B is the return of the benchmark, it is possible to rewrite it as

$$\begin{gathered} \min_{w} \quad Var\Big[w'R_p - R_B\Big] = \min_{w} \quad Var\Big[w'R_p - w'\mathbf{1}R_B\Big] = \\ \min_{w} \quad Var\Big[w'(R_p - \mathbf{1}R_B)\Big] = \min_{w} \quad Var\Big[w'P\Big] \end{gathered} \tag{4.3}$$

where P corresponds exactly to the matrix containing the weekly deviations in returns between each constituent and the benchmark. The latter equation can be in this way assimilated to the one in 2.1 with R being substituted with P.

At this point, without repeating all the passages, but only keeping in mind what has been proposed by Fan, the minimization of the TEV can be rewritten as:

$$\min_{w_{-n} \in Z_n} Var(Z_n - w'Z^*) = \min_{w_{-n} \in R_n} \mathbf{E}[Z_n - w'Z^*]^2 \quad . \tag{4.4}$$

In this case, Z_n represents the vector of deviations between a randomly chosen numeraire asset and the benchmark returns, while Z^* constitutes the matrix containing the difference between the $n-1$ deviations from the benchmark returns and the deviation for the numeraire asset (Z). The solution of this regression problem provides the optimal portfolio weights that minimize the Tracking Error Volatility, except for the one of the numeraire asset. The optimal allocation for the latter is then obtained by imposing the satisfaction of the budget constraint:

$$w_n = 1 - \sum_{i=1}^{n-1} w_i. \tag{4.5}$$

Once this optimization problem has been formulated, keeping in mind that we want to reduce transaction costs and avoid illiquid positions, Equation 4.4 has to be adapted in order to perform a sparse index tracking strategy. On the lights of what has been described in Section 3.3, this objective can be easily reached using a penalized regression model.

Finally, the weights of the sparse portfolio that minimize the Tracking Error Volatility are obtained by solving the following minimization problem:

$$(i) \qquad w_{n-1} = \min_{w_{-n} \in R_n} \mathbf{E}\Big[Z_n - w'Z^*\Big]^2 + k||w||_q \qquad \text{s.t} \quad 0 < w_{n-1} < ub \tag{4.6}$$

$$(ii) \qquad w_n = 1 - \sum_{i=1}^{n-1} w_i \tag{4.7}$$

where k represents the tuning parameter that regulates the degree of penalization of the model and $||w||_q = \sum_{i=1}^{n-1} w^q + |1 - \sum_{i=1}^{n-1} w^q|$ identify the ℓ-q norm constraint. Notice that Equation 4.7 guarantees the satisfaction of the budget constraint and that the portfolio weights are forced to be both positive, in such a way to avoid short-selling, and lower than a fixed threshold *ub*, in order to prevent them to be concentrated in only one or few assets, ensuring in this way an acceptable degree of diversification.

The problem proposed above is solved multiple times using the Matlab code available in Appendix C.2, iterating the GlobalSearch function for different values of q and k. This procedure allows us to retrieve a wide spectrum of possible optimal solutions with different degrees of diversification and sparsity[4].

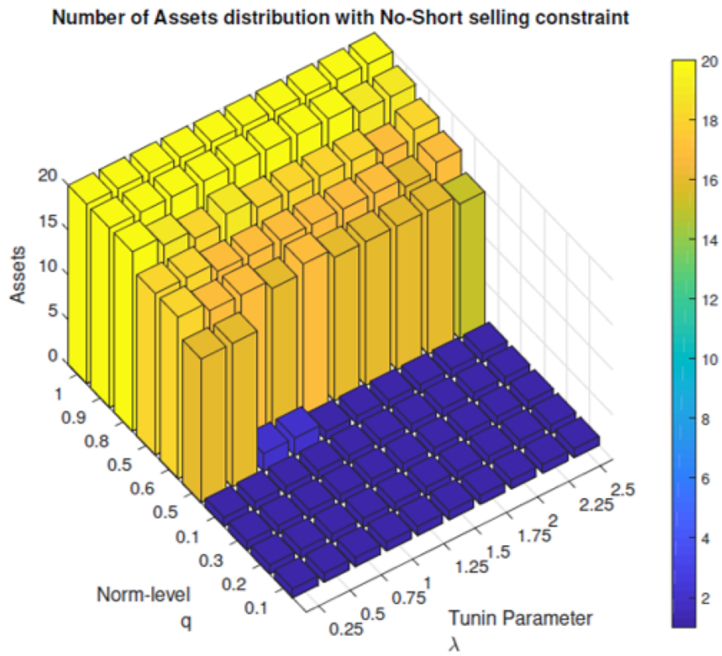


Figure 4.2: Number of assets in the tracking portfolios across q and k. Each bar represents one specific optimal tracking portfolio retrieved after the minimization procedure.

[4]In particular we solve the minimization problem for q that goes from 0 to 1 with a step of 0.1 and for k that goes from 0.25 to 2.5 with a step of 0.25. At the end of the cited procedure, the solutions were represented by a 3 dimensional array 104x10x10.

As shown in Figure 4.2, solving Equation 4.6 for low values of q and high values of k, provides portfolios having almost all assets weights set to zero, since the regression problem results to be strongly penalized and its parameters rapidly converge to zero (Figure 2.7). In the opposite case, instead, the wealth allocation for the tracking portfolio is distributed more homogeneously among the assets.

In particular, in the example proposed, for all the values of k with q lower than 0.5, the initial endowment is concentrated in a single asset (blue bars), because of the high magnitude of the k parameter. If a more smooth path wants to be observed, it is enough to consider smaller values of k (e.i in the order of 10^{-3}). In this case, however, too many iterations have to be repeated and this will dramatically increase the computational time.

After the optimization problem has been solved, we are stuck with a 3-dimensional array containing the weights of a set of portfolios that minimize the Tracking Error Volatility for different values of q and k, but we have no conditions on the value of the portfolio returns. Indeed, it is possible that a portfolio with very low TEV, in reality, has much greater returns than the benchmark portfolio. In order to solve this issue, since the aim of this dissertation is to replicate the returns of the index, we need to select, among all the optimal portfolios, the one that exhibits a return that is almost equal to the one of the benchmark.

For doing so, after the array containing the solution path has been restricted, excluding the tracking portfolios that allocate the wealth on too many assets (high transaction cost) and the ones which include very few concentrated positions (low diversification), we use a relative performance indicator, called Information Ratio, to select the portfolio that more accurately reproduces the returns of the benchmark.

The Information Ratio (IR), calculated as the Tracking Error over Tracking Error Volatility,

$$IR = \frac{TE}{TEV} = \frac{\mathbf{E}\left[R_p - R_B\right]}{\sigma[R_p - R_B]} \tag{4.8}$$

indicates the extra return generated or disrupted by the constructed portfolio, relatively to the benchmark, for each additional unit of risk. In the case in which its value is positive, it represents the extra return generated on average with respect to the benchmark. In the case it is negative, it represents the value disrupted on average by the investment strategy.

In our specific case, as shown in Figure 4.3, we calculate the IR for each portfolio that minimizes the TEV and among all portfolios at our disposal, since the objective is again to closely replicate the benchmark, we select the one having the IR closer to zero[5]. Indeed, if $IR = 0$, then $\mathbf{E}[R_p] = \mathbf{E}[R_B]$, meaning that the return of the tracking portfolio in on average equals to the benchmark return.

[5]If otherwise we want to consider the portfolio that guarantees the higher level of extra return, we just have to consider the portfolio that maximize the Information Ratio.

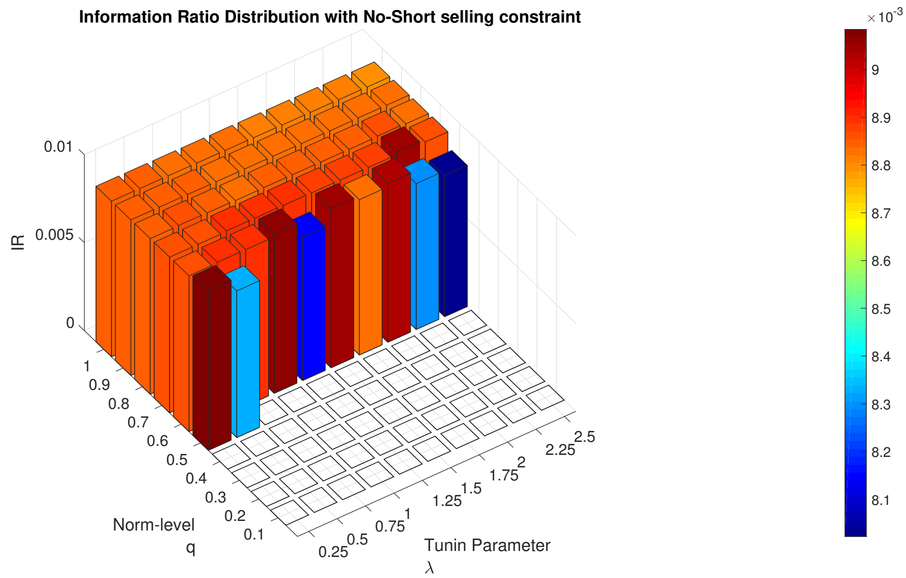


Figure 4.3: Distribution of the portfolio IRs across q and k. Each bar represents one specific optimal tracking portfolio retrieved after the minimization procedure. [6]

Following the procedure described until now, it has been possible to implement some Buy and Hold investment strategies that replicate the *S&P 100 Global Index* from January 2008 until December 2017[7]. However, even if these strategies permit to obtain acceptable performances over time and simultaneously contain transaction costs, since the portfolio allocation has been performed only at the beginning of the investment period, they seem to be not feasible in real applications. Indeed, the implementation of a Buy an Hold strategy does not take into account two main problems that usually a portfolio manager have to address:

1. first, the benchmark composition changes over time. In fact, periodically some assets are excluded and some others are included in the financial index. For this reason, if a Buy and Hold strategy is carried out, we have to deal with the high probability of having periods in which the tracking portfolio is not able to closely replicate the target benchmark returns. This phenomenon is due to the exclusion from the initial investment universe of some assets that have been subsequently included within the index constituents;

2. second, the manager of a passive fund have to deal with client capital flows. Indeed, during a long investment period, like the one we are considering, might happen that some investors decide to cash out or cash in from the investment fund. In these cases, the portfolio has to be properly rebalanced to compensate for the withdrawal of funds or to make use of the newly available resources.

[7]Namely the procedure has been repeated 3 times considering different levels of upper bound for each portfolio weights. In particular we consider 1, 0.2 and 0.1 as upper bounds.

With the purpose of solving these issues, in the second part of the project, we decide to implement a dynamic approach, that allows the portfolio manager to periodically reallocate the wealth at its disposal and modify the portfolio weights. In order to do so, we consider January 2013 as our starting point, with a rolling time windows of 260 observation (5 years). That stated we determine, exploiting the same procedure used in the static approach, the optimal portfolio weights for that specific period, considering as inputs the weekly returns of the index constituents for the prior 5 years. From that point onward, we update the portfolio on a periodical basis until the end of the investment period set at December 2017.

The dynamic approach is performed considering two different reallocation strategies. The first one plans to readjust the portfolio weights when the benchmark constituents change, leaving otherwise the portfolio weights unaltered. This type of strategy permits to obtain more reliable and accurate replication of the benchmark index and, in addition, to hold in check transaction costs.

The second one, instead, executes the reallocations of portfolio weights on a monthly basis, in such a way to simulate the behavior of a passive portfolio manager that is forced to deal also with the problem of capital inflows and outflows. Each month the minimization is repeated: if the index composition at time t is equal to the one at time $t-1$, then the value of q calibrated in the previous period is considered, iterating the minimization only for different values of k; otherwise, if the benchmark composition changes between the two periods, the minimization is again performed from the beginning[8]. This type of strategy is the one that, presumably, better simulates the condition in which a portfolio manager operates, but also the one that produces higher transaction costs.

4.4 Performance indicators

Once the strategies have been implemented, both following the Buy and Hold and the dynamic approaches, we proceed with an evaluation of the portfolios performances. In fact, after we obtain various results, setting different constraints on weights as No-Short selling, upper bound at 0.2 and upper bound at 0.1; we want to retrieve which is the strategy that guarantees the best tracking of the benchmark index. Thus, that said, our aim it is not to evaluate the tracking portfolio performances in absolute terms, but relative to the main index.

In order to do so, the first indicator we consider is the cumulative return of the portfolio. Once the optimal portfolio weights for each strategy are obtained from the GlobalSearch algorithm[9], we calculate the weekly realized returns of the portfolios by multiplying the vector

[8] This procedure allow us to avoid unnecessary iteration that slow down the computation.

[9] At a given end of month t, we have to determine the optimal allocation for time $t+1$ considering only the information available up to time t

of the assets weights w for the vector of assets returns r. Then from the realized returns of the tracking portfolios R_i we calculate their cumulated returns as:

$$Cum.\ Ret_t = \Big[\prod_{i=1}^{t}(1+R_i)\Big] - 1 \tag{4.9}$$

The cumulated return of each strategy represents the aggregate amount an investment has gained or lost over time, without considering the time period involved. In our specific case, we graphically analyze the cumulated returns of the constructed portfolios, comparing them with the benchmark cumulated return. The closer and the more similar the cumulated return paths are, the better the strategy replicates the benchmark.

The second indicator that we use in order to identify the best tracking portfolio, is the annualized standard deviation of the weekly portfolio realized returns

$$Ann.\ Volatility_t = \sum_{i=1}^{52}(R_i - \overline{R_t})^* \sqrt{52} \tag{4.10}$$

where $\overline{R_t}$ is simply the mean of the weekly returns of year t:

$$\overline{R_t} = \frac{1}{52}\sum_{i=1}^{52} R_i \tag{4.11}$$

This quantity refers to the amount of uncertainty or risk assigned to a given portfolio. A high level of volatility (standard deviation of portfolio returns) indicates that the price of the security can change dramatically within the considered time period in either directions. A low level of volatility, instead, testify that the value of the fund does not fluctuate too much, tending to be steady. For our purpose, considering portfolios with the same returns or similar cumulated returns, the one with lower volatility is preferable.

Concentrating now on the dynamic approach, the first relevant piece of information for strategies comparison is offered by the analysis of the portfolio weights. Indeed, we are immediately able to investigate their evolution over time by simply looking at the area plots of their values. We can state whether the wealth is too much concentrated in a specific asset, making the portfolio greatly exposed to single stock risk, or if the portfolio composition variates too much between subsequent periods. Thereafter, the latter investigation can be intensively analyzed by having a look at the so-called *Approximated Turnover*[10], representing the sum of the deviations

[10]The term *Approximated* is used because the weights assigned to a specific asset class at the beginning of each period, are not the same weights we observe at the end of the month just before the implementation of the allocation for the following month

in weight over two subsequent periods in absolute terms, divided by 2:

$$Appr.\ Turnover = \frac{1}{2}|w_{t-1} - w_t|. \tag{4.12}$$

This indicator offers a more complete understanding of which is the portfolio rotation over time and of the trading activity the portfolio manager has to be involved in, in order to adjust the portfolio composition. In the formulation above, the turnover has always a value included between 0 (in case of no stock's sale or purchase) and 1 (in the special case the portfolio completely changes) and it is able to monitor the changing in portfolio weights over time. However, another interesting formulation of the approximated turnover might be obtain if we drop the division by two, retrieving in this way the amount of trades executed: in this second form, the approximated turnover can be utilize to directly calculate the transaction cost in which the manager incurs while implementing a strategy. To do so, we simply have to multiply the turnover of each period for a fixed quantity representing the "fees" applied for the weekly change in the portfolio weights, in our case set at 20 bps[11]:

$$Transaction\ Cost_t = |w_{t-1} - w_t|^{*}0.0020. \tag{4.13}$$

Subtracting this quantity to the realized portfolio return we can then obtain the cost adjusted returns[12]

$$Adjusted\ Ret. = R_t - Transaction\ Cost_t. \tag{4.14}$$

Since the purpose of the dissertation is again the one of creating a tracking portfolio that keeps under control transaction costs, this is a key indicator we should carefully take into consideration.

Finally, in order to have a more complete overview of the results we obtained, the IR and Semi-IR for the constructed portfolios are computed. In fact, if in the static case the calculation of IR is obtained straightforwardly from the solution of the TEV minimization problem, in the former case it has to be calculated once the final portfolio returns are obtained. Each time the reallocation is performed, the portfolio tends to have a different IR due to portfolio rotation.

Spending few words also on the Semi-IR indicator, it is nothing more than the Information

[11]This amount covers transaction costs to execute and process the trades (include brokerage commissions and bid-ask spreads).

[12]It has to be specified that in this case the transaction costs are inputed at the end of each month in which the reallocation takes place. Since the returns R_t are instead calculated on weekly basis, the transaction costs are going to be converted and subtracted to the return of the last week of each month.

ratio, but computed using a TEV calculated only on downside deviations:

$$Semi-IR = \frac{TE}{Semi-TEV} = \frac{\mathbf{E}\left[R_p - R_B\right]}{\sigma\left[R_p - R_B|\ (R_p - R_B < 0)\right]}. \tag{4.15}$$

Indeed, the standard deviation has the great drawback of treating all deviations from the mean in the same way (positive and negative). However, investors are generally more concerned with negative divergences than with positive ones: using downside deviation we can solve this issue by focusing only on downside risk.

Chapter 5

Empirical Evidence

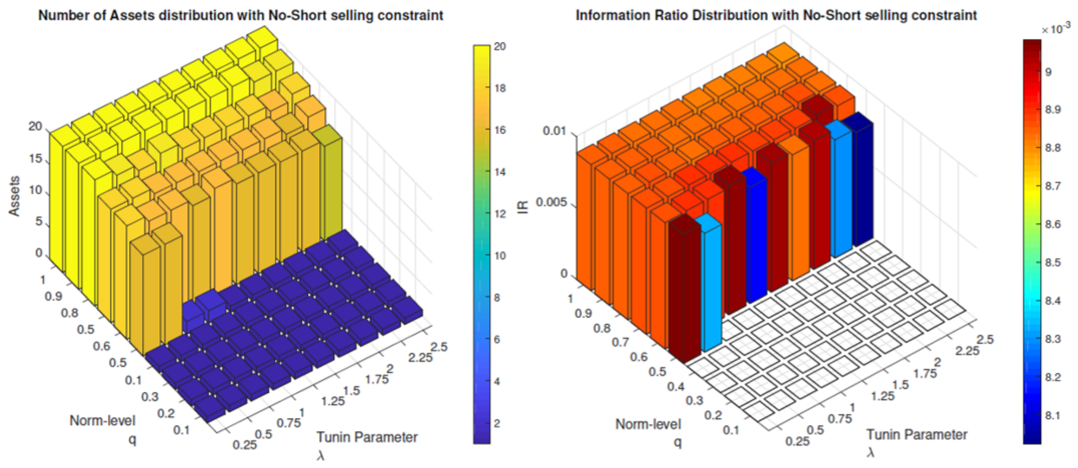
In this Chapter, the empirical results are going to be presented and analyzed, considering both the Buy and Hold strategies and the dynamic approaches. Comparison of cumulated returns, volatilities, weights composition over time and performance indicators between different strategies are examined, in order to determine which is the portfolio that most closely tracks the *S&P 100 Global Index*.

5.1 Results for the static approach

As already anticipated, the first approach we propose consists in the implementation of some Buy and Hold strategies considering, for instance, the No-Short selling constraint and two different type of upper bound constraints (UB) fixed, respectively, at 2% and 1% of the total portfolio weights. This first simple approach, even if we already know it is not implementable in the real world, gives us some important remarks and hints about what we are going to obtain when the working environment becomes more complicated.

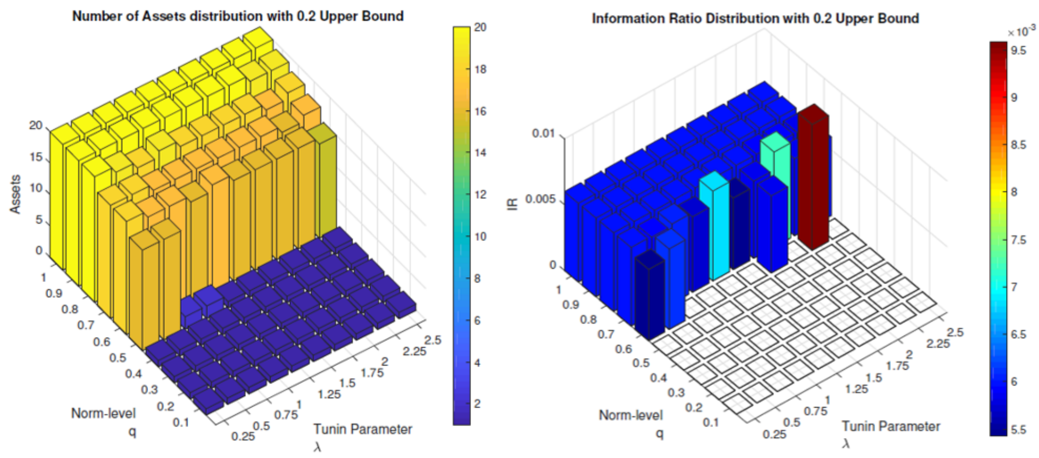
The first important piece of information we get from the minimization procedure refers to the number of selected assets included in the portfolios. As already specified, the GlobalSearch has been run without setting any bounds on the maximum number of assets acceptable in each portfolio, however, in all three cases, the maximum number of stocks converges to 20. This result underlines the goodness of penalized regression in inducing sparsity within the portfolio weights and in cutting transaction costs: a portfolio composed by 20 active positions at maximum is much more convenient and manageable compared to one portfolio containing 100 assets.

Focusing again on the distribution of the number of assets across different values of q and k, in particular comparing the No-Short selling and the 0.2 upper bound cases shown in Figure 5.1.a and Figure 5.2.a, we can appreciate how they appear to be quite similar at first sight. In reality, the distributions of the weights are dramatically different: considering two portfolios with the same number of open positions, in the case of No-Short selling we have the possibility of observing high concentration in some assets (up to 67% of the total wealth) and negligible weights is some others (e.i 0.012%). In the 0.2 upper bound case, instead, even if the selected assets are approximately the same, the amount of wealth is smoothly distributed among different stocks.



(a) (b)

Figure 5.1: Number of assets and IR distributions for portfolios minimizing the TEV with No-Short selling constraint.

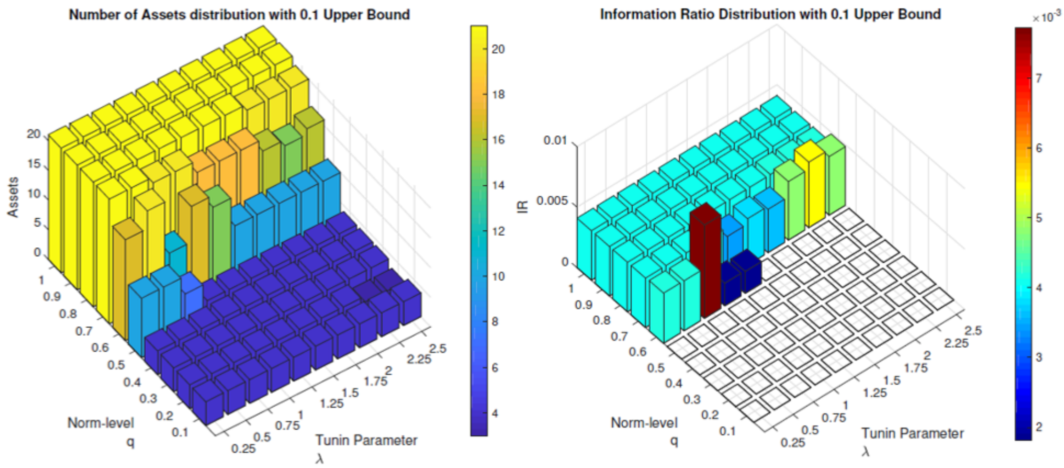


(a) (b)

Figure 5.2: Number of assets and IR distributions for portfolios minimizing the TEV with 0.2 upper bound.

The differences in composition between the two portfolios, can be better emphasized by looking at the distribution of the IR. Indeed, observing Figures 5.1.b and 5.2.b, we can appreciate how, basically for all q and k, the values for the 0.2 UB exhibits a much lower IR with respect to the No-Short selling case.

Moving now to the case of 0.1 UB (Figure 5.3), we can see how the number of assets included in the portfolios are slightly different with respect to the other two cases, most of all for values of q lower than 0.6. In fact, in this circumstance we do not have any portfolio concentrated in one single asset, but the less diversified portfolio is composed by at minimum 5 stocks. Moreover, considering that the Information Ratios are actually lower than in the previous two cases, we might expect that this will be the strategy providing us with the portfolio that better replicates the benchmark index.



(a) (b)

Figure 5.3: Number of assets and IR distributions for portfolios minimizing the TEV with 0.1 upper bound.

Even if, the distribution of the Information Ratios and the number of active positions give us a first general idea of what we obtained, other important factors have to be taken into account in order to determine which is the preferable portfolio. Indeed, if on one side we have already remarked the ability of penalized regression and sparse index tracking to reduce transaction costs, on the other side the inclusion of only 20 assets within the tracking portfolios might raise serious questions about their reliability in term of diversification. In order to remove these doubts, after the portfolio with minimum IR has been chosen for each of the three strategies, a much more detailed analysis about portfolio weights is carried out.

Table 5.1 lists the name of the stocks included in the portfolio with minimum IR for each strategy and their respective weights. As you can notice, the selected securities are for the most part the same: 11 assets appear in all the portfolios with almost the same weights, 5 appear in two of the three strategies and only 5 exclusively characterize one of them. Indeed, this evidence proves the consistency of the GlobalSearch in picking the assets that better describe the benchmark index since, changing the imposed constraints, we induce some notable changing in the weights allocation, but not in the selected stocks.

Tracking No-Short selling		**Tracking with 0.2 upper bound**		**Tracking with 0.1 upper bound**	
BHP BILLITON	0.12	BHP BILLITON	0.13	ALPHABET 'C'	0.07
BP	0.05	BP	0.05	BANCO SANTANDER	0.01
BRISTOL MYERS SQUIBB	0.04	BRISTOL MYERS SQUIBB	0.04	BHP BILLITON	0.10
CITIGROUP	0.00	DEUTSCHE TELEKOM	0.06	BP	0.07
DEUTSCHE TELEKOM	0.06	DOWDUPONT	0.20	BRISTOL MYERS SQUIBB	0.07
DOWDUPONT	0.31	ENGIE	0.03	DEUTSCHE TELEKOM	0.07
ENGIE	0.03	GLAXOSMITHKLINE	0.01	DOWDUPONT	0.10
LINDE	0.14	KIMBERLY-CLARK	0.01	KIMBERLY-CLARK	0.04
L'OREAL	0.03	LINDE	0.20	LINDE	0.10
NATIONAL GRID	0.02	L'OREAL	0.03	L'OREAL	0.06
PROCTER & GAMBLE	0.02	NATIONAL GRID	0.03	NATIONAL GRID	0.04
TEXAS INSTRUMENTS	0.08	PROCTER & GAMBLE	0.03	PROCTER & GAMBLE	0.09
VIVENDI	0.03	SCHNEIDER ELECTRIC SE	0.00	SCHNEIDER ELECTRIC SE	0.04
VODAFONE GROUP	0.03	TEXAS INSTRUMENTS	0.08	VODAFONE GROUP	0.05
WALMART	0.02	UNILEVER DUTCH CERT.	0.01	WALMART	0.08
		VIVENDI	0.04		
		VODAFONE GROUP	0.04		
		WALMART	0.03		

Table 5.1: Summary table for optimal portfolio weights considering the Buy and Hold strategies across different upper bounds settings. The weights are expressed in percentage and sum to one for each strategy.

At first glance, if we would like to select among the three allocations the one exhibiting the higher diversification, the choice will for sure fall to the 0.2 UB strategy because it contains more stocks, 18 assets, compared to the 15 of the others. However, diversification is not only a matter of how many assets are included in the portfolio, but also the nature of these securities has to be taken into account. For example, a portfolio containing 30 stocks coming from the same industry or the same geographical area, might result to be less diversified compared to a portfolio containing only 15 assets, but uniformly spread among different industries and nations.

Thus, in order to clarify this point, we retrieve for each stock some information about their industry and their geographical area, using the classification again provided by *DataStream-Thomson Reuters*. What we obtain is that the portfolio containing 18 stocks, in reality, does not add anything in term of diversification with respect to the other two. Quite the opposite, if we consider the classification by sectors, the 0.2 UB is for the 83% concentrated in the Industrial Sector against the 67% of the others (Appendix B.4). On the other hand, if we consider the classification of stocks across geographical areas, all the portfolios exhibit approximately the same level of diversification with almost 40% of the wealth concentrated in US stocks.

This particular distribution of portfolio weights, for instance, reflects the one observed from the beginning in the benchmark index. In fact, the *S&P 100 Global* is, per se, highly concentrated in Industrial USA companies and, since we are trying to replicate this index, we can expect nothing but a proportional diversification for the constructed portfolios.

Moving forward in our analysis, probably the most important and glaring indicator of each tracker performance, considered relatively to the benchmark index, is the cumulated return. Indeed, looking at the trends of the cumulated returns within the considered investment period (Figure 5.4), we can notice how not only all the constructed portfolios closely follow the path of the *S&P 100 Global*, but they also guarantee slightly better cumulated returns in all the 10 years horizon. Thus, a double accomplishment has been reached: selecting less than 20 stocks we have been able not only to closely replicate the benchmark but also to obtain a small extra return.

Comparing the trends for the tracking portfolios, we can appreciate how, initially, they are basically the same for all the three of them and only after January 2012, the tracker with No-short selling exhibits a higher cumulated return, with a consequently lower tracking ability.

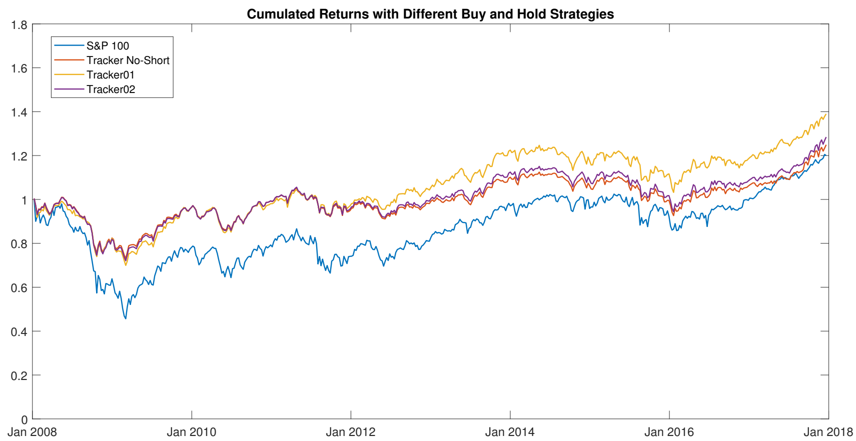


Figure 5.4: Graphical representation of the trackers cumulated returns with No-Short selling constraint, 0.2 upper bound and 0.1 upper bound.

Another phenomenon that we want to underline, is the fact that, visually, the three cumulated returns tend to get closer and closer with respect to the benchmark, when reaching the end of the period. We might suppose this behavior is strongly linked to the fact that approaching the end of the investment period, the assets included into the benchmark entirely matched the 104 stocks we considered for the implementation of the Buy and Hold strategy (Index composition at December 2017). Instead, for periods far from the cited date, some other constituents, that in the past has been included in the *S&P 100 Global* and that we exclude from our investment universe, lead to unavoidable discrepancies in cumulated returns.

Both the 0.2 and 0.1 UB strategies guarantee almost the same cumulated returns provided by the benchmark index and, in addition, they do it presenting a much lower level of volatility. Indeed, analyzing the annualized volatility in Figure 5.5 we can appreciate how the constructed portfolios generate much lower risk for a hypothetical investor, in particular in the first part of the investment period.

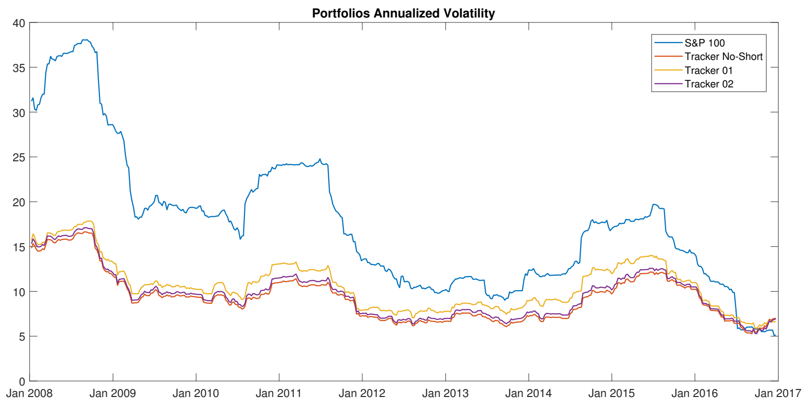


Figure 5.5: Graphical representation of the annualized volatility for the trackers with No-Short Selling constraint, 0.2 upper bound and 0.1 upper bound.

So, we can conclude that the constructed sparse index trackers are able not only to closely replicate the benchmark index and provide superior cumulated returns, but they indeed do so also reducing the risk at which the investor is exposed. Moreover, if we were to choose one single portfolio among the three proposed, considering the reflection discussed above on diversification and number of active positions and, in addition, looking the summary statistics shows in Table 5.2, we might finally conclude that the better alternative is represented by the tracking portfolio with 0.1 UB. It, indeed, guarantees the best TE, equals to 0.004, a positive IR of 0.0054 and a degree of diversification that is in line with the one delivered by the benchmark index.

	Mean	**Median**	**q1**	**q5**	**q95**	**q99**	**StDev**	**Min**	**Max**	**IR**	**Cum. Ret.**
S&P 100	0,071	0.16	-8.18	-4.13	4.07	7.25	2.67	-14.54	13.90	0	1.200
TrackerNS	0,052	0.09	-3.84	-2.27	2.37	3.40	1.37	-5.98	5.03	0,0080	1,249
Tracker02	0,058	0.10	-3.95	-2.38	2.41	3.51	1.42	-6.27	5.28	0,0054	1,284
Tracker01	0,075	0.08	-4.27	-2.52	2.60	4.17	1.53	-6.80	5.34	0,0018	1.390

Table 5.2: Summary table containing some performance indicators enabling to identify the best strategy among the ones proposed.

5.2 Results for the dynamic approaches

As already mentioned, even if the Buy and Hold approach guarantees good performances, it is not implementable in the real world since it considers the allocation procedure from a backward viewpoint, it does not take into account the variation in the benchmark index constituents and it does not consider the possible client capital flow during the investment period. Hence, using two different dynamic approaches, we try to overtake these issues by simulating the behavior of a portfolio manager that on a periodical basis, for example at time *t*, have to determine, knowing all the information about stocks returns up to time *t-1*, the optimal portfolio weights.

5.2.1 Portfolio reallocation at index constituents variation

In the first case, we decide to repeat the optimization procedure and reallocate the wealth among the portfolio weights each time the index composition varies. Since, in this circumstance, the analysis of the weights is much more complicated, due to their changing over time, for sick of simplicity we graphically present the portfolio rotation using area plots.

Figure 5.6 shows the legends of all the area plots provided below, in order to guarantee a more clear understanding of which assets are included in the various portfolios. As you can see, as in the Buy and Hold case, the three approaches are characterized by high robustness in term of selected stocks and in term of diversification: all the portfolios are mainly concentrated in the industrial sector. Furthermore, the No-Short selling portfolio is the one that in the five years has wield the higher number of assets, 34 to be exact, against the 32 of the 0.2 UB and the 31 of the 0.1 case.

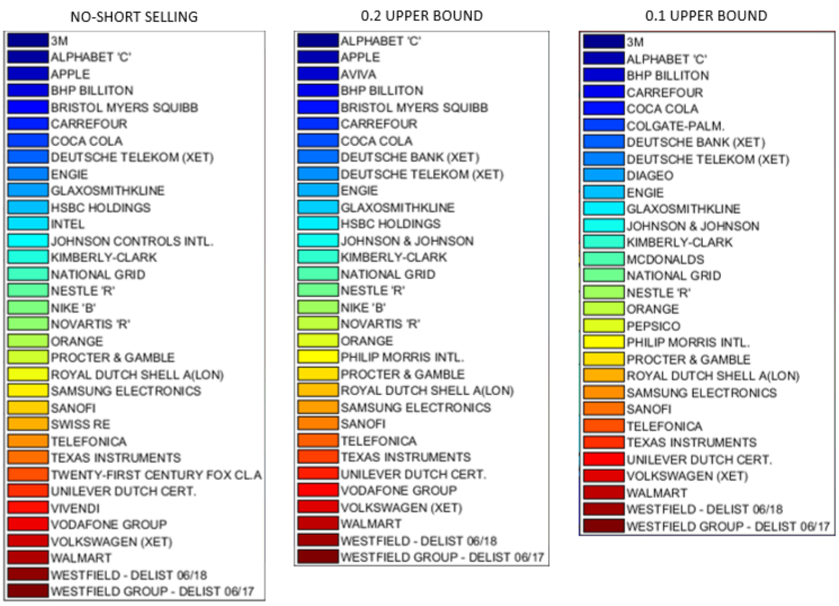


Figure 5.6: Legends of the area plots containing the portfolio rotations over the investment horizon.

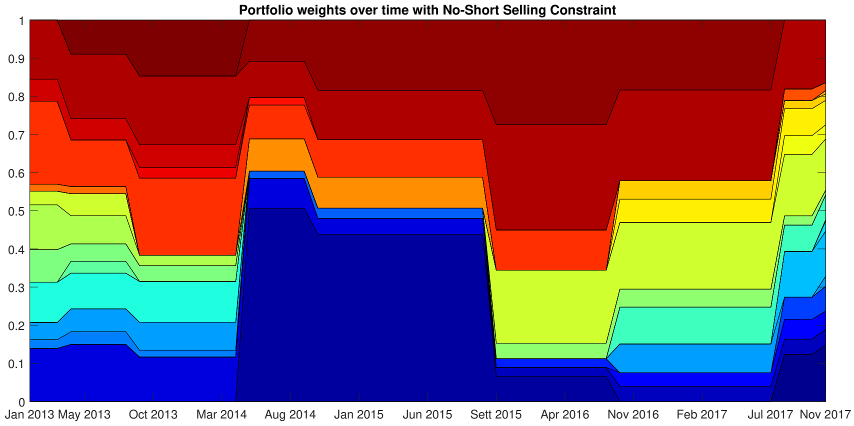


Figure 5.7: Rotation of portfolio weights with No-Short selling constraint

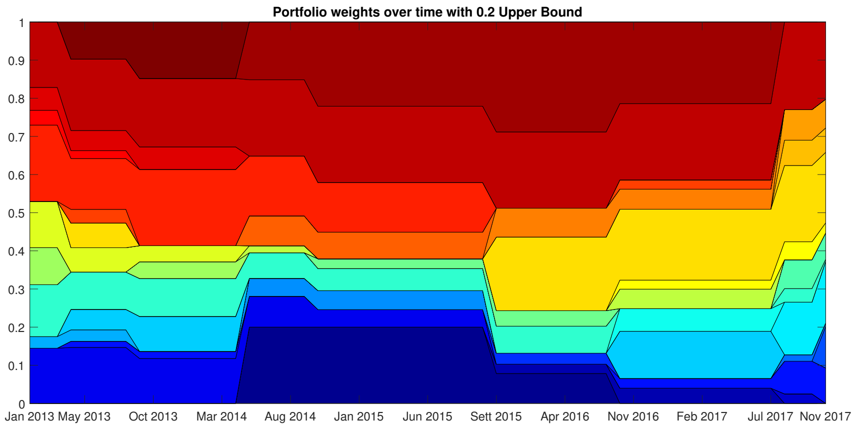


Figure 5.8: Rotation of portfolio weights with 0.2 upper bound

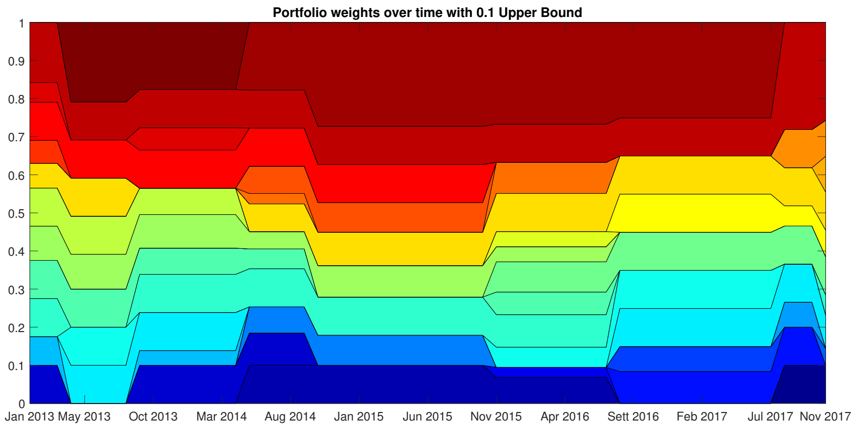


Figure 5.9: Rotation of portfolio weights with 0.1 upper bound

Moving to the graphical representations of the portfolio weights over time (Figure 5.7, Figure 5.8 and Figure 5.9) we have, first of all, to analyze the particular and very similar trends followed by the three portfolios. These paths, by construction, present constant weights for fairly long periods and spikes concurrently with the variation of the index constituents. Indeed, especially in Figure 5.7 and Figure 5.8, great spikes might be observed around April 2014, September 2015 and July 2017. Only to give an example, in the first case, 7 assets vary in total: some stocks have been removed from the index, namely *FUJIFILM HOLDINGS*, *NOKIA* and *WESTFIELD GROUP*; and some others as *ALPHABET A*, *ALPHABET C*, *APPLE* and *WESTFIELD* has been added. This variation induces, mainly in the No-Short selling case, a great level of turnover due to the fact that the portfolio weights have to be adapted considering the new index composition. Furthermore, the new allocation for April 2014 exhibits almost 50% of the invested wealth concentrated into the just included *ALPHABET C*. The high concentration in one specific asset is, perhaps, a source of impracticality for the aforementioned portfolio that, however, might be easily overtaken by imposing the satisfaction of upper bound constraints.

Another important point we want to stress, considering the strategies with upper bounds, regards the weights assigned to the numeraire assets, randomly chosen during the set-up of the minimization problem. In fact, since the weight for these particular stocks have been retrieved subtracting to 1 the sum of the other *n-1* assets, in order to respect the budget constraint, both in Figure 5.8 and Figure 5.9 they do not satisfy the upper bound constraints: *WESTFIELD* and *WALMART* have weights greater than the 20% of the total portfolio.

As done for the Buy and Hold case, Figure 5.10 shows the path of the portfolio cumulated returns.

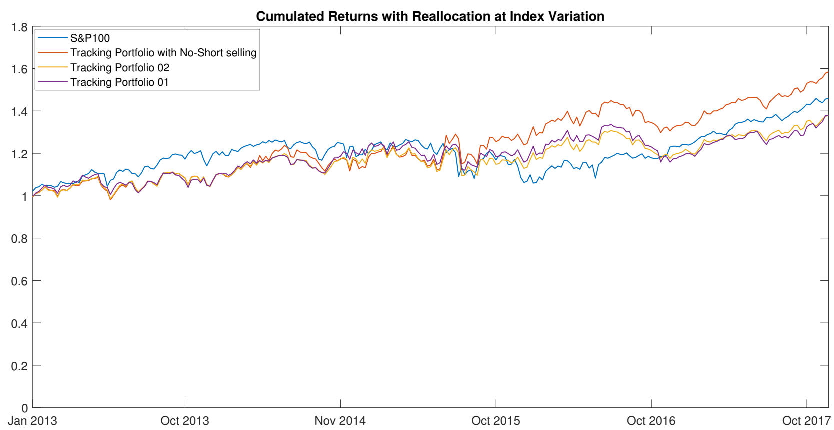


Figure 5.10: Graphical representation of the trackers cumulated returns with No-Short selling constraint, 0.2 upper bound and 0.1 upper bound when the reallocation is performed at index variation.

As expected, the tracking portfolios are able to replicate the benchmark index even better, compared to the previous case. Besides, differently from what we can observe in Figure 5.4, they are not able to guarantee superior cumulated returns neither during the whole period nor at the end of it, apart for the case with No-Short selling. Even though the latter portfolio provides advantages in term of cumulated returns, however it poorly replicates the benchmark index and it forces the investors to be exposed to highly concentrated positions.

Looking at the Annualized Volatilities in Figure 5.11, we can undoubtedly appreciate how the Buy and Hold strategies guarantee a much lower level of risk. Indeed, the volatilities of the trackers considered in this section are much more in line with the one of the benchmark index, apart from the 0.1 UB case that, in the first period, presents a slightly greater standard deviation.

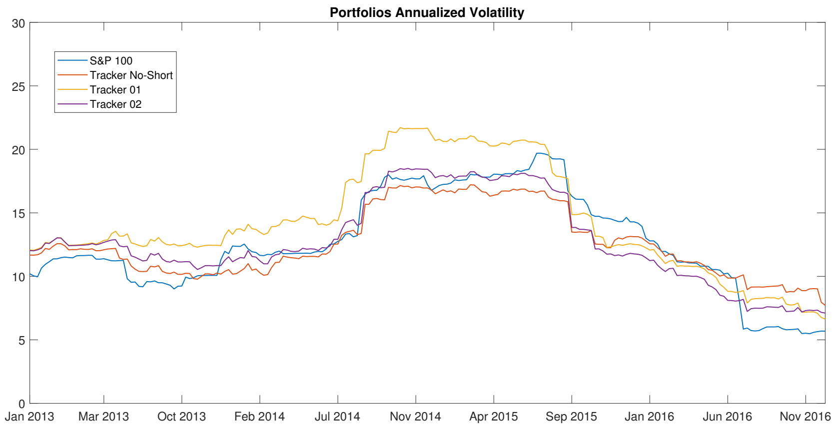


Figure 5.11: Graphical representation of the annualized volatility for the trackers with No-Short selling constraint, 0.2 upper bound and 0.1 upper bound when reallocation is performed at index variation.

As final remarks, to reinforce what has been shown from a graphical viewpoint, we look at some numerical values presented in Table 5.3, containing the summary statistics for both the tracking and the benchmark portfolios. In this case, only the tracker with No-Short selling constraint provides, on average, a superior return compared to the benchmark. Instead, considering both the IR and the Semi-IR the tracker with 0.2 UB guarantees the best performances, closely replicating the benchmark with an IR of 0.0041 and a cumulated return at the end of the period that is 0,0789 lower than the one generated by the *S&P 100 Global*.

	Mean	Median	q1	q5	q95	q99	StDev	Min	Max	IR	SemiIR	Cum.Ret
S&P 100	0,16209	0,31584	-5,27394	-3,01053	2,68516	4,27131	1,71568	-9,28902	5,85810	0	0	1,4597
TrackerNS	0,19758	0,19949	-4,15697	-3,03246	2,98226	5,33730	1,92270	-9,39407	8,25371	0,00564	0,01912	1,5842
Tracker02	0,14031	0,17521	-3,89735	-2,68165	2,86988	4,27434	1,73078	-8,86516	6,70975	-0,00412	-0,01261	1,3798
Tracker01	0,13932	0,10018	-4,13437	-2,82683	2,88152	4,20227	1,69616	-8,30595	5,76603	-0,00450	-0,01358	1,3783

Table 5.3: Summary table containing performance indicators gross of transaction costs, enabling to identify the best strategy among the ones proposed.

The results proposed until now, are the ones obtained without taking into account the impact of transaction costs on the portfolios realized returns. In order to give a more faithful picture of the reality, we should indeed include in the analysis the costs at which the portfolio management is exposed during the implementation of the strategies. In doing so, we consider the approximated turnovers plotted in Figure 5.12.

As you can see, the portfolio variation are quite rare and highly concentrated in the point in time in which the index composition changes. The paths for the three strategies are, obviously very similar, with extremely high turnover concurrently in April 2014, November 2015 and July 2017.

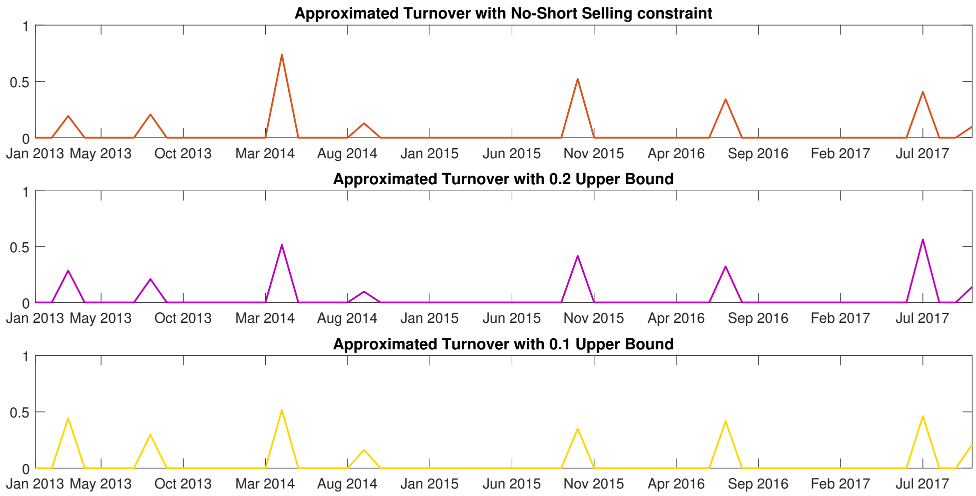


Figure 5.12: Approximated turnover considering reallocation at index variation for the three strategies proposed.

From these values, considering a cost of 20 bps for each portfolio weight adjustment, we calculate the adjusted returns and retrieve again all the summary statics, subtracting the transaction costs to the portfolio realized returns. The values obtained, reported in Table 5.4, are quite the same of the ones observed in the previous case. In fact, the impact that transaction costs have on tracking portfolios are very tiny and homogeneous for all the strategies: at the end of the period they account only for 0.010% of the initial investment.

	Mean	**Median**	**q1**	**q5**	**q95**	**q99**	**StDev**	**Min**	**Max**	**IR**	**SemiIR**	**Cum.Ret**	**Costs**
S&P 100	0,16209	0,31584	-5,27394	-3,01053	2,68516	4,27131	1,71568	-9,28902	5,85810	0	0	1,4597	0
TrackerNS	0,19754	0,19949	-4,15697	-3,03246	2,98226	5,33730	1,92268	-9,39407	8,25371	0,00564	0,01910	1,5840	0,010
Tracker02	0,14027	0,17521	-3,89735	-2,68165	2,86988	4,27434	1,73077	-8,86516	6,70975	-0,00413	-0,01263	1,3797	0,010
Tracker01	0,13927	0,10018	-4,13437	-2,82683	2,88152	4,20227	1,69616	-8,30595	5,76603	-0,00451	-0,01361	1,3782	0,011

Table 5.4: Summary table containing some performance indicators calculated using returns including transaction costs.

5.2.2 Portfolio reallocation on a monthly basis

In the second case, differently from what has been previously done, we repeat the optimization procedure and reallocate the constructed portfolios on a monthly basis, in order to be able to improve the replication ability of the trackers and to adjust their weights consequently to the hypothetical client capital flow.

Figure 5.13 shows all the stocks that, at least for once, have been included in the tracking portfolios. Considering the various strategies, we can still appreciate great similarities among the selected stocks, but also observe, considering them relatively to the ones coming from the previous dynamic approach, how they exhibit higher portfolio rotation. Indeed, reallocating weights distributions on a monthly basis, a much larger number of stocks is included at least once in the trackers. In particular, the No-Short selling portfolio contains 46 stocks against the 34 of the previous case, the 0.1 UB portfolio include 17 securities more with respect to the 31 considered above and the 0.2 UB tracker rotate 62 assets against the 32 analyzed earlier.

This evidence clearly indicate how the last approach is the one that wields the higher number of stocks among the 134 taken as inputs of the minimization problem and, thus, let us strongly believe that it will also be the best in term of replication power.

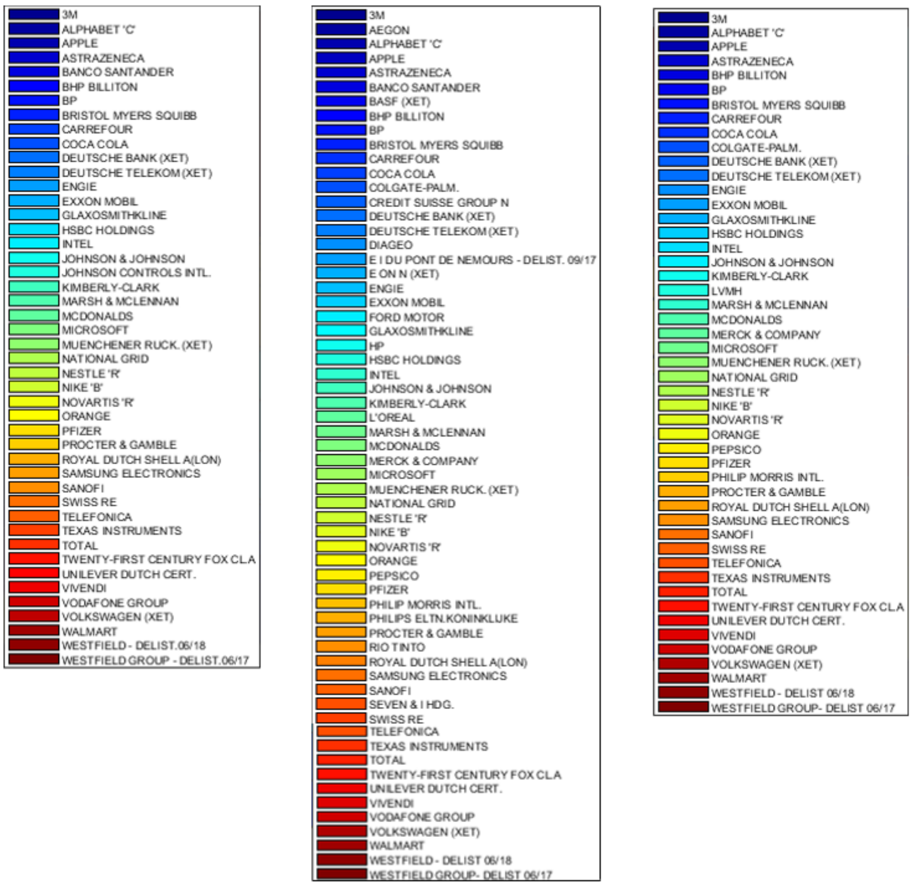


Figure 5.13: Legends of the area plots containing the monthly portfolio rotations.

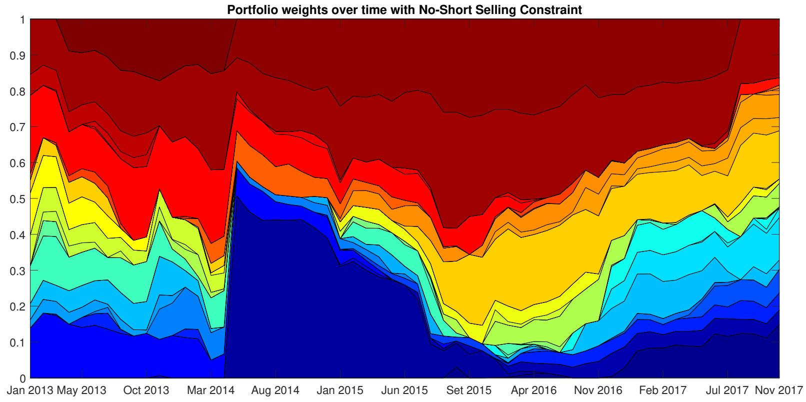


Figure 5.14: Monthly rotation of portfolio weights with No-Short selling.

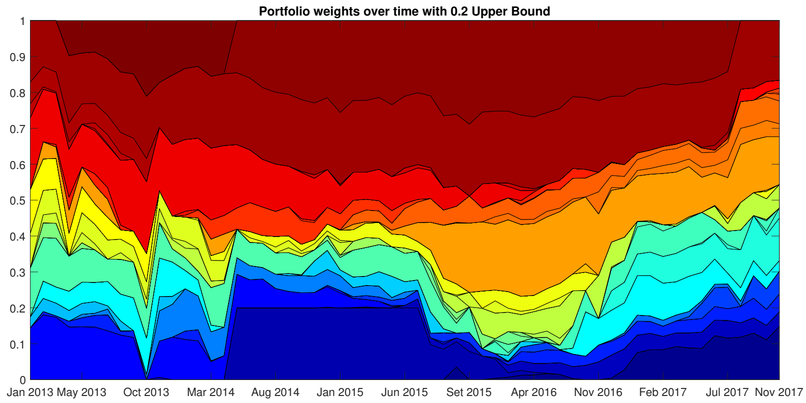


Figure 5.15: Monthly rotation of portfolio weights with 0.2 upper bound.

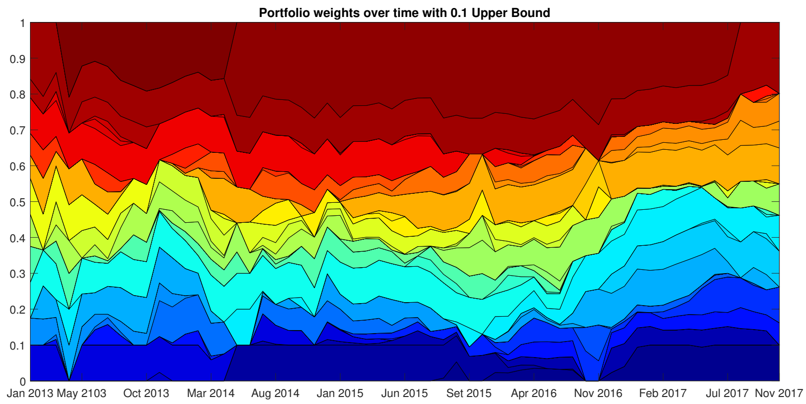


Figure 5.16: Monthly rotation of portfolio weights with 0.1 upper bound.

Looking at the graphical representations of portfolio weights over time, presented in the area plots in Figure 5.14, Figure 5.15 and Figure 5.16, we can appreciate how the results are partially different compared to the ones obtained in the dynamic approach with reallocation at index variation. In fact, in the latter figures, the volatilities of the weights are much higher because of the more frequent portfolio reallocation, but at the same time their paths appear to be smoother over time. The weights, in this case, slightly change time by time at a monthly basis, but they still adjust the most concurrently with the variation of index constituents.

For instance, if we individually analyze the portfolio rotations, we can observe how in the No-Short selling case the huge concentration in *ALPHABET C* still persists in the period following April 2014. Although, if the evidence in Figure 5.7 shows a sudden drop from 0.5 to 0.1 of its weight, in the latter case we can appreciate a gradual reduction and stabilization of it.

On the other hand, considering the 0.2 and 0.1 UB cases, we have to remark how they homogeneously distribute the wealth among an acceptable number of assets, thus enabling to diversify the risk among different securities. In the case of 0.2 UB the active positions registered go from a minimum of 9 in January 2013 to a maximum of 31 in May 2014, with an average of 14 open positions in each period. On the other hand, in the 0.1 UB case, the tracker contains from a minimum of 9 assets in the periods May-June 2014 and August-September 2016 up to a maximum of 24 securities in March-April 2016 with a slightly higher mean of 16 open positions.

Deepening now into the analysis of cumulated returns, we want to stress the fact that, as expected, proceeding with a monthly reallocation, we are able to obtain the best performances for what concerns the benchmark replication.

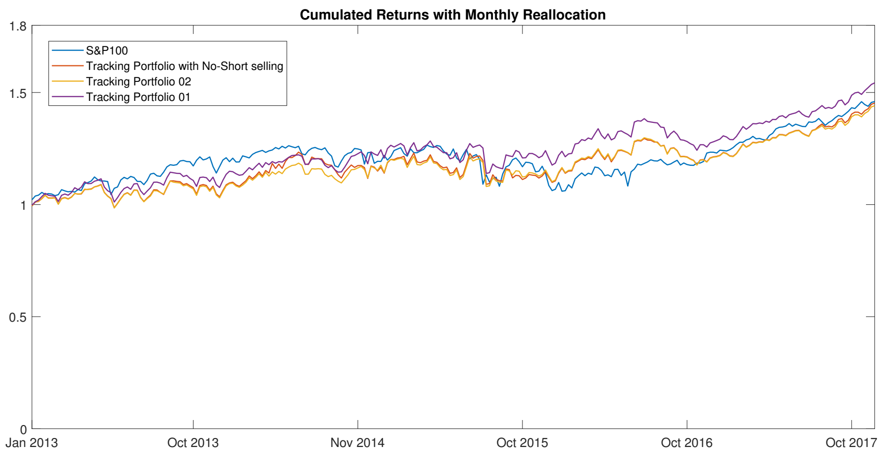


Figure 5.17: Graphical representation of the trackers cumulated returns with No-Short selling constraint, 0.2 upper bound and 0.1 upper bound when the reallocation is performed on a monthly basis.

As shown in Figure 5.17, the 0.1 UB strategy provides, from July 2015, superior cumulated returns with respect to the benchmark and at December 2017, it guarantees a cumulated return equals to 1,543, 0.0083 higher than the one of the index, while exhibiting also a good level of diversification. Considering, instead, the tracking portfolios with 0.2 UB and No-Short selling constraint, they are characterized by almost identical patterns, apart from the months between April 2014 to November 2014: the period in which the No-Short selling portfolio was highly concentrated in *ALPHABET C*. The performances of the two strategies appear to be very close to the one of the index, so as to provide, at the end of the period, a cumulated return of 1,453 and 1,441 respectively. These two, at the of the story, represent the tracking portfolios that most closely replicate the returns of the *S&P 100 Global*: they guarantee cumulated returns that are only 0.0118 and 0.0069 lower than the benchmark one, indeed maintaining for the greatest part of the period a slightly lower annualized volatility (Figure 5.18)

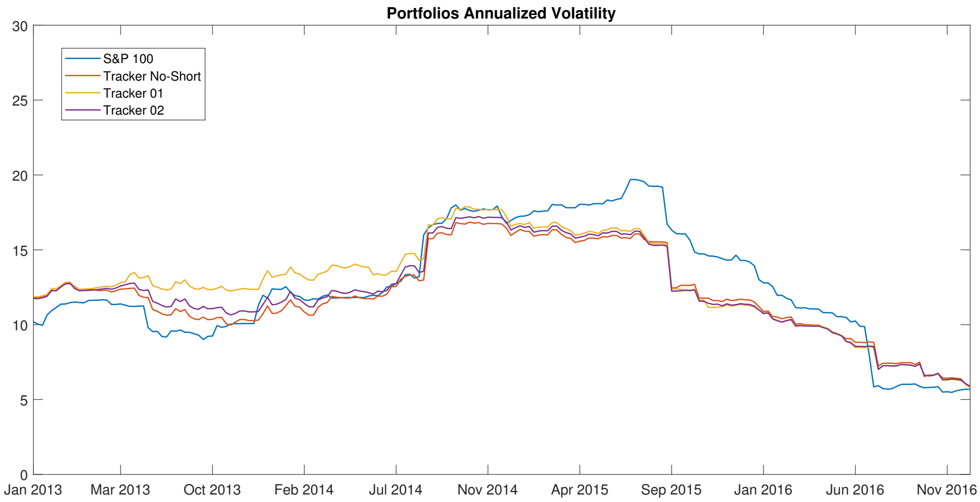


Figure 5.18: Graphical representation of the trackers annualized volatility with No-Short selling constraint, 0.2 upper bound and 0.1 upper bound when the reallocation is performed on a monthly basis.

These conclusions, reached through a graphical analysis of the results, are confirmed and reinforced if we look at the numerical values in Table 5.5. Looking at the values for the mean and the cumulated returns, we can appreciate how, as already visually noticed, only the tracker with No-Short selling is able to guarantee superior average returns compared to the benchmark.

	Mean	**Median**	**q1**	**q5**	**q95**	**q99**	**StDev**	**Min**	**Max**	**IR**	**SemiIR**	**Cum.Ret**
S&P 100	0,16209	0,31584	-5,27394	-3,01053	2,68516	4,27131	1,71568	-9,28902	5,85810	0	0	1.4597
TrackerNS	0,16006	0,15741	-3,98840	-2,73932	2,77436	4,83406	1,71024	-8,90282	4,98092	-0,00038	-0,00115	1,4528
Tracker02	0,15581	0,16139	-3,64431	-2,72945	2,43528	4,51733	1,64637	-8,96708	5,02749	-0,00126	-0,00370	1,4410
Tracker01	0,18211	0,21336	-3,56997	-2,65214	2,59906	4,42746	1,62478	-9,07260	5,31109	0,00418	0,01176	1,5429

Table 5.5: Summary table containing some performance indicators for the monthly adjusted portfolios.

Moreover, considering the values assigned to the Median, we have to underline a crucial difference between the behavior of the benchmark and the ones of the tracking portfolios. Indeed, if the values of the second quartile are very similar to the ones of the mean for the constructed portfolios, that is not the case if we consider the benchmark index, for which the mean has a value that is almost half of the median. So, in general, this testify the presence of some outliers in the very low end of the distribution: half of it is concentrated around 0,31, but there are some very low returns that drag down the average performance of the entire period. For the trackers instead, as shown also in Figure 5.18, the volatility of the returns is reduced simultaneously with the presence of outliers in the left tale, guaranteeing a much more stable and centered distribution of returns. This evidence can be also appreciated considering the values for the *q1* and *q5* quantiles: they are indeed much lower with respect to the ones of the benchmark.

Spending few words on the Information Ratio and Semi-IR, yet they confirm what has been said during graphical analysis of the cumulated returns: the monthly reallocation allows us to obtain the trackers that most closely replicate the returns of the *S&P 100 Global* with IRs of -0.00038 and -0.00126.

As we already said before, the results exposed until now do not take into account the impact that the transaction costs have on the realized returns of the tracking portfolios. Differently from what we observed using the previous approach, their effect on trackers performances are way more important: the approximated turnovers are higher and, most of all, more frequent (Figure 5.19). For the first two strategies the trends of the approximated turnovers are basically the same but with a small difference in April 2014, where the turnover of the No-Short selling portfolio is slightly higher. On the other side, the 0.1 UB case presents higher spikes toward the end of the period and so presumingly a higher level of transaction costs.

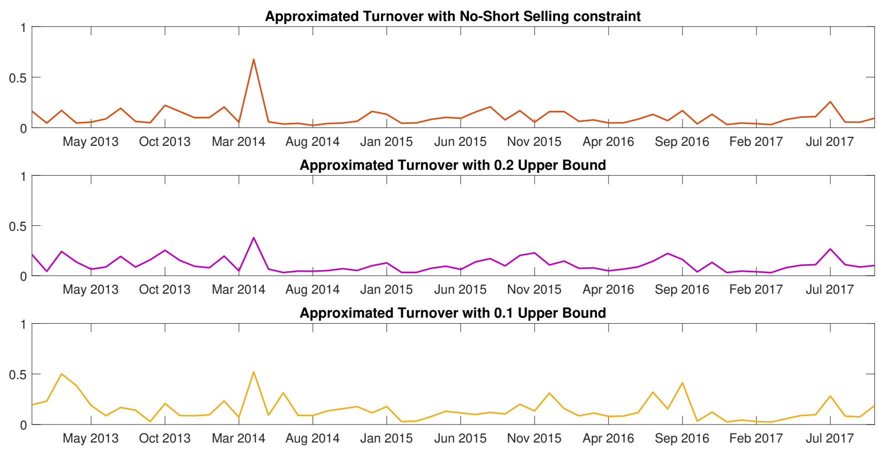


Figure 5.19: Approximated turnover for strategies with monthly reallocation.

Thus, in the monthly reallocation case, the transaction costs have a much greater impact on the performance of the trackers as appreciable in Table 5.6. In the case of No-Short selling and 0.2 UB, the transaction costs are more than doubled, accounting respectively for 0.0024% and 0.0026% of the amount initially invested. In the 0.1 UB case, they are even higher, reaching an amount of 0.0034%.

	Mean	Median	q1	q5	q95	q99	StDev	Min	Max	IR	SemiIR	Cum. Ret.	Costs
S&P 100	0,16209	0,31584	-5,27394	-3,01053	2,68516	4,27131	1,71568	-9,28902	5,85810	0	0	1.4597	0
TrackerNS	0,15997	0,15741	-3,98860	-2,73975	2,77429	4,83406	1,71021	-8,90344	4,98066	-0,00040	-0,00120	1,4524	0,024
Tracker02	0,15571	0,16139	-3,64431	-2,73000	2,43528	4,51710	1,64636	-8,96763	5,02749	-0,00128	-0,00376	1,4406	0,026
Tracker01	0,18198	0,21306	-3,56997	-2,65214	2,59906	4,42678	1,62476	-9,07299	5,31109	0,00415	0,01168	1,5424	0,034

Table 5.6: Summary table containing some performance indicators for the monthly adjusted portfolios net of transaction cost.

Despite of the impact of transaction costs, the 0.2 UB strategy still exhibit the best performances among all the constructed portfolios, with excellent values for both IR, equals to -0.00128, and a Semi-IR equals to -0.00376.

Conclusions and future works

Sparse index tracking is a passive portfolio management strategy that aims to replicate the performance of an index by using only a small fraction of its constituents, in such a way to reduce transaction costs and avoid illiquid positions. The problem is usually formulated as an optimization problem with the objective to minimize a given distance measure between the tracking portfolio and a benchmark index.

In this dissertation, starting from the seminal contribution of Fan et al. (2012) that illustrates the connection between the risk minimization problem and a penalized least-square problem, we propose a methodology that allows to minimize a given tracking error measure by imposing a constraint on the q-norm ($0 < q < 1$) of the portfolio weights. Such penalized regression model, well-established in the statistical community as a way for performing a simultaneous model selection, permits to promote sparsity and, therefore, to select and estimate only few non-zero coefficients of explanatory variables, controlling indeed the degree of diversification in the tracking portfolio.

Considering as distance measure the Tracking Error Volatility, we solve the minimization problem working on the *S&P 100 Global* and implementing three different approaches: a Buy and Hold strategy, a dynamic approach involving portfolio reallocations at index variation and a dynamic approach in which the weights have been adjusted on a monthly basis.

From the empirical evidence obtained, we have analyzed, for each of the cited approaches, the performance of the tracking portfolio with minimum IR imposing, one at the time, the No-Short selling constraint, the 0.2 UB and the 0.1 UB constraints. In particular, we obtain in all cases fairly good performances both in term of tracking ability, the worst portfolio is characterized by an IR of 0.008, and in terms of diversification. In fact, even if the latter is not excellent in absolute terms, due to the fact that from the beginning the *S&P 100 Global* is highly concentrated in USA Industrial stocks, however it preserves in all cases a strong proportionality in the sectoral and geographical distribution of the stocks with respect to the benchmark.

Considering the trends of cumulated returns, we observe how all the strategies closely follow the path of the benchmark index, showing or slightly higher or slightly lower cumulated returns at the end of the investment period, but always guaranteeing a relatively lower level of volatility. Despite everything, the approach involving the monthly reallocation ended to be not only the one that better simulates the conditions in which portfolio managers operate in the real world, considering the periodical variation of the index constituents and the possible clients capital flow, but also the one providing the best results, even taking into account the impact that the transaction costs have on the final return.

Thus, in this dissertation, some important results have been obtained. First, we have shown how posing the q-norm constraint allows to regularize the index tracking problem, determining in one single step the number of active weights and their optimal values. Nevertheless, repeating the optimization procedure for different values of the tuning parameter k and different values of the q-norm, we have been able to retrieve a wide set of portfolios giving us the opportunity to choose among various allocation strategies with different characteristics in term of concentration and number of open positions.

Second, we have highlighted the extraordinary results obtained solving the minimization problem through the GlobalSearch function, both in term of transaction costs, but also and above all in terms of tracking accuracy. Indeed, in the proposed cases, we have been able to closely replicate the performance of the *S&P 100 Global* by using only approximately one third of the total index constituents. In particular, from the empirical evidence, one portfolio could be picked as potential "winner": the portfolio with 0.2 upper bound in the case of monthly reallocation. Indeed, it guarantees in each moment a homogeneous distribution of weight across different assets, it delivers a more than acceptable level of turnover, exposing the portfolio manager to restraint transaction costs, and it generates a return that is in mean and net of transaction costs lower but very close to the one of the benchmark index.

Even if the results coming from the empirical evidence are good in term of tracking accuracy and low transaction costs, the proposed approach evaluates the latter only after the minimization procedure has been implemented, without setting in advance any bounds on the level of approximated turnover. In order to take into consideration also this aspect, in future works we might consider not only the possibility of implementing the discussed approach referring to other financial market indexes, but also the one of including a turnover constraint within the GlobalSearch settings. In order to do so, since the constraints have to be expressed in a linear form, the minimization problem should be reformulated using a pairwise representation of portfolio weights. For instance, at time t the portfolio weights might be expressed as the weights at time $t\text{-}1$ plus the stocks we bought (b) and minus the stocks we sold (s) within the two periods. In this case, setting some bounds on the absolute value of the difference between b and s, we might be able to keep under control from the beginning the variation of portfolio weights and, thus, the transaction costs.

Appendix A

Markovitz Optimization

Markovitz optimization problem:

$$\begin{aligned} \min_{w} \quad & w'\Sigma w \\ s.t. \quad & w'\mu = \mu_p \\ & w'1_N = 1 \end{aligned} \tag{A.1}$$

In order to solve this problem, the method of Lagrange multipliers to convex optimization (minimization) problem subjected to linear constrained is applied.

- Define the Lagrangian:

$$L(w,\lambda_1,\lambda_2) = w'\Sigma w + \lambda_1(\mu_p - w'\mu) + \lambda_2(1 - w'1_N) \tag{A.2}$$

- Derive the First Order Condition:

$$\begin{cases} \frac{\delta L}{\delta w} = 0 & \rightarrow \quad \Sigma w - \lambda_1\mu - \lambda_2 1_N = 0 \\ \frac{\delta L}{\delta \lambda_1} = 0 & \rightarrow \quad \mu_p - w'\mu = 0 \\ \frac{\delta L}{\delta \lambda_2} = 0 & \rightarrow \quad 1 - w'1_N = 0 \end{cases}$$

- Solving for w in term of λ_1 and λ_2 from the first equation in the system gives:

$$w = \lambda_1\Sigma^{-1}\mu + \lambda_2\Sigma^{-1}1_N \tag{A.3}$$

- Solving for λ_1 and λ_2 by substituting (9) in the second and third equations of the system above the following is obtained:

$$\mu_p = w'\mu = \lambda_1(\mu'\Sigma^{-1}\mu) + \lambda_2(\mu'\Sigma^{-1}1_N)$$
$$1 = w'1_N = \lambda_1(\mu'\Sigma^{-1}1_N) + \lambda_2(1_N'\Sigma^{-1}1_N)$$

that written is matrix form become:

$$\begin{bmatrix} \mu p \\ 1 \end{bmatrix} = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} \lambda 1 \\ \lambda 2 \end{bmatrix} \tag{A.4}$$

where $a = (\mu'\Sigma^{-1}\mu)$, $b = (\mu'\Sigma^{-1}1_N)$ and $c = (1'_N\Sigma^{-1}1_N)$
With given values of $lambda_1$ and λ_2, the solution portfolio has minimum variance equals to:

$$\sigma^2 = w'\Sigma w = \lambda_1^2(\mu'\Sigma^{-1}\mu) + 2\lambda_1\lambda_2(\mu'\Sigma^{-1}1_N) + \lambda_2^2(1'_N\Sigma^{-1}1_N) \tag{A.5}$$

or in Matrix form:

$$\sigma 2 = \begin{bmatrix} \lambda_1 \\ \lambda 2 \end{bmatrix}' \begin{bmatrix} a & b \\ b & c \end{bmatrix}^{-1} \begin{bmatrix} \lambda_1 \\ \lambda 2 \end{bmatrix} \tag{A.6}$$

- At this point if $\lambda_1$and λ_2 are isolated from (11) and substituted in (12), the expression for the Efficient Frontier can be retrieved:

$$\sigma 2 = \begin{bmatrix} \mu p \\ 1 \end{bmatrix}' \begin{bmatrix} a & b \\ b & c \end{bmatrix}_{-1} \begin{bmatrix} \mu p \\ 1 \end{bmatrix} = \frac{1}{ac - b^2}(c\mu_p^2 - 2b\mu_p + a)$$

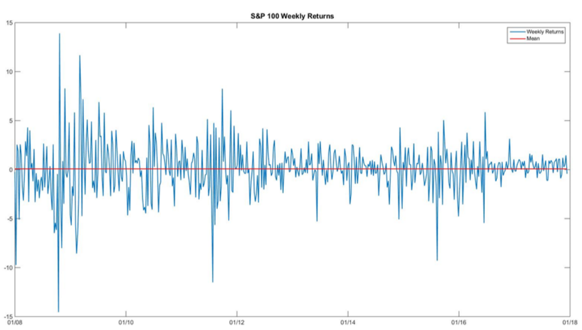
Appendix B

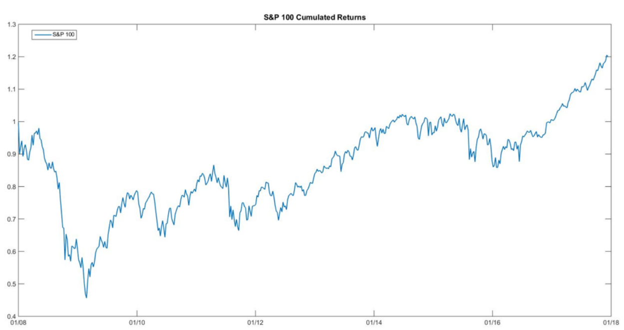
Figures and Tabels

B.1 Summary table for regularization methods

	PENALTY FUNCTION	ADVANTAGES	DISADVANTAGES
Ridge	$\left(\sum_{j=1}^{p} w_j^2\right)^{1/2} < t^2$	• Achieves numerical stability • High predictive performances • Shrinks parameters toward zero	• No variable selection is performed
LASSO	$\sum_{j=1}^{p} \lvert w_j \rvert < t$	• Increases interpretability of the model • Generates sparse model	• If p>n it selects at most n variables • If n>p it has worst performances than Ridge • No grouped selection effects
Elastic Net	$\sum_{j=1}^{p} \left(\alpha \lvert w_j \rvert + (1-\alpha) w_j^2\right) < t$	• Grouped selection is performed • Automatic variable selection	• High computational costs
q-norm	$\sum_{j=1}^{p} w_j^q < t \quad \text{with } 0 < q < 1$	• Promotes sparsity • Reduces the Bias of Lasso	• It represents a no-convex optimization

B.2 Data descriptive analysis



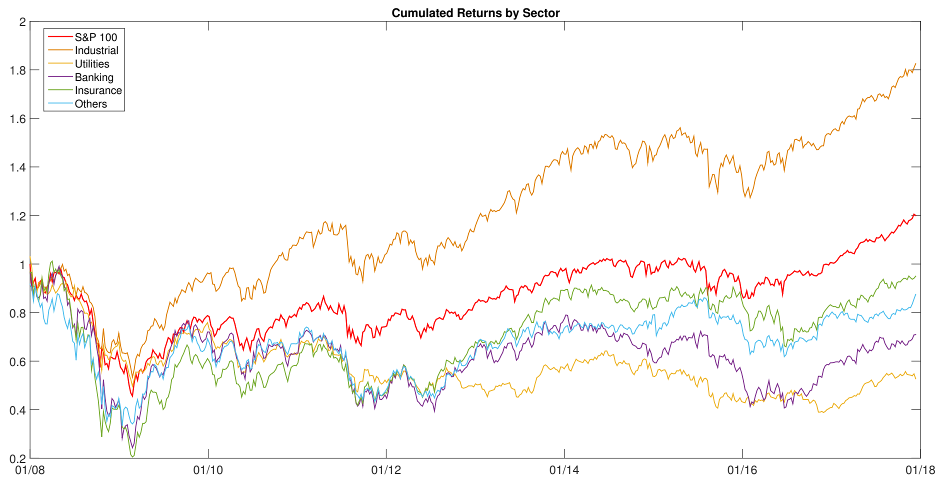


B.3 Summary statistics for all stocks

In order to give an brief overview about the behavior of the constituents composing the Benchmark, since their number is too large, a split by Sector and Nation has been performed. This has been done simply grouping all the securities appertaining to a given super-sector or geographical area and then averaging them out without considering their market capitalization. In such a way the mean return over time for each category is obtained.

B.3.1 Grouped by super-sector

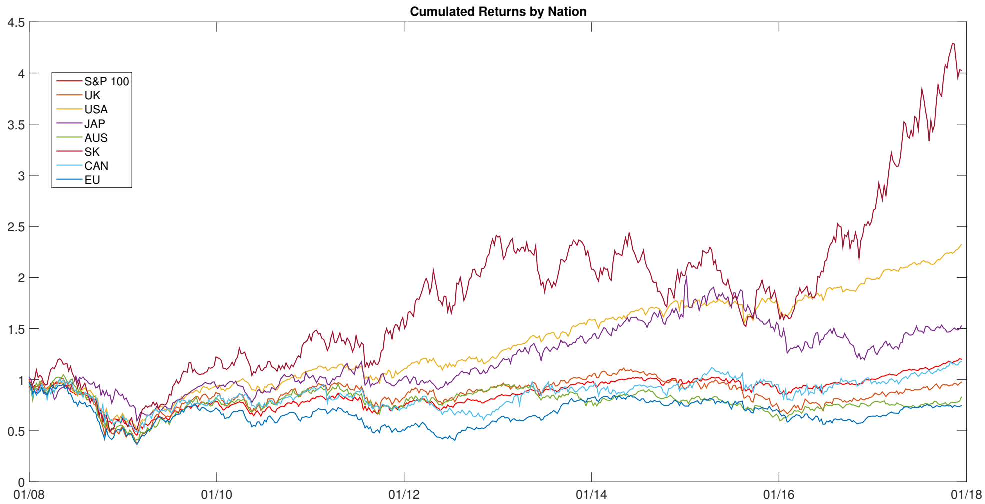
	Mean	Median	StDev	Min	Max	Skew	Kurt
Industrial	0.15	0.28	2.42	-10.58	12.72	-0.27	6.64
Utilities	-0.08	0.03	2.97	-11.70	17.96	0.12	6.05
Banking	0.06	0.02	5.05	-28.40	29.70	0.24	9.69
Insurance	0.10	0.06	4.72	-21.03	35.87	0.68	12.62
Others	0.06	0.06	4.04	-18.70	20.06	-0.05	8.48



The figure above shows the trends of the realized cumulated returns in the investment period considered. Notice that apart from the Industrial sector, containing the vast majority of the stocks included in the index (93 out of 134), all other sectors have shown poor performances that presumably drag down the overall performance of the benchmark.

B.3.2 Grouped by geographical area

	Mean	**Median**	**StDev**	**Min**	**Max**	**Skew**	**Kurt**
UK	0.06	0.06	3.48	-14.43	21.65	0.23	7.14
USA	0.19	0.35	2.35	-12.45	12.34	-0.66	8.17
JAP	0.14	0.24	3.33	-17.40	19.05	-0.34	7.15
AUS	0.02	0.05	3.36	-21.90	17.81	-0.22	8.59
SK	0.36	0.14	4.29	-16.30	12.70	-0.06	3.81
CAN	0.12	0.34	4.15	-23.66	17.91	-0.31	6.81
EU	0.01	0.09	3.63	-12.31	23.27	0.29	7.18



Analyzing the realized cumulated returns split by Geographical area, it has to be underlined that almost all the groups follow the same trend with respect to the Benchmark. The only strange behaviors is showed in the case of South Korea. Indeed, the only stock included in this group is the one of *Samsung Eletronics* and probably this explain the outstanding performance during last two year considered.

B.4 Buy and Hold portfolios composition

	Industrial	Utility	Savings
Tracker UC	66,67%	26,67%	6,67%
Tracker 0.2	83%	17%	0
Tracker 0.1	73%	20%	7%

	UK	USA	AUSTRALIA	FRANCE	GERMANY	NETHERLANDS	SPAIN
Tracker UC	20%	40%	6.67%	29%	13.33%	0	0
Tracker 0.2	17%	33%	6%	22%	17%	6%	0
Tracker 0.1	20%	40%	7%	13%	13%	0	7%

Appendix C

Matlab Codes

C.1 Example 1

```
clear;
clc;
%  matrix that contains the average returns of assets
MM= [0.04 0.05 0.06 0.07; 0.05 0.05 0.06 0.07; 0.04 0.06 0.06 0.07;
    0.04 0.05 0.07 0.07; 0.04 0.05 0.06 0.08];
% vector containing assets variances
Var= [0.13 0.15 0.16 0.17];
stdev= sqrt(Var);
corr= 0.6;
% covariance matrix
for j=1:4
    for i=1:4
            MV(i,j)=stdev(1,i)'*corr*stdev(1,j);
            if i==j
                MV(i,j)=Var(i)
            end
    end
end
% VARIATION IN MEAN OF ASSETS
% Maximum trade-off portfolios: weights, mean
wTAN=zeros(5,4);
rTAN=zeros(5,1);

for i=1:5
wTAN(i,:)=((MV)\MM(i,:)')/sum((MV)\MM(i,:)');
rTAN(i,:)=wTAN(i,:)*MM(i,:)';
end
```

```
% VARIATION IN VARIANCE OF ASSETS
Var1= [0.14 0.15 0.16 0.17];
stdev1=sqrt(Var1);
for j=1:4
    for i=1:4
           MV1(i,j)=stdev1(1,i)'*corr*stdev1(1,j);
           if i==j
               MV1(i,j)=Var1(i)
           end
    end
end

Var2= [0.13 0.16 0.16 0.17];
stdev2=sqrt(Var2);
for j=1:4
    for i=1:4
           MV2(i,j)=stdev2(1,i)'*corr*stdev2(1,j);
           if i==j
               MV2(i,j)=Var2(i)
           end
    end
end

Var3= [0.13 0.15 0.17 0.17];
stdev3=sqrt(Var3);
for j=1:4
    for i=1:4
           MV3(i,j)=stdev3(1,i)'*corr*stdev3(1,j);
           if i==j
               MV3(i,j)=Var3(i)
           end
    end
end

Var4= [0.13 0.15 0.16 0.18];
stdev4=sqrt(Var4);
for j=1:4
    for i=1:4
           MV4(i,j)=stdev4(1,i)'*corr*stdev4(1,j);
           if i==j
               MV4(i,j)=Var4(i)
           end
    end
end
```

```
% Maximum trade-off portfolios with different variance: weights, mean
wTAN1=((MV1)\MM(1,:)')/sum((MV1)\MM(1,:)');
rTAN1=sum(MM(1,:)*wTAN1(1,:)');

wTAN2=((MV2)\MM(1,:)')/sum((MV2)\MM(1,:)');
rTAN2=sum(MM(1,:)*wTAN2);

wTAN3=((MV3)\MM(1,:)')/sum((MV3)\MM(1,:)');
rTAN3=sum(MM(1,:)*wTAN3(1,:));

wTAN4=((MV4)\MM(1,:)')/sum((MV4)\MM(1,:)');
rTAN4=sum(MM(1,:)*wTAN4(1,:));

wTANvar=[wTAN(1,:); wTAN1'; wTAN2'; wTAN3'; wTAN4'];
rTANvar=[rTAN(1,:); rTAN1'; rTAN2'; rTAN3'; rTAN4'];
%Variation of Weights in percentage due to Mean Variation
for a= 2:5
    VariationWeMean(a-1,:)= ((wTAN(a,:)- wTAN(1,:))./wTAN(1,:));
end
TotVarWeMean= sum(abs(VariationWeMean'));
%Variation of Weights in percentage due to Variance Variation
for a= 2:5
    VariationWeCov(a-1,:)= (wTANvar(a,:)- wTANvar(1,:))./wTAN(1,:);
end
TotVarWeCov= sum(abs(VariationWeCov'));

figure
bar(1:4,[TotVarWeMean' TotVarWeCov']);
title(' Figure 4: Sensibility of Portfolio Weights to Variation in Inputs Estimation');
legend('Variation of Mean','Variation of Variance');
```

C.2 GlobalSearch with static approach

```
%The aim of this code is to produce an optimization using the
%GLOBALSEARCH
clear
clc

[data,text]=xlsread('S&P 100.xlsx','S&P 100 Weekly Mono $');
Dnum=text(4:end,1);
DailyPrices=data(2:end,:);
R=((DailyPrices(2:(size(DailyPrices,1)),:)./DailyPrices(1:(size(
```

```
    DailyPrices,1)-1),:))-1)*100;
R(isnan(R))=0;
Y=R(:,end);
R=R(:,1:end-1);
Stocks=text(1,2:end-1);
%GLOBALSEARCH IMPLEMENTATION
%initialization of GLOBALSEARCH
q=(0.1:0.1:1);
k=(0.25:0.25:2.5);
w0=ones(size(R,2)-1,1)/(size(R,2)-1);
%pre-allocation
wfminTEV=zeros(size(R,2),length(k));
wpenTEV_tot=zeros(size(R,2),length(k),length(q));
%iterate GLOBALSEARCH for all values of q and k
for i=1:length(q)
for j=1:length(k)
    problem = createOptimProblem('fmincon','objective',@(w)
        fminsearchTEVGS_q(Y,R,w,k(j),q(i)),'x0',w0,'options',...
    optimoptions(@fmincon,'Algorithm','sqp','Display','off'));

    problem.lb= -ones(size(R,2)-1,1);
    problem.ub= zeros(size(R,2)-1,1);

   out= fmincon(problem);
   gs = GlobalSearch('Display','iter','NumTrialPoints',600);
   rng(14,'twister') % for reproducibility
   out = run(gs,problem);
   out=-out;
   wReg=[out; 1-sum(out)];
   wpenTEV_q(:,j)=wReg;
end
wpenTEV_tot(:,:,i)=wpenTEV_q;
end
% At this point, starting from the whole array constructed, we want
     to restrict the working sample in order to consider only the
    portfolios with an acceptable level of diversification. At that
    those portfolios have not to contain too many assets in order to
     reduce transaction costs and illiquid positions. The assets
    weights different from zero for each k and q are collected in a
    matrix representing the number of assets in which we are
    investing
for i=1:length(q)
    Nassets=sum(wpenTEV_tot(:,:,i)~=0);
    NAssets(i,:)=Nassets;
```

```
end
%In order to select the assets we consider only the portfolio
   containing only a given range of assets. A logical operator 1 is
    used in order to indicate the accettable portfolio.
%pre-allocation
 Filter= NAssets>10 & NAssets<50;

 Filter2=zeros(size(wpenTEV_tot,1),size(wpenTEV_tot,2));
 Filter3=zeros(size(wpenTEV_tot,1),size(wpenTEV_tot,2), size(
    wpenTEV_tot,3));

 for j=1:size(Filter,1)
     for i=1:size(Filter,2)
         Filter1=Filter(j,i)*ones(size(wpenTEV_tot,1),1);
         Filter2(:,i)=Filter1;
     end
 Filter3(:,:,j)=Filter2; %create an Array with the same size of the
     solution of fminserach
 end
 %Exclude the portfolios not contained in the selected range
 w_fminRestricted=wpenTEV_tot.*Filter3;

 P=R-Y*ones(1,size(R,2));

 %calculation of the IR for each portfolio
 %pre-allocation
 IR2=zeros(1,length(k));
 IR3=zeros(length(q),length(k));
%iteration for all portfolios where P*w_fminRestricted=Portfolio
   Return
 for i=1:length(q)
     for j=1:length(k)
         TE=mean(P*(w_fminRestricted(:,j,i)));
         TEV=sqrt(var(P*(w_fminRestricted(:,j,i))));
         IR=TE/TEV;
         IR2(:,j)=IR;
     end
     IR3(i,:)=IR2;
 end
 IR3=abs(IR3);
 %At this point, starting from the matrix with the IR for all
    selected portfolios, we pick the one with the maximum IR saving
     its coordinates
[min_IR_posit,posmin_IRposit]=min(IR3(:));
```

```
[rowmin_IR_posit,colmin_IR_posit]=ind2sub(size(IR3),posmin_IRposit)

w_optmin_posit=w_fminRestricted(:,colmin_IR_posit,rowmin_IR_posit);
```

C.3 GlobalSearch with dynamic approach: index variation.

```
%load Data
%The aim of this code is to produce an optimization using the
    GlobalSearch
clear
clc

[data,text]=xlsread('S&P 100.xlsx','AllCost Weekly Prices $');
[Filter,text2]=xlsread('S&P 100.xlsx','Filter');
Prices=data(2:end,1:end);
R=((Prices(2:(size(Prices,1)),:)./Prices(1:(size(Prices,1)-1),:))
    -1)*100;
WeeklyDates=text(3:end-1,1);
DMonth=datenum(text2(3:end,1),'dd/mm/yyyy');
R(isnan(R))=0;

Y=R(:,end);
R=R(:,1:end-1);

% window size for estimation/allocation
w=60;    % 5 years
Ww=260;
DWeek=datenum(WeeklyDates(1:end,1),'dd/mm/yyyy');
DWeek=datestr(DWeek,'mm/yy');
Dweek1=datenum(DWeek);
DMonth=datestr(DMonth,'mm/yy');
DMonth1=datenum(DMonth);

for j=w+1:length(DMonth1)
    k = find(DMonth1(j,1)==Dweek1,1);
    k1(j,1)=k;
end
k1=k1(w+1:end-1,1);

for s=w:(size(Filter,1)-1)
PresenceIndicator=Filter(s,:);
PresenceIndicatorTot(s,:)=PresenceIndicator;
```

```
end
PresenceIndicatorTot=PresenceIndicatorTot(w+1:end,:);

y=zeros((size(Filter,1)-w),1);
r=zeros((size(Filter,1)-w),size(Filter,2));
for n=1:size(PresenceIndicatorTot,1)                    % r-1 as we use data up to r-1 to allocate 1-step-ahead
r=R(k1(n)-Ww-1:k1(n)-1,:).*PresenceIndicatorTot(n,:);
rTot(:,:,n)=r;
y=Y(k1(n)-Ww-1:k1(n)-1,:);
yTot(:,:,n)=y;
end
%GLOBALSEARCH IMPLEMENTATION
%initialization of GlobalSearch for each slice of the array
q=(0:0.1:1);
k=(0.25:0.25:2.5);
wopt_UNC=zeros(size(rTot,2),(size(k1,1)));

%Calibration of optimal q and k for the first allocation
 s=1
    R=rTot(:,:,s);
    R(:,all(~(R),1))=NaN;
    R( :, all( isnan( R ), 1 ) ) = [];
    Y=yTot(:,:,s);
    Presence=PresenceIndicatorTot(s,:);

    wfminTEV=zeros(size(R,2),length(k));
    wpenTEV_tot=zeros(size(R,2),length(k),length(q));
    IRtot=zeros(length(q),length(k));
    w0=ones(size(R,2)-1,1)/(size(R,2)-1);

   for i=1:length(q)
   for j=1:length(k)
  problem = createOptimProblem('fmincon','objective',@(w) fminsearchTEVGS_q(Y,R,w,k(j),q(i)),'x0',w0,'options',...
    optimoptions(@fmincon,'Algorithm','sqp','Display','off'));

    problem.lb= -ones(size(R,2)-1,1);
    problem.ub= zeros(size(R,2)-1,1);

   out= fmincon(problem);
   gs = GlobalSearch('Display','iter','NumTrialPoints',600);
   rng(14,'twister') % for reproducibility
   out = run(gs,problem);
```

```
    out=-out;
    wpenTEV_q=[out; 1-sum(out)];
    wfminTEV(:,j)=wpenTEV_q;
    end
wpenTEV_tot(:,:,i)=wfminTEV;
    end
% At this point, starting from the whole array constructed, we want to restrict the working sample in order to consider only the portfolios with an acceptable level of diversification.At that those portfolios have not to contain too many assets in order to reduce transaction costs and illiquid positions The assets weights different from zero for each k and q are collected in a matrix representing the number of assets in which we are investing
    for i=1:length(q)
     Nassets=sum(wpenTEV_tot(:,:,i)~=0);
     NAssets(i,:)=Nassets;
    end
%In order to select the assets we consider only the portfolio containing only a given range of assets. A logical operator 1 is used in order to indicate the acceptable portfolio.
%pre-allocation
 Filter= NAssets>8 & NAssets<60;

 Filter2=zeros(size(wpenTEV_tot,1),size(wpenTEV_tot,2));
 Filter3=zeros(size(wpenTEV_tot,1),size(wpenTEV_tot,2), size(wpenTEV_tot,3));

     for j=1:size(Filter,1)
         for i=1:size(Filter,2)
          Filter1=Filter(j,i)*ones(size(wpenTEV_tot,1),1);
          Filter2(:,i)=Filter1;
         end
 Filter3(:,:,j)=Filter2; %create an Array with the same size of the solution of fminserach
     end

 %Exclude the portfolios not contained in the selected range
 w_fminRestricted=wpenTEV_tot.*Filter3;

 P=R(:,:,1)-Y(:,:,1)*ones(1,size(R,2));

 %calculation of the IR for each portfolio
 %pre-allocation
```

```
 IR2=zeros(1,length(k));
 IR3=zeros(length(q),length(k));

%iteration for all portfolios where P*w_fminRestricted=Portfolio Return
    for i=1:length(q)
        for j=1:length(k)
         TE=mean(P*(w_fminRestricted(:,j,i)));
         TEV=var(P*(w_fminRestricted(:,j,i)));
         IR=TE/TEV;
         IR2(:,j)=IR;
        end
     IR3(i,:)=IR2;
    end
    IR4=abs(IR3);
 %At this point, starting from the matrix with the IR for all selected portfolios, we pick the one with the maximum IR saving its coordinates
[min_IR,pos_IR]=min(IR4(:));
[row_IR,col_IR]=ind2sub(size(IR4),pos_IR);

h=1;
Wopt=zeros(size(Presence,2),1);
Ropt=zeros(size(R,1),size(Presence,2));
h2=1;
for n=1:size(Presence,2)
    if Presence(1,n)==1
       Ropt(:,h2)=R(:,h);
       Wopt(h2,1)=w_fminRestricted(h,col_IR,row_IR);
       h=h+1;
    end
    h2=h2+1;
end
wopt_UNC(:,s)=Wopt;
ropt(:,:,s)=Ropt;

clearvars Wopt Ropt
%We want now to modify the port weights when the index composition changes. In order to do so we consider the variation of index composition with respect to the previous period. So if the composition is unchanged we keep the portfolio weights unaltered from previous allocation. Otherwise we restart all the process repeating the calibration for q and k.
for s=2:size(rTot,3)
```

```
     Wopt=zeros(size(Presence,2),1);
     Ropt=zeros(size(R,1),size(Presence,2));

     R=rTot(:,:,s);
     R(:,all(~(R),1))=NaN;
     R( :, all( isnan( R ), 1 ) ) = [];
     Y=yTot(:,:,s);
     Presence=PresenceIndicatorTot(s,:);
     q=(0:0.1:1);
     k=(0.25:0.25:2.5);

    if PresenceIndicatorTot(s-1,:)==PresenceIndicatorTot(s,:)
        q=q(1,row_IR);
        k=k(1,col_IR);
        wopt_UNC(:,s)=wopt_UNC(:,s-1);
        ropt(:,:,s)=ropt(:,:,s-1);
    else
         q=(0:0.1:1);
         k=(0.25:0.25:2.5);
         wfminTEV=zeros(size(R,2),length(k));
         wpenTEV_tot=zeros(size(R,2),length(k),length(q));
         w0=ones(size(R,2)-1,1)/(size(R,2)-1);
           for i=1:length(q)
           for j=1:length(k)
               problem = createOptimProblem('fmincon','objective',@(w) fminsearchTEVGS_q(Y,R,w,k(j),q(i)),'x0',w0,'options',...
               optimoptions(@fmincon,'Algorithm','sqp','Display','off'));

               problem.lb= -ones(size(R,2)-1,1);
               problem.ub= zeros(size(R,2)-1,1);

               out= fmincon(problem);
               gs = GlobalSearch('Display','iter','NumTrialPoints',600);
               rng(14,'twister') % for reproducibility
               out = run(gs,problem);
               out=-out;
               wpenTEV_q=[out; 1-sum(out)];
               wfminTEV(:,j)=wpenTEV_q;
           end
               wpenTEV_tot(:,:,i)=wfminTEV;
           end
```

```

for i=1:length(q)
Nassets=sum(wpenTEV_tot(:,:,i)~=0);
NAssets(i,:)=Nassets;
end

%In order to select the assets we consider only the
   portfolio containing only a given range of assets. A
   logical operator 1 is used in order to indicate the
   accettable portfolio.
%pre-allocation
Filter= NAssets>8 & NAssets<60;

Filter2=zeros(size(wpenTEV_tot,1),size(wpenTEV_tot,2));
Filter3=zeros(size(wpenTEV_tot,1),size(wpenTEV_tot,2),
   size(wpenTEV_tot,3));

for j=1:size(Filter,1)
    for i=1:size(Filter,2)
        Filter1=Filter(j,i)*ones(size(wpenTEV_tot,1),1);
        Filter2(:,i)=Filter1;
    end
    Filter3(:,:,j)=Filter2; %create an Array with the
       same size of the solution of fminserach
end

%Exclude the portfolios not contained in the selected
   range
w_fminRestricted=wpenTEV_tot.*Filter3;

P=R(:,:,1)-Y(:,:,1)*ones(1,size(R,2));

%calculation of the IR for each portfolio
%pre-allocation
IR2=zeros(1,length(k));
IR3=zeros(length(q),length(k));
%iteration for all portfolios where P*w_fminRestricted=
   Portfolio Return
for i=1:length(q)
    for j=1:length(k)
    TE=mean(P*(w_fminRestricted(:,j,i)));
    TEV=sqrt(var(P*(w_fminRestricted(:,j,i))));
    IR=TE/TEV;
    IR2(:,j)=IR;
```

```
                end
                IR3(i,:)=IR2;
            end
            IR4=abs(IR3);
            %At this point, starting from the matrix with the IR for all selected
            %portfolios, we pick the one with the maximum IR saving its coordinates
            [min_IR,pos_IR]=min(IR4(:));
            [row_IR,col_IR]=ind2sub(size(IR4),pos_IR);

            h=1;
            Wopt=zeros(size(Presence,2),1);
            Ropt=zeros(size(R,1),size(Presence,2));
            h2=1;
            for n=1:size(Presence,2)
                if Presence(1,n)==1
                   Ropt(:,h2)=R(:,h);
                   Wopt(h2,1)=w_fminRestricted(h,col_IR,row_IR);
                   h=h+1;
                end
                h2=h2+1;
            end
           wopt_UNC(:,s)=Wopt;
           ropt(:,:,s)=Ropt;
    end

    clearvars h h2 Wopt wpenTEV_q wpenTEV_tot wfminTEV
    ss=s+1
end
```

C.4 GlobalSearch with dynamic approach: monthly reallocation.

The beginning of the code is the same to the previous one until line 156 . From that point onward the code change, since this time the reallocation is performed each month.

```
     %Impose the condition that index composition at time t is equal to the one at t-1
    if PresenceIndicatorTot(s-1,:)==PresenceIndicatorTot(s,:)

        q=q(1,row_IR_posit);
        k=(0.25:0.25:2.5);
```

```
        w0=ones(size(R,2)-1,1)/(size(R,2)-1);
        wfminTEV=zeros(size(R,2),length(k));
        wpenTEV_tot=zeros(size(R,2),length(k),length(q));
           for i=1:length(q)
           for j=1:length(k)
               problem = createOptimProblem('fmincon','objective',@(
                   w) fminsearchTEVGS_q(Y,R,w,k(j),q(i)),'x0',w0,'
                   options',...
               optimoptions(@fmincon,'Algorithm','sqp','Display','
                   off'));

               problem.lb= -ones(size(R,2)-1,1);
               problem.ub= zeros(size(R,2)-1,1);

               out= fmincon(problem);
               gs = GlobalSearch('Display','iter','NumTrialPoints'
                   ,600);
               rng(14,'twister') % for reproducibility
               out = run(gs,problem);
               out=-out;
               wpenTEV_q=[out; 1-sum(out)];
               wfminTEV(:,j)=wpenTEV_q;
           end
               wpenTEV_tot(:,:,i)=wfminTEV;
           end

           %The assets weights different from zero for each k and q
               are collected in a
           %matrix representing the number of assets in which we are
                investing
           for i=1:length(q)
           Nassets=sum(wpenTEV_tot(:,:,i)~=0);
           NAssets(i,:)=Nassets;
           end

           %In order to select the assets we consider only the
               portfolio containing only a given range of assets. A
               logical operator 1 is used in order to indicate the
               acceptable portfolio.
           %pre-allocation
           Filter= NAssets>10 & NAssets<60;

           Filter2=zeros(size(wpenTEV_tot,1),size(wpenTEV_tot,2));
```

```
        Filter3=zeros(size(wpenTEV_tot,1),size(wpenTEV_tot,2),
            size(wpenTEV_tot,3));

        for j=1:size(Filter,1)
            for i=1:size(Filter,2)
                Filter1=Filter(j,i)*ones(size(wpenTEV_tot,1),1);
                Filter2(:,i)=Filter1;
            end
            Filter3(:,:,j)=Filter2; %create an Array with the
                same size of the solution of fminserach
        end

        %Exclude the portfolios not contained in the selected
            range
        w_fminRestricted=wpenTEV_tot.*Filter3;

        P=R(:,:,1)-Y(:,:,1)*ones(1,size(R,2));

        %calculation of the IR for each portfolio
        %pre-allocation
        IR2=zeros(1,length(k));
        IR3=zeros(length(q),length(k));
        %iteration for all portfolios where P*w_fminRestricted=
            Portfolio Return
        for i=1:length(q)
            for j=1:length(k)
            TE=mean(P*(w_fminRestricted(:,j,i)));
            TEV=sqrt(var(P*(w_fminRestricted(:,j,i))));
            IR=TE/TEV;
            IR2(:,j)=IR;
            end
            IR3(i,:)=IR2;
        end
        IR4=abs(IR3);
        %At this point, starting from the matrix with the IR for
            all selected
        %portfolios, we pick the one with the maximum IR saving
            its coordinates
        [min_IR_posit,pos_IR_posit]=min(IR4(:));
        [row_IR_posit,col_IR_posit]=ind2sub(size(IR4),
            pos_IR_posit);

    h=1;
    h2=1;
```

```
        for n=1:size(Presence,2)
         if Presence(1,n)==1
         Ropt(:,h2)=R(:,h);
         Wopt(h2,1)=w_fminRestricted(h,col_IR_posit,row_IR_posit);
         h=h+1;
         end
         h2=h2+1;
        end

         wopt_posit(:,s)=Wopt;
         ropt_posit(:,:,s)=Ropt;
    else
         %Recalibration of q and k with selection of optimal weights
         q=(0:0.1:1);
         k=(0.25:0.25:2.5);
         w0=ones(size(R,2)-1,1)/(size(R,2)-1);
         wfminTEV=zeros(size(R,2),length(k));
         wpenTEV_tot=zeros(size(R,2),length(k),length(q));
            for i=1:length(q)
            for j=1:length(k)
                problem = createOptimProblem('fmincon','objective',@(
                   w) fminsearchTEVGS_q(Y,R,w,k(j),q(i)),'x0',w0,'
                   options',...
                optimoptions(@fmincon,'Algorithm','sqp','Display','
                   off'));

                problem.lb= -ones(size(R,2)-1,1);
                problem.ub= zeros(size(R,2)-1,1);

                out= fmincon(problem);
                gs = GlobalSearch('Display','iter','NumTrialPoints'
                   ,600);
                rng(14,'twister') % for reproducibility
                out = run(gs,problem);
                out=-out;
                wpenTEV_q=[out; 1-sum(out)];
                wfminTEV(:,j)=wpenTEV_q;
            end
                wpenTEV_tot(:,:,i)=wfminTEV;
            end

            for i=1:length(q)
            Nassets=sum(wpenTEV_tot(:,:,i)~=0);
            NAssets(i,:)=Nassets;
```

```
        end

        Filter= NAssets>10 & NAssets<60;
    Filter2=zeros(size(wpenTEV_tot,1),size(wpenTEV_tot,2));
        Filter3=zeros(size(wpenTEV_tot,1),size(wpenTEV_tot,2),size(wpenTEV_tot,3));

        for j=1:size(Filter,1)
            for i=1:size(Filter,2)
                Filter1=Filter(j,i)*ones(size(wpenTEV_tot,1),1);
                Filter2(:,i)=Filter1;
            end
            Filter3(:,:,j)=Filter2; %create an Array with the same size of the solution of fminserach
        end

        %Exclude the portfolios not contained in the selected range
        w_fminRestricted=wpenTEV_tot.*Filter3;

        P=R(:,:,1)-Y(:,:,1)*ones(1,size(R,2));

        %calculation of the IR for each portfolio
        %pre-allocation
        IR2=zeros(1,length(k));
        IR3=zeros(length(q),length(k));
        %iteration for all portfolios where P*w_fminRestricted=Portfolio Return
        for i=1:length(q)
            for j=1:length(k)
            TE=mean(P*(w_fminRestricted(:,j,i)));
            TEV=sqrt(var(P*(w_fminRestricted(:,j,i))));
            IR=TE/TEV;
            IR2(:,j)=IR;
            end
            IR3(i,:)=IR2;
        end
        IR4=abs(IR3);
        %At this point, starting from the matrix with the IR for all selected portfolios, we pick the one with the maximum IR saving its coordinates
        [min_IR_posit,pos_IR_posit]=min(IR4(:));
        [row_IR_posit,col_IR_posit]=ind2sub(size(IR4),pos_IR_posit);
```

```
            h=1;
            Wopt=zeros(size(Presence,2),1);
            Ropt=zeros(size(R,1),size(Presence,2));
            h2=1;
            for n=1:size(Presence,2)
                if Presence(1,n)==1
                    Ropt(:,h2)=R(:,h);
                    Wopt(h2,1)=w_fminRestricted(h,col_IR_posit,
                        row_IR_posit);
                    h=h+1;
                end
                h2=h2+1;
            end
           wopt_posit(:,s)=Wopt;
           ropt_posit(:,:,s)=Ropt;
    end

    clearvars h h2 Wopt wpenTEV_q wpenTEV_tot wfminTEV NAssets IR1
        IR2 IR3 IR4
    ss=s+1
end

wopt_posit=wopt_posit';
woptP_posit=wopt_posit.*(wopt_posit>0);
woptN_posit=wopt_posit.*(wopt_posit<0);
```

Bibliography

Awoye, O. A. (2016). *Markowitz minimum variance portfolio optimization using new machine learning methods* (Unpublished doctoral dissertation). (UCL) University College London.

Beasley, J. E., & Meade. (2003). An evolutionary heuristic for the index tracking problem. *European Journal of Operational Research*, *148*(3), 621–643.

Behr, P., Guettler, A., & Truebenbach, F. (2012). Using industry momentum to improve portfolio performance. *Journal of Banking & Finance*, *36*(5), 1414–1423.

Benidis, K., Feng, Y., Palomar, D. P., et al. (2018). Optimization methods for financial index tracking: From theory to practice. *Foundations and Trends® in Optimization*, *3*(3), 171–279.

Black, F., & Litterman, R. (1992). Global portfolio optimization. *Financial analysts journal*, *48*(5), 28–43.

Bonaccolto, G., Caporin, M., & Paterlini, S. (2018). Asset allocation strategies based on penalized quantile regression. *Computational Management Science*, *15*(1), 1–32.

Boyd, S., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press.

Broadie, M. (1993). Computing efficient frontiers using estimated parameters. *Annals of Operations Research*, *45*(1), 21–58.

Brodie, J., Daubechies, I., De Mol, C., Giannone, D., & Loris, I. (2009). Sparse and stable markowitz portfolios. *Proceedings of the National Academy of Sciences*, *106*(30), 12267–12272.

Bruce, P., & Bruce, A. (2017). *Practical statistics for data scientists: 50 essential concepts*. " O'Reilly Media, Inc.".

Bruder, B., Gaussel, N., Richard, J.-C., & Roncalli, T. (2013). Regularization of portfolio allocation.

Chan, L. K., Karceski, J., & Lakonishok, J. (1999). On portfolio optimization: Forecasting covariances and choosing the risk model. *The review of Financial studies*, *12*(5), 937–974.

Chavez-Bedoya, L., & Birge, J. R. (2014). Index tracking and enhanced indexation using a parametric approach. *Journal of Economics Finance and Administrative Science*, *19*(36), 19–44.

DeMiguel, V., & Garlappi. (2009). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science*, *55*(5), 798–812.

DeMiguel, V., Martin-Utrera, A., & Nogales, F. J. (2013). Size matters: Optimal calibration of shrinkage estimators for portfolio selection. *Journal of Banking & Finance*, *37*(8), 3018–3034.

DeMiguel, V., & Nogales, F. J. (2009). Portfolio selection with robust estimation. *Operations Research*, *57*(3), 560–577.

Dickinson, J. P. (1974). The reliability of estimation procedures in portfolio analysis. *Journal of Financial and Quantitative Analysis*, *9*(3), 447–462.

Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., et al. (2004). Least angle regression. *The Annals of statistics*, *32*(2), 407–499.

Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, *96*(456), 1348–1360.

Fan, J., Zhang, J., & Yu, K. (2012). Vast portfolio selection with gross-exposure constraints. *Journal of the American Statistical Association*, *107*(498), 592–606.

Fastrich, B., Paterlini, S., & Winker, P. (2014). Cardinality versus q-norm constraints for index tracking. *Quantitative Finance*, *14*(11), 2019–2032.

Fastrich, B., Paterlini, S., & Winker, P. (2015). Constructing optimal sparse portfolios using regularization methods. *Computational Management Science*, *12*(3), 417–434.

Fernandes, M., Rocha, G., & Souza, T. (2012). *Regularized minimum-variance portfolios using asset group information.*

Fernholz, R., Garvy, R., & Hannon, J. (1998). Diversity-weighted indexing. *The Journal of Portfolio Management*, *24*(2), 74–82.

Frank, L. E., & Friedman, J. H. (1993). A statistical view of some chemometrics regression tools. *Technometrics*, *35*(2), 109–135.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1) (No. 10). Springer series in statistics New York, NY, USA:.

Frino, A., & Gallagher, D. (2001). Tracking s&p 500 index funds. *Journal of portfolio Management*, *28*(1).

Frost, P. A., & Savarino, J. E. (1988). For better performance: Constrain portfolio weights. *The Journal of Portfolio Management*, *15*(1), 29–34.

Gasso, G., Rakotomamonjy, A., & Canu, S. (2009). Recovering sparse signals with a certain family of nonconvex penalties and dc programming. *IEEE Transactions on Signal Processing*, *57*(12), 4686–4698.

Gunes, F. (2015). Penalized regression methods for linear models in sas/-stat®. In *Proceedings of the sas global forum 2015 conference. cary, nc: Sas institute inc. http://support. sas. com/rnd/app/stat/papers/2015/penalizedregression_linearmodels. pdf.*

Hastie, T., Tibshirani, R., & Wainwright, M. (2015). *Statistical learning with sparsity: the lasso and generalizations*. CRC press.

Hoerl, A., & Kennard, R. (1988). *Ridge regression, in 'encyclopedia of statistical sciences', vol. 8.* Wiley, New York.

Ingersoll, J. E. (1987). *Theory of financial decision making* (Vol. 3). Rowman & Littlefield.

Jagannathan, R., & Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, *58*(4), 1651–1683.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer.

Jobson, J. D., & Korkie, B. (1980). Estimation for markowitz efficient portfolios. *Journal of the American Statistical Association*, *75*(371), 544–554.

John Lu, Z. (2010). The elements of statistical learning: data mining, inference, and prediction. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *173*(3), 693–694.

Jorion, P. (1985). International portfolio diversification with estimation risk. *Journal of Business*, 259–278.

Jorion, P. (1992). Portfolio optimization in practice. *Financial Analysts Journal*, *48*(1), 68–74.

Kassambara, A. (2018). *Machine learning essentials: Practical guide in r*. STHDA.

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.

Markowitz, H. (1952). Portfolio selection. *The journal of finance*, *7*(1), 77–91.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of financial economics*, *8*(4), 323–361.

Michaud, R. O. (1989). The markowitz optimization enigma: Is 'optimized'optimal? *Financial Analysts Journal*, *45*(1), 31–42.

Roncalli, T. (2013). *Introduction to risk parity and budgeting*. CRC Press.

Scherer, B. (2002). *I portfolio construction and risk budgeting*. London: Risk Books.

Schmidt, M. (2005). Least squares optimization with l1-norm regularization. *CS542B*

Project Report, 14–18.

Scott-Forman Roe. (n.d.). *Understanding the Bias-Variance trade-off.* http://scott.fortmann-roe.com/docs/BiasVariance.html, note = Online accessed 07 November 2018 , year=2012,.

Sheikh, A. Z., & Qiao, H. (2009). Non-normality of market returns https://am.jpmorgan.com/blobcontent/1383169198442/83456/11_438.pdf. *JP Morgan Asset Management research paper*.

Strub, O., & Baumann, P. (2018). Optimal construction and rebalancing of index-tracking portfolios. *European journal of operational research*, *264*(1), 370–387.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288.

Trevor, H., Robert, T., & JH, F. (2009). *The elements of statistical learning: data mining, inference, and prediction.* New York, NY: Springer.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(2), 301–320.