Computers & Operations Research 142 (2022) 105701

Contents lists available at ScienceDirect

Computers and Operations Research

journal homepage: www.elsevier.com/locate/cor

A combinatorial optimization approach to scenario filtering in portfolio selection

Justo Puerto [a,b], Federica Ricca [c], Moisés Rodríguez-Madrena [a,b,*], Andrea Scozzari [d]

[a] Institute of Mathematics University of Seville (IMUS), Seville, Spain
[b] Department of Statistics and OR, Universidad de Sevilla, Seville, Spain
[c] MEMOTEF, Faculty of Economics, Università degli Studi di Roma La Sapienza, Italy
[d] Faculty of Economics, Università degli Studi Niccolò Cusano Roma, Italy

ARTICLE INFO

Keywords:
Mean–Variance optimization
Portfolio selection
Filtering methods
Mixed integer quadratic programming

ABSTRACT

Recent studies stressed the fact that covariance matrices computed from empirical financial time series appear to contain a high amount of noise. This makes the classical Markowitz Mean–Variance Optimization model unable to correctly evaluate the performance associated to selected portfolios. Since the Markowitz model is still one of the most used practitioner-oriented tool, several filtering methods have been proposed in the literature to overcome the problem. Among them, the two most promising ones refer to the Random Matrix Theory and to the Power Mapping strategy. The basic idea of these methods is to transform the estimated correlation matrix before applying the Mean–Variance Optimization model. However, experimental analysis shows that these two strategies are not always effective when applied to real financial datasets.

In this paper we propose a new filtering method based on Quadratic Programming. We develop a Mixed Integer Quadratic Programming model, which is able to filter those observations that may affect the performance of the selected portfolio. We discuss the properties of this new model and test it on some real financial datasets. We compare the out-of-sample performance of our portfolios with the one of the portfolios provided by the two above mentioned alternative filtering methods giving evidence that our method outperforms them. Although our model can be solved efficiently with standard optimization solvers, the computational burden increases for large datasets. To solve also these problems, we propose a heuristic procedure, which, on the basis of our empirical results, shows to be both efficient and effective.

1. Introduction and motivation

The Mean–Variance Optimization (MVO) approach, introduced by Markowitz in 1952 (Markowitz, 1952), has dominated the asset allocation process for more than 50 years. The Markowitz model minimizes the variance of the returns of a portfolio under the requirement of getting at least a fixed expected return level. In spite of its success, MVO received many criticisms, in particular, related to the fact that when returns are not Normally distributed the corresponding sampled correlation matrix is biased. Empirical studies have established that the distribution of speculative assets' returns tends to have *fatter tails* than the Gaussian distribution (see, e.g., Fama, 1965; Jansen and de Vries, 1991; Mandelbrot, 1963). Fat tails are typical of high *kurtosis* distributions in which extreme events, characterized by a small probability in theory, empirically occur more frequently than what the Normal distribution predicts. Including observations of the extreme events in the surveyed data produces noise in the estimation process that may affect the correlation matrix evaluation. After the recent financial crisis of 2008, many investors moved their attention to how to deal with the risk associated to these extreme events.

In a paper by Stoyanov et al. (see Stoyanov et al., 2011 and the references therein), the authors discuss and compare some popular methods for fat tails modeling based on full distribution modeling and extreme value theory. They conclude that the best approach should be to extend the Gaussian model incorporating methods for handling fat tails, and then testing their performance on real financial datasets. In any case, there is no evidence on which is the best among all the families of fat tailed models, as there may be different families of fat tailed distributions which are statistically equivalent.

An alternative analysis on the effects of noise in the estimated correlation matrix in the MVO model is provided by Schafer et al. (2010). The basic idea is that the bulk of eigenvalues of the covariance

* Corresponding author.
E-mail addresses: puerto@us.es (J. Puerto), federica.ricca@uniroma1.it (F. Ricca), madrena@us.es (M. Rodríguez-Madrena), andrea.scozzari@unicusano.it (A. Scozzari).

https://doi.org/10.1016/j.cor.2022.105701
Received 3 December 2021; Received in revised form 5 January 2022; Accepted 6 January 2022
Available online 5 February 2022

matrix that are small (or even zero) may produce portfolios of stocks that have nonzero returns but extremely low or even vanishing risk; such portfolios are related to estimation errors in real data. For this reason, in Schafer et al. (2010) the authors propose two filtering techniques in order to eliminate the problem of small eigenvalues in the sampled covariance matrix. According to the first technique, which relies on *Random Matrix Theory* (RMT) (Laloux et al., 2000), after the diagonalization of the correlation matrix, only the h highest eigenvalues are preserved, while the others are set to zero. Then, a new filtered covariance matrix is obtained by using the filtered spectrum and the original eigenvectors. The second method, called *Power Mapping*, transforms each element of the correlation matrix by raising its absolute value to some power q, while preserving the sign. The transformed matrix is used in the application of the MVO model. The authors observe that particular attention must be paid to the value of q that is modeled as a function of the ratio $\frac{n}{T}$, where n is the number of assets and T the number of observation time periods. The first method was also considered in a series of previous papers (Pafka and Kondor, 2002, 2003, 2004) where the authors apply a filter to the covariance matrix based on the RMT approach and introduce a measure defined as the ratio between the risk associated to an empirical portfolio and the risk associated with the *true* optimal portfolio. In the above mentioned papers, the model minimizes the return's variance without the constraint on the fixed expected return level. The true portfolio is formed for two special cases, namely when the true covariance matrix is equal to the unit matrix, and when it is randomly generated. Hence, in this study, the performance of noise filtering techniques is tested in a extremely controlled setting.

From a risk management viewpoint, natural quantities to estimate are the *realized* (or out-of-sample) return and risk of a portfolio. Actually, the novelty in Schafer et al. (2010) is that the authors compare the two proposed filtering techniques applied to the MVO model with and without short-sellings, both in a Monte-Carlo simulation framework and, for the first time, on a real dataset. However, as the authors write, in their simulation framework good realized return and risk are observed, while in the real setting and without short-selling the improvement provided by the two filtering methods is not so evident w.r.t. the non-filtered MVO model (see Schafer et al., 2010 pages 116–117 Section 5.2.2). Moreover, they observe experimentally that the power method approach is quite sensitive to the value of q, and thus to the ratio $\frac{n}{T}$, and they find that a good value for q should be approximatively equal to $\sqrt{\frac{n}{T}}$. The poor behavior of the RMT approach, in particular for the no short-selling case, is also confirmed in El Karoui (2013) where the author observes that, when short-selling is not allowed, the problem of biased sampled covariance matrices is mitigated, and naive estimates of the risk are already very close to the realized risk.

In the classical MVO framework it is crucial to have good estimates of the expected returns and covariances for all securities in the considered market. In this regard, a number of alternative approaches emerged in the literature in order to suppress noise in the data. These procedures include single- and multi-factor models, (for a review, see, e.g., Elton and Gruber (1995) and Elton et al. (2006)) and Bayesian estimators (see Jorion, 1986; Ledoit and Wolf, 2003). Alternatively, some authors propose to construct a portfolio by solving the classical Markowitz model in which the original correlation matrix is replaced by a correlation based clustering matrix with the aim of providing portfolios that are quite robust with respect to noise deriving from the finiteness of the sample size (see, among others, Onnela et al. (2004), Puerto et al. (2020), Tola et al. (2008), Watts and Strogatz (1998) and the references therein). Some of these methods were compared to the RMT approach (see, e.g., Tola et al., 2008), but with the conclusion that the alternative filtering procedures provide different portfolio results each characterized by its specific strengths and weaknesses. Hence, it is difficult to compare the different methods and definitively state which is the best one.

Finally, different mathematical models based on the minimization of *downside risk* measures have also been proposed in order to mitigate the effects of extreme events. These measures are, in fact, functions of skewness and kurtosis of the returns distributions. CVaR (Rockafellar and Uryasev, 2002) provides an example of this approach where the downside risk is taken into account by controlling the probability of the losses that might be encountered in the tail of the distribution.

Important contributions in the literature on portfolio selection and asset management support the belief that, in a MVO framework, uncertainty from estimation errors in expected returns tends to have more influence than the one in the covariance matrix (Best and Grauer, 1991a,b; Kolm et al., 2014). As a matter of fact, these papers affirm that the relative importance of the two errors depends on the investor's risk aversion, and, as the risk tolerance decreases, the impact of errors in expected returns relative to errors in the covariance matrix becomes small. In general, both errors should be controlled in order to get good estimates of the portfolio parameters. In Kolm et al. (2014) the authors present some of the most common techniques for mitigating estimation errors. Among others, they suggest introducing constraints on portfolio weights, diversification measures, incorporating higher moments in a mathematical model or developing robust optimization approaches. The reader is referred to Kolm et al. (2014) for a more detailed description of these techniques.

To summarize, at the moment, although numerous alternatives to the MVO model have appeared in the literature in order to cope with estimate errors when managing historical data, no definitive clear leader emerged up to now. In conclusion, despite of all the flaws affecting the MVO model, at this time it is still the most used tool for portfolio selection.

In this paper we put ourselves from the practitioner point of view with the aim of applying the MVO model to obtain from real financial datasets a portfolio with good out-of-sample values for expected return and variance. We propose a new filtering technique that can be seen also as a refinement of the classical MVO model. The approach relies on the idea that estimation errors may arise for the presence of extremely rare observations (outliers) in the historical assets' return series. Therefore, removing such observations from the dataset helps in filtering only reliable information and provides a way of obtaining portfolios with good out-of-sample performance. This is an already known approach which typically establishes a percentage of observations to remove on the tails of the return distribution, and then computes truncated means. The covariance matrix is computed on the same filtered distribution. After this filtering, the classical MVO model can be applied to select the portfolio. The difficult step in this approach is to detect outliers, and decide how many — and which — observations must be removed from the observed series. We also point out that there are many assets in the financial market for which returns are observed simultaneously, and removing a time observation means removing that observation for all the considered assets. Therefore, the problem cannot be faced asset by asset, but there is the need of finding a systemic rule, taking into consideration that deleting a time observation corresponding to a rare value for a given asset, implies the removal of the corresponding time observations for *all* the other assets. Then, the convenience of removing or not observations at a time t should be carefully evaluated, and the way in which this is actually performed is a relevant methodological issue. We claim that the problem can be actually managed in a systematic approach via combinatorial optimization and mathematical programming.

In this paper, we propose a new filtering methodology based on a revision of the classical MVO Markowitz model in which the optimal filtering decision, related to how many and which observations have to be removed from the dataset, is incorporated in the optimization model which minimizes the variance of a portfolio guaranteeing a minimum level of expected return. The power of the model is that it is able to manage in a single step two decisions, filtering observations and selecting the optimal portfolio. Additional (integer) variables, mapping which time observations must be removed and which not, are included in the model. The optimization procedure outputs simultaneously the best choice for the vector of filtering variables and for the vector of portfolio weights. We point out that in this approach establishing the number of removed observations is part of the decision, so that also it must be considered as a decision variable.

It is important to highlight that, acting on the possible removal of time observations, implies that both errors in expected return and in covariances estimates are managed together in the same process. In this way, no preference is given a priori about which among these two errors is more important; the task of controlling both of them is then left to the model and to the optimization procedure, the only guiding criterion being to minimize the final portfolio variance. Finally, we observe that our filtering strategy, based on the MVO model, can be seen as an alternative way of dealing with the fat tails issue to the CVaR approach. In the following sections we show that, in general, the two approaches are not comparable, but, under some assumptions, both fall into a more general framework consisting in optimizing filtered statistical moments.

Since our proposal falls into the broad class of filtering methods, we compare it in terms of performance with some of the most common filtering techniques found in the literature. In particular, we consider the two filtered procedures presented in Schafer et al. (2010) which share with our method the idea of maintaining the MVO approach, but with a manipulation of the data feeding the model. Both approaches have the common objective to select portfolios with good future (out-of-sample) performance, so that how much robust and effective they are can be established empirically by verifying which produces the best solutions for real financial datasets. It must be pointed out that there is a clear characteristic aspect which distinguishes our approach from the compared ones: filtering and portfolio optimization are conducted in synergy by solving one single model. This is a typical strength of combinatorial optimization, and this is, in fact, the innovative contribution of our approach to filtering MVO for portfolio selection.

The results are encouraging since with our model we obtain better performance results than those of the two competing filtering methods. This in spite of the computational effort required by our model which is a Mixed Integer Quadratic Programming (MIQP) problem. We show that standard solvers are able to find solutions in reasonable times at least for small and medium size financial datasets. For larger size we can still apply our approach by proposing a heuristic procedure that, from a computational viewpoint, has proven to be very efficient in practice.

The paper is organized as follows. Section 2 describes the two filtered MVO models presented in the literature based on the Random Matrix Theory and the Power Mapping techniques. In Section 3 we introduce our Scenario Filtering approach and formulate our general filtered MVO model as a nonlinear programming problem. We also discuss some complexity issues of this problem and analyze the relation between CVaR and our filtered variance approach in Section 3.1. In Section 3.2 we show how to reformulate it as a MIQP problem easy to solve with standard optimization software at least for small and medium size financial datasets. In this section we also derive a set of valid inequalities for the quadratic problem. A heuristic approach for solving large size instances is proposed in Section 3.3. Section 4 presents an extensive experimental analysis where we compare the out-of-sample performance of all the approaches on five real-world financial datasets. Finally, some conclusions and further research are outlined.

2. The MVO filtering models: Notation and definitions

The Mean–Variance Markowitz problem (Markowitz, 1952) is a classic portfolio optimization problem in finance, where, given an amount of money, the aim is to select a portfolio of n assets through two criteria: the expected return and the risk due to the variability of the returns. In the standard framework, the risk is measured by the variance of the portfolio returns. In the ideal case, the covariance and the mean of the returns are known, and the problem is: finding the proportion x_j of capital invested in each asset j, $j=1,\dots,n$, in a stock market in order to minimize the variance of the portfolio for a required level of expected return μ_0. Let μ_j be the expected return of asset j, and σ_{ij} be the covariance of returns of asset i and asset j. The Markowitz model is formulated as the following convex quadratic program:

$$\min \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j \sigma_{ij} \qquad \text{(Markowitz)}$$

s.t.

$$\sum_{j=1}^{n} \mu_j x_j \geq \mu_0$$

$$\sum_{j=1}^{n} x_j = 1$$

$$x_j \geq 0 \quad j=1,\dots,n.$$

The above model refers to the case where short-selling is not allowed in the optimization, i.e. the fractions x_j must be nonnegative. As observed in Schafer et al. (2010), imposing this constraint has the direct and rather limiting consequence that positive correlations between assets' returns cannot be used to reduce the portfolio risk. As we will see later, this may limit the potential for lowering the risk through the application of noise reduction methods based on spectral filtering. When short-selling is allowed, the fractions x_j are not restricted to be nonnegative. In any case, we observe that short-selling models are considered rather unrealistic in the specialized literature, as pointed out in Kondor et al. (2007), for legal and liquidity reasons.

In a real financial market, we have $T+1$ different observations (scenarios) for the prices of n given assets. Let P_{jt} be the price of asset j, $j=1,\dots,n$, at time t, $t=0,1,\dots,T$. For a financial risk manager it is of crucial interest to have good estimates for the returns and correlations between stocks. Following (Schafer et al., 2010), we compute the T rates of return of asset j, r_{jt}, as

$$r_{jt} = \frac{P_{jt}-P_{jt-1}}{P_{jt-1}}, \quad j=1,\dots,n, \quad t=1,\dots,T.$$

We assume that, for each asset j, the observed return r_{jt} has associated a probability p_t, $t=1,\dots,T$, with $\sum_{t=1}^{T} p_t = 1$. Under the hypothesis of no further information, we assume $p_t = \frac{1}{T}$, $t = 1\dots,T$. The average rate of return of asset j is $\mu_j = \sum_{t=1}^{T} p_t r_{jt}$, $j=1,\dots,n$. We estimate the covariance σ_{ij} between the rate of returns of assets i and j by computing

$$\sigma_{ij} = \frac{1}{T}\sum_{t=1}^{T}(r_{it}-\mu_i)(r_{jt}-\mu_j), \quad i,j=1,\dots,n.$$

The correlation matrix C is given by $C_{ij} = \frac{\sigma_{ij}}{\sigma_i\sigma_j}$, $i,j=1,\dots,n$, where the volatility of asset j is measured by $\sigma_j = +\sqrt{\sigma_{jj}}$, $j=1,\dots,n$. The rate of return at time t of a portfolio $x=(x_1,\dots,x_n)$ is

$$y_t(x) = \sum_{j=1}^{n} r_{jt}x_j, \qquad t=1,\dots,T,$$

the expected portfolio rate of return is

$$\mu(x) = \sum_{t=1}^{T} p_t y_t(x) = \sum_{j=1}^{n} \mu_j x_j,$$

and the portfolio risk can be measured by

$$V(x) = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j \sigma_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j \sigma_i \sigma_j C_{ij}.$$

It is important to note that the correlation matrix depends on the time horizon T and, for a finite T, the correlations obtained from historical data are affected by a considerable amount of noise, which leads to a substantial error in the estimation of the portfolio risk (Fama, 1965; Jansen and de Vries, 1991; Mandelbrot, 1963; Schafer et al., 2010).

The Random Matrix Theory introduced by Laloux et al. (2000) is helpful to identify the noise in correlation matrices, and also shows a way to reduce this noise. Actually, the idea is that, after diagonalization of the correlation matrix $C = U^{-1}\Lambda U$, only the h highest eigenvalues in the diagonal of Λ must be considered, while the remaining ones are set to zero. Let Λ^f be the filtered eigenvalues diagonal matrix and U be the original eigenvector matrix. Then, the new filtered correlation matrix C^f is computed

$$C^f = U^{-1}\Lambda^f U.$$

The normalization of the elements on the diagonal of C^f to 1 is restored by setting $C^f_{jj} = 1$ for all $j = 1,\dots,n$. This method is capable of removing the noise for uncorrelated assets completely (see Schafer et al., 2010 for a complete description of the method). Finally, the MVO model filtered according to the RMT becomes:

$$\begin{aligned} \min \ & \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j \sigma_i \sigma_j C^f_{ij} \qquad \text{(RMT)}\\ \text{s.t.} \ & \\ & \sum_{j=1}^{n} \mu_j x_j \geq \mu_0 \\ & \sum_{j=1}^{n} x_j = 1 \\ & x_j \geq 0 \quad j = 1,\dots,n. \end{aligned}$$

The weakness of the RMT method is that it cuts off all the information contained in the discarded eigenvalues. An alternative method developed in order to reduce noise was introduced in Guhr and Kälber (2003). It is the so-called Power Mapping technique. It takes each element of the correlation matrix and raises its absolute value to some power q, while preserving the sign. With this method one obtains a new correlation matrix whose elements are:

$$C^{(q)}_{ij} = \operatorname{sign}(C_{ij})|C_{ij}|^q.$$

The idea behind this method is that the effect of the noise, which typically arises in the small correlations, can be broken in this way, with an effect which is stronger as the value of q increases. Thus, the problem is to choose the right value for q also taking into account that a byproduct effect of q is produced on all correlations. The MVO model filtered according to the Power Mapping is then formulated as follows:

$$\begin{aligned} \min \ & \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j \sigma_i \sigma_j C^{(q)}_{ij} \qquad \text{(Power Mapping)}\\ \text{s.t.} \ & \\ & \sum_{j=1}^{n} \mu_j x_j \geq \mu_0 \\ & \sum_{j=1}^{n} x_j = 1 \\ & x_j \geq 0 \quad j = 1,\dots,n. \end{aligned}$$

The drawback of the two above methods is that, generally, to reduce noise, they, in fact, eliminate too much information contained in the observed time series of the assets prices, and this inevitably affects the estimation of the portfolio return and risk.

3. The scenario filtering approach

In this paper we propose an alternative refinement of the MVO model that incorporates some additional constraints that allows the model to eliminate observations (outliers) in order to obtain a good estimation of return and risk. The idea is that, in particular in the long run, few extreme events in the distribution of assets prices could extremely affect the volatility and performance of a portfolio in the future. With our method we maintain the bulk of the data while dropping only those extreme observations caused by a distribution with fat tails or more simply by measurement errors.

More precisely, we want to model the problem of filtering (eliminating) a fixed number K of the T observed scenarios for the assets rate of returns, $K < T$, while simultaneously solving the MVO problem. We refer to this problem as Problem (P). The selection of the filtered scenarios can be modeled with binary variables:

$$z_t = \begin{cases} 1 & \text{if scenario } t \text{ is filtered} \\ 0 & \text{otherwise} \end{cases} \qquad t = 1,\dots,T.$$

Under the equiprobability assumption, for a filtering realization $z = (z_1,\dots,z_T)$ of $K < T$ scenarios, the return distribution of a portfolio $x = (x_1,\dots,x_n)$ is $y_t(x)$ with probability $\tilde{q}_t(z_t) = \frac{1}{T-K}(1 - z_t)$. Then, for each $t = 1,\dots,T$ we denote by $q_t = \frac{1}{T-K}$ and, recalling that $y_t(x) = \sum_{j=1}^{n} r_{jt}x_j$, the *filtered expected return* of the portfolio is

$$\tilde{\mu}(x,z) = \sum_{t=1}^{T} \tilde{q}_t(z_t) y_t(x) = \sum_{t=1}^{T}\sum_{j=1}^{n} q_t r_{jt} x_j (1 - z_t), \tag{1}$$

and the corresponding *filtered variance* is

$$\tilde{V}(x,z) = \sum_{t=1}^{T} \tilde{q}_t(z_t)(y_t(x) - \tilde{\mu}(x,z))^2 = \sum_{t=1}^{T} q_t (y_t(x) - \tilde{\mu}(x,z))^2 (1 - z_t). \tag{2}$$

Problem (P) can be stated as:

$$\begin{aligned} \min \ & \tilde{V}(x,z) \qquad \text{(P)}\\ \text{s.t.} \ & \\ & \tilde{\mu}(x,z) \geq \mu_0 \\ & \sum_{j=1}^{n} x_j = 1 \\ & \sum_{t=1}^{T} z_t = K \\ & x_j \geq 0 \quad j = 1,\dots,n \\ & z_t \in \{0,1\} \quad t = 1,\dots,T. \end{aligned}$$

Problem (P) is a Quadratically Constrained Mixed Integer Nonlinear Programming problem that falls into the class of considerably difficult NP-hard problems (Floudas and Visweswaran, 1995). We also show that testing feasibility of problem (P) is NP-hard in two cases: (i) when we require that the portfolio expected return be exactly equal to μ_0 (see e.g. Cesarone et al., 2013; Schafer et al., 2010); (ii) when we add *threshold constraints* on investment $\ell_j \leq x_j \leq u_j$, $j = 1,\dots,n$, which impose limitations to the assets shares in the portfolio (see Mansini et al., 2007). Indeed, we note that in this second case we only need to assume upper bounds $u_j > 0$, $j = 1,\dots,n$.

We prove these complexity results via polynomial reduction to the *Partition* problem that can be stated as follows: Given a finite set $A = \{a_1,\dots,a_n\}$ with $a_i \in \mathbb{Z}_+$ for all $i \in I = \{1,\dots,n\}$, the Partition problem asks for the existence of an index subset $I' \subseteq I$ such that

$$\sum_{i \in I'} a_i = \sum_{i \in I \setminus I'} a_i.$$

The Partition problem is NP-complete and remains NP-complete even if we require $|I'| = \frac{n}{2}$ (see Garey and Johnson, 1979).

Proposition 1. *Testing feasibility of Problem (P) is NP-hard when $\tilde{\mu}(x,z)=\mu_0$ is required.*

Proof. In Problem (P), set $T=n$, $K=\frac{n}{2}$, $\mu_0=\frac{1}{n}\sum_{i=1}^{n}a_i$, and the assets rate of returns $r_{jt}=a_t$ for all $j,t\in I=\{1,\dots,n\}$. Then, the rate of return at time $t\in I$ of any feasible portfolio x is

$$y_t(x)=\sum_{j=1}^{n}a_t x_j=a_t,$$

and the corresponding filtered expected return becomes

$$\tilde{\mu}(x,z)=\frac{2}{n}\sum_{t=1}^{T}a_t(1-z_t).$$

In Problem (P) we assume that, given a feasible solution, if $z_t=1$ then $t\in I'\subset I$, otherwise $t\in I\setminus I'$. Furthermore, we also observe that when in the Partition problem $|I'|=\frac{n}{2}$ is required, it implies that

$$\sum_{i\in I'}a_i=\sum_{i\in I\setminus I'}a_i=\frac{\sum_{i\in I}a_i}{2}.$$

Hence, the answer to the Partition problem is "yes" if and only if Problem (P) is feasible. □

Proposition 2. *Testing feasibility of Problem (P) is NP-hard when $x_j\le u_j$, with $u_j>0$, $j=1,\dots,n$, is required.*

Proof. We refer to a modified Partition problem, that is, we assume $A\subseteq\mathbb{Q}_+$ and $\sum_{i=1}^{n}a_i=2$. This assumption implies that if $a_i>1$ for some $i\in I$, then the answer to the problem is trivially "no". Thus, we assume $a_i\le 1$ for all $i\in I$. We observe that this modified problem is equivalent to the (standard) Partition problem if we multiply the numbers in the set A by a proper factor. Hence, the modified Partition problem is NP-complete, as well.

In Problem (P), set $T=n$, $K=\frac{n}{2}$, $\mu_0=\frac{n}{2}$, $u_j=a_j$ for all $j=1,\dots,n$, and the assets rate of returns $r_{jt}=\frac{1}{a_j}$ if $t\neq j$, and $r_{jt}=-(K-1)\frac{1}{a_j}-K\varepsilon$ if $t=j$, for all $j,t\in I=\{1,\dots,n\}$ and any $\varepsilon>0$. Then, the filtered average rate of return of asset $j\in I$ is

$$\tilde{\mu}_j(z)=\frac{1}{K}\sum_{t=1}^{T}r_{jt}(1-z_t).$$

Note that $\tilde{\mu}_j(z)=\frac{1}{a_j}$ if $z_j=1$, and $\tilde{\mu}_j(z)=-\varepsilon$ if $z_j=0$. Thus, it follows that $\tilde{\mu}_j(z)=\frac{1}{a_j}z_j-\varepsilon(1-z_j)$ for all $j\in I$. Hence, the filtered expected return of the portfolio can be expressed as

$$\tilde{\mu}(x,z)=\sum_{j=1}^{n}\tilde{\mu}_j(z)x_j=\sum_{j=1}^{n}\frac{1}{a_j}z_jx_j-\varepsilon\sum_{j=1}^{n}(1-z_j)x_j.$$

Finally, the feasible region of Problem (P) becomes

$$\sum_{j=1}^{n}\frac{1}{a_j}z_jx_j-\varepsilon\sum_{j=1}^{n}(1-z_j)x_j\ge K$$
$$\sum_{j=1}^{n}x_j=1$$
$$\sum_{j=1}^{n}z_j=K$$
$$0\le x_j\le a_j \quad j=1,\dots,n$$
$$z_j\in\{0,1\} \quad j=1,\dots,n.$$

Therefore, with the same reasoning of the previous proposition, the answer to the (modified) Partition problem is "yes" if and only if Problem (P) is feasible. □

In the light of the above propositions, it is important to provide efficient solution methods to solve Problem (P). In the following, we propose both an exact method and a heuristic procedure.

3.1. Relation between filtered variance and CVaR

We recall that for each natural number s, the sth order *raw* moment of a discrete random variable, in particular the rate of return of a portfolio x, is

$$\mu_s^{\text{raw}}(x)=\sum_{t=1}^{T}p_t\left(y_t(x)\right)^s.$$

On the other hand, the sth order central moment is

$$\mu_s^{\text{cent}}(x)=\sum_{t=1}^{T}p_t\left(y_t(x)-\mu(x)\right)^s.$$

In Mansini et al. (2015) it is shown that when equally probable scenarios (i.e., $p_t=\frac{1}{T}$) are considered and a threshold $\beta=\frac{m}{T}$ is taken, with m the number of worst realizations, the corresponding Conditional Value-at-Risk of a portfolio x is

$$\text{CVaR}_{\frac{m}{T}}(x)=\frac{1}{m}\sum_{\ell=1}^{m}y_{t_\ell}(x),$$

where $y_{t_1}(x),\dots,y_{t_m}(x)$ are the m worst realizations for the portfolio rates of return. In this case we observe that the CVaR measure can be interpreted as the minimum first moment of the rate of return of a portfolio x when filtering $T-m$ scenarios, that is:

$$\tilde{\mu}_{1,T-m}^{\text{raw,min}}(x)=\min_{\substack{z_1,\dots,z_T\in\{0,1\}:\\ \sum_{t=1}^{T}z_t=T-m}}\left\{\frac{1}{m}\sum_{t=1}^{T}y_t(x)(1-z_t)\right\}.$$

In light of the above interpretation, we can define higher order *minimum filtered moments*. Let us consider the second central moment of the rate of return of a portfolio x filtering $T-m$ scenarios, we obtain:

$$\tilde{\mu}_{2,T-m}^{\text{cent,min}}(x)=\min_{\substack{z_1,\dots,z_T\in\{0,1\}:\\ \sum_{t=1}^{T}z_t=T-m}}\left\{\frac{1}{m}\sum_{t=1}^{T}\left(y_t(x)(1-z_t)-\frac{1}{m}\sum_{t=1}^{T}y_t(x)(1-z_t)\right)^2\right\}. \tag{3}$$

To find the portfolio that minimizes the filtered second central moment (3) for $m=T-K$ is equivalent to find the portfolio that minimizes the filtered variance $\tilde{V}(x,z)$ (i.e. Problem (P) without considering the minimum expected return constraint). Hence, we remark that although we propose a variation of the mean–variance portfolio model of Markowitz, we also have developed a portfolio selection model that can be seen as an alternative to CVaR but operating in the same, more general, framework of optimizing filtered statistical moments.

In any case, we stress the fact that despite the above relation between CVaR and our filtered model, each measure has a different purpose. Actually, CVaR is used to identify the worst scenarios of a portfolio with the aim of maximizing them, while in the minimization of the filtering variance the idea is to eliminate those scenarios that increase the variance treating them as outliers. Thus, in general these two measures are not fully comparable.

3.2. The mixed integer quadratic programming model

In this section we present a reformulation of Problem (P) as a MIQP problem that, even if it still belongs to the class of difficult NP-hard problems, it can be efficiently solved by using some commercial or free optimization solvers for general MIQP models.

We consider the filtered variance of the portfolio return (2):

$$\tilde{V}(x,z)=\sum_{t=1}^{T}q_t|y_t(x)-\tilde{\mu}(x,z)|^2(1-z_t).$$

Since in Problem (P) we minimize the filtered variance we can apply the McCormick linearization (McCormick, 1976) to the products of squared absolute values and $(1-z_t)$ terms. Therefore, the filtered variance can be equivalently written as

$$\sum_{t=1}^{T} q_t d_t^2 \tag{4}$$

whenever the following set of constraints is satisfied

$$d_t \geq |y_t(x) - \tilde{\mu}(x,z)| - z_t M_t \quad t = 1,\dots,T$$

for nonnegative variables $d_t \geq 0$ and big enough constants $M_t > 0$, $t = 1,\dots,T$. In addition, since $|y_t(x) - \tilde{\mu}(x,z)| = \max\{y_t(x) - \tilde{\mu}(x,z), -(y_t(x) - \tilde{\mu}(x,z))\}$, the above set of constraints can be replaced by the following one:

$$\begin{aligned} &d_t \geq y_t(x) - \tilde{\mu}(x,z) - z_t M_t^+ \quad t = 1,\dots,T \\ &d_t \geq -(y_t(x) - \tilde{\mu}(x,z)) - z_t M_t^- \quad t = 1,\dots,T \end{aligned} \tag{5}$$

where the constants $M_t^+, M_t^- > 0$ are big enough for each $t = 1,\dots,T$. We can also linearize the quadratic terms $x_j(1-z_t)$ in $\tilde{\mu}(x,z)$ by adding a new set of variables $\tilde{x}_{jt} \geq 0$, $j = 1,\dots,n$, $t = 1,\dots,T$, such that $\tilde{x}_{jt} = x_j(1-z_t)$ by imposing the following constraints:

$$\begin{aligned} &\tilde{x}_{jt} \leq (1-z_t) \quad j = 1,\dots,n, t = 1,\dots,T \\ &\tilde{x}_{jt} \leq x_j \quad j = 1,\dots,n, t = 1,\dots,T \\ &\tilde{x}_{jt} \geq x_j - z_t \quad j = 1,\dots,n, t = 1,\dots,T. \end{aligned} \tag{6}$$

Constraints (6) refer to a standard way of linearizing the product of a binary variable and a continuous nonnegative one. They can be included, together with (5), in a reformulation of Problem (P) in which the objective function (4) is minimized.

It is easy to see that, when integrated in our model, constraints (6) are implied by the following ones:

$$\begin{aligned} &\sum_{j=1}^{n} \tilde{x}_{jt} = 1 - z_t \quad t = 1,\dots,T \\ &\tilde{x}_{jt} \leq x_j \quad j = 1,\dots,n, t = 1,\dots,T. \end{aligned} \tag{7}$$

The above constraints lead to an equivalent formulation that in our experiments showed to require a smaller computational effort in the solution process. Thus, in our model we replace (6) by (7).

From the above discussion, we propose the following MIQP model:

$$\min \sum_{t=1}^{T} q_t d_t^2 \tag{Scenario Filtering}$$

s.t.

$$\begin{aligned}
&d_{t'} \geq \sum_{j=1}^{n} r_{jt'} x_j - \sum_{t=1}^{T}\sum_{j=1}^{n} q_t r_{jt} \tilde{x}_{jt} - z_{t'} M_{t'}^+ \quad t' = 1,\dots,T \\
&d_{t'} \geq -\sum_{j=1}^{n} r_{jt'} x_j + \sum_{t=1}^{T}\sum_{j=1}^{n} q_t r_{jt} \tilde{x}_{jt} - z_{t'} M_{t'}^- \quad t' = 1,\dots,T \\
&\sum_{j=1}^{n} \tilde{x}_{jt} = 1 - z_t \quad t = 1,\dots,T \\
&\tilde{x}_{jt} \leq x_j \quad j = 1,\dots,n, t = 1,\dots,T \\
&\sum_{t=1}^{T}\sum_{j=1}^{n} q_t r_{jt} \tilde{x}_{jt} \geq \mu_0 \\
&\sum_{j=1}^{n} x_j = 1 \\
&\sum_{t=1}^{T} z_t = K \\
&x_j \geq 0 \quad j = 1,\dots,n \\
&z_t \in \{0,1\} \quad t = 1,\dots,T \\
&d_t \geq 0 \quad t = 1,\dots,T \\
&\tilde{x}_{jt} \geq 0 \quad j = 1,\dots,n, t = 1,\dots,T.
\end{aligned}$$

The above model is a MIQP model that uses big-M constants. It is well-known that big-M constants often produce large gaps between the continuous relaxation of the model and the MIP objective values, which can induce a poor performance of the model from the computational time point of view. Thus, it is important to provide *tight* values for these constants. In our case, for a given $t' = 1,\dots,T$, we can set the values as follows:

$$M_{t'}^+ = \max_{\substack{x_1,\dots,x_n \geq 0 \\ z_1,\dots,z_T \in \{0,1\}}} \left\{ \sum_{j=1}^{n} x_j \left(r_{jt'} - \sum_{t=1}^{T} q_t r_{jt}(1-z_t) \right) : \sum_{j=1}^{n} x_j = 1, \sum_{t=1}^{T} z_t = K, z_{t'} = 1 \right\}$$

and

$$M_{t'}^- = \max_{\substack{x_1,\dots,x_n \geq 0 \\ z_1,\dots,z_T \in \{0,1\}}} \left\{ \sum_{j=1}^{n} x_j \left(-r_{jt'} + \sum_{t=1}^{T} q_t r_{jt}(1-z_t) \right) : \sum_{j=1}^{n} x_j = 1, \sum_{t=1}^{T} z_t = K, z_{t'} = 1 \right\}.$$

It is not difficult to see that the solution of the above maximization problems can be easily computed as

$$M_{t'}^+ = \max_{j=1,\dots,n} \left\{ r_{jt'} + B_{jt'}^+ \right\} \quad \text{and} \quad M_{t'}^- = \min_{j=1,\dots,n} \left\{ -r_{jt'} + B_{jt'}^- \right\}$$

where $B_{jt'}^+$ is the sum of the $K-1$ greater numbers of the set $\{-q_1 r_{j1}, \dots, -q_T r_{jT}\} \setminus \{-q_{t'} r_{jt'}\}$ and $B_{jt'}^-$ is the sum of the $K-1$ greater numbers of the set $\{q_1 r_{j1}, \dots, q_T r_{jT}\} \setminus \{q_{t'} r_{jt'}\}$, for each $j = 1,\dots,n$. We observe that, although alternative MIQP models without big-M constants can be obtained by using different reformulations of Problem (P), the advantage of our MIQP model reformulation is that its objective function is convex. Due to this property, the continuous relaxation of the problems arising from a Branch and Bound strategy are convex quadratic programs, for which solvers provide more reliable solutions than the ones obtained for nonconvex problems. Reliability of solutions is crucial when one takes into account the numerical issues that occur when optimizing a MIQP model (see Bienstock, 1996), and also when one consider the possible effect of these numerical issues on the portfolios (out-of-sample) performance. In this way, as observed in our computational experiments, our MIQP model guarantees the reliability of the solutions found without worsening the computational times too much. This is due to the tightness of the big-M constants. Moreover, we observed that when additional variables are used to reformulate our MIQP model avoiding the use of big-M constants, the computational time saved does not compensate for the additional time obtained by augmenting the problem's dimension.

Our MIQP model admits a set of valid inequalities related with some others presented in the literature for similar problems. In particular, we note the similarity of Problem (P) with the problem studied in Bienstock (1996). In the historical data approach assumed in the present work, the set of considered stocks and the set of available observations determine the input for the MVO problem. Roughly speaking, the problem in Bienstock (1996) corresponds to solving the MVO problem by selecting a cardinality constrained subset of the stocks. In our case, we are interested in solving the MVO problem but we want to select a subset of time-observations. Therefore, in the former problem there is a cardinality constraint over the number of assets included in the portfolio, while in our model the cardinality constraint is on the number of considered observations. Of course, the constraints are similar but not the same. As in Bienstock (1996) we derive a set of valid inequalities for our problem that are inspired by the cover inequalities for the knapsack problem (Nemhauser and Wolsey, 1988; Wolsey, 1998).

We have derived our MIQP model by reformulating Problem (P). In that process, we have introduced variables $\tilde{x}_{jt} = x_j(1-z_t)$, $j = 1,\dots,n, t = 1,\dots,T$, originally not defined in Problem (P). We show now that indeed Problem (P) can be conceptually stated using only $\tilde{x}_{jt}$ variables. We start by noting that the filtered expected return (1) can be expressed using $\tilde{x}_{jt}$ variables as

$$\tilde{\mu}(\tilde{x}) = \sum_{t=1}^{T}\sum_{j=1}^{n} q_t r_{jt}\tilde{x}_{jt}.$$

In the same way, and taking into account that $\tilde{V}(x,z) = \sum_{t=1}^{T} \tilde{q}_t(z_t)(y_t(x))^2 - (\tilde{\mu}(x,z))^2$, we can write the corresponding filtered variance (2) as

$$\tilde{V}(\tilde{x}) = \sum_{t=1}^{T} q_t(\sum_{j=1}^{n} r_{jt}\tilde{x}_{jt})^2 - (\tilde{\mu}(\tilde{x}))^2.$$

We also need to rewrite the minimum expected return requirement constraint

$$\sum_{t=1}^{T}\sum_{j=1}^{n} q_t r_{jt}\tilde{x}_{jt} \geq \mu_0$$

in the equivalent form with positive coefficients and positive right-hand side

$$\sum_{t=1}^{T}\sum_{j=1}^{n} a_{jt}\tilde{x}_{jt} \geq \beta,$$

where $a_{jt} = q_t(r_{jt} + \alpha) > 0$, $j = 1,\dots,n, t = 1,\dots,T$, $\beta = \mu_0 + \alpha > 0$, and $\alpha \geq 0$ is taken big enough. The equivalence between the constraints above follows from the fact that $\sum_{t=1}^{T}\sum_{j=1}^{n} q_t \alpha \tilde{x}_{jt} = \alpha \frac{1}{T-K}\sum_{t=1}^{T}(1 - z_t)\sum_{j=1}^{n} x_j = \alpha$ in the setting of Problem (P). Finally, Problem (P) can be re-stated as

$$\min \ \tilde{V}(\tilde{x}) \tag{$P_{\tilde{x}}$}$$

s.t.

$$\sum_{t=1}^{T}\sum_{j=1}^{n} a_{jt}\tilde{x}_{jt} \geq \beta$$

$$\left|\left\{t \in \{1,\dots,T\} : \tilde{x}_{jt} > 0 \text{ for some } j = 1,\dots,n\right\}\right| = T - K$$

$$\sum_{j=1}^{n} \tilde{x}_{jt} = 1 \quad t = 1,\dots,T : \tilde{x}_{jt} > 0 \text{ for some } j = 1,\dots,n$$

$$\tilde{x}_{jt} = \tilde{x}_{jt'} \quad j = 1,\dots,n, t, t' = 1,\dots,T : x_{jt}, x_{jt'} > 0$$

$$0 \leq \tilde{x}_{jt} \leq 1 \quad j = 1,\dots,n, t = 1,\dots,T.$$

Using the same terminology as in Bienstock (1996), we say that a set $S \subseteq \{1,\dots,T\}$ is *critical* if:

for every $R \subseteq S$ with $|R| = T - K$, $\sum_{t\in R}\sum_{j=1}^{n} a_{jt} < \beta$.

Then, the following results hold.

Proposition 3. *If S is a critical set, then the set of inequalities*

$$\sum_{t\in S} \tilde{x}_{jt} \leq T - K - 1 \quad j = 1,\dots,n$$

are valid for Problem ($P_{\tilde{x}}$).

Proof. Note that the definition of critical set S implies that in Problem ($P_{\tilde{x}}$) at most $T - K - 1$ indices $t \in S$ can have associated nonzero x_{jt} variables, otherwise the minimum expected return requirement constraint is not satisfied. Since Problem ($P_{\tilde{x}}$) also forces the consistency of the 'actives' scenarios in the sets $\{\tilde{x}_{j1},\dots,\tilde{x}_{jT}\}$, $j = 1,\dots,n$, the result follows. □

Corollary 4. *The set of inequalities given in Proposition 3 are also valid for our MIQP model.*

We remark that each critical set in Bienstock (1996), or cover set in Wolsey (1998) (also called dependent set in Nemhauser and Wolsey (1988)), has a naturally associated valid inequality for the problem considered in each reference. Given the particularity of our Problem ($P_{\tilde{x}}$), each critical set provides n valid inequalities for this problem.

As we will see in the experimental results in Section 4, our MIQP model is efficiently solvable for small and medium size financial datasets. For larger size instances, computational times are considerably longer. For this reason, in the following section we propose a heuristic approach for solving Problem (P).

3.3. The heuristic algorithm

In this subsection we present our heuristic procedure. We note that solving Problem (P) implies to decide which K scenarios have to be filtered from the T observed ones. This corresponds to make a decision among $\binom{T}{K}$ possible ones. Thus, it is reasonable to think that the computational effort to solve Problem (P) grows with the cardinality of the decision set, from $K = 1$ up to $K = \frac{T}{2}$. Indeed, this was confirmed by our computational experiments. Based on this consideration, we design a heuristic algorithm that exploits a nested solutions strategy. Let Problem (P_k) be our Problem (P) when $k \leq K$ scenarios have to be filtered, and let $z(k) = (z_1(k),\dots,z_t(k),\dots,z_T(k))$ be the filtering decision variables of Problem (P_k), $k = 1,\dots,K$. Then, at each step we solve Problem (P_{k+1}) but keeping the best scenarios filtering $z^*(k)$ found by the heuristic for the previous Problem (P_k), $k = 1,\dots,K-1$. The computational effort to perform each step equals the effort for solving Problem (P) when only one scenario has to be filtered. We show the basic pseudocode below.

Algorithm 1 Heuristic

1: **begin**
2: solve Problem (P_1)
3: let $z^*(1)$ be the scenarios filtering obtained
4: **for** $k = 2,\dots,K$ **do**
5: solve Problem (P_k) fixing $z_t(k) = 1$ if $z_t^*(k-1) = 1$, $t = 1,\dots,T$
6: let $z^*(k)$ be the scenarios filtering obtained
7: **return** x^* the best portfolio found
8: **end**

It is clear that, in Algorithm 1 the solution found at the first step (line 2) is optimal for $k = 1$. On the contrary, from step 2 on, we obtain suboptimal solutions for $k \geq 2$. In spite of this, in our experiments the heuristic showed to be effective in finding good quality solutions. To implement Algorithm 1 we can solve our MIQP model for $K = 1$ in line 2, and our MIQP model for $K = k$ in line 5 but fixing appropriately the z_t variables. We call Version 1 this implementation of Algorithm 1. An alternative implementation, which we call Version 2, is described below.

Let $\mathcal{T} = \{1,\dots,T\}$. Suppose that we want to solve the MVO problem but considering only a given subset of the observations $R \subseteq \mathcal{T}$. From the results in Section 3.2, it is straightforward to see that such a problem can be formulated as follows:

$$\min \sum_{t\in R} \hat{p}_t(R) d_t^2 \tag{R-MVO}$$

s.t.

$$d_{t'} \geq \sum_{j=1}^{n} r_{jt'}x_j - \sum_{t\in R}\sum_{j=1}^{n} \hat{p}_t(R) r_{jt} x_j \quad t' \in R$$

$$d_{t'} \geq -\sum_{j=1}^{n} r_{jt'}x_j + \sum_{t\in R}\sum_{j=1}^{n} \hat{p}_t(R) r_{jt} x_j \quad t' \in R$$

$$\sum_{t\in R}\sum_{j=1}^{n} \hat{p}_t(R) r_{jt} x_j \geq \mu_0$$

$$\sum_{j=1}^{n} x_j = 1$$

$$x_j \geq 0 \quad j = 1,\dots,n$$

$$d_t \geq 0 \quad t \in R$$

where $\hat{p}_t(R) = \frac{1}{|R|}$, $t \in R$. Problem R-MVO is a simple convex quadratic program and therefore it can be solved quite efficiently. Now, note that line 2 in Algorithm 1 can be performed by solving Problem R-MVO for each set $R = \mathcal{T} \setminus \{t\}$, $t \in \mathcal{T}$, and returning the best solution found. In the case of line 5, it can be performed by solving Problem R-MVO for each set $R = (\mathcal{T} \setminus \{t \in \mathcal{T} : z_t^*(k-1) = 1\}) \setminus \{t'\}$, $t' \in \{t \in \mathcal{T} : z_t^*(k-1) = 0\}$, and returning the best solution found. This procedure is our Version 2 of the implementation of Algorithm 1. The reader may notice that this procedure can be replicated using the original MVO model, which is also a convex quadratic program, using the corresponding covariance matrix in function of the considered observations. However, this alternative way of performing the procedure requires the computation of the covariance matrix for each MVO model, which charges an unnecessary additional computational effort.

Which among the two versions of Algorithm 1 is more efficient depends on the number of observations T and the size n of the set of assets. In the conditions of our computational experiments, for a fixed T, Version 1 of the Algorithms performs better than Version 2 for small size datasets, while Version 2 performs quite better than Version 1 for medium/large size datasets.

4. Experimental results

In this section we present an empirical analysis on real stocks market data with the aim of evaluating the out-of-sample performance of the portfolios selected by the MVO model when our Scenario Filtering approach is applied. In addition, we compare the performances of our filtered portfolios to those obtained by the MVO model filtered by the RMT and the Power Mapping techniques, and also to the portfolios selected by the classical Markowitz model.

We test all the above portfolio selection strategies on some real datasets belonging to the major stock markets across the world. We consider the following datasets[1] that were also used in Bruni et al. (2017), Puerto et al. (2020):

1. DJIA (Dow Jones Industrial Average, USA), containing 28 assets and 1353 weekly price observations (period: 07/05/1990–04/04/2016);
2. EUROSTOXX50 (Europe's leading blue-chip index, EU), containing 49 assets and 729 weekly price observations (period: 22/04/2002-04/04/2016);
3. FTSE100 (Financial Times Stock Exchange, UK), containing 83 assets and 625 weekly price observations (period: 19/04/2004-04/04/2016);
4. SP500 (Standard & Poor's, USA), containing 442 assets and 573 weekly price observations (period: 18/04/2005-04/04/2016).

To evaluate the performance of the models in practice, we divide the observations in two sets, where the first one is regarded as the past (in-sample window), and so it is considered known, and the rest is regarded as the future (out-of-sample window), supposed unknown at the time of portfolio selection. The in-sample window is used for selecting the portfolio, while the out-of-sample one is used for testing the performance of the selected portfolio. In particular, in our experiments we use a *rolling time window* scheme allowing for the possibility of rebalancing the portfolio composition during the holding period, at fixed intervals. We observe that a rolling window method is able to better capture intertemporal effects than single-period portfolio choice policies. A deeper discussion about the usefulness of multi-period optimization can be found in Kolm et al. (2014) and the references therein.

Following Jegadeesh and Titman (2001), Mansini et al. (2007), Puerto et al. (2020), for each dataset we adopt a period of 52 weeks (one year) as in-sample window and we consider 12 weeks (three months) as out-of-sample, with rebalancing allowed every 12 weeks.

For each dataset, in all the portfolio selection models considered we set μ_0 equal to the average of the market portfolio return in the in-sample period. When considering the RMT filtering method we choose the 5 largest eigenvalues ($h = 5$), since, as reported in Schafer et al. (2010), this choice yielded the best results in their experimental framework. For the Power Mapping method we set $q = 1.25$ since this value in Schafer et al. (2010) provided the best results in the case without short-selling. In the case of our Scenario Filtering approach we decided to remove up to 5 scenarios of each in-sample period, in order to not increase the computational burden of our method and, at the same time, to not distort the dataset too much. Since each in-sample period contains 52 observations, the above decision corresponds to remove up to the 10% (approximately) of the observations, which seems reasonable if one takes into account the frequency of occurrence of extreme events in the fat tail distributions that characterize returns distributions in financial markets. From the portfolio value performance viewpoint, we show that, in fact, this choice of K is enough to outperform the other competing filtering models.

In the following tables we consider some classic out-of-sample performance measures described below:

1. *Average return* (AvReturn): it is defined as the average $E[\mu^{\text{out}}(x)]$ of the out-of-sample returns of a portfolio x. The larger is the value of the index, the better is the corresponding portfolio performance.
2. *Out-of-sample Variance* (V-Out): it is the variance $\sigma^2(\mu^{\text{out}}(x))$ of the returns of the out-of-sample portfolios. The smaller is the value of the index, the better is the corresponding portfolio performance.
3. *Sharpe Ratio* (Sharpe) (Sharpe, 1966, 1996): consider the difference between the average of the returns of the out-of-sample portfolios, $\mu^{\text{out}}(x)$, and a constant risk-free rate of return r_f (i.e., the *expected extra return*). The Sharpe index is given by the ratio between the expected extra return and the standard deviation of the returns of the out-of-sample portfolios:
$$\frac{E[\mu^{\text{out}}(x)] - r_f}{\sigma(\mu^{\text{out}}(x))}.$$
In the bi-criteria optimization approach of the classical MVO model, the larger is the value of the index, the better is the portfolio performance. In our analysis we set $r_f = 0$.
4. *Sortino Ratio* (Sortino) (Sortino and Satchell, 2001): it is defined as the ratio between the average of expected extra return and a measure of the portfolio downside risk, namely:
$$\frac{E[\mu^{\text{out}}(x)] - r_f}{\sigma(\min\{(\mu^{\text{out}}(x) - r_f), 0\})}.$$
The larger is the value of the index, the better is the portfolio performance.

We note that, when a rolling time window scheme is adopted, the out-of-sample portfolio x is rebalanced in each successive in-sample period. Therefore, the above indices evaluate the global performance of a portfolio selection method by averaging over all the out-of-sample portfolios.

In addition, we include in the tables the following information:

5. *Mean number of assets* (MeanAssets): it is the mean of the numbers of assets selected in portfolio x in each in-sample period. We consider that an asset j is selected in portfolio x if $x_j \geq 0.01$, $j = 1, \ldots, n$.
6. *Mean time* (MeanTime): it is the mean of the CPU times for solving the considered model in each in-sample period. In the case of the heuristic procedure, it is the mean of the running times. In each in-sample period we set a time limit of 7200 s for solving the

[1] The datasets are available upon request.

Table 1
Out-of-sample risk ($\cdot 10^{-4}$) with a 12 weeks rebalancing when the expected return of the portfolio is set to μ_0.

	DJIA	EUROSTOXX	FTSE100	SP500
Markowitz	4.518	6.465	4.629	4.188
RMT	4.583	6.389	4.631	4.215
Power Mapping	4.375	6.398	4.398	4.052
Scenario Filtering				
$K = 1$	4.416	**5.424**	**4.280**	**3.976**
$K = 2$	**4.307**	5.627	4.491	4.233
$K = 3$	4.408	5.648	4.311	4.462
$K = 4$	4.328	5.909	4.342	4.521
$K = 5$	4.352	5.803	4.397	4.554

model. This time limit is reached by our MIQP model only in some instances of the SP500 dataset. In these cases in which the solver may not have been able to find an optimal solution within the time limit, we also provide a measure of the gap obtained.

7. *Mean gap* (MeanGap): it is the mean of the relative gaps, in percentage, in each in-sample period. As commented above, it only applies to our MIQP model in some instances of the SP500 dataset.
8. *Mean Relative Error* (MRE): it is the mean of the relative errors, in percentage, in each in-sample period. It only applies to our heuristic procedure as a measure of the relative difference between the optimal solution obtained by solving our MIQP model and the best heuristic solution found. Small MRE values indicate that the solutions provided by the heuristic are close to the corresponding optimal ones. For the SP500 dataset, MRE uses the best (possibly optimal) solutions found by our MIQP model within the time limit.

The models have been implemented in MATLAB R2018a and they make calls to XPRESS solver version 8.5 for solving the MIP problems. All experiments were run in a computer DellT5500 with a processor Intel(R) Xeon(R) with a CPU X5690 at 3.75 GHz and 48 GB of RAM memory.

We start by showing the potential of our approach for lowering the out-of-sample risk (i.e., V-Out). We will only use for this demonstration our MIQP model, as it corresponds to the exact implementation of our Scenario Filtering approach and given that it is enough for our illustrative purpose. Although similar results are obtained for our heuristic algorithm, a more detailed analysis of this method will be provided in the sequel. We show in Table 1 the portfolio out-of-sample risk regardless of the effect of the portfolio return. This allows to better evaluate only the realized risk of optimized portfolios. To do this, in all models we require that the portfolio expected return be exactly equal to μ_0 (see, e.g., Kondor et al., 2007). Table 1 shows the portfolio out-of-sample risk values for all the portfolio selection models and datasets considered. Best values are reported in bold.

From the above table there seems to be a slight preference for our model. In fact, on all datasets, there always exists a value of K by which we are able to produce portfolios with lower out-of-sample risk than the others. We observe that our MIQP model always provides the best realized risk values (in bold). We also remark that for three of the four datasets this result was obtained with $K = 1$. The choice $K = 1$ corresponds removing about only the 2% of the observations in each in-sample period. This shows the impact of extreme observations in the data.

On the other hand, the classical Markowitz model, from which all the approaches presented in this paper originate, is a bi-criteria optimization model, and, therefore, the comparison must be performed on the basis of both values of realized risk and return. To this aim, Tables 2–5 report the complete out-of-sample analysis based on all the performance measures introduced at the beginning of this section. In all the models the portfolio expected return constraint is modeled as an inequality, so that, it is possible to jointly assess the effect of return and risk in the selected portfolios. For the sake of completeness, in the following tables and for each dataset we also include the values of the performance measures of the market portfolio, labeled as "Market".

In Tables 2–5 best values are in bold. We also underline the best values of the performance measures when the comparison is restricted only to the filtering methods.

We observe that our MIQP model always provides portfolios having (on average) the best out-of-sample performance in terms of return. On the other hand, except for the FTSE100 dataset, the realized risk of our portfolios is worse than the realized risk provided by the portfolios found by the two alternative filtering models. However, in a bi-criteria framework and from an investor viewpoint, the indices which evaluate the compromise between the return of a portfolio and the risk that the investor is affording are the most significant. In our experiments, the Sharpe and the Sortino ratios of our portfolios outperform those of the Power Mapping and RMT models (see underlined values) for all the datasets. Here one has to remember that in our approach K is treated as a decision variable, and the model is solved for each of the different values $K = 1, 2, 3, 4, 5$ with the precise aim of finding the best value for K. Therefore, in Tables 2–5 the final result of our method should be read as the best among those reported in the five rows of the "Scenario Filtering" block.

It can be observed in the Tables that the Market portfolio performance is generally worse than all the other portfolios. In particular, this confirms that filtering portfolios — with any method — is an effective tool to improve the portfolio realized return and risk.

About the "MeanAssets" column, in Tables 2–5 we also note that the average number of selected stocks provided by our approach is the same as the one provided by the classical Markowitz model, while the RMT and Power Mapping models tend to select slightly more assets than our model, especially for large datasets. Limiting the number of selected stocks is often a requirement that come from real-world practice where the administration of a portfolio made up of a large number of assets, possibly with very small holdings for some of them, is clearly not desirable because of transactions costs, minimum lot sizes, complexity of management, or due to specific policies of the asset management companies.

As observed, and as evident from the above tables, our MIQP model is hard to solve up to proven optimality especially for large financial datasets. In general, computational times grow w.r.t. to the number of assets and the parameter K. From a practitioner viewpoint, there is the need of computing portfolios having a good out-of-sample performance without too much waste of time. Hence, in the following Tables 6–9 we report the same experimental analysis by applying the heuristic procedures introduced in Section 3.3. Regarding the two versions of the heuristic described in Section 3.3, the CPU times reported in the tables have been obtained by implementing the Version 1 for DJIA and EUROSTOXX50 datasets, and the Version 2 for FTSE100 and SP500. Version 2 is therefore more suitable for large datasets in our setting.

We can see in Tables 6–9 that the performance results of our heuristic are in line with the ones obtained for the exact model. As expected, the computational times have been dramatically reduced while the relative error is on average no larger than the 1%. This shows the effectiveness of the heuristic procedure. We also note that in Table 9 for $K = 5$ the heuristic was able to find solutions with a better value than the values provided by the exact model (see the negative MRE), pointing out that the larger the value of parameter K, the harder is to find an optimal solution with the exact approach. Indeed, the computational times needed to solve the exact model grow exponentially with K, while this growth is linear in the case of the heuristic. The CPU time needed to run the heuristic seems reasonable (less than one minute on average), and indeed it is lower than the one required by the Power Mapping filtering technique in the case of

Table 2
Out-of-sample performances for DJIA ($n = 28$) with a 12 weeks rebalancing.

	AvReturn ($\cdot 10^{-3}$)	V-Out ($\cdot 10^{-4}$)	Sharpe ($\cdot 10^{-2}$)	Sortino	MeanAssets	MeanTime (sec.)
Market	1.657	5.460	7.090	0.117	28	–
Markowitz	1.908	4.036	9.500	0.157	9.9	0.031
RMT	1.842	4.039	9.164	0.152	9.7	0.038
Power Mapping	1.930	**3.893**	9.780	0.163	11.9	0.035
Scenario Filtering						
$K = 1$	1.915	4.136	9.418	0.158	9.8	1.969
$K = 2$	1.907	4.096	9.420	0.157	9.7	2.806
$K = 3$	1.888	4.099	9.325	0.156	9.8	7.033
$K = 4$	1.924	4.166	9.428	0.158	9.8	22.402
$K = 5$	**2.032**	4.245	**9.864**	**0.167**	9.8	63.690

Table 3
Out-of-sample performances for EUROSTOXX ($n = 49$) with a 12 weeks rebalancing.

	AvReturn ($\cdot 10^{-3}$)	V-Out ($\cdot 10^{-4}$)	Sharpe ($\cdot 10^{-2}$)	Sortino	MeanAssets	MeanTime (sec.)
Market	0.799	8.732	2.703	0.043	49	–
Markowitz	1.778	4.902	**8.029**	**0.124**	8.9	0.036
RMT	1.666	4.899	7.527	0.115	9.1	0.041
Power Mapping	1.705	**4.785**	7.793	0.120	11.1	0.038
Scenario Filtering						
$K = 1$	1.724	5.250	7.526	0.116	9.6	3.253
$K = 2$	1.810	5.388	7.798	0.121	9.7	4.502
$K = 3$	1.790	5.500	7.633	0.117	9.9	7.951
$K = 4$	1.840	5.546	7.812	0.120	9.6	21.540
$K = 5$	**1.887**	5.606	**7.971**	**0.123**	9.9	61.881

Table 4
Out-of-sample performances for FTSE100 ($n = 83$) with a 12 weeks rebalancing.

	AvReturn ($\cdot 10^{-3}$)	V-Out ($\cdot 10^{-4}$)	Sharpe ($\cdot 10^{-2}$)	Sortino	MeanAssets	MeanTime (sec.)
Market	0.762	6.656	2.954	0.047	83	–
Markowitz	1.852	4.339	8.892	0.144	12.7	0.042
RMT	1.577	4.333	7.578	0.121	14.4	0.046
Power Mapping	1.749	4.091	8.646	0.139	16.4	0.047
Scenario Filtering						
$K = 1$	**2.406**	4.015	**12.001**	**0.197**	12.7	8.365
$K = 2$	2.152	**3.944**	10.840	0.175	13.2	13.877
$K = 3$	1.905	4.176	9.324	0.149	13.3	31.997
$K = 4$	1.873	4.134	9.212	0.149	13.1	120.866
$K = 5$	1.837	4.096	9.075	0.146	13.4	367.511

Table 5
Out-of-sample performances for SP500 ($n = 442$) with a 12 weeks rebalancing.

	AvReturn ($\cdot 10^{-3}$)	V-Out ($\cdot 10^{-4}$)	Sharpe ($\cdot 10^{-2}$)	Sortino	MeanAssets	MeanTime (sec.)	MeanGap (%)
Market	1.292	7.544	4.704	0.076	442	–	–
Markowitz	1.560	3.603	8.220	0.127	16.4	0.230	–
RMT	1.619	**3.347**	8.847	0.140	21.2	0.263	–
Power Mapping	1.506	3.398	8.170	0.129	21.8	69.195	–
Scenario Filtering							
$K = 1$	1.273	3.706	6.613	0.103	17.6	131.609	0
$K = 2$	1.161	3.766	5.981	0.093	17.4	319.337	0
$K = 3$	1.320	3.924	6.662	0.105	17.2	1055.618	0
$K = 4$	1.455	4.165	7.132	0.112	17.9	4106.864	15.488
$K = 5$	**2.016**	4.302	**9.720**	**0.156**	18.1	6174.948	58.857

Table 9. We conjecture that this increase in the computational time of the Power Mapping method is due to the fact that the correlation matrix after the transformation may be indefinite. To conclude this analysis of our heuristic, we note that the performance measures of the heuristic in Tables 6–9 sometimes improved the corresponding results of the exact method in Tables 2–5. The possible capacity of the heuristic of avoiding *overfitting* effects explains this fact.

Finally, to emphasize the out-of-sample performance of our approach, in the following figures we show the weekly out-of-sample portfolio values. For each dataset, K is the number of filtered observations corresponding to the values in bold in column "AvReturn" in Tables 2–5. The red and dark blue lines report the weekly values of the portfolios obtained with our MIQP model ("Scenario Filtering") and our heuristic algorithm, respectively. Note that in Fig. 1(c) the red

Table 6
Out-of-sample performances for DJIA ($n = 28$) with a 12 weeks rebalancing applying the heuristic procedure.

	AvReturn ($\cdot 10^{-3}$)	V-Out ($\cdot 10^{-4}$)	Sharpe ($\cdot 10^{-2}$)	Sortino	MeanAssets	MeanTime (sec.)	MRE (%)
Market	1.657	5.460	7.090	0.117	28	–	–
Markowitz	1.908	4.036	9.500	0.157	9.9	0.031	–
RMT	1.842	4.039	9.164	0.152	9.7	0.038	–
Power Mapping	1.930	**3.893**	9.780	0.163	11.9	0.035	–
Heuristic							
$K = 1$	1.915	4.136	9.418	0.158	9.8	1.985	0
$K = 2$	1.897	4.103	9.367	0.156	9.7	4.206	0.073
$K = 3$	1.957	4.106	9.658	0.161	9.8	6.557	0.294
$K = 4$	1.980	4.176	9.690	0.162	9.8	8.870	0.453
$K = 5$	**2.069**	4.245	<u>10.043</u>	<u>0.169</u>	9.8	11.155	0.575

Table 7
Out-of-sample performances for EUROSTOXX ($n = 49$) with a 12 weeks rebalancing applying the heuristic procedure.

	AvReturn ($\cdot 10^{-3}$)	V-Out ($\cdot 10^{-4}$)	Sharpe ($\cdot 10^{-2}$)	Sortino	MeanAssets	MeanTime (sec.)	MRE (%)
Market	0.799	8.732	2.703	0.043	49	–	–
Markowitz	1.778	4.902	**8.029**	**0.124**	8.9	0.036	–
RMT	1.666	4.899	7.527	0.115	9.1	0.041	–
Power Mapping	1.705	**4.785**	7.793	0.120	11.1	0.038	–
Heuristic							
$K = 1$	1.724	5.250	7.526	0.116	9.6	3.259	0
$K = 2$	1.810	5.388	7.798	0.121	9.7	7.303	0
$K = 3$	1.803	5.502	7.686	0.118	9.9	11.711	0.011
$K = 4$	**1.862**	5.541	<u>7.909</u>	<u>0.122</u>	9.8	16.110	0.160
$K = 5$	1.840	5.620	7.763	0.119	9.9	20.493	0.247

Table 8
Out-of-sample performances for FTSE100 ($n = 83$) with a 12 weeks rebalancing applying the heuristic procedure.

	AvReturn ($\cdot 10^{-3}$)	V-Out ($\cdot 10^{-4}$)	Sharpe ($\cdot 10^{-2}$)	Sortino	MeanAssets	MeanTime (sec.)	MRE (%)
Market	0.762	6.656	2.954	0.047	83	–	–
Markowitz	1.852	4.339	8.892	0.144	12.7	0.042	–
RMT	1.577	4.333	7.578	0.121	14.4	0.046	–
Power Mapping	1.749	4.091	8.646	0.139	16.4	0.047	–
Heuristic							
$K = 1$	**2.406**	4.015	**<u>12.008</u>**	**<u>0.197</u>**	12.7	7.939	0
$K = 2$	2.149	**3.951**	10.811	0.175	13.3	15.569	0.074
$K = 3$	1.876	4.176	9.179	0.147	13.2	22.999	0.206
$K = 4$	1.899	4.147	9.325	0.151	13.0	30.215	0.465
$K = 5$	1.992	4.107	9.830	0.159	13.2	37.339	0.498

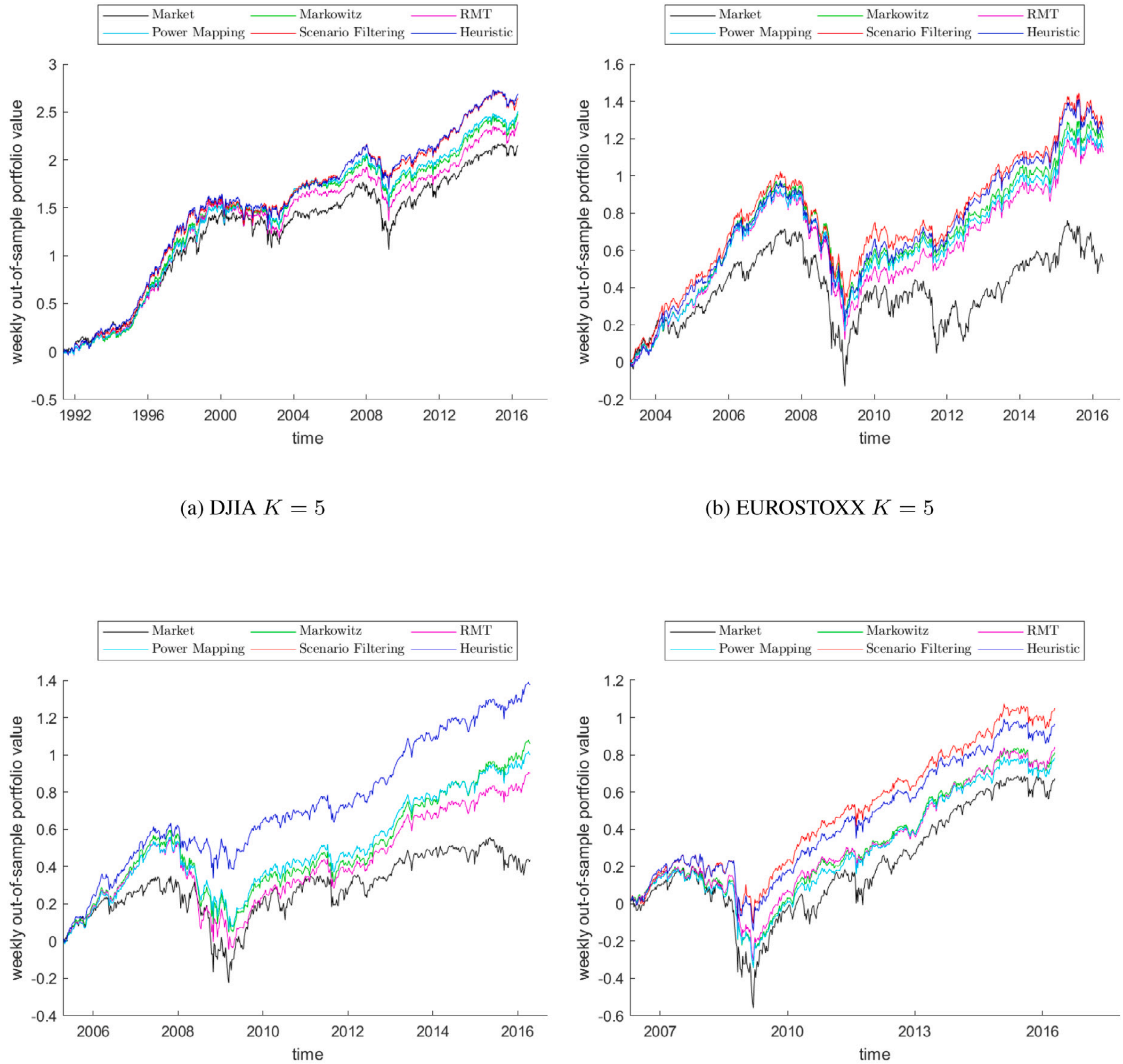
Table 9
Out-of-sample performances for SP500 ($n = 442$) with a 12 weeks rebalancing applying the heuristic procedure.

	AvReturn ($\cdot 10^{-3}$)	V-Out ($\cdot 10^{-4}$)	Sharpe ($\cdot 10^{-2}$)	Sortino	MeanAssets	MeanTime (sec.)	MRE (%)
Market	1.292	7.544	4.704	0.076	442	–	–
Markowitz	1.560	3.603	8.220	0.127	16.4	0.230	–
RMT	1.619	**3.347**	8.847	0.140	21.2	0.263	–
Power Mapping	1.506	3.398	8.170	0.129	21.8	69.195	–
Heuristic							
$K = 1$	1.273	3.707	6.611	0.103	17.6	9.018	0
$K = 2$	1.202	3.828	6.144	0.096	17.4	17.641	0.113
$K = 3$	1.364	4.008	6.814	0.107	17.3	26.040	0.097
$K = 4$	1.472	4.250	7.142	0.112	17.8	34.249	0.518
$K = 5$	**1.860**	4.265	**<u>9.006</u>**	**<u>0.143</u>**	18.1	42.331	−0.743

and dark blue lines coincide since, as pointed out in Section 3.3, the solutions of the exact and heuristic methods coincide for $K = 1$. From our computational experiments, one can conclude that our exact and heuristic algorithms outperform the alternative filtering methods on the analyzed real-world datasets.

4.1. Daily prices dataset

In order to validate our approach also in a set-up with high-frequency observations, we replicate the analysis above using daily prices data (see e.g. Jegadeesh and Titman, 2001; Onnela et al., 2004; Tola et al., 2008). In particular we consider the following dataset that was used for the first time in Scozzari (2021):



(a) DJIA $K = 5$ (b) EUROSTOXX $K = 5$

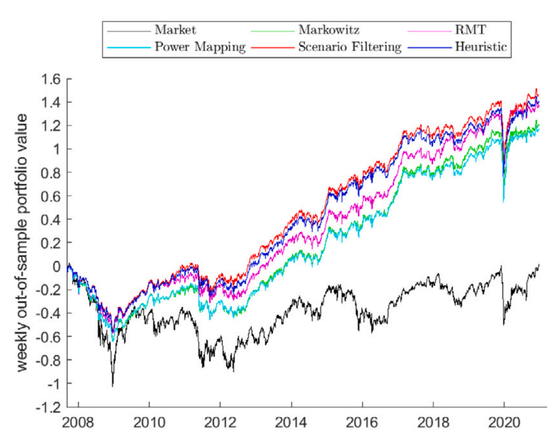
(c) FTSE100 $K = 1$ (d) SP500 $K = 5$

Fig. 1. Weekly out-of-sample portfolio values for the four weekly prices datasets.

1. FTSEMIB (Financial Times Stock Exchange Milano Indice Borsa, Italy), containing 23 assets and 3432 daily price observations (period: 03/09/2007–12/03/2021).

Similar to the analysis made for the four weekly prices datasets, we use a rolling time window scheme for the FTSEMIB dataset adopting a period of 52 days as in-sample window and we consider 12 days as out-of-sample, while the rebalancing is performed every 12 days. Table 10 reports the out-of-sample analysis, based on the previous considered performance measures, of the different exact methods. In Table 11 it is also shown the results obtained with the heuristic procedure. The daily out-of-sample portfolio values obtained with the different methods is shown in Fig. 2, where K is chosen as the number of filtered observations that provides the best value (in bold) in the column "AvReturn" in Table 10.

The results shown in Table 10, Table 11 and Fig. 2, are in line with the ones obtained considering weekly prices observations, so our approach is also valid when high-frequency observations are considered.



FTSEMIB $K = 3$

Fig. 2. Daily out-of-sample portfolio values for the FTSEMIB dataset.

Table 10
Out-of-sample performances for FTSEMIB ($n = 23$) with a 12 days rebalancing.

	AvReturn ($\cdot 10^{-3}$)	V-Out ($\cdot 10^{-4}$)	Sharpe ($\cdot 10^{-2}$)	Sortino	MeanAssets	MeanTime (sec.)
Market	0.004	2.845	0.023	0.001	23	–
Markowitz	0.353	1.270	3.129	0.048	7.6	0.033
RMT	0.401	1.304	3.508	0.054	7.0	0.053
Power Mapping	0.343	1.261	3.058	0.046	9.0	0.059
Scenario Filtering						
$K = 1$	0.419	**1.242**	3.763	**0.058**	7.6	1.847
$K = 2$	0.393	1.292	3.456	0.053	7.6	2.452
$K = 3$	**0.432**	1.267	**3.836**	**0.058**	7.7	4.033
$K = 4$	0.406	1.279	3.591	0.055	7.7	9.016
$K = 5$	0.362	1.288	3.192	0.049	7.6	21.654

Table 11
Out-of-sample performances for FTSEMIB ($n = 23$) with a 12 days rebalancing applying the heuristic procedure.

	AvReturn ($\cdot 10^{-3}$)	V-Out ($\cdot 10^{-4}$)	Sharpe ($\cdot 10^{-2}$)	Sortino	MeanAssets	MeanTime (sec.)	MRE (%)
Market	0.004	2.845	0.023	0.001	23	–	–
Markowitz	0.353	1.270	3.129	0.048	7.6	0.033	–
RMT	0.401	1.304	3.508	0.054	7.0	0.053	–
Power Mapping	0.343	1.261	3.058	0.046	9.0	0.059	–
Heuristic							
$K = 1$	0.419	**1.242**	**3.763**	**0.058**	7.6	1.836	0
$K = 2$	**0.428**	1.293	3.759	**0.058**	7.6	3.752	0.088
$K = 3$	0.413	1.275	3.660	0.056	7.7	5.743	0.113
$K = 4$	0.393	1.283	3.471	0.053	7.7	7.797	0.242
$K = 5$	0.368	1.290	3.237	0.049	7.7	9.838	0.410

5. Conclusions

Due to the presence of outliers in the distribution of assets' returns, covariance matrices typically incorporate a huge amount of noise. Therefore, their use in portfolio selection may be misleading. Several methods have been presented in the literature to cope with this problem ranging from extending the Gaussian model incorporating methods for handling outliers to models based on the minimization of downside risk measures.

In this paper we propose a new approach based on Quadratic Programming that fits in the MVO framework, which is still at the basis of the most popular portfolio selection models used by practitioners. We provide a new MIQP model and apply it to some real-world financial datasets. We show that it is able to eliminate outliers in order to lower the in-sample variance and obtaining a good out-of-sample performance of the portfolios. We compare our approach with some popular outliers filtering procedures provided in the literature. In order to solve large size financial datasets we also provide a heuristic procedure based on the same MIQP. We show that our new approach for filtering is effective in hitting the goal of eliminating noise in the observed data. From a computational viewpoint, the two possible (exact and heuristic) strategies are able to find optimal or near optimal solutions in reasonable times.

Finally, under some assumptions on the tolerance level β, we show that the CVaR_β of a portfolio can be interpreted as the minimum first moment of a portfolio rate of return, while the optimal value of our filtered variance corresponds to the minimum second central moment of a portfolio rate of return. These observations lead one to consider how it may be possible to model higher order minimum filtered moments like filtered skewness and filtered kurtosis, which is worth investigating in future lines of research.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research has been partially supported by Spanish Ministry of Education and Science/FEDER grant number PID2020-114594GB-(01-02), and projects FEDER-US-1256951, CEI-3-FQM331, P18-FR-1422 and *NetmeetData*: Ayudas Fundación BBVA a equipos de investigación científica 2019.

References

Best, M., Grauer, R., 1991a. The analytics of sensitivity analysis for mean variance portfolio problems. Int. Rev. Financial Stud. 4, 315–342.
Best, M., Grauer, R., 1991b. On the sensitivity of mean–variance-efficient portfolios to changes in asset means: Some analytical and computational results. Rev. Financ. Stud. 4, 315–342.
Bienstock, D., 1996. Computational study of a family of mixed-integer quadratic programming problems. Math. Program. 74, 121–140.
Bruni, R., Cesarone, F., Scozzari, A., Tardella, A., 2017. On exact and approximate stochastic dominance strategies for portfolio selection. European J. Oper. Res. 259, 322–329.
Cesarone, F., Scozzari, A., Tardella, F., 2013. A new method for mean–variance portfolio optimization with cardinality constraints. Ann. Oper. Res. 205, 213–234.
El Karoui, N., 2013. On the realized risk of high-dimensional Markowitz portfolios. SIAM J. Financial Math. 4, 737–783.
Elton, E.J., Gruber, M.J., 1995. Modern Portfolio Theory and Investment Analysis. John Wiley & Sons.
Elton, E.J., Gruber, M.J., Spitzer, J., 2006. Improved estimates of correlation coefficients and their impact on optimum portfolios. Eur. Financial Manag. 12, 303–318.
Fama, E., 1965. The behavior of stock-market prices. J. Bus. 38, 34–105.
Floudas, C.A., Visweswaran, V., 1995. Quadratic optimization. In: Handbook of Global Optimization. Springer, Boston, MA, pp. 217–269.
Garey, M.R., Johnson, D.S., 1979. Computers and Intractability, Vol. 174. freeman., San Francisco.
Guhr, T., Kälber, B., 2003. A new method to estimate noise in financial correlation matrices. J. Phys. A 36, 3009–3032.
Jansen, D.W., de Vries, C.G., 1991. On the frequency of large stock returns: putting booms and busts into perspective. Rev. Econ. Stat. 73, 18–24.
Jegadeesh, N., Titman, S., 2001. Profitability of momentum strategies: An evaluation of alternative explanations. J. Finance 56, 699–720.
Jorion, P., 1986. Bayes-Stein estimation for portfolio analysis. J. Financ. Quant. Anal. 21, 279–292.

Kolm, P.N., Tutuncu, R., Fabozzi, F.F., 2014. 60 Years of portfolio optimization: Practical challenges and current trends. European J. Oper. Res. 234, 356–371.
Kondor, I., Pafka, S., Nagy, G., 2007. Noise sensitivity of portfolio selection under various risk measures. J. Bank. Financ. 31, 1545–1573.
Laloux, L., Cizeau, P., Bouchaud, J.P., Potters, M., 2000. Random matrix theory and financial correlations. Int. J. Theor. Appl. Finance 3, 391–397.
Ledoit, O., Wolf, M., 2003. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. J. Empir. Financ. 10, 603–621.
Mandelbrot, B.B., 1963. The variation of certain speculative prices. J. Bus. 36, 392–417.
Mansini, R., Ogryczak, W., Speranza, M.G., 2007. Conditional value at risk and related linear programming models for portfolio optimization. Ann. Oper. Res. 152, 227–256.
Mansini, R., Ogryczak, W., Speranza, M.G., 2015. EURO: The Association of European Operational Research Societies: Linear and Mixed Integer Programming for Portfolio Optimization. Springer, Cham, Switzerland.
Markowitz, H., 1952. Portfolio selection. J. Finance 7, 77–91.
McCormick, G.P., 1976. Computability of global solutions to factorable nonconvex programs: Part I – convex underestimating problems. Math. Program. 10, 147–175.
Nemhauser, G.L., Wolsey, L.A., 1988. Integer and Combinatorial Optimization. Wiley.
Onnela, J.P., Kaski, K., Kertesz, J., 2004. Clustering and information in correlation based financial networks. Eur. Phys. J. B 38, 353–362.
Pafka, S., Kondor, I., 2002. Noisy covariance matrices and portfolio optimization. Eur. Phys. J. 27, 277–280.
Pafka, S., Kondor, I., 2003. Noisy covariance matrices and portfolio optimization II. Physica A 319, 487–494.
Pafka, S., Kondor, I., 2004. Estimated correlation matrices and portfolio optimization. Physica A 343, 623–634.
Puerto, J., Rodríguez-Madrena, M., Scozzari, A., 2020. Clustering and portfolio selection problems: A unified framework. Comput. Oper. Res. 117, 104891.
Rockafellar, R.T., Uryasev, S., 2002. Conditional value-at-risk for general loss distributions. J. Bank. Financ. 26, 1443–1471.
Schafer, R., Nilsson, N.F., Guhr, T., 2010. Power mapping with dynamical adjustment for improved portfolio optimization. Quant. Finance 10, 107–119.
Scozzari, A., 2021. A network-based analysis of the dynamics of the Italian stock market. Chaos Complex. Lett. 15, 35–53.
Sharpe, W.F, 1966. Mutual fund performance. J. Bus. 39, 119–138.
Sharpe, W.F., 1996. The sharpe ratio. J. Portfolio Manag. 21, 49–58.
Sortino, F.A., Satchell, S., 2001. Managing Downside Risk in Financial Markets. Batterworth-Heinemann Finance, Oxford.
Stoyanov, S.V., Rachev, S.T., Racheva-Yotova, B., Fabozzi, F.J., 2011. Fat-tailed models for risk estimation. J. Portfolio Manag. 37, 107–117.
Tola, V., Lillo, F., Gallegati, M., Mantegna, R.N., 2008. Cluster analysis for portfolio optimization. J. Econom. Dynam. Control 32, 235–258.
Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of small-worldnetworks. Nature 393 (6684), 440.
Wolsey, L.A., 1998. Integer Programming. John Wiley & Sons.