# A new model for robust portfolio optimization

**S. Benati and E. Conde***

February 4, 2020

**Abstract**

A *robust optimization* model to find an investment portfolio is analyzed according to twofold risk sources: the random nature of the returns given an economic scenario which is itself unknown. Our model combines measures of deviation, risk and regret to find a solution ensuring aceptable expected returns while we are hedged against the market *volatility*. Several mathematical formulations are stated and numerically tested. Using duality relations we obtain bounds on the optimal objective value of the problem. Furthermore, these bounds are integrated into an iterative numerical procedure, developed as an alternative to exact formulations. We check, by means of experimental financial data, that our solution deals with volatility and economic uncertainty by seeking more diversification while expected returns are preserved.

## 1 Introduction

The choice of an appropriate investment portfolio is a challenge involving multiple uncontrollable elements. Financial markets are inherently uncertain environments where prices, interest rates or availability of resources define economic factors that are usually modeled through *random variables*. Here, we can distinguish, among others, two main sources of uncertainty. One is linked to the lack of information about the own nature of the buying/selling operations of securities in financial systems. The other one is related to the *economic scenario* (economic growth, recession, military conflicts, trade war, ...) under which these operations take place.

The research activities have been supported by projects FQM-331 (Junta de Andalucía) and MTM2016-74983-C2-1-R (MICINN, Spain)

*Department of Statistic and Operations Research
Faculty of Mathematics. Campus Reina Mercedes, 41011.
University of Seville. Seville. Spain.
E-mail: educon@us.es

In the literature, the problem of finding *good* portfolio policies has been deeply analyzed in terms of a *corpus* of properties characterizing risk measures (see e.g. [?]), properties of coherence and compatibility with stochastic dominance principles (see e.g. [?,?,?]) or theory of choice principles (see e.g. [?]).

Most of the existing approaches (see, e.g. [?,?,?,?,?,?,?]) need the explicit knowledge of probability distributions of the involved economic factors, which is, by itself, a challenging task. Furthermore, fixing a specific probability distribution for the involved economic factors means that the conditions defining the economic scenario have also been fixed, discarding the above mentioned second source of uncertainty. An *optimal* investment under these conditions may not be appropriate since the decision maker remains exposed to any change in the economic behavior of the financial system.

Some existing optimization models address this situation by defining a family of *possible* probability distributions and seeking a compromise solution with a good performance whatever the distribution is chosen. These are the so-called *robust portfolios* [?,?,?]. In this case, different probability distributions model different economic scenarios and a solution is found in order to maximize its performance under each considered scenario, sometimes measuring the discrepancy between its performance and the optimal one under each scenario, what is called *regret*.

In this paper, we propose a *robust optimization* model in order to find an investment portfolio by analyzing its twofold risk: the random nature of the returns given an economic scenario and the fact that this economic scenario is itself unknown. For the first task the average return and the *conditional Value-at-Risk*, [?], are used, while for the second one the minimization of the maximum regret respect to the optimal performance under each scenario is considered. In this way, our model combines several measures of deviation, risk and regret to find a feasible portfolio ensuring certain level of expected returns while decision makers are hedged against the market *volatility* represented by a set of possible economic scenarios.

In our approach, we avoid using theoretical probability distributions by just working with the empirical distributions obtained from data. Hence, we will assume that a set of historical data of economic indicators is available. For instance, a vector of intra-day price variations for a given set of assets. Usually, this data will not represent a numerical sample of our vector of indicators, in the sense of being independent and identically distributed realizations of that vector. Although the independence could be assumed since this data represents intra-day variations, the economic conditions, that is, the underlying probability distributions, usually change over time, reflecting different market behaviors such as expansion, peak or contraction periods. We will assume that our data-set is the union of *subsets of observations* -identified with scenarios- obtained during specific time periods (typically *time windows*). If these time windows can be associated to different market conditions, with possibly the same underlying probability distribution, our empirical measures (sample mean, variance,...) will approach the corresponding theoretical measures (under known distributions). On the other hand, if there is not a direct relation amongst scenarios and different economic behaviours, these time windows may still capture different economic patterns or changes in the economic cycle of the market against whose adverse effects, the decision maker may want be hedged.

The idea of considering a family of subsets of the historical data set according to scenarios to model the behavior of a system has been used recently, [?], in the context of a dynamic pricing problem. In this case, a firm wants to determine a dynamic pricing strategy under uncertainty about the demand for the product. In that paper, the authors shown by means of a numerical experiment that *regret-based optimization* models deliver a more *robust* revenue performance as compared to the sample-average approximation and also in comparison with a Bayesian approach. In fact, the sample-average approximation only had a better average revenue performance when a large sample of data was provided from just one underlying distribution which will not be the case in our model where the underlying economic conditions of the market are assumed to change over time according to different scenarios.

The paper is structured as follows. In Section **??**, we discuss the properties of the conditional value at risk when estimated from historical data. As will be seen, our definition differs slightly from the established literature and therefore some of its properties must be reassessed. In Section **??**, we show that the Mean-CVaR linear model can be solved by its dual reformulation, or even by constraint generation. Those two methods have been largely overlooked when one deals with a single period problem, but are useful when the model contains scenarios, as can be seen in Section **??**. There, we formulate three alternative linear programming approaches to the min-max regret portfolio. The first one requires knowledge about the optimal portfolio under each considered scenario whilst the second approach avoid the need for that by using valid dual relations of the nominal optimization problem. Finally, the third model formulates the restriction linked to the conditional value at risk through an exponential number of constraints. They are the basis of an iterative procedure in which only a few of these constraints are generated and used through a separation subroutine.

Some numerical properties and bounds on the optimal minmax regret objective value are presented in Section **??**. The bounding procedure will be a part of the numerical procedure proposed in this section to solve our problem. This procedure solves a sequence of relaxed formulations of moderate size in order to approach the optimal solution. In Section **??** a numerical experiment is conducted in order to compare the iterative procedure with the resolution of different formulations of the problem. Financial features of our robust portfolio are also analyzed revealing the strength of potential applications. In particular, we find that our solution represents investments, more conservative than the ones obtained by considering single-scenario models, but also more diversified. This may set out the rules for less risky investments, with less turnover than the standard ones, while the ex-post returns is still maintained to the same level, or even higher than those obtained from single-scenario models. Finally, in Section **??**, we suggest some guidelines to improve the performance of our solution in practical applications and give possible ideas for future contributions.

## 2 Sample conditional value at risk

The conditional value at risk (also known in finance as expected shortfall [?,?]), $\mathrm{CVaR}_s(x)$ for a given portfolio $x \in X$ (and scenario $s \in S$) is defined by Rockafellar and Uryasev, [?], as the conditional expectation of losses above the $\beta$-quantile

$\alpha_\beta(x)$ of the losses. In order to define properly this value for a sample of returns we will introduce some notation:

- A set of time periods $T$ and a set of assets $A$.
- A data set $R = [r_{ta}; t \in T; a \in A]$, in which $r_{ta}$ is the return of the asset $a$ at the time period $t$.
- A set of possible scenarios $S$, where each scenario $s \in S$ is defined as a subset of time periods $T_s \subseteq T$.
- The portfolio vector $x = [x_a; a \in A]$, where $x_a$ is the proportion of wealth invested in the asset $a$.
- A feasible set of portfolios $X$ which is supposed to be independent of the considered scenario.

Our optimization model is built on the basis of the following two measures:

- $E_s(x)$, the average reward of the portfolio $x$ under the scenario $s \in S$. This measure is computed using the vectors of returns $r_{ta}, a \in A$ corresponding to every time period $t \in T_s$ and represents an estimator of the expected reward of the portfolio under the considered scenario.
- $\text{CVaR}_s(x)$, the conditional value-at-risk of the portfolio $x$, a risk measure computed as the average of a given proportion of the largest losses borne by the investor.

The first measure is defined in terms of an average of rewards and the second one, in terms of an average of losses. Hereafter, we will use both terms interchangeably, where losses are negative rewards.

In [**?**], Rockafellar and Uryasev assumed that a absolutely continuous probability distribution of losses is known and shown several properties of CVaR under these conditions. In [**?**] the maximization of the portfolio return subject to a constraint on the CVaR is proposed for a known discrete distribution. Here, we follow the same idea and state the CVaR in terms of a sample of historical data.

First, the $\beta$-quantile of the sample of losses for a given portfolio is computed, that is, given the sample

$$\left\{ -\sum_{a \in A} r_{ta} x_a \,:\, t \in T_s \right\}. \tag{1}$$

the corresponding $\beta$-quantile for the scenario $s \in S$ is

$$\alpha^s_\beta(x) = \text{argmin}\left\{ \alpha : \left| \{ t \in T_s : -\sum_{a \in A} r_{ta} x_a \leq \alpha \} \right| \geq \lceil \beta |T_s| \rceil \right\}, \tag{2}$$

where $|\bullet|$ stands for the cardinality of a set. The sample $\beta$-quantile $\alpha^s_\beta(x)$ is strongly consistent respect to the $\beta$-quantile of the underlying unknown probability distribution under mild conditions [**?**]. Here, we will use the value $\alpha^s_\beta(x)$ as a cutpoint dividing the sample of ordered losses (**??**) into two sets,

- the first of them corresponding to the subset $T_s^-$ of the $\lceil \beta |T_s| \rceil$ time periods corresponding to the smallest losses and
- its complementary, $T_s^+$ with the $\lfloor (1-\beta)|T_s| \rfloor$ time periods of the highest losses.

Note that, in the presence of repeated values of the same losses, one or several of the repeated values could be in $T_s^-$ and the others, in $T_s^+$.

In the continuous case, this quantile is also called $\beta -$ VaR, Value at Risk, [?], where $\beta \in (0, 1)$ is a parameter defined by the investor according to her attitude to risk (*risk level*). The larger the $\beta$-level is, the more aversion to risk the decision maker has.

According to this definition, the average loss in the latter set $T_s^+$ is what we will call the conditional value at risk for the portfolio $x$ under the scenario $s \in S$,

$$\mathrm{CVaR}_s(x) = -\frac{1}{\lfloor (1-\beta)|T_s| \rfloor} \sum_{t \in T_s^+} \sum_{a \in A} r_{ta} x_a. \tag{3}$$

The right-hand side of equation (**??**) represents a conditional mean for the sample of losses under scenario $s \in S$. Taking into account that $\alpha_\beta^s(x)$ is the $\lceil \beta |T_s| \rceil$ order statistic of that sample of losses, its strong consistency ([**?**]) respect to the $\beta$-quantile of the underlying unknown probability distribution of losses, ensures this same convergence for CVaR respect to the true conditional mean of losses under mild conditions on such an unknown distribution (for instance, absolute continuity).

Following Rockafellar and Uryasev, [**?**] we can state some useful properties of $\alpha_\beta^s(x)$ and $\mathrm{CVaR}_s(x)$ respect to the function

$$F_\beta(x, \alpha) = \alpha + \frac{1}{\lfloor (1-\beta)|T_s| \rfloor} \sum_{t \in T_s} [-\sum_{a \in A} r_{ta} x_a - \alpha]^+, \tag{4}$$

where $[a]^+ = \max\{a, 0\}$.

Let us observe here that $F_\beta(x, \alpha)$ is not exactly the same function defined in (27) of [**?**]. They both differ if $\lfloor (1-\beta)|T_s| \rfloor \neq (1-\beta)|T_s|$. This is a consequence of the definition of CVaR in (**??**) which does not coincide with expression (25) of [**?**] for the CVaR of a discrete distribution of losses. Neither does it coincide with $\mathrm{CVaR}^+$ (*upper CVaR*) nor with $\mathrm{CVaR}^-$ (*tail VaR* or *lower CVaR*) as defined in [**?**] (Definition 4), which are alternative definitions to CVaR in discrete distributions, the so-called *scenario model* following [**?**]. Moreover we will maintain the expression (**??**) since with this slight modification all the losses above $\beta$-VaR are weighted with the same value (conditional probability) which is not true using the expression (25) of [**?**] according to their definition of the conditional distribution function (eq. (8) of [**?**]). This property is needed in order to ensure that the minimum value of $F_\beta(x, \alpha)$ for a fixed $x$ is reached at $\mathrm{CVaR}_s(x)$. Moreover, the expression (**??**) preserves the same property shown in Theorem 10 of [**?**] which is called there the *fundamental minimization formula* stated as the following

**Proposition 1** *The function $F_\beta(x, \alpha)$ is a convex piecewise linear function of $(x, \alpha)$ verifying*

$$CVaR_s(x) = F_\beta(x, \alpha_\beta^s(x)) = \min_\alpha F_\beta(x, \alpha).$$

*Proof* From the expression (**??**), it directly follows that $F_\beta(x, \alpha)$ is a convex piecewise linear function of $(x, \alpha)$ for all $\beta \in (0, 1)$.

Let $\alpha < \alpha^s_\beta(x)$, by its definition (**??**), $\left|\{t \in T_s : -\sum_{a\in A} r_{ta}x_a \le \alpha\}\right| < \lceil \beta |T_s| \rceil$ which implies that $\left|\{t \in T_s : -\sum_{a\in A} r_{ta}x_a > \alpha\}\right| > |T_s| - \lceil \beta|T_s|\rceil = \lfloor (1-\beta)|T_s| \rfloor$, that is, $F_\beta(x, \bullet)$ is strictly decreasing in the linear piece corresponding to $\alpha$.

On the other hand, if $\alpha \ge \alpha^s_\beta(x)$, one has that $\left|\{t \in T_s : -\sum_{a\in A} r_{ta}x_a > \alpha\}\right| \le \left|\{t \in T_s : -\sum_{a\in A} r_{ta}x_a > \alpha^s_\beta(x)\}\right| = |T_s| - \left|\{t \in T_s : -\sum_{a\in A} r_{ta}x_a \le \alpha^s_\beta(x)\}\right| \le \lfloor (1-\beta)|T_s| \rfloor$, that is, $F_\beta(x, \bullet)$ is non-decreasing in the linear piece corresponding to $\alpha$. Hence $F_\beta(x, \alpha^s_\beta(x)) = \min_\alpha F_\beta(x, \alpha)$, in fact, if there is an interval where this minimum is reached, $\alpha^s_\beta(x)$ would be the lower bound of this interval. □

Another consequence of the definition of CVaR as it is done in (**??**) is that it becomes a coherent measure whereas CVaR$^-$ and CVaR$^+$ are not, [**?**].

Let us denote by $y$ the uniform discrete random variable of losses corresponding to the choice of a given portfolio $x$ under the scenario $s \in S$, that is, $y$ takes the value $y_t = -\sum_{a\in A} r_{ta}x_a : t \in T_s$ with probability $1/|T_s|$. Let $z$ be the corresponding random variable of losses for another given portfolio $x'$ under the same scenario $s \in S$. In our context, the axioms in Artzner et al. [**?**] for coherence of a risk measure $\rho$ amount to the requirement that $\rho$ be sublinear,

$$\rho(y+z) \le \rho(y) + \rho(z) \quad \text{and} \quad \rho(\lambda z) = \lambda \rho(z), \qquad \text{for all } y, z \text{ and } \lambda \ge 0,$$

and in addition satisfy

$$\rho(y) = c, \qquad \text{when } y = c \text{ (constant)},$$

along with

$$\rho(y) \le \rho(z), \qquad \text{when } y \le z,$$

where the inequality $y \le z$ refers to first-order stochastic dominance.

**Proposition 2** *$\rho(x) = CVaR_s(x)$ defined by (**??**) is a coherent risk measure.*

*Proof* By definition, it is ease to see that CVaR can be written as the maximum of a Binary Linear Programming problem as follows

$$\rho(y) = \frac{1}{\lfloor (1-\beta)|T_s| \rfloor} \max_{\omega \in W} \sum_{t=1}^{|T_s|} y_t \omega_t \tag{5}$$

where $W = \{\omega \in \{0,1\}^{|T_s|} : \sum_{t=1}^{|T_s|} \omega_t = \lfloor (1-\beta)|T_s| \rfloor\}$.

Taking into account that

$$\max_{\omega \in W} \sum_{t=1}^{|T_s|} (y_t + z_t)\omega_t \le \max_{\omega \in W} \sum_{t=1}^{|T_s|} y_t\omega_t + \max_{\omega \in W} \sum_{t=1}^{|T_s|} z_t\omega_t,$$

one has $\rho(y+z) \le \rho(y) + \rho(z)$. Property $\rho(\lambda y) = \lambda \rho(y)$ for any $\lambda \ge 0$ follows directly from

$$\max_{\omega \in W} \sum_{t=1}^{|T_s|} \lambda y_t \omega_t = \lambda \max_{\omega \in W} \sum_{t=1}^{|T_s|} y_t \omega_t,$$

and when $y = c$ (constant) one has that $y_t = c$ for all $t \in |T_s|$ hence, from (**??**), $\rho(y) = c$.

Finally, the first-order stochastic dominance $y \leq z$ can be written in terms of the order statistics $y_{(t)} \leq z_{(t)}$, where $y_{(t)}$ is the loss occupying the $t$-th position in the sorted sample of losses in increasing order. Hence, the monotonicity of CVaR can be verified as follows

$$\rho(y) = \max_{\omega \in W} \sum_{t=1}^{|T_s|} \lambda y_{(t)} \omega_t \leq \lambda \max_{\omega \in W} \sum_{t=1}^{|T_s|} z_{(t)} \omega_t = \rho(z).$$

□

## 3 Optimizing the performance of the portfolio under controlled conditional value at risk

Properties as the one stated in Proposition **??**, in a more general context of probability distribution for losses, have been used in the literature (see e.g. [**?**,**?**]) to obtain the Linear Programming formulations of portfolio problems that can be efficiently solved using off-the-shelf optimization software. A generic problem including a constraint where the incurred conditional value at risk is upper bounded could have the following appearance

$$\begin{aligned} z_s^* = & \max p(s,x) \\ \text{s.t.} & \\ & \text{CVaR}_s(x) \leq c_s, \\ & x \in X. \end{aligned} \tag{6}$$

Here, we will assume that the performance of the portfolio $x \in X$, under a given scenario $s \in S$, is measured as $p(s,x)$. This measure can be an indicator of the efficiency of the portfolio or a measure of the risk of the investment that, consequently, should be minimized in order to hedge to the decision maker against the uncertain effects of the market.

Some reasonable choices of $p(s,x)$ in (**??**) could be the following ones:

– **maximization of the average return** [**?**], when

$$p(s,x) = E_s(x) = \frac{1}{|T_s|} \sum_{t \in T_s} \sum_{a \in A} r_{ta} x_a,$$

– **maximization of the Shannon entropy** (see [**?**]), when

$$p(s,x) = -\frac{1}{|T_s|} \sum_{t \in T_s} ln(\sum_{a \in A} r_{ta} x_a),$$

including the constraint $E_s(x) = \mu_s$ in the set of feasible portfolios $X$,

- **minimization of the conditional value at risk** (see e.g. [?] and the references therein), known also as Mean-CVaR approach [?],

$$p(s,x) = -c_s,$$

including the constraint $E_s(x) = \mu_s$ in the set of feasible portfolios $X$,
- **minimization of the total absolute deviation from the average return** [?,?], when

$$p(s,x) = -\sum_{t\in T_s} |\sum_{a\in A} r_{ta}x_a - E_s(x)|$$

- **minimization of the modified absolute deviation from the average return** [?,?], when

$$p(s,x) = -\sum_{t\in T_s}\left(|\sum_{a\in A} r_{ta}x_a - E_s(x)| - 2\sum_{a\in A} r_{ta}x_a\right)$$

- **minimization of the total absolute deviation from the median return** [?], when

$$p(s,x) = -\min_{\theta\in\Re}\sum_{t\in T_s} |\sum_{a\in A} r_{ta}x_a - \theta|$$

When a given scenario $s \in S$ is fixed, a reasonable model to configure the portfolio $x \in X$ studied in the literature (see [?] and the references therein) is the maximization of the average return subject to a constraint on the CVaR,

$$\begin{aligned} z_s^* &= \max E_s(x) \\ \text{s.t.}& \\ &\text{CVaR}_s(x) \le c_s, \\ &x \in X. \end{aligned} \tag{7}$$

In the formulation (**??**), the set of feasible portfolios $X$ is able to include a rich range of possible technical constraints as those considered in [?] related to transaction costs, diversification, changes in individual positions (liquidity constraints) or bounds on positions. These constraints can frequently be formulated as linear inequalities on the decision variables of the optimization model which does not have a significant influence in the numerical method used to solve the problem. Hence, in order to ease the following formulations we will assume that $X$ is given as

$$X = \left\{x \ge 0 : \sum_{a\in A} x_a \le 1\right\}, \tag{8}$$

and we will denote by $z_s$ the average return for the portfolio $x$ under the scenario $s \in S$ appearing in (**??**), that is,

$$z_s = E_s(x) = \frac{1}{|T_s|}\sum_{t\in T_s}\sum_{a\in A} r_{ta}x_a = \sum_{a\in A} \overline{r}_{as}x_a, \tag{9}$$

where

$$\bar{r}_{as} = \frac{1}{|T_s|} \sum_{t \in T_s} r_{ta}.$$

The CVaR model (??) is a case of the so-called Coherent Measures of Risk ([?]), that are risk measures satisfying, among others, the principle of convexity. In [?] it is shown how to optimize a mean-coherent risk model in the cases that random variables, e.g. risks, have a discrete support set, including the case in which they are described by temporal observations $t \in T_s$. In the following, let $\mathcal{Q}_s = \{Q | Q \subseteq T_s, |Q| = \lfloor (1-\beta)|T_s| \rfloor\}$:

$$\begin{aligned} z_s^* = \max & \sum_{a \in A} \bar{r}_{as} x_a \\ \text{s.t.} & \\ & r_t = \sum_{a \in A} r_{at} x_a, \text{ for all } t \in T_s \\ & \frac{1}{\lfloor (1-\beta)|T_s| \rfloor} \sum_{t \in Q} (-r_t) \le c_s, \text{ for all } Q \in \mathcal{Q}_s, \\ & \sum_{a \in A} x_a \le 1, \\ & x_a \ge 0, \text{ for all } a \in A. \end{aligned} \tag{10}$$

Observe that (??) contains an exponential number of constraints, one for every $Q \in \mathcal{Q}_s$. However, as discussed in [?], it is not necessary to solve a problem with all constraints. Rather, one can solve an incomplete formulation of (??) and then insert constraints only if they are needed, e.g., when they are violated by the incumbent solution. To do this, the separation subroutine, listed in Algoritm **??**, checks whether an incumbent solution $x$ is feasible or not. If yes, the solution is optimal, otherwise, the subroutine finds a violated inequality to insert in the model.

**Algorithm 1** Finding a valid inequality of (??)

1: For the incumbent solution $x_a, a \in A$, calculate $r_t = \sum_{a \in A} r_{at} x_a$, for all $t \in T_s$.
2: Rank $r_t$ in increasing order, to obtain $r_{i(1)}, \ldots, r_{i(j)}, \ldots, r_{i(|T_s|)}$. Let $C = \{t_{i(j)} \in T_s | j = 1, \ldots, \lfloor (1-\beta)|T_s| \rfloor\}$.
3: **if** $\frac{1}{\lfloor (1-\beta)|T_s| \rfloor} \sum_{t \in C} (-r_t) > c_s$ **then**
4: Add valid inequality $\frac{1}{\lfloor (1-\beta)|T_s| \rfloor} \sum_{t \in C} (-r_t) \le c_s$ to (??).
5: **else**
6: $x_a$ is optimal.

However, Proposition **??** can be used to linearize the problem (??) under each scenario $s \in S$, without the need of adding an exponential number of constraints as in (**??**), as follows

$$
\begin{aligned}
z_s^* = &\max \sum_{a\in A} \bar{r}_{as} x_a \\
\text{s.t.}& \\
&\alpha_s + \frac{1}{\lfloor (1-\beta)|T_s| \rfloor} \sum_{t\in T_s} \xi_{ts} \le c_s, \\
&\xi_{ts} \ge -\sum_{a\in A} r_{ta} x_a - \alpha_s, \ \text{for all } t \in T_s, \\
&\sum_{a\in A} x_a \le 1, \\
&\xi_{ts} \ge 0, \ \text{for all } t \in T_s, \\
&x_a \ge 0, \ \text{for all } a \in A.
\end{aligned}
\tag{11}
$$

This new formulation is a Linear Programming (LP) problem with a compact set of feasible portfolios $X$, then its optimal objective value is finite and coincides with the one of its following dual formulation

$$
\begin{aligned}
z_s^* = &\min \lfloor (1-\beta)|T_s| \rfloor c_s u_{0s} + v_s \\
\text{s.t.}& \\
&-\sum_{t\in T_s} r_{ta} u_{ts} + v_s \ge \bar{r}_{as}, \ \text{for all } a \in A \\
&\lfloor (1-\beta)|T_s| \rfloor u_{0s} - \sum_{t\in T_s} u_{ts} = 0, \\
&u_{0s} - u_{ts} \ge 0, \ \text{for all } t \in T_s, \\
&v_s, u_{0s}, u_{ts} \ge 0, \ \text{for all } t \in T_s.
\end{aligned}
\tag{12}
$$

Here, the dual variable $u_{0s}$ is the one associated to the first primal constrain of (**??**), $u_{ts}$, $t \in T_s$ the set of dual variables for the second block of constraints and $v_s$, the one corresponding to the last constraint of (**??**).

The formulation (**??**) ensures that, to upper bound the optimal value of the averaged reward under each scenario $s \in S$, it is only needed a feasible dual solution $(u_{0s}, u_{ts:\, t\in T_s}, v_s)$. This will be used in the general formulation of the minmax regret model stated in the next section.

## 4 Minmax Regret Mean Return Portfolio Problem

We will propose in this section an optimization model, the *Minmax Regret Portfolio* (MRP) problem, in which a *compromise* configuration of the investments is sought with an efficiency *as close as possible* to the performance of an optimal portfolio under any one of the considered scenario of $S$. This optimization model can be written as

$$
\begin{aligned}
& R^*(S) = \min_x R(x,S) := \max_s \{z_s^* - E_s(x)\} \\
& \text{s.t.} \\
& \qquad \text{CVaR}_s(x) \le c_s, \ \text{for all } s \in S, \\
& \qquad x \in X,
\end{aligned}
\tag{MRP}
$$

Here, we will call $\mathcal{X}(S)$ to the feasible set of portfolios corresponding to a given set of scenarios $S$, that is,

$$
\mathcal{X}(S) = \{x \in X : \text{CVaR}_s(x) \le c_s, \ s \in S\}. \tag{13}
$$

Observe that in this minmax regret model the set of scenarios defines the regrets considered in the objective function but also the set of feasible portfolios since all the solutions must fulfill the upper bounds on the conditional value at risk under each scenario ($c_s$).

The formulation (**??**) can be written as an LP problem as follows

$$
\begin{aligned}
R^*(S) = \ & \min \rho \\
\text{s.t.} \ & \\
& \rho \ge z_s^* - \sum_{a \in A} \bar{r}_{as} x_a, \ \text{for all } s \in S, \\
& \alpha_s + \frac{1}{\lfloor (1-\beta)|T_s| \rfloor} \sum_{t \in T_s} \xi_{ts} \le c_s, \ \text{for all } s \in S, \\
& \xi_{ts} \ge -\sum_{a \in A} r_{ta} x_a - \alpha_s, \ \text{for all } s \in S, \, t \in T_s, \\
& \sum_{a \in A} x_a \le 1, \\
& \xi_{ts} \ge 0, \ \text{for all } s \in S, \, t \in T_s, \\
& x_a \ge 0, \ \text{for all } a \in A.
\end{aligned}
\tag{MRP-F1}
$$

Now, taking $z_s$, the average reward for the portfolio $x$ defined in (**??**), one can write the first block of constraints of (**??**) as $z_s^* \le \rho + z_s$ for all $s \in S$ which is fulfilled if and only if there exists at least a feasible solution $(u_{0s}, u_{ts:\, t \in T_s}, v_s)$ for the dual LP problem (**??**) verifying

$$
\lfloor (1-\beta)|T_s| \rfloor c_s u_{0s} + v_s \le \rho + z_s \ \text{for all } s \in S.
$$

Hence, (**??**) can be reformulated as the following LP problem

$$
\begin{aligned}
R^*(S) = \ & \min \rho \\
\text{s.t.} \\
& \lfloor (1-\beta)|T_s| \rfloor c_s u_{0s} + v_s - \sum_{a \in A} \bar{r}_{as} x_a \leq \rho \text{ for all } s \in S, \\
& -\sum_{t \in T_s} r_{ta} u_{ts} + v_s \geq \bar{r}_{as}, \text{ for all } s \in S,\ a \in A \\
& \lfloor (1-\beta)|T_s| \rfloor u_{0s} - \sum_{t \in T_s} u_{ts} = 0, \text{ for all } s \in S, \\
& u_{0s} - u_{ts} \geq 0, \text{ for all } s \in S,\ t \in T_s, \\
& \alpha_s + \frac{1}{\lfloor (1-\beta)|T_s| \rfloor} \sum_{t \in T_s} \xi_{ts} \leq c_s, \text{ for all } s \in S, \\
& \xi_{ts} \geq -\sum_{a \in A} r_{ta} x_a - \alpha_s, \text{ for all } s \in S,\ t \in T_s, \\
& \sum_{a \in A} x_a \leq 1, \\
& \xi_{ts} \geq 0, \text{ for all } s \in S,\ t \in T_s, \\
& x_a \geq 0, \text{ for all } a \in A, \\
& v_s, u_{0s}, u_{ts} \geq 0, \text{ for all } s \in S,\ t \in T_s.
\end{aligned}
\tag{MRP-F2}
$$

Formulation (**??**) has $O(\sum_{s \in S} |T_s|)$ variables and $O(\sum_{s \in S} |T_s|)$ constraints which could be difficult to be managed if the number of scenarios and the number of periods contained in each one of them are large enough. In order to avoid this drawback we will examine an iterative procedure in which the feasible set of (**??**) is initially relaxed, a tentative portfolio is obtained by solving the relaxation and after checking a stopping rule, a cut for the relaxation is derived in order to continue the iterative process.

The iterative procedure relies to the following formulation, that is the straightforward extension of Model (**??**) to the case in which the CVaR constraints must be satisfied in multiple scenarios.

$$
\begin{aligned}
R^*(S) = \; & \min \rho \\
\text{s.t.} \; & \\
& \rho \geq z_s^* - \sum_{a \in A} \bar{r}_{as} x_a, \text{ for all } s \in S, \\
& r_t = \sum_{a \in A} r_{at} x_a, \text{ for all } t \in \textstyle\bigcup_{s \in S} T_s \\
& \frac{1}{\lfloor (1-\beta)|T_s| \rfloor} \sum_{t \in Q} (-r_t) \leq c_s, \text{ for all } Q \in \mathcal{Q}_s, \, s \in S \\
& \sum_{a \in A} x_a \leq 1, \\
& x_a \geq 0, \text{ for all } a \in A.
\end{aligned}
\tag{MRP-F3}
$$

## 5 Numerical procedures

Here, we will use the dual problem (**??**) together with the formulations proposed in the above section in order to bound the optimal objective value of the minmax regret model (**??**). These bounds are used in the development of numerical procedures to solve our problem which will be computationally tested in the following section.

### 5.1 Deriving lower bounds on the optimal average rewards under each scenario

Let $x^0 \in X$, the set of feasible portfolios defined in (**??**), and suppose that $\mathrm{CVaR}_s(x^0) \leq c_s$ for a given scenario $s \in S$, then one has trivially

$$\sum_{a \in A} \bar{r}_{as} x_a^0 \leq z_s^*$$

since $z_s^*$ is the optimal value of Problem (**??**).

Suppose now that $x^0 \in X$ verifies $\mathrm{CVaR}_s(x^0) > c_s$ then, by definition of CVaR one has

$$\mathrm{CVaR}_s\left(\frac{c_s}{\mathrm{CVaR}_s(x^0)} x^0\right) = c_s,$$

and, taking into account the definition (**??**) of the set of feasible portfolios $X$, one also have

$$\frac{c_s}{\mathrm{CVaR}_s(x^0)} x^0 \in X.$$

Therefore, for any given $x^0 \in X$ we can define a lower bound on the optimal reward under the scenario $s$ by

$$\frac{c_s}{\max\{c_s, \mathrm{CVaR}_s(x^0)\}} \sum_{a \in A} \bar{r}_{as} x_a^0 \leq z_s^*. \tag{14}$$

### 5.2 Deriving upper bounds on the optimal average rewards under each scenario

In order to have upper bounds on $z_s^*$, the optimal value of (**??**), we can construct feasible solutions of the dual formulation (**??**). Here, we propose such a feasible dual solution on the basis of some properties that can be expected to be verified by a primal optimal solution.

Using the formulation (**??**) and Proposition **??** there exists at least an optimal solution of Problem (**??**) for which at most $\lfloor(1-\beta)|T_s|\rfloor$ indices $t \in T_s$ verify

$$-\sum_{a\in A} r_{ta}x_a - \alpha_s > 0. \tag{15}$$

Let us assume that there exists an optimal portfolio $x \in X$ for the problem (**??**) in which

1. for every $t \in T_s^+$ one has $-\sum_{a\in A} r_{ta}x_a - \alpha_s > 0$ and
2. for every $t \in T_s^-$ one has $-\sum_{a\in A} r_{ta}x_a - \alpha_s < 0$.

Under this assumption, we can construct an optimal solution for the dual problem (**??**) from the sets of time periods $T_s^+$ and $T_s^-$.

By complementary slackness conditions, we know the following relations are fulfilled,

$$u_{0s} = u_{ts}, \text{ for all } t \in T_s^+,$$
$$u_{ts} = 0, \text{ for all } t \in T_s^-.$$

Then, given a nonnegative value for $u_{0s}$, a feasible solution of the dual problem (**??**) can be obtained by taking

$$v_s = \max\left\{0, \bar{r}_{as} + u_{0s}\sum_{t\in T_s^+} r_{ta}, \forall a \in A\right\},$$

that is, the objective value of the dual problem (**??**) at this solution can be written as

$$\varphi_{T_s^+}(u_{0s}) := \max\left\{\lfloor(1-\beta)|T_s|\rfloor c_s u_{0s}, \bar{r}_{as} + u_{0s}\sum_{t\in T_s^+}(c_s + r_{ta}), \forall a \in A\right\}. \tag{16}$$

Hence, the optimal objective value of the dual problem (**??**), $z_s^*$, should coincide with the minimum of $\varphi_{T_s^+}(u_{0s})$ in $u_{0s} \geq 0$. When our choice of the sets $T_s^-$ and $T_s^+$ does not correspond to any optimal solution of (**??**) or does not fulfill assumptions (**??**) and (**??**), this minimum will achieve just an upper bound on the optimal value $z_s^*$ since the dual solution we have constructed is always feasible for the problem (**??**), that is,

$$z_s^* \leq \min_{u_{0s}\geq 0} \varphi_{T_s^+}(u_{0s}). \tag{17}$$

**Remark 1** All the slopes of the linear functions defining $\varphi_{T_s^+}$ in (**??**) can be considered different. Otherwise, one of the two linear functions sharing a common slope can be deleted since it does not affect to the maximum defining $\varphi_{T_s^+}$.

**Algorithm 2** Finding an optimum of (**??**)

**procedure** Initialization

Let $\bar{r}_{a_0 s} = 0$, $\delta_0 = \lfloor(1-\beta)|T_s|\rfloor c_s$ and $\delta_i = \sum_{t \in T_s^+}(c_s + r_{ta_i})$ where $a_i$ is the i-th element of $A$.

Arrange the components of $\delta$ in increasing order. Let $\delta_{(i)} : i = 0, 1, \dots, |A|$ be the ordered components (all of them different according to the remark **??**) where $(i)$ is the index occupying the position $i+1$ in the ordered sequence.

Let $k := \arg\max\{\bar{r}_{a_i s} : i = 0, 1, \dots, |A|\}$ and $u_s^* := 0$ and $t \in \{0, 1, \dots, |A|\}$ the index for which $(t) = k$, that is, $\delta_k$ occupies the position $t+1$ in the sequence of ordered $\delta$-values.

**while** $\delta_k < 0$ **do**

$$u_s^* := \min\left\{ \frac{\bar{r}_{a_k s} - \bar{r}_{a_{(j)} s}}{\delta_{(j)} - \delta_k} : j = t+1, \dots, |A| \right\}$$

$$k := \arg\min\left\{ \frac{\bar{r}_{a_k s} - \bar{r}_{a_{(j)} s}}{\delta_{(j)} - \delta_k} : j = t+1, \dots, |A| \right\},$$

and $t \in \{0, 1, \dots, |A|\}$ the index for which $(t) = k$.

Now we will consider a simple numerical example in order to clarify the way in which Algorithm **??** works and to show the number of operations needed to bound its worst-case running time.

*Example 1* Let us take $\beta = \frac{2}{3}$, $|T_s| = 10$, $\lfloor(1-\beta)|T_s|\rfloor = 3$, $c_s = 0.1$ and the assets $A = \{a_1, a_2, a_3, a_4\}$. We fix a given portfolio $x = (x_a)_{a \in A} = (0.10, 0.25, 0.25, 0.4)$. Consider the table **??** of individual returns under the scenario $s$ corresponding to ten given time periods.

| $r_{t,a}$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $a_1$ | -0.83 | -0.22 | -0.28 | -0.83 | -0.60 | 0.02 | -0.60 | -0.28 | -0.22 | -0.69 |
| $a_2$ | -0.60 | -0.68 | 0.00 | -0.53 | -0.32 | 0.30 | -0.43 | 0.13 | 0.09 | -0.01 |
| $a_3$ | -0.22 | 0.38 | 0.48 | 0.65 | -0.30 | 0.55 | -0.27 | 0.67 | 0.69 | 0.36 |
| $a_4$ | 0.28 | 0.09 | 0.55 | 0.55 | 0.30 | 0.04 | -0.08 | -0.02 | 0.46 | 0.33 |
| $-\sum_{a \in A} r_{ta} x_a$ | 0.18 | 0.06 | -0.31 | -0.17 | 0.09 | -0.23 | 0.27 | -0.16 | -0.36 | -0.15 |

Table 1: Values of $r_{ta}$ for $t \in T_s$ and $a \in A$.

In order to apply Algorithm **??** it is first needed to compute the values $\delta_i$ and $\bar{r}_{a_i s}$ of Table **??** with a $O(|A|\lfloor(1-\beta)|T_s|\rfloor)$ number of elementary operations.

| $i$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $\delta_i$ | 0.30 | -1.73 | -1.05 | -0.49 | 0.81 |
| $\bar{r}_{a_i s}$ | 0.00 | -0.45 | -0.21 | 0.30 | 0.25 |

Table 2: Values $\delta_i$ and $\bar{r}_{a_i s}$ for $i = 0, 1, \dots, |A|$.

Figure **??** depicts the function $\varphi_{T_s^+}(u_{0s})$ for the values corresponding to Table **??**.

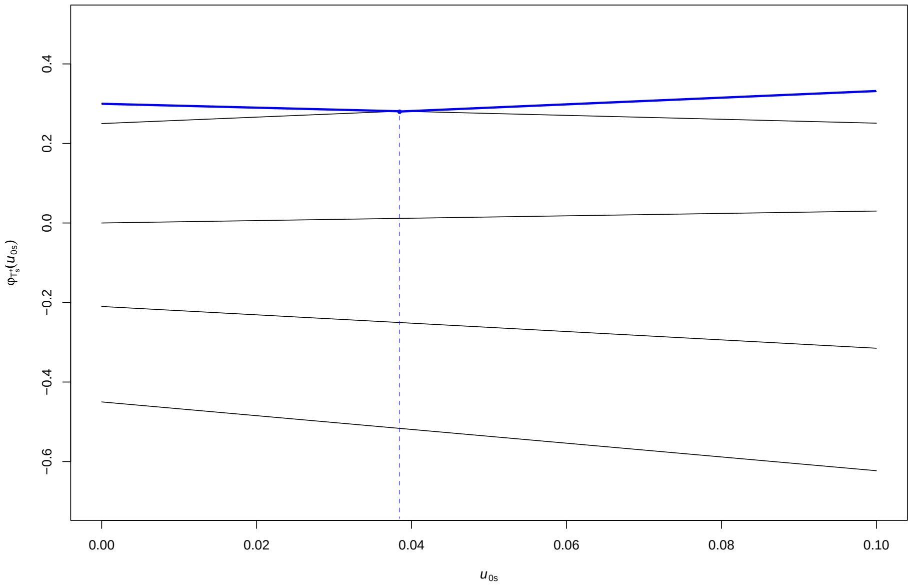



Fig. 1: The function $\varphi_{T_s^+}(u_{0s})$

From Figure **??** one has that the step 5 of Algorithm **??** just will be executed once since:

- In the step 4, $k = 3$, that is, $\overline{r}_{a_3 s} = 0.30$ is the maximum of the $\overline{r}_{a_i s}$ values, then $t = 2$ because the slope $\delta_3 = -0.49$ occupies the third position in the ordered sequence of slopes in increasing order.
- The value $u_s^*$ of step 5 is given by just computing two intersection values, the one corresponding to the index $i = (3) = 0$ and the one corresponding to $i = (4) = 4$. The minimum of these values are reached for the index $j = 4$, hence $k$ is actualized to the value $k = 4$ and $t$ to the value $t = 4$.
- Taking into account that $\delta_4 = 0.81 > 0$, the stopping rule is verified.

The following results states de worst-case running time of the above numerical procedure. In order to simplify notation we will assume that $|A| < |T_s|$, which seems to be natural in real applications.

**Proposition 3** *The worst-case complexity of Algorithm* **??** *is* $O(|A||T_s|)$.

*Proof* Step 2 requires $|A|\lfloor(1-\beta)|T_s|\rfloor$ sums. Step 3 has a worst-case running time of $O(|A|log(|A|))$ that can be achieved by algorithms like `Merge-sort` and `Heapsort` (note that any ordering procedure has a lower bound of $\Omega(|A|log(|A|))$ on its asymptotic worst-case running time [?]). Step 4 needs $O(|A|)$ operations while Step 5 needs $O(|A|^2)$ in a worst-case. These upper bounds, together with the assumption $|A| < |T_s|$, amount to the overall worst-case running time stated in the proposition. □

**Remark 2** In [?] it is shown that the optimization problem in the right-hand-side of (**??**) can be written as a linear problem with $|A| + 2$ variables and $2|A| + 1$

constraints which is solved in $O(|A|)$ time since it is the dual of a linear multiple-choice knapsack problem [?]. However, the computation of the coefficients (slopes and constant coefficients) of the linear functions defining $\varphi_{T_s^+}(u)$ in (??) requires $\lfloor(1-\beta)|T_s|\rfloor|A|$ elementary operations. Hence, the most efficient way of computing the upper bound given by (??) has $O(\lfloor(1-\beta)|T_s|\rfloor|A|)$ running time complexity. In particular, for a set of assets $A$ with $|A| = O(\lfloor(1-\beta)|T_s|\rfloor|A|)$, the algorithm ?? reaches the best worst-case running time bound.

To end this section we will show how can the proposed bounds be integrated together with the formulation (??) into an iterative procedure to solve our problem. The main advantage of this procedure is that just a subset of values $z_s^*$ needs to be computed and only a subset of the constraints of (??), linked to sets of time periods $\mathcal{Q}_s$ have to be explicitly generated.

### 5.3 Iterative procedure

Solving the LP problem (??) requires having the optimal values $z_s^*$ of the nominal problem (??) for each $s \in S$ and the exponential number of CVaR constraints to define the feasible solutions. Both requirements can be very expensive in terms of computational time and space. However, it is very likely that a large amount of these constraints were redundant since they are not really bounding the objective function nor constraining the feasible set. So, there is no need to formulate the problem (??) in its extensive form, that is, with all CVaR constraints and the maximum return under each scenario. Rather, we can formulate and solve a relaxed problem using only some scenarios and constraints. Then we can test whether the incumbent solution is unfeasible, and, in case, adding a separating constraint to the model. In this way, we can implement an iterative procedure in which LP problems of much more moderate sizes are solved in each iteration, until optimality can be certified. ore formally, let $S^0 \subseteq S$ and $Q^0 \subseteq \bigcup_s \mathcal{Q}_s$ be scenarios and CVaR subsets of constraints of (??). Then, the relaxed formulation of the (??) model can be stated as follows

$$
\begin{aligned}
R^*(S^0, Q^0) = \ & \min \rho \\
\text{s.t.} \ & \\
& \rho \geq z_s^* - \sum_{a\in A} \bar{r}_{as} x_a, \ \text{for all } s \in S^0, \\
& r_t = \sum_{a\in A} r_{at} x_a, \ \text{for all } t \in \textstyle\bigcup_{s\in S} T_s \\
& \frac{1}{\lfloor(1-\beta)|T_s|\rfloor} \sum_{t\in Q} (-r_t) \leq c_s, \ \text{for all } Q \in Q^0 \\
& \sum_{a\in A} x_a \leq 1, \\
& x_a \geq 0, \ \text{for all } a \in A.
\end{aligned}
\tag{MRP-R}
$$

At the beginning, we use only one scenario, $S^0 = \{s\}$, and no CVaR constraint, $Q^0 = \emptyset$. After solving the relaxed formulation (??), we obtain the first unfeasible solution $x_{begin}$. Applying the separation subroutine of Algorithm (**??**) to all $s \in S$, we progressively introduce new constraints in $Q^0$, until we find a solution $x$ satisfying the CVaR constraints for all scenario $s \in S$. Then, we calculate the regret upper bound for all $s \in S - S_0$ using Algorithm **??** and, if for some $s$ the upper bound is greater than the incumbent regret, we update $S_0$ with $s$. Model (**??**) is optimized again and the new solution $x$ is checked for CVaR feasibility first, and regret bound then, until optimality can be certified.

The proposed scheme has four main advantages:

1. It is not required solving (**??**) for each $s \in S$.
2. It is not required to formulate all CVaR constraints, that even though are polynomial in (**??**) or (**??**), they are of a huge size.
3. The optimality of a tentative portfolio can be certified before all the scenarios of $S$ are included in $S^0$.
4. In each iteration it is updated an upper bound on the gap on the maximum regret incurred if the approximate portfolio were implemented. This allows us ending the computations when a reasonable accuracy is reached.
5. The algorithm is a natural way of handling the continuous incoming of economic information in the format of new scenarios of $S$.

Here we describe how we calculate the bounds on the regret under any scenario $s \in S - S_0$.

**Proposition 4** *After solving the problem (**??**) for a subset of scenarios $S^0 \subseteq S$ and a family of subsets of indices $Q^0$ one has a valid lower bound on the optimal objective of the problem (**??**), that is,*

$$R^*(S^0, Q^0) \leq R^*(S), \quad \forall S^0 \subseteq S.$$

*Proof* First, if $S^0 \subseteq S$ one has $R^*(S^0, Q^0) \leq R^*(S^0)$ since (**??**) is a relaxation of (**??**). On the other hand, as

$$\max_{s \in S^0}\{z_s^* - E_s(x)\} \leq \max_{s \in S}\{z_s^* - E_s(x)\}, \quad \forall x,$$

and $\mathcal{X}(S^0) \supseteq \mathcal{X}(S)$ it directly follows that $R^*(S^0) \leq R^*(S)$. □

Now we will upper bound the optimal objective of the problem (**??**) by using the bounds of the subsection **??**. The following expression is used later

$$\gamma_s^0 = \begin{cases} z_s^* & \text{if } \ s \in S^0, \\ \min\limits_{u_{0s} \geq 0} \varphi_{T_s^+}(u_{0s}) & \text{if } \ s \notin S^0, \end{cases} \tag{18}$$

where $x^0$ is an optimal solution of Problem (**??**) and $T_s^+$ is the set of $\lfloor (1-\beta)|T_s| \rfloor$ time periods of the highest losses for the portfolio $x^0$ under the scenario $s \in S$ (defined in Section **??**).

**Remark 3** Computing the value $\gamma_s^0$ for a scenario $s \notin S^0$ just requires a computational effort of $O(|A||T_s|)$ by Proposition **??**. □

The following result states an upper bound on the optimal value of the problem (**??**).

**Proposition 5** *Let $x^0$ be an optimal solution of Problem (**??**), the following inequality holds*

$$\overline{R}(S^0,Q^0) := \max_{s\in S}\left\{\gamma_s^0 - \min_{\bar{s}\in S}\left\{\frac{c_{\bar{s}}^*}{\max\{c_{\bar{s}}^*,\, CVaR_{\bar{s}}(x^0)\}}\right\}\sum_{a\in A}\bar{r}_{as}x_a^0\right\} \geq R^*(S),$$

*where $\gamma_s^0$ is defined by* (**??**).

*Proof* First note that (**??**) implies the feasibility of the portfolio

$$\min_{\bar{s}\in S}\left\{\frac{c_{\bar{s}}^*}{\max\{c_{\bar{s}}^*, \mathrm{CVaR}_{\bar{s}}(x^0)\}}\right\}x^0 \in \mathcal{X}(S),$$

then, using (**??**) one has that $\gamma_s^0$ is an upper bound on the maximum return under $s \in S$, which implies the result. □

Propositions **??** and **??** give us a procedure to bound the optimal objective value of our minmax regret problem (**??**) from a subset of scenarios $S^0$. Furthermore, if the upper bound $\overline{R}(S^0,Q^0)$ is reached at a scenario $s^0$ in $S^0$ and $x^0$ is a feasible portfolio of $\mathcal{X}(S)$ we can certify the optimality of $x^0$. This is true due to the following chain of inequalities

$$R^*(S^0,Q^0) \leq R^*(S) \leq R(x^0,S) = \overline{R}(S^0,Q^0) = z_s^* - \sum_{a\in A}\bar{r}_{as}x_a^0 \leq R^*(S^0,Q^0).$$

- The first inequality is true from the fact that (**??**) is a relaxation of (**??**),
- the second one is due to the feasibility of $x^0$ and
- the last inequality follows from the fact that $s^0$ is supposed to belong to $S^0$ and the structure of the first block of constraints of (**??**).

Hence, as the first and last values of the above chain of inequalities are the same we have the optimality of $x^0$ which is stated in the following

**Proposition 6** *Given a subset of scenarios $S^0 \subseteq S$, a subset of CVaR constraints $Q^0$and $x^0$ an optimal solution of Problem (**??**) if the upper bound $\overline{R}(S^0,Q^0)$ is reached at a scenario $s^0$ in $S^0$ and $x^0$ is a feasible portfolio of $\mathcal{X}(S)$ we can certify the optimality of $x^0$.*

To end this section we will propose a numerical procedure whose main operations are stated as Algorithm **??**.

This is an iterative procedure in which the problem (**??**) will be sequentially solved for an increasing set of constraints, labeled by $Q^0$ and separated in Line 7 until the portfolio is feasible. Then scenarios of $S^0$ are increased, until the stopping rule of Proposition **??** can be applied (line 10 of Algorithm **??**).

Our procedure combines

- a partial description of the feasible set of the minmax regret portfolio problem (**??**) by using the formulation (**??**)
- and the bounds on the optimal value of that problem (**??**) obtained by means of the dual (**??**) of the linearization (**??**).

Formulation (**??**) allows us to find *good* candidate portfolios by solving a formulation not as heavy as the linear formulation (**??**). After checking optimality of that candidate according to Proposition **??**, the algorithm is stopped or it continues solving a new version of (**??**) in which $S^0$ and $Q^0$ are augmented. In both cases we use Proposition **??** where feasibility is contrasted as in Algorithm **??**. Since in each iteration the set $S^0$ and $Q^0$ are strictly augmented and the cardinality of both sets is bounded, it is obvious that this procedure ends in a finite number of iterations with the optimal solution.

**Algorithm 3** Iterative procedure

Initialize the set $S^0 \leftarrow \{s\}$.
Initialize the set $Q^0 \leftarrow \emptyset$.
Initialize $x^0$ as unfeasible.
**while** $x^0$ is not optimal **do**
    Solve the problem (**??**) and let $x^0$ be the optimal solution.
    **if** **search_cvar_violation**(q) = True **then**
        $Q^0 \leftarrow Q^0 \cup q$ ($q$ is a violated constraint).
    **else**
        **if** **search_regret_violation**(s) = True **then**
            $S^0 \leftarrow S^0 \cup s$ ($s$ is a violated scenario).
        **else**
            $x^0$ is optimal.
**procedure search_cvar_violation**($q$)
    **search_cvar_violation** $\leftarrow$ False.
    **for** $s \in S$ **do**
        Apply separation subroutine with input $T_s, c_s$:
        **if** $q$ is violated constraint: **then**
            **search_cvar_violation** $\leftarrow$ True.
            **return** $q$.
**procedure search_regret_violation**($s$)
    **search_regret_violation** $\leftarrow$ False.
    **for** $s \in S - S_0$ **do**
        $\gamma_s \leftarrow \min_{u_{0s} \geq 0} \varphi_{T_s^+}(u_{0s})$
        $\overline{R}_s \leftarrow \gamma_s - \sum_{a \in A} \bar{r}_{as^0} x_a^0$.
        **if** $\overline{R}_s - R^*(S^0) > \epsilon$ **then**
            **search_regret_violation** $\leftarrow$ True
            **return** $s$.

## 6 Computational results

Computational experiments are twofold. Firstly, we want to compare numerical procedures to solve the models (**??**), (**??**) and (**??**) to establish if there exists one of them outperforming clearly the other methods. Secondly, we want to apply an optimal minmax-regret portfolio to real financial data in order to check features as the diversification of the investment expecting to hedge the decision maker's exposure to future event risk. This should be guaranteed in those cases in which these future events share, in some way, the economic pattern with at least one of the scenarios considered in the model.

We compare the computational aspects of three numerical procedures to solve:

- Model (**??**), (labeled as LP 1 in following tables): The characteristic of this procedure is that it works in two phases. In the first phase, $z_s^*$ are computed for all scenarios $s \in S$, using formulation (**??**), and then the LP formulation (**??**) is solved.

- Model (**??**) (labeled as LP 2 in following tables): Its characteristic is that it is a compact formulation. Here, optimal $z^s$'s are replaced by their dual representation, so that a single LP problem is to be solved. The method avoids to solve explicitly the LP models associated to the scenarios, but at the cost of requiring more constraints and variables.
- Model (**??**) (labeled as LP 3 in following tables): Its characteristic is that scenarios and constraints are inserted as Benders cuts only when effectively needed by the formulation, as described in Algorithm **??**. Since not all CVaR constraints and all scenarios are necessary to bound the objective function, the model can save constraints and variables.

Tests are run using Julia 1.0.3 and its package JuMP for algebraic modeling, GuRoBi 8.0.1 as linear solver, in a PC equipped with an Intel Core i5-6200U CPU. We considered financial problems with $|T| = 200$ past observations of some asset set $A$. Scenarios $T_s$ consist of $l$ consecutive time periods that begin on $t_i^u$ and end on $t_i^l$. The size of the considered scenarios, that is, $t_i^l - t_i^u + 1$, is a random value chosen between 50 and 100.

There are several reasons by which we have made this naive selection of the scenarios. First, our main goal in this section is the analysis of the computational behaviour of different numerical procedures developed in previous section in order to solve the minmax regret problem (**??**). In this sense, we try to design *neutral* instances of the problem by selecting subsets of time periods with a random mechanism instead of taking advantage of the specific nature of the economic data in order to select a set of scenarios which can benefit to the computational performance of the numerical algorithms. On the other hand, this way of selecting scenarios could be a type of *blind selection* which may be done by an inexpert decision maker in order to try to capture different economic patterns or changes in the economic cycle of the market against which she want to be hedged.

The benchmark that we used to calculate the CVaR bounds $c_s$ is the 1/N portfolio: On time $t$, the 1/N portfolio benchmark return is $r_t^b = 1/N(\sum_{a\in A} r_{at})$, so that $c_s = CVaR[r_t^b | t \in T_s]$ for all $s \in S$. The use of the $1/N$ benchmarks has been already used in [**?**] and motivated by the fact that the $1/N$ portfolio is a financial benchmark that obtain good returns even through maximum diversification, see [**?**].

The return data $R$ is simulated by using one of the two following procedures. In the first one, returns $r_{ta}$ are random uniform values between -0.5 and 0.5. In the second procedure, $r_{ta}$ are real daily stock market returns. We motivated this choice because past experiments showed that the two simulated sets can have different computational properties, see [**?**]. The tolerance threshold has been fixed to $\epsilon = 10^{-4}$, as this value is sufficient to calculate the same optimal portfolio in all three model versions.

In Table (**??**), we report some computational features corresponding to the three models when the input are random returns, while in Table (**??**) and (**??**) we report results obtained when the inputs are real returns. To compare the methods, we considered the total numbers of LP iterations needed to find the optimum (including the ones necessary to compute $z_s^*$ in LP1 and LP3). We also show the overall computational times, expressed in seconds, and finally, in the columns headed *size*, the pair $(m, n)$ reports the number of variables $m$ and constraints $n$ that have been necessary to formulate the LP models.

We will analyze how computation time and memory requirements change with the increase of the number of scenarios, ranging from $s = 20$ to $s = 500$ and fixing the number of assets to $|A| = 5$ and $|A| = 30$. When we look at the computational results on random data in Table (**??**), Model (**??**) is the one requiring less iterations to calculate the optimum, and this happens for all scenarios size. However, the result is not confirmed when using an input of real financial return, see Tables **??** and **??**, in which the Model (**??**) is the one using less iterations instead. When we look at the computational times, we can see that the compact formulation of Model (**??**) is the fastest method even though sometimes it takes more iterations and its corresponding formulation has the largest size. It can be the case that this model take advantage of delegating all computational efforts to the LP solver, avoiding the computational overhead of communicating processes. In any case, all the instances are solved in a reasonable time, less than 8 seconds of CPU time even when the number of scenarios is as large as $s = 500$.

It is interesting to compare the number of constraints and variables that were required by the three LP models, as this correspond to the most efficient use of the PC memory. As can be seen, the size of LP1 and LP2 is very large, its numerical magnitude is $10^4$ in the largest models, while the LP3 formulation remains very compact, with order $10^2$. The number of variables does not depend on the number of scenarios, and only a very small number of Benders cuts are needed to describe the feasible region.

### 6.1 Application to real financial data

In this section, we apply the regret model to real financial data and decision-making, to see what are the main features of the minmax regret model, compared to the standard CVaR model, in which optimization is carried out on a single scenario. We considered the data coming from the Milan Stock market, that have been previously used to compare different models in [**?**]. These data are made up of returns calculated from daily market prices of the 60 main Italian companies, that have been traded in the period from March 2003 to March 2008, for a whole of 1305 trading days. We simulate a decision maker that re-balances its portfolio every 20 days (called the rolling window), with the purpose of maximizing her future wealth, and we compare the financial results of the plain CVaR model with the models in which there are multiple scenarios.

The CVaR model of this simulation is run with $\alpha = 0.10$ (corresponding to $\beta = 0.90$), and in which the CVaR has been estimated through $t = 120$ past observations. Then, as scenarios, the minmax regret uses the subsets composed of $t$ past consecutive returns, evenly spaced by $l$ periods. More formally, let $T$ be the index corresponding to the most recent observed returns, then a scenario $s$ is composed of returns ranging from $t_{ls} = T - t + 1 - (s-1)l$ to $t_{us} = T - (s-1)l$, with $s = 1, \ldots, S$. For $s = 1$, we have the plain CVaR model.

As in the previous subsection, to calculate CVaR bounds $c_s$ we use the 1/N portfolio, for which in time $t$ the return is $r_t^b = 1/N(\sum_{a \in A} r_{at})$. Then, we calculate $c_s = CVaR[r_t^b | t \in T_s]$ for all $s \in S$.

Table 3: Computational results with random returns, $|A| = 5$.

| | Iterations | | | Times | | | Sizes | | |
|---|---|---|---|---|---|---|---|---|---|
| Scenarios | LP1 | LP2 | LP3 | LP1 | LP2 | LP3 | LP1 | LP2 | LP3 |
| 20 | 438 | 616 | 272 | 0.108 | 0.044 | 0.181 | 2026, 1441 | 6066, 2961 | 206, 232 |
| 40 | 809 | 1140 | 462 | 0.200 | 0.099 | 0.320 | 4046, 3001 | 12126, 6161 | 206, 260 |
| 60 | 1199 | 1270 | 588 | 0.291 | 0.132 | 0.476 | 6066, 4421 | 18186, 9081 | 206, 257 |
| 80 | 1557 | 2093 | 422 | 0.437 | 0.271 | 0.504 | 8086, 6071 | 24246, 12461 | 206, 238 |
| 100 | 1867 | 2505 | 924 | 0.528 | 0.258 | 0.783 | 10106, 7581 | 30306, 15561 | 206, 293 |
| 120 | 2202 | 2580 | 846 | 0.605 | 0.285 | 0.869 | 12126, 8931 | 36366, 18341 | 206, 285 |
| 140 | 2325 | 2990 | 722 | 0.731 | 0.384 | 1.046 | 14146, 10331 | 42426, 21221 | 206, 285 |
| 160 | 3248 | 3698 | 807 | 0.809 | 0.398 | 1.021 | 16166, 12041 | 48486, 24721 | 206, 312 |
| 180 | 2229 | 3869 | 1537 | 0.902 | 0.449 | 1.500 | 18186, 13571 | 54546, 27861 | 206, 340 |
| 200 | 3312 | 4411 | 984 | 0.978 | 0.524 | 1.198 | 20206, 14651 | 60606, 30101 | 206, 291 |
| 220 | 2722 | 4326 | 1770 | 0.944 | 0.595 | 1.687 | 22226, 16621 | 66666, 34121 | 206, 384 |
| 240 | 3035 | 5133 | 1779 | 1.034 | 0.631 | 1.685 | 24246, 17621 | 72726, 36201 | 206, 368 |
| 260 | 4325 | 6284 | 1604 | 1.187 | 0.737 | 1.687 | 26266, 19221 | 78786, 39481 | 206, 343 |
| 280 | 4929 | 5980 | 1878 | 1.272 | 0.805 | 1.950 | 28286, 20941 | 84846, 43001 | 206, 368 |
| 300 | 4353 | 6991 | 1420 | 1.304 | 0.838 | 1.751 | 30306, 22591 | 90906, 46381 | 206, 318 |
| 320 | 5332 | 7750 | 1611 | 1.451 | 0.898 | 1.898 | 32326, 23791 | 96966, 48861 | 206, 333 |
| 340 | 5420 | 8358 | 1111 | 1.509 | 0.998 | 1.837 | 34346, 25231 | 103026, 51821 | 206, 328 |
| 360 | 6453 | 7535 | 1771 | 1.611 | 1.045 | 2.254 | 36366, 26761 | 109086, 54961 | 206, 388 |
| 380 | 6221 | 8093 | 2049 | 1.711 | 1.182 | 2.392 | 38386, 28231 | 115146, 57981 | 206, 380 |
| 400 | 6482 | 8488 | 1865 | 1.841 | 1.275 | 2.409 | 40406, 29721 | 121206, 61041 | 206, 377 |
| 420 | 5287 | 8393 | 2853 | 1.825 | 1.236 | 3.165 | 42426, 31231 | 127266, 64141 | 206, 435 |
| 440 | 7142 | 9817 | 1962 | 1.937 | 1.487 | 2.857 | 44446, 32671 | 133326, 67101 | 206, 377 |
| 460 | 6735 | 10154 | 2255 | 2.061 | 1.396 | 2.910 | 46466, 34121 | 139386, 70081 | 206, 412 |
| 480 | 7721 | 10464 | 1997 | 2.169 | 1.634 | 2.849 | 48486, 35821 | 145446, 73561 | 206, 377 |
| 500 | 7961 | 11028 | 1825 | 2.289 | 1.679 | 2.911 | 50506, 37231 | 151506, 76461 | 206, 391 |

In Figures (**??**), it can be seen the temporal behavior of the CVaR and the minmax regret portfolios with respect to the benchmark of the $1/N$ portfolio. As can be seen, and as already remarked in [**?**], the CVaR model (the active strategy) obtains higher returns than the 1/N model (the passive strategy). The novelty here is that, when we use the minmax regret, again we find that this model still outperforms the 1/N portfolio for all scenarios $|S|$, as can be seen in Figures (**??**) (b), (c), and (d). Quantitative comparison between the models are reported in Table (**??**, in which results for models with $|S|$ scenarios are labeled $|S|$_sc. As can be seen, increasing the number of scenarios does not modify significantly the mean and the median of the realized profits, but the variability of the realized profits decreases sharply. As can be seen, the maximum portfolio loss decreases from -6.02% to -5.42% and the maximum portfolio gain decreases as well, from 7.15% to 4.46%. Quartile measures (the 1st and the 3rd) decreases too.

Next we analyze the effect of the minmax regret model on the portfolio diversification and turnover. In our previous tests, [**?**], we measured the portfolio diversification through the Herfindahl-Hirschman (HH) and the Max index, and the smaller the values, the most the portfolio is diversified. In Table **??**, it can be seen that the effect of the minmax regret is to increase the portfolio diversification, an effect that is readily observable even when we are using just two scenarios. Then, as we increase the number of scenarios, we obtain more diversified portfolios, but the effect is less significative. Those numbers can be compared with the ones obtained using the median portfolio, [**?**], where the diversification was slightly

Table 4: Computational results with real returns, $|A| = 5$

| | Iterations | | | Times | | | Sizes | | |
|---|---|---|---|---|---|---|---|---|---|
| Scenarios | LP1 | LP2 | LP3 | LP1 | LP2 | LP3 | LP1 | LP2 | LP3 |
| 20 | 813 | 477 | 502 | 0.102 | 0.048 | 0.183 | 2026, 1481 | 6066, 3041 | 206, 251 |
| 40 | 1441 | 1103 | 932 | 0.215 | 0.162 | 0.364 | 4046, 2781 | 12126, 5721 | 206, 282 |
| 60 | 2197 | 1300 | 1590 | 0.345 | 0.251 | 0.586 | 6066, 4691 | 18186, 9621 | 206, 332 |
| 80 | 2648 | 1534 | 1912 | 0.408 | 0.314 | 0.698 | 8086, 5951 | 24246, 12221 | 206, 337 |
| 100 | 3170 | 1796 | 2360 | 0.585 | 0.255 | 0.908 | 10106, 7421 | 30306, 15241 | 206, 374 |
| 120 | 3881 | 1951 | 2976 | 0.737 | 0.403 | 1.189 | 12126, 8901 | 36366, 18281 | 206, 414 |
| 140 | 4373 | 2488 | 3237 | 0.991 | 0.451 | 1.261 | 14146, 10351 | 42426, 21261 | 206, 400 |
| 160 | 5031 | 3217 | 3662 | 1.010 | 0.494 | 1.391 | 16166, 12141 | 48486, 24921 | 206, 395 |
| 180 | 5919 | 2983 | 4210 | 1.031 | 0.525 | 1.722 | 18186, 13511 | 54546, 27741 | 206, 468 |
| 200 | 6182 | 3757 | 4615 | 1.489 | 0.744 | 1.939 | 20206, 14801 | 60606, 30401 | 206, 473 |
| 220 | 7426 | 4284 | 5191 | 1.242 | 0.832 | 2.324 | 22226, 16411 | 66666, 33701 | 206, 501 |
| 240 | 8652 | 4209 | 5603 | 1.245 | 0.677 | 2.302 | 24246, 17821 | 72726, 36601 | 206, 508 |
| 260 | 7972 | 4910 | 6135 | 1.422 | 1.168 | 2.500 | 26266, 19841 | 78786, 40721 | 206, 530 |
| 280 | 8166 | 4680 | 6353 | 1.645 | 1.080 | 2.836 | 28286, 20601 | 84846, 42321 | 206, 542 |
| 300 | 8955 | 4223 | 6931 | 1.627 | 1.149 | 2.956 | 30306, 22371 | 90906, 45941 | 206, 578 |
| 320 | 9349 | 6931 | 6984 | 1.693 | 1.244 | 3.136 | 32326, 23821 | 96966, 48921 | 206, 569 |
| 340 | 10623 | 6163 | 7895 | 1.833 | 1.668 | 3.439 | 34346, 25561 | 103026, 52481 | 206, 602 |
| 360 | 10404 | 5877 | 8072 | 1.955 | 1.586 | 3.716 | 36366, 26641 | 109086, 54721 | 206, 648 |
| 380 | 11637 | 6299 | 8453 | 1.917 | 1.571 | 4.021 | 38386, 28351 | 115146, 58221 | 206, 625 |
| 400 | 11745 | 7962 | 8957 | 1.955 | 1.803 | 4.171 | 40406, 29611 | 121206, 60821 | 206, 672 |
| 420 | 13208 | 8624 | 9293 | 2.126 | 1.828 | 4.294 | 42426, 31201 | 127266, 64081 | 206, 655 |
| 440 | 13110 | 7139 | 9850 | 2.251 | 2.286 | 4.595 | 44446, 32781 | 133326, 67321 | 206, 687 |
| 460 | 13750 | 8062 | 10329 | 2.398 | 2.419 | 4.838 | 46466, 34251 | 139386, 70341 | 206, 702 |
| 480 | 14233 | 6673 | 10955 | 2.544 | 2.250 | 5.263 | 48486, 35771 | 145446, 73461 | 206, 738 |
| 500 | 14771 | 6964 | 11246 | 2.621 | 2.135 | 5.450 | 50506, 37211 | 151506, 76421 | 206, 770 |

over. Next, we analyze the average of the portfolio turnover, that is the percentage of the portfolio that is sold in every re-balancing period. Its value can range from 0, when there is no re-balancing, to 1, when all the portfolio is sold. For the same level of return, usually portfolio managers prefer less turnover, as it implies less transaction costs. In Table **??**, we can see that the minmax regret portfolio is a more conservative strategy than CVaR: The latter required that more than 50% of the portfolio is reallocated on average and in every re-balancing period, while the minmax regret investment decreases this quantity to 36% on average.

In conclusion, the experimental results on financial data are suggesting that the minmax regret decreases the return variability while maintaining the realized profits approximately to the same level. Moreover, the minmax regret is characterized by more diversification and less turnover than the CVaR. It is worth to note that this results has been obtained using a very simple rule-of-thumb to generate scenarios, that is, using a rolling window on the past days and without any attempt to detect the adverse scenarios that the decision maker may wish to hedge.

Table 5: Computational results with real returns, $|A| = 30$

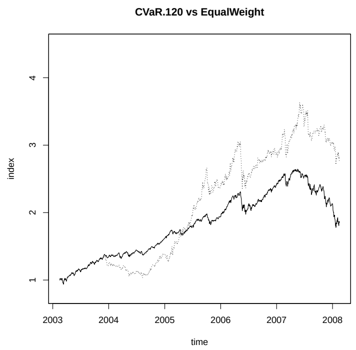
| | Iterations | | | Times | | | Sizes | | |
|---|---|---|---|---|---|---|---|---|---|
| Scenarios | LP1 | LP2 | LP3 | LP1 | LP2 | LP3 | LP1 | LP2 | LP3 |
| 20 | 754 | 547 | 564 | 0.148 | 0.101 | 0.289 | 2051, 1481 | 6091, 3541 | 231, 309 |
| 40 | 1066 | 961 | 765 | 0.269 | 0.270 | 0.453 | 4071, 2781 | 12151, 6721 | 231, 332 |
| 60 | 2125 | 1526 | 1663 | 0.503 | 0.511 | 0.932 | 6091, 4691 | 18211, 11121 | 231, 476 |
| 80 | 2341 | 1501 | 1758 | 0.575 | 0.606 | 1.045 | 8111, 5951 | 24271, 14221 | 231, 384 |
| 100 | 3123 | 1643 | 2274 | 0.847 | 0.752 | 1.297 | 10131, 7421 | 30331, 17741 | 231, 459 |
| 120 | 3901 | 1983 | 2876 | 0.989 | 0.769 | 1.623 | 12151, 8901 | 36391, 21281 | 231, 557 |
| 140 | 4383 | 1814 | 2843 | 1.262 | 0.933 | 1.746 | 14171, 10351 | 42451, 24761 | 231, 450 |
| 160 | 6061 | 2425 | 3345 | 1.500 | 1.074 | 2.105 | 16191, 12141 | 48511, 28921 | 231, 463 |
| 180 | 6411 | 2172 | 3761 | 1.726 | 1.245 | 2.379 | 18211, 13511 | 54571, 32241 | 231, 578 |
| 200 | 6080 | 3016 | 4152 | 1.755 | 1.431 | 2.696 | 20231, 14801 | 60631, 35401 | 231, 585 |
| 220 | 8471 | 2722 | 4323 | 1.928 | 1.459 | 2.855 | 22251, 16411 | 66691, 39201 | 231, 570 |
| 240 | 10006 | 3663 | 4715 | 2.156 | 1.963 | 3.140 | 24271, 17821 | 72751, 42601 | 231, 601 |
| 260 | 8229 | 3736 | 4841 | 2.321 | 1.843 | 3.319 | 26291, 19841 | 78811, 47221 | 231, 614 |
| 280 | 10079 | 3109 | 5175 | 2.390 | 1.992 | 3.615 | 28311, 20601 | 84871, 49321 | 231, 641 |
| 300 | 10051 | 3625 | 5818 | 2.589 | 2.183 | 4.108 | 30331, 22371 | 90931, 53441 | 231, 688 |
| 320 | 11138 | 4362 | 5633 | 2.935 | 2.318 | 4.514 | 32351, 23821 | 96991, 56921 | 231, 634 |
| 340 | 12741 | 4663 | 5883 | 2.981 | 2.534 | 4.476 | 34371, 25561 | 103051, 60981 | 231, 653 |
| 360 | 14492 | 4416 | 6193 | 3.376 | 2.996 | 4.906 | 36391, 26641 | 109111, 63721 | 231, 673 |
| 380 | 13344 | 5442 | 6429 | 3.347 | 3.151 | 5.091 | 38411, 28351 | 115171, 67721 | 231, 696 |
| 400 | 14506 | 4658 | 6755 | 3.544 | 3.019 | 5.389 | 40431, 29611 | 121231, 70821 | 231, 704 |
| 420 | 13403 | 5142 | 7211 | 3.597 | 3.381 | 5.917 | 42451, 31201 | 127291, 74581 | 231, 744 |
| 440 | 15268 | 5885 | 7739 | 4.088 | 3.728 | 6.315 | 44471, 32781 | 133351, 78321 | 231, 777 |
| 460 | 16594 | 5360 | 7755 | 4.279 | 3.591 | 6.643 | 46491, 34251 | 139411, 81841 | 231, 736 |
| 480 | 16135 | 5825 | 7318 | 4.612 | 3.808 | 6.667 | 48511, 35771 | 145471, 85461 | 231, 731 |
| 500 | 17138 | 8010 | 8678 | 4.656 | 4.792 | 7.356 | 50531, 37211 | 151531, 88921 | 231, 819 |

Table 6: Portfolio return distribution as the number of scenarios increases.

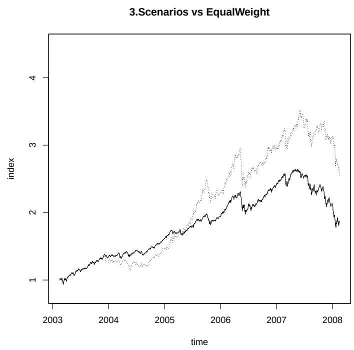
| Model | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|---|---|---|---|---|---|---|
| CVaR | -6.99 | -0.54 | 0.08 | 0.09 | 0.72 | 7.15 |
| 2_sc | -6.02 | -0.48 | 0.10 | 0.09 | 0.71 | 5.00 |
| 3_sc | -6.00 | -0.46 | 0.07 | 0.08 | 0.68 | 4.75 |
| 4_sc | -6.23 | -0.45 | 0.09 | 0.08 | 0.69 | 4.96 |
| 5_sc | -5.70 | -0.43 | 0.09 | 0.09 | 0.69 | 4.61 |
| 6_sc | -5.44 | -0.42 | 0.09 | 0.09 | 0.70 | 4.47 |
| 7_sc | -5.48 | -0.42 | 0.11 | 0.10 | 0.70 | 4.32 |
| 8_sc | -5.48 | -0.40 | 0.10 | 0.09 | 0.68 | 4.46 |

Table 7: Portfolio indexes as the number of scenarios increases.

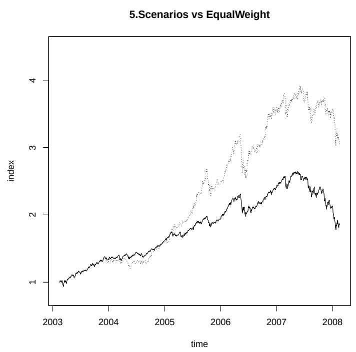
| Model | H index | Max index | Turn over |
|---|---|---|---|
| CVaR | 0.435 | 0.546 | 0.524 |
| 2_sc | 0.287 | 0.399 | 0.461 |
| 3_sc | 0.279 | 0.381 | 0.413 |
| 4_sc | 0.259 | 0.361 | 0.391 |
| 5_sc | 0.244 | 0.351 | 0.387 |
| 6_sc | 0.229 | 0.335 | 0.357 |
| 7_sc | 0.221 | 0.325 | 0.367 |
| 8_sc | 0.217 | 0.322 | 0.369 |



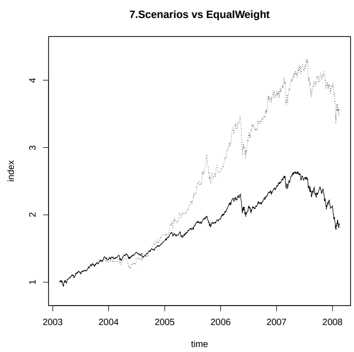

(a) Comparison between the CVaR and the Equal Weight portfolios




(b) Comparison between the 3-scenarios and the Equal Weight portfolios




(c) Comparison between the 5-scenarios and the Equal Weight portfolios




(d) Comparison between the 7-scenarios and the Equal Weight portfolios

Fig. 2: Comparison between the Scenarios (dotted lines) and the Equal Weight portfolios (dark line)

## 7 Concluding remarks

In this paper a new portfolio optimization model is studied in which the decision maker is hedged against market uncertainties by using two different safeguards:

- The minimization of a discrepancy measure (regret) between our portfolio performance and the optimal one under each possible of the considered market scenarios.
- A set of constraints on the allowed *conditional Value-at-Risk* under this set of scenarios.

In the literature we can find enough examples about how the CVaR measure can be adapted to the decision maker attitude to risk. As was said in the paper, the parameter $\beta \in (0, 1)$ can be defined by the investor under the premise that the larger the $\beta$-level is, the more aversion to risk will be transferred to the optimization model.

More novelty implications are carried out by including the minmax regret safeguard to the model in particular in what is referred to the choice of the set of scenarios. Here, the decision maker can use her previous knowledge on the market conditions to construct the set of scenarios in a strategic way. In our paper we just use a plain rule-of-thumb of a constant rolling window on past data to define these scenarios. It would be interesting to explore the possibility of modeling scenarios under different technical criteria about the market conditions and compare the financial properties transferred to the corresponding proposed portfolios.

Another interesting idea is the automatization of that process of selecting scenarios from data. A number of clustering techniques formalize the notion of clusters through the probability distribution of the corresponding sample of economic variables. These so-called model-based tools (see [?] and the references therein) will classify our data set of economic indicators into a finite set of scenarios in which the corresponding set of data values may be considered as a sample of independent and identically distributed random vectors. In our paper, these economic indicators have been given through variations of the asset prices during specific trading periods (returns). In this way, for any given portfolio, we can obtain random samples of our specific effectiveness measure allowing us to approximate theoretical (under known random distributions) measures as means, medians, variances or quantiles using the corresponding sample estimators and the law of large numbers.

Regarding the computational applications of our model to real data, we have analyzed three different Linear Programming formulations by conducting several empirical experiments. While the numerical results show that all of them can be solved efficiently under large sets of historical data and a considerable number of scenarios, we have also highlighted several differences. The observed differences mainly refer to the number of iterations needed by the solver, the amount of memory used due to the size of the different formulations and the CPU times. This may suggest that different formulations should be use in different situations according to our computational resources or the speed of action needed. In any case, what has been clearly evidenced by our study is that the proposed portfolios improve the experimental ex-post risk, as they are more diversified than the standard portfolios. This opens new research lines in which the model can be extended to other risk measures, such that standard deviation, maximum loss, Value-at-Risk or Gini index, and other efficiency measures, such as the median instead of the mean.

## References

1. Acerbi, C., Nordio, C., and Sirtori, C. Expected Shortfall as a Tool for Financial Risk Management. *https://arxiv.org/abs/cond-mat/0102304v1* (03 2001).
2. Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. Coherent measures of risk. *Mathematical Finance 9*, 3 (1999), 203–228.
3. Balbás, A., Balbás, R., and Mayoral, S. Portfolio choice and optimal hedging with general risk functions: A simplex-like algorithm. *European Journal of Operational Research 192*, 2 (2009), 603–620.
4. Benati, S. The optimal portfolio problem with coherent risk measure constraints. *European Journal of Operational Research 150*, 3 (2003), 572–584.
5. Benati, S. The computation of the worst conditional expectation. *European Journal of Operational Research 155* (2004), 414–425.
6. Benati, S. Heuristic methods for the optimal statistic median problem. *Computers and Operations Research 38*, 1 (2011), 379–386.

7. Benati, S. Using medians in portfolio optimization. *Journal of the Operational Research Society 66*, 5 (2015), 720–731.
8. Benati, S., and Rizzi, R. A mixed integer linear programming formulation of the optimal mean/Value-at-Risk portfolio problem. *European Journal of Operational Research 176*, 1 (2007), 423–434.
9. Bouveyron, C., and Brunet-Saumard, C. Model-based clustering of high-dimensional data: A review. *Computational Statistics & Data Analysis 71* (2014), 52 – 78.
10. Cohen, M., Lobel, R., and Perakis, G. Dynamic Pricing through Data Sampling. *Production and Operations Management 27*, 6 (2018), 1074–1088.
11. Cormen, T.H., Leiserson, C.E., Rivest, R.L., and Stein, C. Introduction to algorithms. *3rd ed. Cambridge: The MIT Press.* (2009).
12. DeMiguel, V., Garlappi, L., and Uppal, R. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *Review of Financial Studies 22*, 5 (2009), 1915–1953.
13. Grechuk, B., Molyboha, A., and Zabarankin, M. Maximum entropy principle with general deviation measures. *Mathematics of Operations Research 34*, 2 (2009), 445–467.
14. Grechuk, B., Molyboha, A., and Zabarankin, M. Mean-Deviation Analysis in the Theory of Choice. *Risk Analysis 32*, 8 (2012), 1277–1292.
15. Grechuk, B., and Zabarankin, M. Sensitivity analysis in applications with deviation, risk, regret, and error measures. *SIAM Journal on Optimization 27*, 4 (2017), 2481–2507.
16. Iyengar, G., and Ma, A. Fast gradient descent method for Mean-CVaR optimization. *Annals of Operations Research 205*, 1 (2013), 203–212.
17. Konno, H., Akishino, K., and Yamamoto, R. Optimization of a long-short portfolio under nonconvex transaction cost. *Computational Optimization and Applications 32*, 1-2 (2005), 115–132.
18. Krokhmal, P., Palmquist, J., and Uryasev, S. Portfolio optimization with conditional Value-at-Risk criterion. *Journal of Risk 4* (2002), 43–68.
19. Lim, A., Shanthikumar, J., and Vahn, G.-Y. Robust portfolio choice with learning in the framework of regret: Single-period case. *Management Science 58*, 9 (2012), 1732–1746.
20. Natarajan, K., Shi, D., and Toh, K.-C. A probabilistic model for minmax regret in combinatorial optimization. *Operations Research 62*, 1 (2014), 160–181.
21. Ogryczak, W. and Tamir, A. Minimizing the sum of the k largest functions in linear time. *Information Processing Letters 85*, 3 (2003), 117–122.
22. Rappoport, P. A new approach: Average shortfall. *Technical report, J.P. Morgan*, (1993).
23. Rockafellar, R., and Uryasev, S. Optimization of conditional value-at-risk. *Journal of Risk 2* (2000), 21–41.
24. Rockafellar, R., and Uryasev, S. Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance 26*, 7 (2002), 1443–1471.
25. Rockafellar, R., Uryasev, S., and Zabarankin, M. Generalized deviations in risk analysis. *Finance and Stochastics 10*, 1 (2006), 51–74.
26. Rockafellar, R., Uryasev, S., and Zabarankin, M. Optimality conditions in portfolio analysis with general deviation measures. *Mathematical Programming 108*, 2-3 (2006), 515–540.
27. Zemel, E. An O(n) algorithm for the linear multiple choice knapsack problem and related problems. *Information Processing Letters 18*, 3 (1984), 123–128.
28. Zieliński, R. Uniform strong consistency of sample quantiles. *Statistics and Probability Letters 37*, 2 (1998), 115–119.
29. Zymler, S., Rustem, B., and Kuhn, D. Robust portfolio optimization with derivative insurance guarantees. *European Journal of Operational Research 210*, 2 (2011), 410–424.