6 WORKING PAPER SERIES
QUADERNI DEL DIPARTIMENTO DI
ECONOMIA AZIENDALE ROMA TRE
DEPARTMENT OF BUSINESS STUDIES
ROMA TRE - "BS RM3"

ALESSANDRA CONGEDO,
ALESSIO DI PAOLO,
CARLO DOMENICO MOTTURA,
JACOPO MARIA RICCI

A RISK-GAIN-SPARSITY OPTIMIZATION APPROACH

RomaTrE-Press
2024

ROMA TRE
UNIVERSITÀ DEGLI STUDI

Dipartimento di Economia Aziendale

QUADERNI
DEL DIPARTIMENTO DI ECONOMIA AZIENDALE ROMA TRE
WORKING PAPER SERIES
DEPARTMENT OF BUSINESS STUDIES ROMA TRE – "BS RM3"

Collana diretta da Ornella Ricci

1. Fractional Volterra integral equations: a neural network approach
   M. Cenci, M.A. Congedo, A.L. Martire, B. Rogo
2. Portfolio Selection using behavioral models
   M.A. Congedo, M. Corradini, A. Gheno, J. Riccioni, C. Sansoni
3. Sensemaking and job crafting for the inclusion of refugees in Universities through EQPR
   L. Marchegiani, F. Ceschel, C. Finocchietti, P. Limata, S. Spitalieri
4. Corporate Culture and Merger Success
   F. Fiordelisi, O. Ricci, F.S. Stentella Lopes
5. Stability of Entropic Risk Measures
   J.R. Ricci
6. A Risk-Gain-Sparsity Optimization Approach
   M.A. Congedo, A. Di Paolo, C.D. Mottura, J.R. Ricci

6 WORKING PAPER SERIES
QUADERNI DEL DIPARTIMENTO DI
ECONOMIA AZIENDALE ROMA TRE
DEPARTMENT OF BUSINESS STUDIES
ROMA TRE – "BS RM3"

ALESSANDRA CONGEDO, ALESSIO DI PAOLO,
CARLO DOMENICO MOTTURA, JACOPO MARIA RICCI

A RISK-GAIN-SPARSITY OPTIMIZATION APPROACH

RomaTrE-Press
2024

QUADERNI
DEL DIPARTIMENTO DI ECONOMIA AZIENDALE ROMA TRE
WORKING PAPER SERIES
DEPARTMENT OF BUSINESS STUDIES ROMA TRE – "BS RM3"
COLLANA DIRETTA DA ORNELLA RICCI

Cura editoriale e impaginazione
teseo editore Roma teseoeditore.it

Elaborazione grafica della copertina
MOSQUITO. mosquitoroma.it

Caratteri grafici utilizzati: Minion Concept Roman; Minion Pro Regular (copertina e frontespizio). Garamond, Cambria Math (testo).

Edizioni *Roma TrE-Press* ©
Roma, giugno 2024
ISBN 979-12-5977-332-6
http://romatrepress.uniroma3.it

L'attività della *Roma TrE-Press* è svolta nell'ambito della
Fondazione Roma Tre-Education, piazza della Repubblica 10, 00185 Roma.

EDITORIAL POLICY E SCOPO DELLA COLLANA

Il principale obiettivo della Collana consiste nel diffondere i risultati della ricerca scientifica del Dipartimento di Economia Aziendale, consentendone tempestiva e ampia diffusione sia in ambito accademico che nel mondo esterno all'Università. In linea con la tradizione interdisciplinare del Dipartimento, la Collana si propone di coprire un'ampia area di studi, accogliendo contributi di Accounting, Management e Finance che adottino i metodi tipici delle scienze economico-aziendali, matematiche e giuridiche.
Coerentemente con il suddetto obiettivo, la Collana è destinata ad accogliere contributi scientifici di elevata qualità, in cui almeno un autore sia appartenente al Dipartimento di Economia Aziendale dell'Università degli Studi Roma Tre. Possono essere proposti contributi sia di natura teorica che empirica, scritti sia in lingua italiana che in lingua inglese.

La Collana si avvale del lavoro di tre diversi organi:
– il Comitato Scientifico ha il compito di indirizzare la politica editoriale e selezionare i membri del Comitato Editoriale;
– il Direttore propone al Comitato Scientifico ogni iniziativa utile a promuovere la Collana e coordina le attività del Comitato Editoriale;
– il Comitato Editoriale ha la funzione di supervisionare il processo di referaggio per la valutazione di merito degli articoli proposti.

I lavori pubblicati nella Collana sono sottoposti a referaggio "doppio cieco" affidato a due revisori esterni, secondo il processo di seguito specificato, improntato a criteri di qualità, efficienza e tempestività.
Gli autori che intendono proporre un contributo lo inviano alla Segreteria della Ricerca di Dipartimento che, d'intesa con il Direttore, lo assegna ad un membro del Comitato Editoriale, individuato in base alla coerenza tra competenze scientifiche e tema affrontato nel lavoro sottomesso. Il membro del Comitato Editoriale responsabile del prodotto procede all'individuazione di due revisori esterni tra esperti del mondo accademico e professionale. Qualora il contributo sia proposto da un membro del Comitato Editoriale, il compito di selezionare i revisori esterni è assolto dal Direttore.
I revisori esterni valutano il contributo sulla base dei criteri di rilevanza, originalità e rigore metodologico. Al termine del processo di revisione, il Direttore esprime al Comitato scientifico il proprio parere sul rilascio dell'autorizzazione alla pubblicazione. Il Comitato scientifico è l'organo competente al rilascio dell'autorizzazione alla pubblicazione del prodotto editoriale.

Il Direttore comunica la decisione agli Autori e conserva l'archivio dei referaggi.

I lavori pubblicati nella Collana sono liberamente accessibili in formato elettronico sul sito dell'editore Roma TrE-Press. La versione a stampa è acquistabile in modalità 'Print on demand'.

Le pubblicazioni hanno una numerazione progressiva e eventuali richiami o citazioni ad essi devono riportare la denominazione estesa del contributo a cui si fa riferimento.

A Risk-Gain-Sparsity Optimization Approach

Maria Alessandra Congedo, Alessio Di Paolo, Carlo Domenico Mottura, Jacopo Maria Ricci

Index

A Risk-Gain-Sparsity Optimization Approach

Maria Alessandra Congedo, Alessio Di Paolo, Carlo Domenico Mottura,
Jacopo Maria Ricci *

Abstract

One of the fundamental principles of portfolio selection models is risk minimization through investment diversification. However, the benefits of diversification are reduced when there is a high correlation between assets. It is well-known that diversifying through the use of larger portfolios is not the best way to achieve an improvement in out-of-sample performance. Moreover, including a large number of positions in the portfolio increases management and transaction costs. While classical portfolio selection models focus on risk minimization and return maximization, the purpose of this work is to include a third objective: the l_1-Norm. This allows for the selection of sparse portfolios, that is, with a limited number of assets, which are easier to manage and allow for good risk-return results.

Our empirical analysis is based on a publicly available dataset often used in the literature.

Keywords: Portfolio Selection; Diversification; Out-of-Sample; Sparsity.

---

*Roma Tre University - Department of Business Studies Roma Tre, "BS RM3", Via Silvio D'Amico, 77, 00145 Roma, Italy.

Corresponding author: alessandra.congedo@uniroma3.it, alessio.dipaolo@uniroma3.it, carlodomenico.mottura@uniroma3.it, jacopomaria.ricci@uniroma3.it

1. Introduction and literature review

The main problem in asset allocation is to select a portfolio with appropriate features in terms of gain and risk. More in detail, the aim is to build a portfolio that maximizes a measure of profit and minimizes a measure of risk. In 1952, Harry Markowitz laid the foundation of Modern Portfolio Theory introducing the mean-variance framework: by solving a quadratic optimization problem, the investor can find the optimal portfolio allocation that minimizes the portfolio expected risk, for a given level of expected return. A major limitation of the mean-variance approach is that the optimized weights are very sensitive to estimation error and the presence of multicollinearity in the inputs. In particular, it is well known that estimating expected returns is more challenging than just focusing on risk minimization and then looking for portfolios with minimum risk, i.e. the so-called global minimum variance portfolios (Jagannathan and Ma, 2003). In addition, one method of achieving positive returns while keeping risk levels stable is portfolio diversification (Markowitz, 1952; Sharpe, 1964). It is based on the principle that several assets may perform differently under various market conditions. However, promoting diversification also means investing in a large number of assets, which implies, in practical terms, high transaction and portfolio management costs. In general, the development of successful asset allocation strategies requires the construction of portfolios that perform well out-of-sample, provide diversification benefits, and are cheap to maintain and monitor. In this perspective, an ideal portfolio has conservative asset weights, which are stable in time while still promoting the right amount of diversification and being able to control the total amount of shorting. Following this rationale, a natural approach is to extend the Markowitz optimization framework by using a penalty function on the weight vector, typically the norm, whose intensity is controlled by a tuning parameter (λ). In this context, one of the most recent successful approach using convex penalty function, that can control the total amount of shorting, while avoiding to invest in the entire asset universe is the Least Absolute Shrinkage and Selection Operator (LASSO) (Tibshirani, 1996). In the field of portfolio selection, the LASSO framework consists on adding to the Markowitz formulation a penalty proportional to the l_1-Norm on the asset weight vector. DeMiguel et al. (2009a) provide a general framework for determining portfolios with superior out-of-sample performance in the presence of estimation error, based on solving the traditional minimum-variance problem subject to the additional constraint that the norm of the portfolio-weight vector be smaller than a given threshold. Brodie et al. (2009) show that LASSO results in constraining the gross exposures and it can be used to account for transaction costs. In this perspective Brodie et al. (2009), DeMiguel et al. (2009b) also consider a portfolio with an l_2-Norm penalty on the weight vector, known in statistical literature as RIDGE. Although, the RIDGE penalty stabilizes the mean-variance optimization, as it controls for multicollinearity, the shape of the penalty does not promote sparsity, leading to portfolios with an undesirably large number of active positions. Despite its appealing properties, the LASSO is ineffective in presence of no short-selling and budget constraints (i.e., $x_i \geq 0$, $\sum_{i=1}^{n} x_i = 1$), as the l_1-Norm is equal to 1. To overcome these issues, non-convex penalties, like the l_q-Norm, the log and the SCAD penalty have gained increased attention in the portfolio literature (Chen et al., 2013; Fastrich et al., 2014; Xing et al., 2014). In this work we propose a novel portfolio selection approach which aims to maximize, as performance measure, the (weighted) geometric mean of the differences between its gain, risk and sparsity and those of a suitable benchmark. In order to consider sparse portfolios, we rely on the l_1-Norm, and, to make it functional, we allow short sales, thus focusing on long-short portfolios. In addition to stabilizing the optimization problem and generalizing no-short-positions–constrained optimization, the l_1 penalty allows the limitation of transaction costs. For large investors, whose principal cost is a fixed bid–ask spread, transaction costs are effectively proportional to the gross market value of the selected portfolio, i.e., to the l_1 penalty term (Brodie et al., 2009). Even for small investors, "overhead" costs (volume-independent) cannot be ignored, so sparse solutions are preferred (sparse portfolios or sparse changes to portfolios). In other words, the goal of this work is to build a profitable and stable portfolio that, by taking advantage of the l_1-Norm, invests in a limited number of assets.

2. Real world problems optimization

Optimization is a branch of applied mathematics that refers to the minimization (or maximization) of a given objective function of one or more decision variables that satisfy functional constraints. A typical optimization model addresses the allocation of scarce resources among possible alternative uses in order to maximize an objective function such as total profit. Decision variables, the objective function, and constraints are three essential elements of any optimization problem. Problem that lack constraints are called *unconstrained optimization* problems, while others are often referred to as *constrained optimization*

problems. Problems with no objective function are called *feasibility* problems. In practice, it is rare that a unique objective is able to summarise the problem that must be faced; generally multi-objective optimization is an efficient technique to use in order to find a set of solutions to several objectives. These problems are often addressed by reducing them to a single-objective optimization problem or a sequence of such problems. If the decision variables in an optimization problem are restricted to integers, or to a discrete set of possibilities, we have an *integer* or *discrete optimization* problem. If there are no such restrictions on the variables, the problem is a *continuous optimization* problem. Of course, some problems may have a mixture of discrete and continuous variables.
A typical example in finance is the portfolio selection problem that aims to find the optimal assets' combination in order to satisfy two different goals, the risk minimization and the return maximization. As we shall see in later sections of this paper, by changing the risk measurement, due to the properties of the analyzed functions, one can convert the portfolio selection problems, for example from mean-variance quadratic programming to a linear programming, which is simpler and easier for being solved.

2.1. Single-objective problems

For the sake of completeness, even if the focus of this work is on multi-objective problems, we can formally define a single-objective problem as:

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{s.t.} \quad & x \in C \end{aligned} \tag{1}$$

where f is the objective function, C is the system of equalities or inequalities constraints that identifies the so-called "feasible set", while $x = (x_1, x_2, \ldots, x_n)$ is the vector of decision variables. If C is empty, the problem is infeasible and it does not admit any solution. If it is possible to find a sequence $x^k \in C$ such that $f(x^k) \to -\infty$ as $k \to +\infty$, then the problem is unbounded. If the problem is neither infeasible nor unbounded, then it is often possible to find a solution $x^* \in C$ that satisfies

$$f(x^*) \leq f(x), \quad \forall x \in C.$$

In most cases, the feasible set C is described explicitly using functional constraints (equalities and inequalities). For example, C may be given as

$$C := \{x : g_i(x) = 0, i \in \mathcal{I} \text{ and } g_i(x) \geq 0, i \in \mathcal{E}\}$$

where $\mathcal{I}$ and $\mathcal{E}$ are the index sets for equality and inequality constraints. Many factors affect the potential efficiency of solving optimization problems. For example, the number n of decision variables and the total number of constraints; other factors are related to the properties of the functions f and g_i that define the problem. One of the most common and easy to solve optimization problems is the linear programming (LP), that consists in optimizing a linear objective function subject to linear equality and inequality constraints. A more general optimization problem is the quadratic programming (QP) problem, where the objective function is a quadratic function of the variables subject to linear equality and inequality constraints. Problems with a linear objective function and linear constraints are easier, as are problems with quadratic objective functions and convex feasible sets. When at least one of the functions f or g_i (with $i \in \mathcal{I}, i \in \mathcal{E}$) is not linear, we are in the case of non-linear programming (NLP). As extensively explained in Cesarone (2020), the main differences between an LP and a NLP problems are:

- in NLP problems, one may find points of local minimum or maximum that can not be global. In LP problems, if a point is a local minimum or maximum, it is also global

- the feasible region of a NLP problem can be any shape, not necessarily a polyhedron as in LP, and the optimal solution can be anywhere, not necessarily on a vertex as in LP

- the algorithms used to solve NLP problems are generally less efficient than those for LP problems. Furthermore, the solutions are not always global, but they could be any stationary points, more precisely Karush-Kuhn-Tucker (KKT) points.

2.2. Multi-objective problems

Multi-objective optimization addresses problems involving multiple conflicting objectives. Its aim is to find the optimal solution while considering multiple objective functions that must be simultaneously optimized. A multi-objective optimization problem can be formulated as follows

$$\begin{aligned} \text{minimize} \quad & (f_1(x), f_2(x), \ldots, f_k(x)) \\ \text{s.t.} \quad & x \in C \end{aligned} \tag{2}$$

where $k \geq 2$ objective functions need to be minimized simultaneously and the feasible set C is defined as in the single objective problem. If there is no conflict between the objective functions, then a solution can be found where every objective function attains its optimum. We assume that there does not exist a single solution that is optimal with respect to every objective function. This means that the objective functions are at least partly conflicting. Anyway, some of the objective vectors can be extracted for examination. Such vectors are those where none of the components can be improved without deterioration to at least one of the other components.

Definition 1 (Pareto optimality) *A decision vector $x^* \in C$ is Pareto optimal if there does not exist another decision vector such that $f_i(x) \leq f_i(x^*)$ for all $i = 1, \ldots, k$ and $f_j(x) < f_j(x^*)$ for at least one index j.*

There are usually a lot (infinite number) of Pareto optimal solutions that define the Pareto optimal set, even called efficient frontier. It is possible to define the efficient frontier (or Pareto optimal set) where all the feasible solutions that cannot be improved w.r.t. both objectives (Pareto optimal points) lie on. The efficient frontier can be a curve in case of a bi-objective problem, a surface with three objective functions and an iper-surface when more. According to the previous definition, it is important to consider the following

Definition 2 (Pareto Dominance) *A vector $\hat{x}$ is said to dominate x^* ($\hat{x} \prec x^*$) if:*

- $f_i(\hat{x}) \leq f_i(x^*) \; \forall i \in [1, \ldots, k]$
- *There is at least one i such that $f_i(\hat{x}) < f_i(x^*)$*

When dealing with multi-objective optimization problems, convexity is a requirement that functions and feasible sets may satisfy or not. We know that

Definition 3 (Multi-objective problem convexity) *The multi-objective optimization problem is convex if all the objective functions and the feasible region are convex.*

Definition 4 (Convex function) *A function $f_i : C \rightarrow R$ is convex if for all $x_1, x_2 \in C$ is valid that $f_i(\beta x_1 + (1-\beta)x_2) \leq \beta f_i(x_1) + (1-\beta) f_i(x_2)$ for all $0 \leq \beta \leq 1$.*

Definition 5 (Convex set) *A subset C of a given vector space R is convex if* $x_1, x_2 \in C$ *and* $\beta \in [0, 1]$ *it is always implied that* $\beta x_1 + (1 - \beta)x_2 \in C$

In other words, a feasible set is convex if for any two given points in the set, the line segment connecting these two points lies entirely in the set. If the previous two conditions are satisfied we are facing a convex multi-objective optimization problem, that is an important characteristic for the optimality of solutions, as follows from the following

Theorem 6 (Optimal solutions) *Let the multi-objective optimization problem be convex. Then every locally Pareto optimal solution is also globally Pareto optimal.*

3. Portfolio selection models

Portfolio selection aims at computing the proportion of capital to allocate among available assets on a given market for the purpose of maximizing the future portfolio return. Resorting to optimization problems it is possible to obtain a portfolio with appropriate features in terms of gain and risk. The base for determining the solution of portfolio selection problems was laid out by Harry Markowitz with his paper "Portfolio Selection" published in 1952 in the Journal of Finance, signing the beginning of the Modern Portfolio Theory. It is still widely used by both academics and practitioners to support investment decisions. However, its success has inevitably drawn many criticism and proposals of alternative or more refined models (King, 1993; Konno and Yamazaki, 1991; Mitra et al., 2003; Rockafellar et al., 2000). The real innovation introduced by Markowitz is the concept of diversification related to the correlation existing between securities. Diversification between uncorrelated asset returns can reduce the portfolio risk, but could not eliminate it. That's why Markowitz talks about a trade-off between risk and gain: if we invest in assets not (perfectly) correlated each other, the sacrifice in terms of portfolio return will be compensated by the reduction of the overall risk. Let us an investment universe composed of n risky assets, with expected returns $\mathbb{E}(R_k) = \mu_k$ and covariances $\sigma_{k,j}$ with $k = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, n$. From a mathematical viewpoint, the mean-variance problem of Markowitz consists of a bi-objective optimization problem, solvable by satisfying two conflicting goals: minimizing variance and maximizing portfolio expected returns. We now indicate with $x = (x_1, x_2, \ldots, x_n)$ the vector of portfolio weights, that are the decision variables of the problem, for which the full investment and no short-selling constraints hold: $\sum_{k=1}^{n} x_k = 1$ and $x_k \geq 0$ respectively. In order to solve the mean-variance problem of Markowitz under these conditions, it has been proposed, using the ϵ-constraint method, to set a required level of portfolio expected return $\mu_p = \eta$. In this way, the Mean-Variance model becomes a convex quadratic programming problem numerically solvable with efficient algorithms:

$$\begin{aligned} \text{minimize} \quad & \sum_{k=1}^{n} \sum_{j=1}^{n} x_k x_j \sigma_{kj} \\ \text{s.t.} \quad & \sum_{k=1}^{n} \mu_k x_k = \eta \\ & \sum_{k=1}^{n} x_k = 1 \\ & x_k \geq 0 \qquad\qquad k = 1, \ldots, n \end{aligned} \tag{3}$$

In order to build the efficient frontier, the procedure has to be iterated by varying μ_p between the expected return of the minimum variance portfolio η_{min} and that of the maximum expected return portfolio η_{max}. In this way, the resulting optimal portfolios define the efficient frontier (those lying on the interval $[\eta_{min}, \eta_{max}]$).

3.1. From a bi-objective to a three-objective model

As already mentioned, after Markowitz's work, other portfolio selection approaches have been developed in order to overcome its main limitations and to reach other objectives. In particular, starting from the portfolio selection model proposed by Cesarone et al. (2019), in which the aim is to select an optimal portfolio that is able to maximize the weighted geometric mean of the distances between its risk and gain values and those of a given benchmark index, the aim of this paper is to extend this approach including a third objective: the l_1 norm. Adding the l_1 penalty term to the objective function has several useful consequences:

- it promotes sparsity, playing a key role in the task of formulating investment portfolios when investors want to be able to limit the number of positions they must create, monitor and liquidate;
- it stabilizes the problem, imposing a penalty on the size of the coefficients of the portfolio vector x, it is possible to reduce the sensitivity of the optimization to possible collinearity between assets;

- it incorporates a proxy for the transaction costs that, in a liquid market, can be modeled by a two-component structure: one that is a fixed "overhead", independent of the size of the transaction, and a second one, given by multiplying the transacted amount with the marketmaker's bid–ask spread applicable to the size of the transaction.

For large investors, the overhead portion can be neglected; in that context, the total transaction cost paid is just $\sum_{i=1}^{n} s_i|x_i|$, the sum of the products of the absolute trading volumes $|x_i|$ and bid-ask spread (s_i) is the same for all assets. For small investors, the overhead portion of the transaction costs is nonnegligible; for a very small investor, this portion may even be the only one worth considering. If the transaction costs are asset-independent, then the total cost is simply proportional to the number of assets selected (i.e., corresponding to nonzero weights). The key point is that holding a large number of assets and frequently changing the portfolio composition (high turnover) could erode profits. On the other hand, it is easy to understand how holding portfolios with a limited number of assets can greatly reduce commissions and trading costs. A sparse portfolio should also promote conservative asset weights (low Turnover), still maintaining diversification benefits. So, the goal is to implement the risk-gain-sparsity model in order to construct a small, profitable and stable portfolio.

3.1.1. A risk-gain dominance maximization approach

The Max-Area approach (Cesarone et al., 2019) tries to bring together the advantages of a typical risk-gain analysis with those of the enhanced index tracking, following a no-preference strategy. This work is based on the idea to outperform the reference index by maximizing the weighted geometric mean between its risk and gain and those of the reference index.
Let us first introduce a generic portfolio selection problem expressed as a bi-objective optimization problem:

$$\begin{aligned} \text{maximize} \quad & \gamma_p(x) \\ \text{minimize} \quad & \rho_p(x) \\ \text{s.t.} \quad & x \in C \end{aligned} \tag{4}$$

where $\gamma_p(x)$ is a continuous concave measure of gain and $\rho_p(x)$ is any continuous convex measure of risk. Let then C be the feasible region of the problem such that $C = [x \in R^n : x \geq l, e^T x = 1]$, considering $e \in R^n$ as a vector of ones and $l \in R^n$ as the vector of lower bounds. Accordingly, C is a nonempty, convex and compact feasible set where it is possible to find all the portfolios satisfying problem constraints. Let us now consider a reference index $(\rho_p^{ref}, \gamma_p^{ref}) \in R^2$ defined in the risk-gain plane. It can be selected, for instance, the Nadir point, for which ρ_p^{ref} is the risk of the portfolio with maximal gain and γ_p^{ref} is the gain of the minimum risk portfolio. Alternatively, any market index can be chosen as a reference point. For any feasible portfolio $x \in C$ such that $\gamma_p(x) \geq \gamma_p^{ref}$ and $\rho_p(x) \leq \rho_p^{ref}$, it can be reformulated that $(\rho_p^{ref} - \rho_p(x)) \geq 0$ and $(\gamma_p(x) - \gamma_p^{ref}) \geq 0$. The idea is to maximize the non-negative quantity that corresponds to the weighted geometric mean of the distances between the portfolio risk and gain and those of the reference index. In terms of risk-gain dominance, the model becomes:

$$
\begin{aligned}
\text{maximize} \quad & (\gamma_p(x) - \gamma_p^{ref})(\rho_p^{ref} - \rho_p(x)) \\
\text{s.t.} \quad & \gamma_p(x) \geq \gamma_p^{ref} \\
& \rho_p(x) \leq \rho_p^{ref} \\
& x \in C
\end{aligned}
\tag{5}
$$

By focusing on the constraints of problem (5), it is worth noticing how the choice of the reference index affects the feasible region of the overall problem. Indeed, ρ_p^{ref} and γ_p^{ref} are the worst values in the risk-gain plane that a feasible portfolio can assume. In other words, the objective function in (5) is the area of a rectangle R_x with height $\gamma_p(x) - \gamma_p^{ref} \geq 0$ and base $\rho_p^{ref} - \rho_p(x) \geq 0$. Therefore, the non-negative objective function will be expressed as $A(x) = (\gamma_p(x) - \gamma_p^{ref})(\rho_p^{ref} - \rho_p(x))$. The maximization of the area of the rectangle R_x, as shown in Figure 1, results in a portfolio x_A that dominates the majority of portfolios dominating the benchmark. As a consequence, x_A is the portfolio dominating the most the reference index $(\rho_P^{ref}, \gamma_p^{ref})$. Moving from the optimal portfolio x_A to another feasible portfolio x, can be observed an enhancement in terms of one of two objectives while a worsening on the other. Specifically, the improvement obtained in terms of an objective is always smaller than the worsening achieved in terms of the other objective. In other terms, it can be demonstrated that if a feasible portfolio $x \in C$ records an improvement in gain such that $\gamma_p(x) - \gamma_p^{ref} = \alpha(\gamma_p(x_A) - \gamma_p^{ref})$ with $\alpha \geq 1$, as a consequence the worsening in terms of risk is exactly $\alpha(\rho_p^{ref} - \rho_p(x)) \leq \rho_p^{ref} - \rho_p(x_A)$. Alternatively, let us suppose that the feasible portfolio x is better in terms of risk in a way that $\rho_p^{ref} - \rho_p(x) = \beta(\rho_p^{ref} - \rho_p(x_A))$ with $\beta \geq 1$, then the worsening in gain will be $\beta(\gamma_p(x) - \gamma_p^{ref}) \leq \gamma_p(x_A) - \gamma_p^{ref}$. Last, by recalling

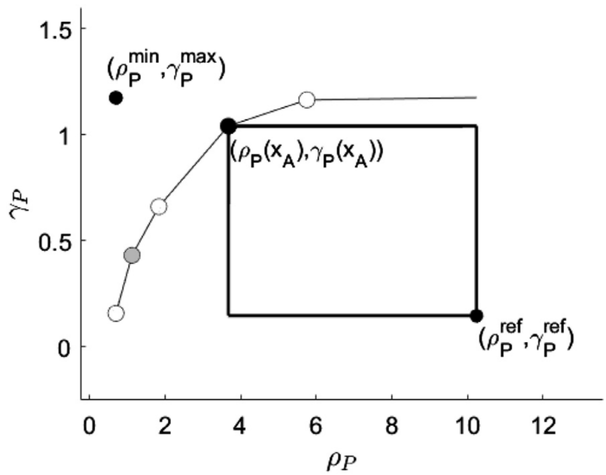

Figure 1: The portfolio entailed by the Max-area approach, using the nadir as reference, is marked in black, while that obtained by the Sharpe ratio maximization in grey. The Markowitz portfolios for 3 different target return levels are marked in white.

that $A(x_A) \geq A(x)$ for every $x \in C$ it is valid that:

$$\alpha = \frac{\gamma_p(x) - \gamma_p^{ref}}{\gamma_p(x_A) - \gamma_p^{ref}} \leq \frac{\rho_p^{ref} - \rho_p(x_A)}{\rho_p^{ref} - \rho_p(x)} = \frac{1}{\beta}.$$

Another distinctive feature of the model is its flexibility: by changing the measures of risk and gain also the selected portfolios will change accordingly.

3.1.2. A risk-gain-sparsity optimization approach

From the birth of Modern Portfolio Theory, Markowitz observed that given the high correlations in the stock market, increasing the size of a portfolio may reduce the benefits of diversification. In fact, when dealing with highly correlated assets, only a limited reduction in portfolio riskiness can be obtained by increasing the number of assets included (Markowitz, 1959). On the other hand, benefits of diversification are more prominent when increasing the amount of securities that are uncorrelated each other. As already mentioned, holding a portfolio with a large number of assets may be expensive from the perspective of monitoring and transaction costs and, overall, these costs may result greater than the benefits of diversification. On the contrary, holding portfolios that include a limited number of assets reduces the respective management expenses and commissions. These relationships can be seen as a trade off between the resulting reduction of risk due to the diversification and the increase of transaction costs (reduction of return) and vice-versa. Another important advantage of small portfolios seems to be that of reducing the estimation errors for variances and covariances thus leading to better out-of-sample performance (Cesarone et al., 2014). At the light of the above considerations, ideal portfolios are those characterized by sparsity, i.e. the capital is invested in a few assets. A sparse portfolio should also promote a low Turnover still maintaining diversification benefits. To achieve this goal we can rely on the portfolio norm. A *Norm* on a real vector space X is a function $\|.\| : X \to R$ that associates to a n-dimension vector a positive length. Considering a vector space X, the norm $\|x\|$ is such that the following conditions are satisfied:

1. $\|x\| \geq 0 \quad \forall x \in X$
2. $\|x\| = 0$ if and only if $x = 0$
3. for any scalar quantity λ it is true that $\|\lambda x\| = \|\lambda\| \|x\|$
4. $\forall\, x, y \in X \; \|x + y\| = \|x\| + \|y\|$

The most common norms are know under the name of *p*-norms or l_p norms family. Considering a vector *x*, the *p*-norm $\|x\|_p$ can be defined as:

$$\|x\|_p = (x_1^p + x_2^p + x_3^p + ... + x_n^p)^{\frac{1}{p}}.$$

It can be also re-written in simplified terms as:

$$\|x\|_p = (\sum_{i=1}^{n} x_i^p)^{\frac{1}{p}} \tag{6}$$

with $p \in [1, +\infty]$.
When $p = 0$ the third of the above introduced axioms does not stand anymore: we are dealing with a pseudo-norm l_0. It is not a real norm, but it can be useful since it counts the number of non-zero elements in a vector and therefore be functional when dealing with a vector of portfolio weights.
The 1-Norm, also called "Manhattan Distance", simply measures the sum of the absolute values of the vector components. When $p = 1$, (6) becomes the sum of

the magnitudes of the vector in the space.

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|$$

The most common is the Euclidean Norm or Norm-2 since it implies that $p = 2$. It computes the shortest distance between two points resulting in a non-negative value.

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$$

Finally, the last to be mentioned is the infinity norm that measures the maximum absolute value in the given vector.

$$\|x\|_\infty = \max |x_i|$$

Following this reasoning, a third objective can be introduced to select the portfolio with a limited number of assets. To make it possible, the l_0 pseudo-norm as additional objective to be minimized is considered. The model thus becomes:

$$\begin{array}{ll} \text{maximize} & \gamma_p(x) \\ \text{minimize} & \rho_p(x) \\ \text{minimize} & \|x\|_0 \\ \text{s.t.} & x \in C \end{array}$$

However, the l_0 pseudo-norm has a characteristic that makes the optimization problem computationally hard to be solved: it is non convex. One way to address a non-convex optimization is to resolve a similar convex optimization problem. It is geometrically demonstrated that l_1 is the best convex approximation to l_0 and that, under some specific conditions, it provides the same solution as the original problem with l_0. Therefore a theoretical model that is able to promote a sparse portfolio can be expressed this way:

$$\begin{array}{ll} \text{maximize} & \gamma_p(x) \\ \text{minimize} & \rho_p(x) \\ \text{minimize} & \|x\|_1 \\ \text{s.t.} & x \in C \end{array}$$

In order to apply the Max-area approach, all the components must be continuously differentiable, but the l_1 is not. In fact, the absolute value function $f(x) = |x|$ is not differentiable at the origin, where a corner point can be identified. The module can be imagined as as the sum of two components in order to overcome the non-differentiability of the function. Let us split the function in two elements:

- the first x_{1+} described by the function

$$\max[0, x_1] = \begin{cases} x_1 & \text{if} \quad x_1 \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

- the second x_{1-} described by the function

$$\max[0, x_1] = \begin{cases} 0 & \text{if} \quad x_1 \geq 0 \\ -x_1 & \text{otherwise} \end{cases}$$

Therefore, we can write:

$$|x_1| = \max[0, x_1] + \max[0, -x_1]$$

with $x_1 = \max[0, x_1] - \max[0, -x_1]$ and $\max[0, x_1] \geq 0$, $\max[0, -x_1] \geq 0$. At the light of the above considerations, the three objectives optimization problem

$$\begin{aligned} \text{minimize} \quad & \sum_{k=1}^{n} \sum_{j=1}^{n} x_k x_j \sigma_{kj} \\ \text{maximize} \quad & \sum_{k=1}^{n} \mu_k x_k \\ \text{minimize} \quad & \sum_{k=1}^{n} |x_k| \\ \text{s.t.} \quad & \sum_{k=1}^{n} x_k = 1 \\ & x_k \geq l_k \qquad\qquad k = 1, \ldots, n \end{aligned} \tag{7}$$

can be reshaped as the equivalent version:

$$
\begin{aligned}
\text{minimize} \quad & \sum_{k=1}^{n}\sum_{j=1}^{n}(u_k - v_k)(u_j - v_j)\sigma_{kj} \\
\text{maximize} \quad & \sum_{k=1}^{n}\mu_k(u_k - v_k) \\
\text{minimize} \quad & \sum_{k=1}^{n}(u_k + v_k) \\
\text{s.t.} \quad & \sum_{k=1}^{n}(u_k - v_k) = 1 \\
& u_k - v_k \geq l_k \\
& u_k \geq 0 \\
& v_k \geq 0 \qquad\qquad k = 1, \dots, n
\end{aligned} \tag{8}
$$

Thus, the original not differentiable model has been turned into a continuously differentiable one, so that the Max-Area approach is now applicable.

$$
\begin{aligned}
\text{maximize} \quad & (\sum_{k=1}^{n}\mu_k(u_k - v_k) - \gamma_p^{ref})(\rho_p^{ref} - \sum_{k=1}^{n}\sum_{j=1}^{n}(u_k - v_k)(u_j - v_j)\sigma_{kj}) \\
& (\delta_P^{ref} - \sum_{k=1}^{n}(u_k + v_k)) \\
\text{s.t.} \quad & \sum_{k=1}^{n}\mu_k(u_k - v_k) \geq \gamma_p^{ref} \\
& \sum_{k=1}^{n}\sum_{j=1}^{n}(u_k - v_k)(u_j - v_j)\sigma_{kj} \leq \rho_p^{ref} \\
& \sum_{k=1}^{n}(u_k + v_k) \leq \delta_P^{ref} \\
& \sum_{k=1}^{n}(u_k - v_k) = 1 \\
& u_k - v_k \geq l_k \\
& u_k \geq 0 \\
& v_k \geq 0 \qquad\qquad k = 1, \dots, n
\end{aligned} \tag{9}
$$

For the sake of completeness, in the risk-gain-sparsity space any feasible portfolio no longer identifies a rectangle (as in the risk-gain space), but an hypercube: what it is going to be maximized is now a volume. The optimization of a nonlinear non convex single-objective problem is then implied: implementing the novel model in MATLAB, the objective of the paper results in finding the portfolio that mostly dominates the selected "reference index".

3.2. The elements of the model

As in the Max-Area approach, the model is implemented considering standard deviation and expected returns as measures of risk and profit, respectively. Furthermore the novel approach is enhanced by a measure of sparsity: the l_1-norm. The aim is to find the optimal values of expected return, volatility and l_1-norm as far as possible from the reference values. The choice of the reference index has played a fundamental role in the entire empirical analysis of this work. Before getting into the details of how the reference point has been computationally achieved, it is necessary to point out that the following analysis, to make the 1-norm functional, does not consider long-only portfolios, but it also involves portfolios with short positions (i.e. long-short portfolios). In practice, it has been possible by setting a negative vector of lower bounds l to the optimization problems solved: $x_k \geq l_k$. For practical implementation, it has been necessary to derive the reference point $(\rho_P^{ref}, \gamma_P^{ref}, \delta_P^{ref})$ in the three-dimensional objective space.

Risk-return reference

Following a similar approach of Cesarone et al. (2019), it has been considered γ_p^{ref} as the gain of the minimum risk portfolio and ρ_p^{ref} as the risk of the portfolio obtained by going short on all the assets $x_i = l$, $\forall i = 1, \dots, n-1$ and long for an amount equal to $x_n = |l|(n-1) + 1$ on the asset with the highest expected return.

For clarity, let us consider an example with 2 assets A and B, lower bound equal to $l = -0.1$ and B with highest expected return. The reference risk in this case is the portfolio obtained by investing $x_A = -0.1$ and $x_B = 1.1$. The same rationale can be extended to the case of n assets.

l_1-Norm reference

The previous point is combined with the worst possible value for the l_1-norm δ_P^{ref}, achieved through a closed-form solution. Let us first remind that the l_1-norm is simply given by the sum of the absolute values of the *n* vector elements as:

$$\|x\|_1 = \sum_{i=1}^{n} |x_i| = |x_1| + |x_2| + ... + |x_n| \tag{10}$$

The development of a closed-form able to maximize the l_1-Norm of a vector starts from the basic case of two assets. As a starting point, the concept of level line has to be introduced: the level line of a function of two variables is defined as the set of points in the plane whose coordinates satisfy the equation $f(x, y) = c$, where c is a real constant. Let us consider this function as the l_1-Norm of a vector with two elements:

$$f(x, y) = \|x\|_1 = |x_1| + |x_2| = c$$

For the purpose of this work, it can be supposed a portfolio selection problem with two assets whose weights must respect the budget constraint. By imposing a negative lower bound to their values, it can be obtained a greater than one level of l_1-Norm, which is the minimum it can get at the light of the required budget constraint. Furthermore, setting a negative lower bound l leverages the advantages of implying short sales: in the trivial case of two assets if one assumes a negative value $x_1 = l$ with $l < 0$, the other has to take a greater than one positive value to respect the budget constraint

$$x_2 = 1 + |l|$$

since $x_1 + x_2 = 1$. In terms of l_1-Norm, this implies that the sum of the absolute values is maximal when one of the two assets is bounded to its minimal feasible value, that in this case is $l < 0$.

This reasoning can be extended to the case of n assets: leaving at the lower bound level $n - 1$ assets (such that $x_i = l, \forall i = 1, \dots, n - 1$), x_n will get a remarkable positive value so as to maximize the l_1-Norm. It is thus possible to develop a closed-form starting from the budget constraint. Since the $n - 1$ assets are bounded to l, we can write:

$$x_n = 1 + (n - 1)|l|$$

since $l_1 + l_2 + \cdots + l_{n-1} + x_n = 1$.

Following the 1-norm formulation:

$$\|x\|_1 = |l| + |l| + ... + |l| + 1 + (n - 1)|\, l\,|$$

In other terms, the maximum 1-norm is:

$$\|x\|_1 = 1 + 2(n - 1)|\, l\,|.$$

4. Empirical Analysis

In this Section, we provide an extensive empirical analysis based on a real world dataset, where we compare three different strategies:

- Global Minimum Variance (**GMV**) portfolio, obtained by including short-selling to the classical minimum variance optimization model

$$
\begin{aligned}
\text{minimize} \quad & \sum_{k=1}^{n}\sum_{j=1}^{n} x_k x_j \sigma_{kj} \\
\text{s.t.} \quad & \sum_{k=1}^{n} x_k = 1 \\
& x_k \geq l \qquad\qquad k = 1, \dots, n
\end{aligned}
\tag{11}
$$

- Max-Area (**Max-A**) portfolio, applying the model developed in Cesarone et al. (2019) including short-selling

$$
\begin{aligned}
\text{maximize} \quad & (\sum_{k=1}^{n} \mu_k x_k - \gamma_p^{ref})(\rho_p^{ref} - \sqrt{\sum_{k=1}^{n}\sum_{j=1}^{n} x_k x_j \sigma_{k,j}}) \\
\text{s.t.} \quad & \sum_{k=1}^{n} \mu_k x_k \geq \gamma_p^{ref} \\
& \sqrt{\sum_{k=1}^{n}\sum_{j=1}^{n} x_k x_j \sigma_{k,j}} \leq \rho_p^{ref} \\
& \sum_{k=1}^{n} x_k = 1 \\
& x_k \geq l \qquad\qquad k = 1, \dots, n
\end{aligned}
\tag{12}
$$

- Max-Volume (**Max-V**) portfolio, based on the risk-gain-sparsity model proposed

$$
\begin{aligned}
\text{maximize} \quad & \Big(\sum_{k=1}^{n}\mu_k(u_k - v_k) - \gamma_p^{ref}\Big)\Big(\rho_p^{ref} - \sqrt{\sum_{k=1}^{n}\sum_{j=1}^{n}(u_k - v_k)(u_j - v_j)\sigma_{kj}}\Big) \\
& \Big(\delta_P^{ref} - \sum_{k=1}^{n}(u_k + v_k)\Big) \\
\text{s.t.} \quad & \sum_{k=1}^{n}\mu_k(u_k - v_k) \geq \gamma_p^{ref} \\
& \sqrt{\sum_{k=1}^{n}\sum_{j=1}^{n}(u_k - v_k)(u_j - v_j)\sigma_{kj}} \leq \rho_p^{ref} \\
& \sum_{k=1}^{n}(u_k + v_k) \leq \delta_P^{ref} \\
& \sum_{k=1}^{n}(u_k - v_k) = 1 \\
& u_k - v_k \geq l_k \\
& u_k, v_k \geq 0 \qquad k = 1, \ldots, n
\end{aligned}
\tag{13}
$$

4.1. Datasets and experimental setup

The experiments have been conducted on a real-world dataset downloaded from Refinitiv Eikon. The data consists of daily linear returns, from October 2006 to December 2020 for a total of $T = 3582$ observations, of assets belonging to the DowJones, that is a price-weighted stock index representing the performance of the top 30 stocks on the New York Stock Exchange (NYSE).

For the out-of-sample performance analysis we adopt a Rolling Time Window (RTW) scheme of evaluation, namely we allow for the possibility of rebalancing the portfolio composition during the holding period at fixed intervals. Here, we choose one financial month ($H = 20$ days) both as a rebalancing interval and as a holding period and two years ($L = 500$ days) as in-sample window. For each portfolio strategy analyzed, this procedure generates $T - L$ out-of-sample portfolio returns, on which several performance measures are computed. In addition, to robustly test the performance of models, the experiments are conducted for several lower bound levels $l = -10\%, -50\%, -100\%$, allowing short-selling in different proportions. All the procedures have been implemented in MATLAB R2022B and have been executed on a PC with an Intel(R) Xeon(R)

CPU E5-2623 v4 @ 2.6 GHz processor and 64,00 GB of RAM. Both the objective functions of problems (12) and (13) are maximized using `fmincon` with `interior-point` algorithm, `optimality-tolerance` of $1e-10$ and `max-iterations` of $1e4$.

4.2. Computational results

We examine here the computational results obtained by implementing the three models, for different levels of lower bounds (l) on the DowJones dataset. As we can observe in Tables 1, 2, 3, the rank of performance results is shown through different colours. More precisely, for each row, the colours range from deep green to deep red, where the former represents the best performance and the latter indicates the worst performance. For all lower bound levels, both in terms of sample mean and sample volatility, the Max-V model displays intermediate results between GMV and Max-A. Interestingly, the proposed model has the best Sharpe and Sortino ratios, thus showing the best gain per unit of risk. Furthermore, the Max-V model has, in all cases, the best Rachev ratio, demonstrating the best potential for extreme positive returns versus the risk of extreme losses (negative returns). The GMV model displays the lowest values of Ulcer index and Maximum Drawdown and it remains stable for all lower bound levels; in these terms, the Max-V approach results are always better than the Max-A, showing strongly increasing values as short-selling increases. The portfolio Turnover, that is a proxy of the amount of trading required to execute the strategies, reveals that the GMV model is the most stable. It is clear that, both the Max-A and the Max-V approaches show high Turnover values, with the difference that, in the second case, there are far fewer assets included in the portfolio, greatly reducing monitoring and transaction costs. The latter is the key point for which, in this work, we decided to extend the Max-Area approach (Cesarone et al., 2019): including the l_1-Norm it is possible to build sparse portfolios with few active positions while maintaining good results from a risk-return perspective.

Table 1: Out-of-sample performance results of DowJones with $l = -10\%$

$l = -10\%$	GMV	Max-A	Max-V
μ^{out}	0.00026	0.00103	0.00099
σ^{out}	0.00970	0.02695	0.02002
Sharpe	0.02720	0.03826	0.04942
Sortino	0.03823	0.05525	0.07334
Ulcer	0.07686	0.20427	0.12434
MaxDD	-0.33163	-0.47479	-0.35403
Rachev 5%	0.93200	1.00480	1.07459
Turnover	0.27173	1.59414	0.98741
Norm	1.96503	5.06765	2.01086
N assets	24	28	10
Long	15	7	4
Short	9	21	6

Table 2: Out-of-sample performance results of DowJones with $l = -50\%$

$l = -50\%$	GMV	Max-A	Max-V
μ^{out}	0.00026	0.00267	0.00216
σ^{out}	0.00974	0.09828	0.06524
Sharpe	0.02695	0.02712	0.03306
Sortino	0.03784	0.03855	0.04840
Ulcer	0.07680	0.98401	0.73558
MaxDD	-0.33679	-0.99978	-0.93200
Rachev 5%	0.93338	0.97914	1.04525
Turnover	0.28508	6.08814	3.61903
Norm	1.98882	21.31086	9.09925
N assets	24	28	12
Long	15	7	3
Short	9	21	9

Table 3: Out-of-sample performance results of DowJones with $l = -100\%$

$l = -100\%$	GMV	Max-A	Max-V
μ^{out}	0.00026	0.00471	0.00371
σ^{out}	0.00974	0.18914	0.12436
Sharpe	0.02695	0.02489	0.02984
Sortino	0.03783	0.03526	0.04350
Ulcer	0.07681	1.00299	0.99230
MaxDD	-0.33681	-1.53042	-1.00000
Rachev 5%	0.93338	0.97449	1.03716
Turnover	0.28510	11.71835	6.98820
Norm	1.98887	41.66965	18.04714
N assets	24	28	13
Long	15	7	3
Short	9	21	9

5. Conclusions

In this paper we extend and further improve the approach proposed in (Cesarone et al., 2019) through a tri-objective model. The effect of including the l_1-Norm is that of obtaining sparse portfolios, i.e. with a small number of assets, but, to make it functional it is necessary to consider long-short portfolios. Markowitz observed that a large-sized portfolio can reduce the benefits of diversification if high correlations exist in the stock market. In addition, it was soon realized that risk-return optimization can be very sensitive to changes in the inputs, due to errors in the estimates of the means and covariances of the assets returns, with the means being more crucial than the covariances (Best and Grauer, 1991; Chopra and Ziemba, 2013). Several methods are proposed in the literature to decrease the influence of estimation errors on portfolio selection. These include approaches developed in the area of statistics, such as shrinkage method (Ledoit and Wolf, 2003), Bayesian approaches (Craig MacKinlay and Pástor, 2000; Pástor and Stambaugh, 2000) and robust optimization procedures (Cornuejols and Tütüncü, 2006; Goldfarb and Iyengar, 2003). According to this evidences, in this work we investigate the possible benefits of portfolios composed of a limited number of assets, i.e. *"small portfolios"*. Our empirical results show that, including the l_1-Norm improves the out-of-sample portfolio performance in terms of risk-return balance and, selecting sparse portfolios, allows to reduce monitoring and transaction costs, that in most cases could erode the investors' profits. Furthermore, short-selling is an attractive but risky choice; however, it can be beneficial and profitable if positions are taken based on proper long-short allocation. These results are in line with the tendency described in DeMiguel et al. (2009a), where an improved out-of-sample performance is often observed for the norm-constrained minimum-variance portfolios. Possible future research could be oriented to investigate the validity of the proposed approach in larger markets and extend it to the enhanced index tracking field by changing the reference point.

References

Best, M. J. and Grauer, R. R. (1991). On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results. *The review of financial studies*, 4(2):315–342.

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

Cesarone, F. (2020). *Computational Finance: MATLAB® Oriented Modeling*. Routledge.

Cesarone, F., Lampariello, L., and Sagratella, S. (2019). A risk-gain dominance maximization approach to enhanced index tracking. *Finance Research Letters*, 29:231–238.

Cesarone, F., Moretti, J., and Tardella, F. (2014). Does greater diversification really improve performance in portfolio selection? *Available at SSRN 2473630*.

Chen, C., Li, X., Tolman, C., Wang, S., and Ye, Y. (2013). Sparse portfolio selection via quasi-norm regularization. *arXiv preprint arXiv:1312.6350*.

Chopra, V. K. and Ziemba, W. T. (2013). The effect of errors in means, variances, and covariances on optimal portfolio choice. In *Handbook of the fundamentals of financial decision making: Part I*, pages 365–373. World Scientific.

Cornuejols, G. and Tütüncü, R. (2006). *Optimization methods in finance*, volume 5. Cambridge University Press.

Craig MacKinlay, A. and Pástor, L. (2000). Asset pricing models: Implications for expected returns and portfolio selection. *The Review of financial studies*, 13(4):883–916.

DeMiguel, V., Garlappi, L., Nogales, F. J., and Uppal, R. (2009a). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management science*, 55(5):798–812.

DeMiguel, V., Garlappi, L., and Uppal, R. (2009b). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The review of Financial studies*, 22(5):1915–1953.

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

Goldfarb, D. and Iyengar, G. (2003). Robust portfolio selection problems. *Mathematics of operations research*, 28(1):1–38.

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

King, A. J. (1993). Asymmetric risk measures and tracking models for portfolio optimization under uncertainty. *Annals of Operations Research*, 45:165–177.

Konno, H. and Yamazaki, H. (1991). Mean-absolute deviation portfolio optimization model and its applications to tokyo stock market. *Management science*, 37(5):519–531.

Ledoit, O. and 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.

Markowitz, H. (1959). Portfolio selection: efficient diversification of investments. *Cowles Foundation for Research in Economics at Yale University*, Monograph 16.

Mitra, G., Kyriakis, T., Lucas, C., and Pirbhad, M. (2003). A review of portfolio planning: Models and systems. *Advances in portfolio construction and implementation*, pages 1–39.

Pástor, L. and Stambaugh, R. F. (2000). Comparing asset pricing models: an investment perspective. *Journal of Financial Economics*, 56(3):335–381.

Rockafellar, R. T., Uryasev, S., et al. (2000). Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3):425–442.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288.

Xing, X., Hu, J., and Yang, Y. (2014). Robust minimum variance portfolio with l-infinity constraints. *Journal of Banking & Finance*, 46:107–117.

One of the fundamental principles of portfolio selection models is risk minimization through investment diversification. However, the benefits of diversification are reduced when there is a high correlation between assets. It is well-known that diversifying through the use of larger portfolios is not the best way to achieve an improvement in out-of-sample performance. Moreover, including a large number of positions in the portfolio increases management and transaction costs. While classical portfolio selection models focus on risk minimization and return maximization, the purpose of this work is to include a third objective: the l1-Norm. This allows for the selection of sparse portfolios, that is, with a limited number of assets, which are easier to manage and allow for good risk-return results.

ALESSANDRA CONGEDO
è ricercatore universitario presso il Dipartimento di Economia Aziendale (Università degli Studi Roma Tre).

ALESSIO DI PAOLO
è dottorando presso il Dipartimento di Economia Aziendale (Università degli Studi Roma Tre).

CARLO DOMENICO MOTTURA
è professore ordinario presso il Dipartimento di Economia Aziendale (Università degli Studi Roma Tre).

JACOPO MARIA RICCI
è assegnista di ricerca presso il Dipartimento di Economia Aziendale (Università degli Studi Roma Tre).

ROMA TRE
UNIVERSITÀ DEGLI STUDI