PORTFOLIO SELECTION MODELS: A REVIEW AND NEW DIRECTIONS

Diana Roman, Gautam Mitra*

November 11, 2008

*CARISMA: The Centre for the Analysis of Risk and Optimisation Modelling Applications, School of Information Systems, Computing and Mathematics, Brunel University, UK and Optirisk Systems, One Oxford Road, Uxbridge, UB9 4DA, UK

Contents

Abstract

Modern Portfolio Theory (MPT) is based upon the classical Markowitz model which uses variance as a risk measure. A generalisation of this approach leads to mean-risk models, in which a return distribution is characterised by the expected value of return (desired to be large) and a "risk" value (desired to be kept small). Portfolio choice is made by solving an optimisation problem, in which the portfolio risk is minimised and a desired level of expected return is specified as a constraint. The need to penalise different undesirable aspects of the return distribution led to the proposal of alternative risk measures, notably those penalising only the downside part (adverse) and not the upside (potential). The downside risk considerations constitute the basis of the Post Modern Portfolio Theory (PMPT). Examples of such risk measures are lower partial moments, Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR). We revisit these risk measures and the resulting mean-risk models. We discuss alternative models for portfolio selection, their choice criteria and the evolution of MPT to PMPT which incorporates: utility maximisation and stochastic dominance.

1 Introduction and Motivation

Portfolio selection is well known as a leading problem in finance; since the future returns of assets are not known at the time of the investment decision, the problem is one of decision-making under risk. A portfolio selection model is an ex-ante decision tool: decisions taken today can only be evaluated at a future time, once the uncertainty regarding the assets' returns is revealed.
Formally, the problem can be stated as follows: given a set of n assets in which we may invest, how to divide *now* an amount of money W amongst these assets, such that, *after a specified period of time* T, to obtain a return on investment as high as possible?
The future returns of the assets are random variables; let us denote these by $R_1, \ldots R_n$. A portfolio is denoted by $x = (x_1, \ldots x_n)$ where x_j is the fraction of the capital invested in asset j, $j = 1 \ldots n$. $(x_1, \ldots x_n$ are called "portfolio weights"; these are the required investment decisions).
To represent a portfolio, the weights $(x_1, \ldots x_n)$ must satisfy a set of constraints that form a set X of feasible decision vectors. The simplest way to define a feasible set is by the requirement that the weights are non-negative and sum to 1 (meaning, buy to hold and and short selling is not allowed). For this basic version of the problem, the set of feasible decision vectors is

$$X = \{(x_1, \ldots x_n) | \sum_{j=1}^{n} x_j = 1, \quad x_j \geq 0, \forall j = 1 \ldots n\}. \tag{1}$$

(In fact, X may be a general linear programming (LP) feasible set given in a canonical form as a system of linear equations, without changing the analysis.)
The return of a portfolio $x = (x_1, \ldots x_n)$ is a random variable, denoted by R_x, and its dependence on individual asset returns is expressed by the relation:
$R_x = x_1 R_1 + \ldots + x_n R_n$.
Three main issues arise, leading to three main research areas:

1. How is the distribution of the random returns represented?
2. How to choose between payoff distributions?
3. Once an investment decision is made, when / how to rebalance it?

It is usual in portfolio selection to represent the random returns as discrete random variables, given by their realisations under a specified number S of states of the world (scenarios), with some corresponding probabilities of occurrence $p_1, \ldots, p_S$.
(We denote the S possible outcomes of the random return R_j, $j = 1 \ldots n$, by $r_{1j}, \ldots, r_{Sj}$. Thus, the return R_x of a portfolio x is a discrete random variable with S possible outcomes $R_{1x}, \ldots R_{Sx}$, occurring with probabilities $p_1, \ldots, p_S$, where $R_{ix} = x_1 r_{i1} + \ldots + x_n r_{in}$, for all $i = 1 \ldots S$.)
How to obtain such a discrete representation is beyond the scope of this paper; the interested reader may

find a good review in [16].

This paper tries sets out to address the question 2 which is posed above. The (discrete joint) distribution of the random returns $R_1, \ldots, R_n$ is considered to be known: we denote by r_{ij} the return of asset j under scenario i, $i = 1 \ldots S$, $j = 1 \ldots n$. We consider only single-period investment problems; thus, the third question posed above is also out of the scope of this paper.
The well established models for choosing among payoff distributions are: mean-risk models and expected utility maximisation / stochastic dominance. They first define a preference relation (choice criterion) among random variables. Based on this choice criterion, the investment decisions are taken by solving optimisation problems. In the rest of the paper we consider in detail formulation of such decision problems.
This paper gives a review of the available models for choice and of the risk measures used in portfolio selection. It is known that some models, although theoretically questionable, are widely used in practice by the finance community. Other models, although popular and justified by the research community, have not been much applied in practice. Lastly, there are models that are theoretically sound, but difficult to implement. We do not intend to plead for a particular approach, but just give an objective summary of what is available in portfolio choice and outline the new directions.

The rest of the paper is structured as follows. Section 2 presents a historical background on models for choice under risk. In Section 3, mean-risk models and risk measures are presented. We review variance and alternative risk measures, their advantages and disadvantages. The corresponding optimisation models are formulated. Section 4 concerns the expected utility approach. In Section 5 stochastic dominance is introduced and its role in portfolio selection is explained. Conclusions and new directions are presented in Section 6.

2 Historical background

The starting point in modeling choice under risk is utility theory, whose basic idea dates from at least 1738 (see [6], translation of a paper originally published in 1738). In [6], Bernoulli assumed that individuals possessed a utility function for outcomes and thus a utility value for random variables can be inferred, which could order the random variables. However, he did not prove it to be a rational criterion for making choices. It was only in the 1940's and in an economic context when it was proved that the expected utility axioms implied the existence of an "expected utility rule" for random variables, which is a rational criterion for ordering them. More than any others, von Neumann and Morgenstern's work ([29]) introduced the expected utility concept to decision theory.
In spite of its theoretical and intuitive appeal, expected utility theory has had little applicability in practice, mostly because the choice of an appropriate utility function is somewhat subjective.
Mean-risk models were first proposed in the early fifties in order to provide a practical solution for the portfolio selection problem. It is much easier to rank random variables and make choices just using a few numbers of attributes (statistics) describing the payoff distributions considered. This idea lies at the heart of mean-risk models, in which distributions are described by just two parameters ("mean", i.e. expected value of return and "risk"). In his seminal work "Portfolio Selection", Markowitz ([25]) proposed variance as a risk measure. Moreover, he introduced it in a computational model, by measuring the risk of a portfolio via the covariance matrix associated with individual asset returns; this leads to a quadratic programming formulation. This was far from being the final answer to the problem of portfolio selection. For example, a natural question is to ask is how meaningful are the solutions obtained with this model and what is the best measure of risk. Variance as a risk measure has been criticized, mostly for its symmetric nature and for the lack of consistency with axiomatic models of choice (stochastic dominance). Since then, alternative risk measures have been proposed.
Asymmetric risk measures have been proposed since symmetric risk measures do not intuitively point to risk as an undesirable result (a "bad outcome"). Symmetric risk measures penalise favourable (upside) deviations from the mean (or from any other target) in the same way they penalise unfavourable (downside) deviations. Aware of these criticisms, Markowitz ([26]) himself proposed the semi-variance as a risk measure. Independently of Markowitz, Roy ([44]) proposed as a risk measure the Safety-First Criterion (which is the probability that the portfolio return falls below a predefined disaster level). This was the first occurrence of

the so-called "below target" risk measures. Lower partial moments were introduced in the 1970's by Bawa and Fishburn ([5], [12]). They constitute the generalized case for "below target" risk measures. Fishburn ([12]) developed the (α, τ) model, which is a mean-risk model using lower partial moments as a risk measure. The target semi-variance is one example of risk measure in the category of lower partial moments.
Nevertheless, symmetric risk measures were not abandoned. Yitzhaki ([53]) introduced and analysed the mean-risk model using the Gini's mean difference as a risk measure. Konno and Yamazaki ([18]) analysed the complete model using mean absolute deviation (MAD) as a risk measure. Using MAD was just one of the attempts to linearize the portfolio optimisation procedure ([47]), following the pioneering work of Sharpe ([46]). Until late seventies and early eighties computational methods for the solution of large scale quadratic programs were not well developed ([34]). Today, though solving quadratic programming problems is no longer difficult, MAD remains an alternative to the mean-variance model.
Asymmetric risk measures were brought into attention again by Ogryczak and Ruszczynski ([31], [32]). They analysed the mean-risk model using central semi-deviations, with semi-absolute deviation and semi-variance as special cases. Central semi-deviations differ from lower partial moments due to the fact that the target for the outcomes is not fixed, but distribution-dependent (the expected value).

An important step was the introduction of risk measures concerned only with extremely unfavourable results, or, in other words, with the left tail of a distribution. In 1993, in the G-30 Report ([30]), Value-at-Risk (VaR) was proposed, with the precise task of answering the question: "how much can be the loss on a given horizon with a given probability?" The report helped shape the emerging field of financial risk management and underlined the importance of measuring risk for regulatory purposes, not only as a parameter in a model of choice. Since then, VaR has become the most widely used risk measure for regulatory purposes, especially since 1994, when JP Morgan introduced its free service "Risk Metrics" ([28]). However, VaR has several drawbacks; for example, it fails to reward diversification (since it is not sub-additive) and it is difficult to optimise (since it is not a convex function of the portfolio weights $x_1, \ldots, x_n$). In order to overcome these difficulties, recent research on risk measures focused on proposing axiomatic characterizations of risk measures and proposing new risk measures that satisfy these axioms.
In 1997, Artzner, Delbaen, Eber and Heath ([4]) introduced the concept of coherence as a set of desirable properties for risk measures concerned with the tail of distribution.
Among the most important coherent risk measures is Conditional Value-at -Risk (CVaR) ([39], [40]), named Expected Shortfall by some authors ([1]). Enhanced CVaR measures were proposed in [24].

Since it has been often argued that describing a distribution by just two parameters involves great loss of information, there have been attempts to improve on classical mean-risk models by using three parameters (instead of two) in order to characterise distributions ([17], [14], [15], [43]).
In spite of the amount of research on risk measures and mean-risk models, the question of which risk measure is most appropriate is still open. Furthermore, the theoretical soundness of the mean-risk models continues to be questioned.

The concept of stochastic dominance (SD) or stochastic ordering of random variables was inspired by earlier work in the theory of majorization ([27]), that is, ordering of real-valued vectors. It has been used since the early 1950's, in the fields of statistics (see for example [22]). In economics, stochastic dominance was introduced in the 1960's; Quirk and Saposnik ([37]) considered the first order stochastic dominance relation and demonstrated the connection to utility functions. Second order stochastic dominance was brought to economics by Hadar and Russel ([13]) and third order stochastic dominance by Whitmore ([50]). A detailed discussion is given in ([19]). Second order stochastic dominance (SSD) is generally regarded as a meaningful choice criterion in portfolio selection, due to its relation to models of risk-averse economic behaviour. The theoretical attractiveness of stochastic dominance, but also its limited applicability due to computational difficulty, has been widely recognised (see [51] and references therein). Generally, stochastic dominance has not been considered able to deal with situations where the number of possible choices (i.e. random variables to be compared) was infinite; thus, until very recently, it has not been considered a valid alternative to portfolio construction but just a powerful concept for analysis and investigation.
During the last few decades, one of the major concerns in decision-making under risk has been finding a way to combine the computational ease of mean-risk models with the theoretical soundness of stochastic dominance. Most of the efforts in this direction have been in finding risk measures consistent with stochastic

dominance rules, meaning that the efficient solutions of the corresponding mean-risk models are also efficient with respect to stochastic dominance rules and thus optimal for important classes of decision-makers. Risk measures consistent with SD were proposed in [12], [31], [32], [53], [52].
In the last few years, new models for portfolio selection that use SSD as a choice criterion (and are computationally tractable) have been proposed ([7], [8], [42], [10]).

3 Mean-risk models and risk measures

With mean-risk models, return distributions are characterised and compared by considering two scalars for each distribution. One scalar is the expected value (mean); large expected returns are desirable. The other scalar is the value of a "risk measure". Loosely speaking, a risk measure is a function that associates to each distribution a number which describes its "riskiness"; obviously, small risk values are desirable. Preference among distributions is defined using a trade-off between mean and risk.
In what follows, $E(\cdot)$ denotes the expected value operator.

Definition 1 *Consider two feasible portfolios $x = (x_1, \dots x_n)$ and $y = (y_1, \dots y_n)$ with returns R_x and R_y respectively. In the mean-risk approach with the risk measure denoted by ρ, the random variable R_x dominates (is preferred to) the random variable R_y if and only if: $E(R_x) \geq E(R_y)$ and $\rho(R_x) \leq \rho(R_y)$) with at least one strict inequality.*

Alternatively, we may say that portfolio x is preferred to portfolio y.

The intention is to find those portfolios that are non-dominated with respect to the preference relation described above; meaning, there is no other feasible portfolio that can improve on both mean return and risk.
Formally: A feasible portfolio $x = (x_1, \dots x_n)$ is non-dominated in the mean-risk approach (risk measure denoted by ρ) if, for any other feasible portfolio y, we cannot have both $E(R_x) \leq E(R_y)$ and $\rho(R_x) \geq \rho(R_y)$, unless $E(R_x) = E(R_y)$ and $\rho(R_x) = \rho(R_y)$.
Thus, a non-dominated portfolio (called also an "efficient" portfolio) has the lowest level of risk, for a given mean return (and the highest mean return, for a given level of risk.)
The efficient portfolios are obtained by solving optimisation problems, that can be formulated in three alternative ways. The most common formulation is based on the use of a specified minimum target on the portfolio's expected return while minimising the portfolio's risk:

$$\begin{array}{rl} \min & \rho(R_x) \\ \text{Subject to:} & E(R_x) \geq d \quad \text{and} \quad x \in X \end{array} \tag{2}$$

where d represents the desired level of expected return for the portfolio (chosen by the decision-maker).
Varying d and repeatedly solving the corresponding optimisation problem identifies the minimum risk portfolio for each value of d[1]. These are the efficient portfolios that compose the efficient set. By plotting the corresponding values of the objective function and of the expected return respectively in a risk-return space, we trace out the efficient frontier.

An alternative formulation, which explicitly trades risk against the return in the objective function, is:

$$\begin{array}{rl} \max & E(R_x) - \lambda\rho(R_x) \qquad (\lambda \geq 0) \\ \text{Subject to:} & x \in X \end{array} \tag{3}$$

Varying the trade-off coefficient λ and repeatedly solving the corresponding optimization problems traces out the efficient frontier.
Alternatively, we can maximise the portfolio expected return while imposing a maximum level of risk.

[1]Solving (2) without the constraint on the expected return identifies the absolute minimum risk portfolio. Denote by d_1 the expected return of this portfolio. On the other hand, the maximum possible expected return of a portfolio formed with the considered assets is obtained by investing everything in the asset with the highest expected return. Denote by d_2 this highest expected return among the component assets. Obviously, d should be varied between d_1 and d_2. Choosing d less than d_1 will only result in the absolute minimum risk portfolio. Choosing d higher than d_2 results in infeasibility.

One of these efficient portfolios should be chosen for implementation.

The specific portfolio depends on the investor's taste for risk: there are low risk - low return portfolios, as well as medium risk - medium return and high risk - high return portfolios.

How to measure risk and the choice of an appropriate risk measure in portfolio selection have been subject of continuous research and much debate. It is obvious that the risk measure used plays an important role in the decision-making process; portfolios chosen using different risk measures can be quite different.

Risk measures can be classified into two categories.
Risk measures that consider the deviation from a target form the first category. (In [2], they are called "risk measures of the first kind"). The target could be fixed (e.g. a minimal acceptable return), distribution-dependent (e.g. the expected value) or even a stochastic benchmark (e.g. an index). Such risk measures can only have positive values. These risk measures can be further divided into: symmetric (two-sided) risk measures and asymmetric (one-sided, downside, shortfall) risk measures. Symmetric risk measures quantify risk in terms of probability-weighted dispersion of results around a pre-specified target, usually the expected value. Measures in this category penalize negative as well as positive deviations from the target. Commonly used symmetric risk measures are variance and MAD. Asymmetric risk measures quantify risk according to results and probabilities below target values. Only the cases in which the outcomes of are less than the target value are penalised. It has been argued that this is much more in accordance with the intuitive idea about risk, as an undesirable result, an "adverse" outcome (see for example the experimental study [49], in which it is concluded that most investors associate risk with failure to attain a specific target). The global idea about downside risk is that the left-hand side of a return distribution involves risk while the right-hand side contains better investment opportunities. Among asymmetric risk measures, lower partial moments and central semi-deviations are of great importance.
The risk measure in the second category consider the overall seriousness of possible losses. They are sometimes called "tail risk measures" because they only consider the left tail of distributions, corresponding to largest losses. Only a certain number of worst outcomes are taken into consideration (this number depends on a specified confidence level α, e.g. if $\alpha = 0.05$ only the worst 5% of the outcomes are considered). Commonly used risk measures in this category are Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR). The values of these risk measures can be both positive and negative .

In what follows, we review some of the most commonly used risk measures.

3.1 Variance

Historically, variance was the first risk measure used in portfolio selection. Markowitz ([25]) introduced the mean-variance approach, which was later generalised to mean-risk models. Moreover, he introduced a computational model to find the efficient portfolios in this approach.
Variance has been widely used in statistics as an indicator for the spread around the expected value. Its square root, the standard deviation, is the most common measure of statistical dispersion.
The variance of a random variable R_x (denoted by $\sigma^2(R_x)$) is defined as its second central moment: the expected value of the square of the deviations of R_x from its own mean ($\sigma^2(R_x) = E[((R_x - E(R_x))^2]$.)
Among the properties of variance, the one used in calculating the variance of a linear combination of random variables is of particular importance:

$$\sigma^2(aR_1 + bR_2) = a^2\sigma^2(R_1) + b^2\sigma^2(R_2) + 2abCov(R_1, R_2) \tag{4}$$

where R_1, R_2 are random variables, a, b are real numbers, and $Cov(R_1, R_2)$ is the covariance of R_1 and R_2: $Cov(R_1, R_2) = E[(R_1 - E(R_1))(R_2 - E(R_2))]$.

Relation (4) above is particularly useful in the context of portfolio optimisation. It allows us to express the variance of the portfolio return $R_x = x_1R_1 + \ldots + x_nR_n$, resulting from choice $x = (x_1, \ldots, x_n)$ as:

$$\sigma^2(R_x) = \sum_{j=1}^{n}\sum_{k=1}^{n} x_j x_k \sigma_{jk} \tag{5}$$

where σ_{jk} denotes the covariance between R_j and R_k.
Thus, the portfolio variance is expressed as a quadratic function of the required decisions $x_1, \ldots, x_n$.

A mean-variance efficient portfolio is found as the optimal solution of the following quadratic program (QP):

$$\begin{aligned} \min \quad & \sum_{j=1}^{n}\sum_{k=1}^{n} x_j x_k \sigma_{jk} \\ \text{such that:} \quad & x \in X, \\ & \sum_{j=1}^{n} x_j \mu_j \geq d \end{aligned} \tag{MV}$$

where $\mu_j = E(R_j)$ is the expected return of asset j, $j = 1, \ldots, n$ and d is a target expected return for the portfolio (chosen by the decision maker). The last constraint in the above QP imposes a minimum expected value d for the portfolio return.
It may be noticed that the mean-variance model does not require the whole set of scenario returns $\{r_{ij}, i = 1 \ldots S, j = 1 \ldots n\}$ as parameters, but only the expected returns and the covariances between the component assets.

3.2 Mean Absolute Deviation

The mean absolute deviation of a random variable R_x is defined as the expected value of the absolute values of the deviations from the mean: $MAD(R_x) = E[|(R_x - E(R_x)|]$.
MAD is also a common measure of statistical dispersion; conceptually, it is very similar to variance. In the practice of portfolio selection, it was introduced mainly because the mean-variance model was difficult to solve for large data sets, due to the quadratic objective function (5). Using MAD instead of variance overcame this problem since the mean-MAD model is a linear program (LP). At present, with modern solvers, large quadratic programming (QP) models can be routinely solved. Nevertheless, using MAD as a risk measure remains an alternative to the classical mean-variance model. For a detailed analysis of the mean-risk model using the MAD measure, see [18]. The mean-MAD model can be formulated as follows:

$$\begin{aligned} \min \quad & \frac{1}{S}\sum_{i=1}^{S} y_i \\ & \text{Subject to:} \\ & \sum_{j=1}^{n} (r_{ij} - \mu_j) x_j \leq y_i, \quad \forall i \in \{1 \ldots S\} \\ & \sum_{j=1}^{n} (r_{ij} - \mu_j) x_j \geq -y_i, \quad \forall i \in \{1 \ldots S\} \\ & y_i \geq 0, \quad \forall i \in \{1 \ldots S\} \\ & \sum_{j=1}^{n} \mu_j x_j \geq d; \quad x \in X. \end{aligned} \tag{MAD}$$

For this model, in addition to the decision variables $x_1, \ldots, x_n$ representing the portfolio weights, there are S decision variables y_i, $i = 1 \ldots S$ representing the absolute deviations of the portfolio return R_x from its

expected value, for every scenario $i \in \{1 \dots S\}$:
$y_i =| \sum_{j=1}^{n} (r_{ij} - \mu_j)x_j |, \forall i \in \{1 \dots S\}$.

3.3 Lower Partial Moments and Central Semideviations

Sometimes, the use of a symmetric risk measure may lead to counter-intuitive results; we illustrate this with an example.

Consider a mean-risk model with a symmetric risk measure such as the variance and two random variables R_x and R_y. R_x has one certain outcome W_0. R_y has 2 outcomes: W_0 and W_1, each with equal probability. Let $W_1 > W_0$. Obviously R_y should be preferred since it yields at least as much as R_x. However, according to a mean-variance model, neither one is preferred, since, although R_x has a smaller expected value, it also has zero risk.

Asymmetric or below-target risk measures provide a better representation of risk as an undesirable outcome. Lower partial moments is a generic name for asymmetric risk measures that consider a fixed target.

Lower partial moments measure the expected value of the deviations below a *fixed* target value τ. They are described by the following formulation:

Let τ be a predefined (investor-specific) target value for the portfolio return R_x and let $\alpha > 0$. *The lower partial moment of order α around τ* of a random variable R_x with distribution function F is defined as:

$$LPM_\alpha(\tau, R_x) = E\{[\max(0, \tau - R_x)]^\alpha\} = \int_{-\infty}^{\tau} (\tau - r)^\alpha \, dF(r) \tag{6}$$

Often the "normalised version" is considered in the literature: $LPM_\alpha(\tau, R_x) = [E\{[\max(0, \tau - R_x)]^\alpha\}]^{\frac{1}{\alpha}}$. The role of τ is unambiguous; every decision-maker sets his own target below which he does not want the return to fall ($\tau = 0$, i.e. no losses, is the most common example). The role of α is not so straightforward. In [12], Fishburn analysed the mean-risk model with the risk measure defined by the lower partial moment of order α around τ, which he called the (α, τ) model. He proved the connection between the choice of α and the decision-maker's feelings about falling short of τ by various amounts:

Proposition 1 *Let $d > 0$, R_x a degenerate random variable having the only possible outcome $(\tau - d)$ with probability 1 and R_y a non-degenerate random variable with the same mean (i.e. $E(R_x) = E(R_y) = \tau - d$) having τ as one of the outcomes. For example, let the possible outcomes for R_y be τ and $\tau - 2d$, each of them with probability 0.5. Then:*

- *R_x is preferred to R_y in the (α, τ) model if and only if $\alpha > 1$.*
- *R_y is preferred to R_x in the (α, τ) model if and only if $\alpha < 1$.*
- *There is indifference between R_x and R_y in the (α, τ) model if and only if $\alpha = 1$.*

In other words, if the main concern is failure to meet the target without particular regard to the amount, the decision maker (DM) is willing to take a risk in order to minimise the chance that the return falls short of τ. In this case, choosing a small α for measuring risk is appropriate. If small deviations below the target are relatively harmless when compared to large deviations, the DM prefers to fall short of τ by a little amount than to take a risk that could result in a big loss. In this case, a larger α is indicated. To summarise, α is a parameter describing the investor's risk-aversion. The larger α, the more risk-averse is the investor.

We provide the formulation of the most commonly used below-target risk measures using the (α, τ) formulation:

1. **Safety First:** when $\alpha \to 0$.

$$SF(R_x) = LPM_{\alpha \to 0}(\tau, R_x) = E\{[\max(0, \tau - R_x)]^{\alpha \to 0}\}$$

   SF is a shortfall probability which measures the chances of the portfolio return falling below some predefined disaster level τ: $SF(R_x) = P(R_x < \tau)$.

2. **Expected Downside Risk:** when $\alpha = 1$.

$$LPM_1(\tau, R_x) = E[\max(0, \tau - R_x)]$$

3. **Target Semi-Variance:** when $\alpha = 2$.

$$LPM_2(\tau, R_x) = E\{[\max(0, \tau - R_x)]^2\}$$

Central semi-deviations are similar to the lower partial moments; however, they measure the expected value of the deviations below the mean (while the lower partial moments measure the expected value of the same deviations below a fixed target value).
The central semi-deviation of order α ($\alpha = 1, 2, \ldots$) of a random variable R_x is defined as:

$$CSD_\alpha(R_x) = E\{[\max(0, E(R_x) - R_x)]^\alpha\}. \tag{7}$$

The most famous risk measures in this category are those obtained for $\alpha = 1$: the absolute semi-deviation, and for $\alpha = 2$: the semi-variance (Semi-variance was first introduced by Markowitz (1959)).

- The absolute semi-deviation: $\delta^-(R_x) = E[\max(0, E(R_x) - R_x)]$
- The standard semideviation: $\sigma^-(R_x) = [E\{[\max(0, E(R_x) - R_x)]^2\}]^{\frac{1}{2}}$

A detailed analysis of these measures is given in [31], [32].

Remark 1 *The sum of the deviations below the mean is equal to the sum of the deviations above the mean, which implies that*
$MAD(R_x) = 2 \cdot \delta^-(R_x)$*. Thus, the absolute semi-deviation measure and the MAD measure are equivalent, in the sense that both measures provide exactly the same ranking of random variables.*

We present below the formulation of the mean-target semivariance model (meaning, the risk measure involved is the lower partial moment of order 2 around a fixed target return τ) . For this model, in addition to the decision variables $x_1, \ldots, x_n$ representing the portfolio weights, there are S decision variables, representing the magnitude of the negative deviations of the portfolio return R_x from τ, for every scenario $i \in \{1 \ldots S\}$ ($\sum_{j=1}^{n} r_{ij}x_j$ is the portfolio return under scenario i):

$$y_i = \begin{cases} \tau - \sum_{j=1}^{n} r_{ij}x_j, & \text{if } \sum_{j=1}^{n} r_{ij}x_j \leq \tau; \\ 0, & \text{otherwise.} \end{cases}$$

The mean-target semivariance formulation is as follows.

$$\min \quad \frac{1}{S}\sum_{i=1}^{S} y_i^2 \tag{M-TSV}$$

Subject to:

$$\tau - \sum_{j=1}^{n} r_{ij}x_j \leq y_i, \quad \forall i \in \{1 \dots S\}$$

$$y_i \geq 0, \quad \forall i \in \{1 \dots S\}$$

$$\sum_{j=1}^{n} \mu_j x_j \geq d; \quad x \in X.$$

Obviously, if the objective function in (M-TSV) is replaced by $1/S\sum_{i=1}^{S} y_i$ we obtain the mean-expected downside risk formulation.

The mean-semivariance formulation is similar to (M-TSV); however, instead of the fixed target τ, there is the distribution dependent target $E(R_x) = \sum_{j=1}^{n} x_j\mu_j$, where μ_j is the expected return of asset j, $j = 1 \dots n$. Thus, for this formulation, in addition to the variables $x_1, \dots x_n$, there are S decision variables, representing the magnitude of the negative deviations of the portfolio return R_x from its expected value, for every scenario $i \in \{1 \dots S\}$:

$$y_i = \begin{cases} \sum_{j=1}^{n}(\mu_j - r_{ij})x_j, & \text{if } \sum_{j=1}^{n} r_{ij}x_j \leq \sum_{j=1}^{n}\mu_j x_j; \\ 0, & \text{otherwise.} \end{cases}$$

$$\min \quad \frac{1}{S}\sum_{i=1}^{S} y_i^2 \tag{M-SV}$$

Subject to:

$$\sum_{j=1}^{n}(\mu_j - r_{ij})x_j \leq y_i, \quad \forall i \in \{1 \dots S\}$$

$$y_i \geq 0, \quad \forall i \in \{1 \dots S\}$$

$$\sum_{j=1}^{n} \mu_j x_j \geq d; \quad x \in X.$$

While in theory the order α can take any positive value, in practice it seems reasonable to implement $\alpha = 1$ (which leads to a linear program) or $\alpha = 2$ (which leads to a quadratic program). Other values of α would result in an optimisation problem that is more difficult to solve.

3.4 Value-at-Risk

Let R_x be a random variable describing the return of a portfolio x over a given holding period and $A\% = \alpha \in (0,1)$ a percentage which represents a sample of "worst cases" for the outcomes of R_x. Values of α close to 0 are of interest (e.g., $\alpha = 0.01 = 1\%$ or $\alpha = 0.05 = 5\%$).
The calculation of Value-at-Risk (VaR) at level α of random variable R_x (or of portfolio x) indicates that, with probability of at least $(1-\alpha)$, the loss[2] will not exceed VaR over the given holding period. At the same time, the probability of loss exceeding VaR is strictly greater than α.

[2] The loss is considered relative to 0: negative returns are considered as positive losses. Thus, the "loss" associated with the random variable R_x is described by the random variable $-R_x$.

Mathematically, the Value-at-Risk at level α of R_x is defined using the notion of α-quantiles. Below there are stated definitions and properties regarding quantiles, which are important in understanding the definition of VaR. In the definitions below, F is the cumulative distribution function of the random variable R_x.

Definition 2 *An* α-quantile *of* R_x *is a real number* r *such that*
$P(R_x < r) \leq \alpha \leq P(R_x \leq r)$.

Definition 3 *The* lower α-quantile *of* R_x*, denoted by* $q_\alpha(R_x)$ *is defined as*
$q_\alpha(R_x) = \inf\{r \in \mathbb{R} : F(r) = P(R_x \leq r) \geq \alpha\}$.

Definition 4 *The* upper α-quantile *of* R_x*, denoted by* $q^\alpha(R_x)$ *is defined as*
$q^\alpha(R_x) = \inf\{r \in \mathbb{R} : F(r) = P(R_x \leq r) > \alpha\}$.

The set of α-quantiles of R_x is denoted by
$Q_\alpha(R_x) = \{r \in \mathbb{R} : P(R_x < r) \leq \alpha \leq P(R_x \leq r)\}$.
Obviously, the set of α-quantiles of R_x may be written as:
$Q_\alpha(R_x) = \{r \in \mathbb{R} : P(R_x \leq r) \geq \alpha, P(R_x \geq r) \geq 1 - \alpha\}$.
The following two results are proven in [21].

Proposition 2 *The set of* α*-quantiles of* R_x *is a closed and non-empty interval whose left and right ends are* $q_\alpha(R_x)$ *and* $q^\alpha(R_x)$ *respectively.*

Proposition 3 $Q_\alpha(R_x) = -Q_{1-\alpha}(-R_x)$,
where, for a general set A*, we define* $-A = \{-a : a \in A\}$.

The Proposition 3 above states that the α-quantiles of R_x are the $(1 - \alpha)$ - quantiles of $-R_x$. From this result it follows immediately that:

$$q_\alpha(R_x) = -q^{1-\alpha}(-R_x) \tag{8}$$
$$q^\alpha(R_x) = -q_{1-\alpha}(-R_x) \tag{9}$$

Definition 5 *The* Value-at-Risk at level α of R_x *is defined as the negative of the upper* α*-quantile of* R_x*:*
$VaR_\alpha(R_x) = -q^\alpha(R_x)$.

The minus sign in the definition of VaR is because $q^\alpha(R_x)$ is likely to be negative, but when speaking of "loss", absolute values are considered.
Considering (9), VaR can be also defined as: $VaR_\alpha(R_x) = q_{1-\alpha}(-R_x)$.
Thus,

$$VaR_\alpha(R_x) = -q^\alpha(R_x) = q_{1-\alpha}(-R_x). \tag{10}$$

In (10), VaR is expressed as a return. It can be expressed in terms of monetary value once the initial monetary value of the portfolio x is given; for example, if $VaR_\alpha(R_x) = 0.1$, then, expressed as a monetary value, VaR is 10% of the intial investment value.

Remark 2 $VaR_\alpha(R_x) = q_{1-\alpha}(-R_x)$ *means that*
$VaR_\alpha(R_x) = \inf\{r \in \mathbb{R} : P(-R_x \leq r) \geq 1 - \alpha\}$.
This explains why VaR is referred to as "the maximum portfolio loss with probability (at least) $(1 - \alpha)$*".*
On the other hand, the lower $(1 - \alpha)$*-quantile* $q_{1-\alpha}(-R_x)$ *can be expressed as:*
$q_{1-\alpha}(-R_x) = \sup\{r \in \mathbb{R} : P(-R_x < r) < 1 - \alpha\}$ *(see [21]). Since* $P(-R_x < r) < 1 - \alpha$*, it follows that* $P(-R_x \geq r) > \alpha$*. Thus, sometimes VaR is referred to as "the minimum loss with probability strictly exceeding* α*."*

If the cumulative distribution function F is continuous and strictly increasing, $q^\alpha(R_x)$ is simply the unique value of r satisfying $F(r) = \alpha$ (see Figure 1).
This is not always the situation. The equation $F(r) = \alpha$ may have no solutions (see Figure 2) or multiple solutions (see Figure 3).

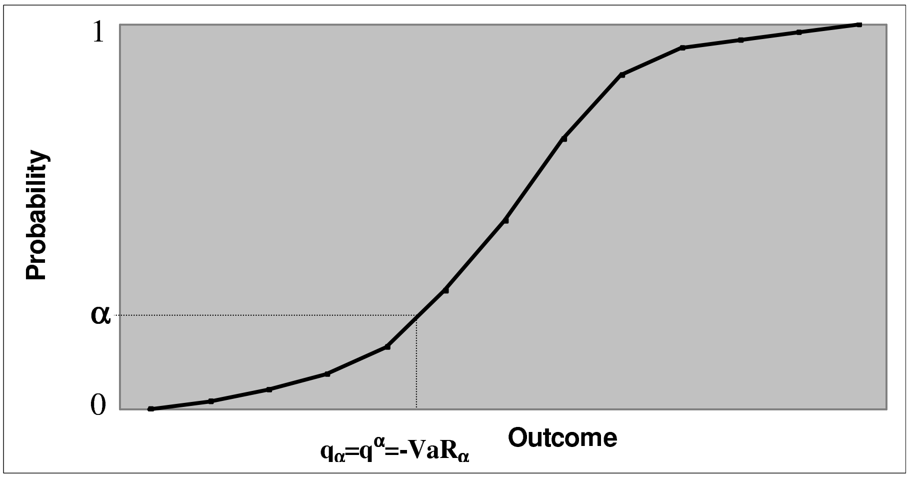

Figure 1: A continuous and strictly increasing cumulative distribution function

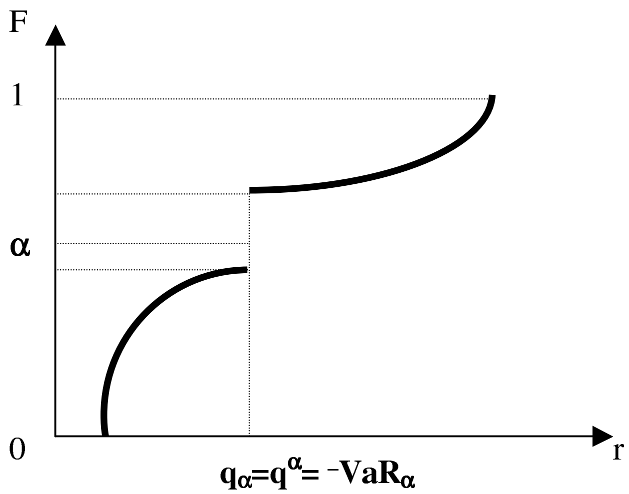

Figure 2: The equation $F(r) = \alpha$ has no solutions.

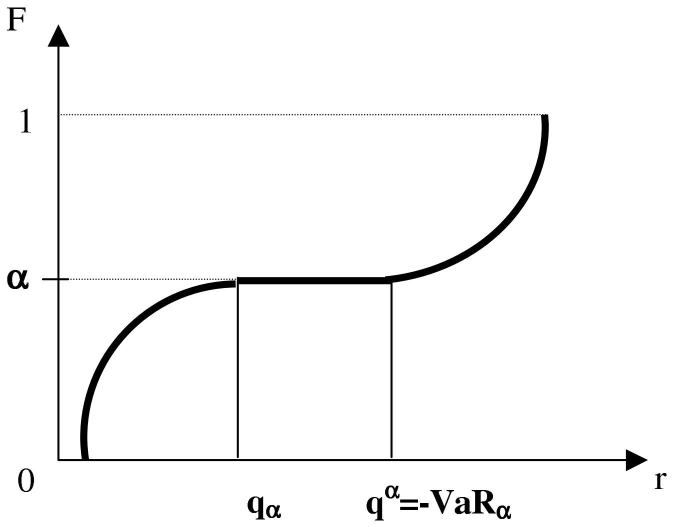

Figure 3: The equation $F(r) = \alpha$ has multiple solutions.

Remark 3 *Although the calculation of VaR_α indicates that with probability at least $\alpha = A\%$ the losses will exceed this value, VaR fails to measure the extent of losses beyond the confidence level. It is incapable of distinguishing between losses that are just a little bit worse and losses that can be overwhelming.*

We consider now the special case when the random variable R_x is discrete, with a finite number of possible outcomes. As presented in the Introduction, this is the representation used for random returns in order to lead to tractable portfolio optimisation models. In this case F (the cumulative distribution function of R_x) is a step function; thus, the equation $F(r) = \alpha$ can have either no solutions or an infinity of solutions (see Figure 4).
Assume we have m possible states (m scenarios) with probabilities $p_1, \dots, p_m$ (strictly positive and their sum equals 1). Then, the distribution of R_x is concentrated in finitely many points $r_1, \dots, r_m$, with probabilities $p_1, \dots, p_m$. Without loss of generality, we can assume that these points are in ascending order, so we have $r_1 \leq r_2 \leq \dots \leq r_m$. Then:
$F(r) = 0$, for $r < r_1$;
$F(r) = p_1 + \dots + p_{k-1}$, for $r \in [r_{k-1}, r_k)$, $\forall k \in \{2 \dots m\}$;
$F(r) = 1$, for $r \geq r_m$.
There are two possible cases:

1. Equation $F(r) = \alpha$ has no solutions. In this case, there is a unique $k \in \{1 \dots m\}$ such that $p_0 + \dots + p_{k-1} < \alpha < p_0 + \dots + p_k$, with the convention $p_0 = 0$. Then, $\{r : P(R_x < r) \leq \alpha\} = \{r_1, \dots, r_k\}$ and $\{r : P(R_x \leq r) \geq \alpha\} = \{r_k, r_{k+1}, \dots, r_m\}$. Thus, there is just one α-quantile: r_k. $VaR_\alpha(R_x) = -r_k$. For example, in Figure 4, $\alpha \in (p_1 + p_2, p_1 + p_2 + p_3)$, so $VaR_\alpha(R_x) = -r_3$.

2. Equation $F(r) = \alpha$ has an infinity of solutions. In this case, there is a unique $k \in \{1 \dots m\}$ such that $\alpha = p_1 + \dots + p_k$. $\{r : P(R_x < r) \leq \alpha\} = \{r_1, \dots, r_{k+1}\}$ and $\{r : P(R_x \leq r) \geq \alpha\} = \{r_k, r_{k+1}, \dots, r_m\}$. Thus, the set of α-quantiles is $[r_k, r_{k+1}]$. The upper α-quantile is r_{k+1} and thus $VaR_\alpha(R_x) = -r_{k+1}$.

The two cases above are briefly described as follows:

Proposition 4 *Consider a discrete random variable R_x with m possible outcomes $r_1 \leq r_2 \leq \dots \leq r_m$ occurring with probabilities $p_1, \dots, p_m$. For every $\alpha \in (0, 1)$, there is an unique index $k_\alpha \in \{1 \dots m\}$ such that*

$$\sum_{k=0}^{k_\alpha - 1} p_k \leq \alpha < \sum_{k=0}^{k_\alpha} p_k.$$

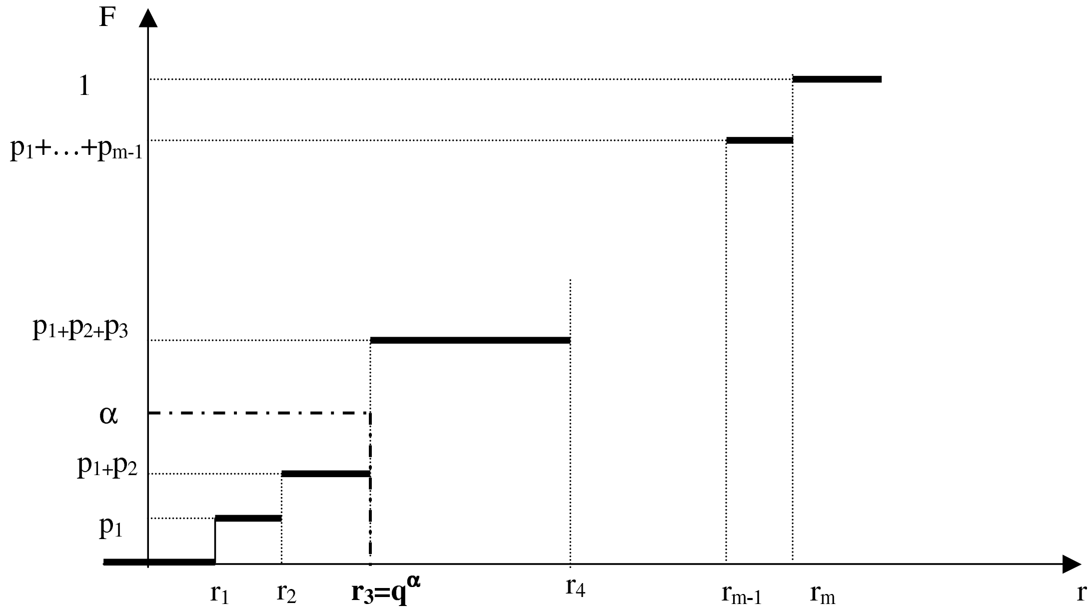

Figure 4: The distribution function of a discrete random variable.

(with the convention $p_0 = 0$).
In this case, $VaR_\alpha(R_x) = -r_{k_\alpha}$.

Remark 4 *Notice the discontinuity in the behaviour of VaR with respect to the choice of confidence level $\alpha \in (0,1)$: it could have big jumps for a slight variation of α. For example, in Figure 4, $VaR_\alpha(R_x)$ is the same for any choice of α in an interval of the form $[p_1+p_2+\ldots+p_{k-1}, p_1+p_2+\ldots+p_k)$ ($VaR_\alpha(R_x) = -r_k$), but, when $\alpha = p_1+p_2+\ldots+p_k$, $VaR_\alpha(R_x) = -r_{k+1}$.*

Thus, calculating the VaR of a given portfolio x with a return R_x can be done as follows:

- If R_x is assumed to have a specific continuous distribution (the most notable case is the normal distribution), VaR is found as the α-quantile of the given distribution.
- If the distribution of R_x is obtained by simulation, thus represented as discrete, with a finite number of outcomes, then VaR is one of the worst possible (depending on the confidence level α) outcomes of R_x.

Although widely used in finance for evaluating the risk of portfolios, VaR has several undesirable properties. It is not a subadditive measure of risk (see Section...)
which means that the risk of a portfolio can be larger than the sum of the stand-alone risks of its components when measured by VaR - see [4] for more details and see [48] for examples showing the lack of subadditivity of VaR. In addition, VaR is not convex with respect to choice x, thus it is difficult to optimise with standard available methods ([20], [23] and references therein). (Convexity is an important property in optimisation and in particular it eliminates the possibility of a local minimum being different from a global minimum ([38])). In spite of considerable research in this regard, optimising VaR is still an open problem. For heuristic algorithms on VaR optimisation, see for example [20], [23] and references therein. Thus, the main use of VaR is for regulatory purposes rather than as a parameter in a model of choice.

3.5 Conditional Value-at-Risk

As before, let R_x denote the random return of a portfolio x over a given holding period and $A\% = \alpha \in (0,1)$ a percentage which represents a sample of "worst cases" for the outcomes of R_x (usually, $\alpha = 0.01 = 1\%$ or $\alpha = 0.05 = 5\%$).

The definition of CVaR at the specified level α is the mathematical transcription of the concept "average of losses in the worst $A\%$ of cases" ([1]), where the "loss" associated with R_x is described by the random variable $-R_x$.
CVaR is approximately equal to the average of losses greater than or equal to VaR (at the same confidence level α); in some cases, the equality is exact.

If there exists a unique r such that $P(R_x \leq r) = \alpha$, the situation is simple: VaR is the negative of r and CVaR is the expected loss given that the loss is greater than VaR. If R_x is a continuous random variable, this is always the situation. (Also in the case of multiple values of r such that $P(R_x \leq r) = \alpha$, CVaR is the conditional expectation of losses beyond VaR ([1]).
In the case that there does not exist a value r such that $P(R_x \leq r) = \alpha$ (see for examples Figure 2 and Figure 4), an α-quantile $q^\alpha(R_x)$ is considered. However, $P(R_x \leq q^\alpha(R_x)) > \alpha$ and thus the "excess probability" $P(R_x \leq q^\alpha(R_x)) - \alpha$, corresponding to $q^\alpha(R_x)$, is extracted from the conditional expectation of outcomes below the α-quantile. The formal definition of CVaR is stated below:

Definition 6 The Conditional Value-at-Risk at level α of R_x *is defined as:*

$$CVaR_\alpha(R_x) = -\frac{1}{\alpha}\{E(R_x 1_{\{R_x \leq q^\alpha(R_x)\}}) - q^\alpha(R_x)[P(R_x \leq q^\alpha(R_x)) - \alpha]\} \tag{11}$$

where

$$1_{\{Relation\}} = \begin{cases} 1, & \textit{if Relation is true;} \\ 0, & \textit{if Relation is false.} \end{cases}$$

(thus, $R_x 1_{\{R_x \leq q^\alpha(R_x)\}}$ *is obtained from* R_x *by considering only the outcomes below the upper* α*-quantile* $q^\alpha(R_x)$*.)*

Remark 5 *In [1] it is shown that in the definition* (11) *above,* $q^\alpha(R_x)$ *may be replaced by any* α*-quantile; the quantity* $CVaR_\alpha(R_x)$ *depends only on the distribution of* R_x *and the level* α *considered.*

In [39] and [40] the Conditional Value-at-Risk at level α of R_x is defined as minus the mean of the α-tail distribution of R_x, where the α-tail distribution is the one with distribution function defined by:

$$F_\alpha(r) = \begin{cases} \frac{F(r)}{\alpha}, & \text{if } r < q^\alpha(R_x); \\ 1, & \text{if } r \geq q^\alpha(R_x) \end{cases} \tag{12}$$

F_α truly is another distribution function, meaning that it is non-decreasing and right continuous, with $F_\alpha(r) \to 0$ as $r \to -\infty$ and $F_\alpha(r) \to 1$ as $r \to \infty$ ([40]); thus, the α-tail is well defined by (12). This definition of CVaR is equivalent to (11). The minus sign is because the tail mean is likely to be negative, but when speaking of "loss", absolute values are considered.

Remark 6 *To be more precise, in [39] and [40], CVaR is defined in a more general case, by considering a random variable* L_x *that describes the loss associated with the decision vector* x*. If we consider losses only as negative outcomes of the original random variable* R_x *and gains only as positive outcomes, i.e.* $L_x = -R_x$ *(the most general case), the same definition* (11) *is obtained.*

Remark 7 *In several papers (for example [1], [48]), the value defined by* (11) *is called the "Expected Shortfall". In [1], a value named Conditional Value-at-Risk is defined as the optimum value of an optimisation problem (see Proposition 6) and it is proved that this value is equal to the Expected Shortfall as defined by* (11)*. Thus, in [1] it is pointed out that Expected Shortfall and Conditional Value-at-Risk are two different names for the same concept of the risk measure.*

Remark 8 *For a discrete random variable* R_x*, the* α*-tail is obtained by considering only the outcomes less than or equal to an* α*-quantile; the probabilities attached to the outcomes strictly less than the* α*-quantile are the original probabilities divided by* α *and the probability attached to the* α*-quantile is such that the sum of all probabilities is equal to one.*

Example 1 *Consider the case presented in Figure 4: a discrete random variable R_x with m possible outcomes $r_1 \leq r_2 \leq \ldots \leq r_m$ having probabilities $p_1, \ldots, p_m$ and $\alpha \in (p_1 + p_2, p_1 + p_2 + p_3)$. In this case, $q_\alpha(R_x) = q^\alpha(R_x) = r_3 = -VaR_\alpha(R_x)$.*
The α-tail is the random variable which has outcomes r_1, r_2 and r_3 with probabilities p_1/α, p_2/α and $(\alpha - p_1 - p_2)/\alpha$ respectively. $CVaR_\alpha(R_x)$ is the negative of the expected value of the α- tail:

$$CVaR_\alpha(R_x) = -\frac{1}{\alpha}[r_1p_1 + r_2p_2 + r_3(\alpha - p_1 - p_2)]$$

The same result is obtained following directly the definition (11) *of CVaR:*
$E(R_x 1_{\{R_x \leq q^\alpha(R_x)\}}) = r_1p_1 + r_2p_2 + r_3p_3$.
$P(R_x \leq q^\alpha(R_x)) = p_1 + p_2 + p_3$.
Thus,

$$\begin{aligned} CVaR_\alpha(R_x) &= -\frac{1}{\alpha}[r_1p_1 + r_2p_2 + r_3p_3 + r_3(\alpha - p_1 - p_2 - p_3)] \\ &= -\frac{1}{\alpha}[r_1p_1 + r_2p_2 + r_3(\alpha - p_1 - p_2)]. \end{aligned}$$

Example 2 *Consider again the case presented in Figure 4, this time with $\alpha = p_1 + p_2 + p_3$. Then, $q_\alpha(R_x) = r_3$ and $q^\alpha(R_x) = r_4$. The α-tail has outcomes r_1, r_2 and r_3 with probabilities p_1/α, p_2/α and p_3/α respectively. The expected value of the α-tail is minus $CVaR_\alpha$:*

$$CVaR_\alpha(R_x) = -\frac{1}{\alpha}[r_1p_1 + r_2p_2 + r_3p_3]$$

Following the definition (11) *of CVaR the same value is obtained. Consider first the lower α-quantile $q_\alpha(R_x)$ in this definition:*
$E(R_x 1_{\{R_x \leq q_\alpha(R_x)\}}) = r_1p_1 + r_2p_2 + r_3p_3$.
$P(R_x \leq q_\alpha(R_x)) = p_1 + p_2 + p_3 = \alpha$.
Thus, $CVaR_\alpha(R_x) = -\frac{1}{\alpha}[r_1p_1 + r_2p_2 + r_3p_3]$.
Consider now the upper α-quantile $q^\alpha(R_x)$ in the definition (11)*:*
$E(R_x 1_{\{R_x \leq q^\alpha(R_x)\}}) = r_1p_1 + r_2p_2 + r_3p_3 + r_4p_4$.
$P(R_x \leq q^\alpha(R_x)) = p_1 + p_2 + p_3 + p_4 = \alpha + p_4$.
Thus, $CVaR_\alpha(R_x) = -\frac{1}{\alpha}[r_1p_1 + r_2p_2 + r_3p_3 + r_4p_4 - r_4p_4] = -\frac{1}{\alpha}[r_1p_1 + r_2p_2 + r_3p_3]$.

In some earlier papers, $CVaR_\alpha$ is defined as the conditional expectation of losses beyond VaR_α. This definition and the definition (11) of CVaR lead to the same value only in particular situations (e.g. for continuous distributions) and not in the general case (see Proposition 5 in [40]). To illustrate this, consider again the case presented in Example 1.

$$\begin{aligned} &- E[R_x|R_x \leq -VaR_\alpha(R_x)] = -\sum_{k=1}^{m} r_k P(R_x = r_k | R_x \leq r_3) \\ &= -\left(r_1 \frac{P(R_x = r_1)}{P(R_x \leq r_3)} + r_2 \frac{P(R_x = r_2)}{P(R_x \leq r_3)} + r_3 \frac{P(R_x = r_3)}{P(R_x \leq r_3)} \right) \\ &= -\frac{r_1p_1 + r_2p_2 + r_3p_3}{p_1 + p_2 + p_3}. \end{aligned}$$

In addition,

$$-E[R_x|R_x < -VaR_\alpha(R_x)] = -\frac{r_1p_1 + r_2p_2}{p_1 + p_2}.$$

None of the two values above is the same with

$$CVaR_\alpha(R_x) = -\frac{1}{\alpha}[r_1p_1 + r_2p_2 + r_3(\alpha - p_1 - p_2)].$$

In [40], the values $-E[R_x|R_x \leq -VaR_\alpha(R_x)]$ and $-E[R_x|R_x < -VaR_\alpha(R_x)]$ are called lower CVaR and upper CVaR respectively. The same paper presents interesting properties of these risk measures and their connections with CVaR.

We can summarize the formulae for VaR_α and $CVaR_\alpha$ in the discrete case as follows:

Proposition 5 (VaR and CVaR for discrete distributions) *Assume that the distribution of R_x is concentrated in finitely many points $r_1, \ldots, r_m$, with probabilities $p_1, \ldots, p_m$. Without loss of generality, we can assume that these points are in ascending order, so we have $r_1 \leq r_2 \leq \ldots \leq r_m$.*
For every $\alpha \in (0,1)$, there is a unique index $k_\alpha \in \{1 \ldots m\}$ such that

$$\sum_{k=0}^{k_\alpha - 1} p_k \leq \alpha < \sum_{k=0}^{k_\alpha} p_k.$$

(with the convention $p_0 = 0$).
In this case,

$$VaR_\alpha(R_x) = -r_{k_\alpha}$$

and

$$CVaR_\alpha(R_x) = -\frac{1}{\alpha}[\sum_{k=0}^{k_\alpha - 1} p_k r_k + r_{k_\alpha}(\alpha - \sum_{k=0}^{k_\alpha - 1} p_k)].$$

An important result, proved by Rockafellar and Uryasev ([40]), is that the CVaR of a portfolio x can be calculated by solving a convex optimisation problem. Moreover, very importantly, CVaR can be optimised over the set of feasible decision vectors (feasible portfolios) and this is also a convex optimisation problem. These results are summarised in Proposition 6 below.

Proposition 6 (CVaR calculation and optimisation) *Let R_x be a random variable depending on a decision vector x that belongs to a feasible set X as defined by (1) and let $\alpha \in (0,1)$. We denote by the $CVaR_\alpha(x)$ the CVaR of the random variable R_x for confidence level α. Consider the function:*

$$F_\alpha(x, v) = \frac{1}{\alpha} E[-R_x + v]^+ - v,$$

where

$$[u]^+ = \begin{cases} u, & \text{if } u \geq 0; \\ 0, & \text{if } u < 0. \end{cases}$$

Then:

1. *As a function of v, F_α is finite and convex (hence continuous) and*

$$CVaR_\alpha(x) = \min_{v \in \mathbb{R}} F_\alpha(x, v).$$

   In addition, the set consisting of the values of v for which the minimum is attained, denoted by $A_\alpha(x)$, is a non-empty, closed and bounded interval (possibly formed by just one point), that contains $-VaR_\alpha(x)$.

2. *Minimising $CVaR_\alpha$ with respect to $x \in \mathcal{P}$ is equivalent to minimising F_α with respect to $(x, v) \in \mathcal{P} \times \mathbb{R}$:*

$$\min_{x \in \mathcal{P}} CVaR_\alpha(x) = \min_{(x,v) \in \mathcal{P} \times \mathbb{R}} F_\alpha(x, v).$$

   In addition, a pair (x^*, v^*) minimises the right hand side if and only if x^* minimises the left hand side and $v \in A_\alpha(x^*)$.

3. *$CVaR_\alpha(x)$ is convex with respect to x and F_α is convex with respect to (x, v).*

Thus, if the set X of feasible decision vectors is convex, minimising CVaR is a convex optimisation problem.

In practical applications, the random returns are usually described by their realisations under various scenarios; thus a portfolio return R_x is a discrete random variable. In this case, calculating and optimising CVaR is even simpler: the two (convex) optimisation problems above become linear programming problems. Indeed, suppose that R_x has S possible outcomes $r_{1x}, \ldots, r_{Sx}$ with probabilities $p_1, \ldots, p_S$, with $r_{ix} = \sum_{j=1}^{n} x_j r_{ij}$, $\forall i \in \{1 \ldots S\}$, as described in the introductory part (r_{ij} is the return of asset j under scenario i). Then:

$$F_\alpha(x, v) = \frac{1}{\alpha} \sum_{i=1}^{S} p_i [v - r_{ix}]^+ - v = \frac{1}{\alpha} \sum_{i=1}^{S} p_i [v - \sum_{j=1}^{n} x_j r_{ij}]^+ - v$$

Thus, in the formulation of the mean-CVaR model, in addition to the decision variables $x_1, \ldots x_n$ representing the portfolio weights, there are $S + 1$ decision variables. The variable v represents the negative of an α-quantile of the portfolio return distribution. Thus, when solving this optimisation problem, the optimal value of the variable v may be used as an approximation for VaR_α, and, in the case of a unique solution, is exactly equal to VaR_α. The other S decision variables represent the magnitude of the negative deviations of the portfolio return from the α-quantile, for every scenario $i \in \{1 \ldots S\}$:

$$y_i = \begin{cases} -v - \sum_{j=1}^{n} r_{ij} x_j, & \text{if } \sum_{j=1}^{n} r_{ij} x_j \leq -v; \\ 0, & \text{otherwise.} \end{cases}$$

The algebraic formulation of the mean-CVaR model is given below:

$$\min \quad v + \frac{1}{\alpha S} \sum_{i=1}^{S} y_i \tag{M-CVaR}$$

Subject to:

$$\sum_{j=1}^{n} -r_{ij} x_j - v \leq y_i, \quad \forall i \in \{1 \ldots S\}$$

$$y_i \geq 0, \quad \forall i \in \{1 \ldots S\}$$

$$\sum_{j=1}^{n} \mu_j x_j \geq d; \quad x \in X.$$

To conclude, CVaR has good theoretical and practical properties:

- It takes into account information about the right tail of the distribution of losses (corresponding to extreme losses);
- It behaves continuously as a function of $\alpha \in (0, 1)$ ([40], [1]);
- It is easy to compute and to optimise; for scenario models, it can be optimised using linear programming techniques;
- It is coherent in the sense of [4] (see Section....)

3.6 Axiomatic properties: coherence

The wide use of risk measures for regulatory and reporting purposes, led to a requirement for a formal specification of the properties that an acceptable risk measure should have.

Coherence is defined as a set of four desirable properties for "tail" risk measures [3] . Let V be a set of

[3]Similar desirable properties are defined for "deviation" risk measures, like variance, in [41].

random variables, e.g. representing future returns of portfolios. A risk measure $\rho : V \to \mathbb{R}$ is said to be *coherent* if it satisfies the following four conditions:

1. *Condition T (translation invariance):*

$$\rho(Z + C) = \rho(Z) - C, \qquad \forall Z \in V, \forall C \text{ constant.}$$

2. *Condition S (subadditivity):*

$$\rho(Y + Z) \leq \rho(Y) + \rho(Z), \qquad \forall Y, Z \in V.$$

3. *Condition PH (positive homogeneity):*

$$\rho(\lambda Z) = \lambda \rho(Z), \qquad \forall Z \in V, \forall \lambda \geq 0.$$

4. *Condition M (monotonicity):*

$$\text{If } Z \leq Y \text{ then } \rho(Y) \leq \rho(Z), \qquad \forall Y, Z \in V.$$

Similar desirable properties are defined for "deviation-type" risk measures, like variance. In ([41]) a new concept, that of "deviation measures", is defined. Formally, a deviation measure ρ is a risk measure of the first kind that satisfies four properties. Two of them are the subadditivity condition and the positive homogeneity condition stated above. The other two conditions are similar to the translation invariance and the monotonicity (stated above), but adapted for the case of deviation-type risk measures:

1. $$\rho(Z + C) = \rho(Z), \qquad \forall Z \in V, \forall C \text{ constant.}$$

2. $$\rho(Z) \geq 0, \qquad \forall Z \in V \text{ non-constant random variable and } \rho(Z) = 0 \text{ otherwise.}$$

4 Expected utility maximisation

In expected utility theory, a single scalar value (the expected utility) is attached to each random variable. Preference is then defined by comparing expected utilities with a larger value preferred.

A *utility function* is a real-valued function defined on real numbers (representing possible outcomes of a random variable, e.g. a portfolio return), which measures the relative value of outcomes.

Expected Utility Theory provides a basis for extending a utility function defined on real numbers (outcomes) to a utility function defined on random variables. A utility value is assigned to each random variable in terms of the utility values of its outcomes and the probabilities associated with these outcomes. The expected utility criterion is applicable to any decision problem under risk, provided that the decision maker is prepared to make decisions consistent with the Expected Utility Axioms ([51]).

Definition 7 *Given a utility function U,* the expected utility *of a random variable R_x is:*

$$E[U(R_x)] = \int_{-\infty}^{\infty} U(w)\, dF(w),$$

where F is the distribution function of R_x.

In the discrete case, if the random variable R_x has the outcomes $w_1, \ldots, w_S$ with probabilities $p_1, \ldots, p_S$, the expected utility of R_x is:
$E[U(R_x)] = p_1 U(w_1) + \ldots + p_S U(w_S)$.

In this approach, a random variable (or a portfolio) is *non-dominated* or *efficient* if and only if its expected utility is maximal (i.e. there is no other random variable with a greater expected utility).

The next issue to be addressed is the (form of) utility function that should be used, in order that the expected utility maximization approach leads to rational decisions. The use of utility functions in the economic context is based on an implied assumption that they reflect the behaviour of investors or decision-makers (DM). There are progressively stronger assumptions about the behaviour of investors, which lead to increasing requirements for the properties of the corresponding utility functions.

The first assumption on investors' behaviour is that they prefer more to less and are hence rational. In order to reflect this non-satiation attitude, the utility function must be nondecreasing. This is the only non-arguable condition for utility functions on wealth since all investors are assumed to be rational.

The second aspect to consider is investors' attitudes towards risk. There are three possible attitudes: risk-aversion, risk-neutrality and risk seeking, which can be illustrated by the following example, in which a choice is required between two random variables with the same expected value. The first random variable is a fair gamble, i.e. it has two possible outcomes w_1 and w_2, each with probability 0.5. The second random variable is a "sure thing", i.e. it has one possible outcome: $(w_1 + w_2)/2$, with probability 1.

If the DM is *risk averse*, he rejects the gamble, since he does not want to take a risk that, even though it could result in a higher outcome w_1, it could also result in a worse result w_2. Thus, the expected utility of the "sure thing" is greater for a decision maker who is risk- averse: $U\left(\frac{w_1+w_2}{2}\right) > \frac{1}{2}[U(w_1)+U(w_2)]$. If this inequality is true for all possible outcomes w_1, w_2, then U is strictly concave, meaning that $U[tw_1 + (1-t)w_2] \geq tU(w_1) + (1-t)U(w_2)$, $\forall t \in [0,1]$, with strict inequality whenever $w_1 \neq w_2$ and $t \in (0,1)$ ([51]).

Thus, a risk-averse decision-maker has a strictly concave utility function.[4] If the appropriate derivatives exist, strict concavity is equivalent to U' being strictly decreasing, or $U'' < 0$ (see [51] and references therein). Strict concavity requires that a segment uniting two points on the graph of U lies below the graph, and this means that the graph of U has a slope that "flattens" with increasing w, i.e. U' strictly decreasing (see Figure 5).

If the DM is *risk seeking*, he prefers the gamble: $U\left(\frac{w_1+w_2}{2}\right) < \frac{1}{2}[U(w_1) + U(w_2)]$. A risk-seeking decision-maker has a strictly convex utility function ([51]). Finally, if the DM is *risk neutral*, he is indifferent between the gamble and the sure thing: $U\left(\frac{w_1+w_2}{2}\right) = \frac{1}{2}[U(w_1) + U(w_2)]$, in which case his utility function is linear.

It is generally assumed that investors are risk averse, which implies a nondecreasing and concave utility function. This means that, as wealth increases, each additional increment is less valuable than the previous one.

A further assumption on the behaviour of rational and risk-averse investors is that they exhibit *decreasing absolute risk-aversion (DARA):* they are willing to take more risk when they are financially secure. A decision-maker manifests DARA if a small gamble becomes more attractive (or at least does not become less attractive) as his wealth increases. Formally, this is expressed using the notion of risk premium ([36]).

To reflect this attitude, the utility function must be such that A is a decreasing function of w, where $A = -\frac{U''}{U'}$ is *the Arrow-Pratt absolute risk aversion coefficient* ([36], [51]).

However, the DARA attitude is not so widely accepted in investment, compared to the risk-aversion attitude ([51]). Thus, it may be argued that A being a decreasing function of w is a strong condition imposed on utility functions.

A weaker condition imposed on utility functions is to have the third derivative positive: $U''' > 0$. It is obvious that this is a necessary condition for decreasing absolute risk aversion. It is not a sufficient

[4] In many papers and textbooks, this is in fact the definition of a risk-averse decision-maker: as one whose utility function is concave.

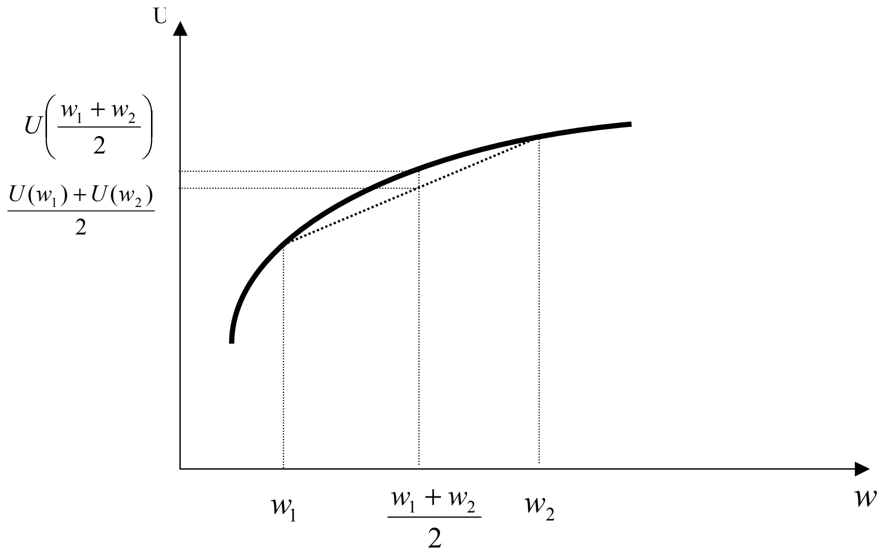

Figure 5: Risk aversion and concave utility functions

condition, however: there are functions with a positive third derivative that do not exhibit decreasing absolute risk aversion ([51]). The effect of $U''' > 0$, together with $U' > 0$ and $U'' < 0$ is to reduce the "concavity" of the utility function while the wealth is increasing. It may be considered as an intermediate stage between risk aversion and decreasing absolute risk aversion. A decision-maker whose utility function has a positive third derivative prefers positively skewed distributions ([3], [45]), thus he manifests what is called a *ruin-averse behaviour* or *prudence* ([33]).

In conclusion, the observed economic behaviour is risk-aversion and in order to represent this, utility functions have to be non-decreasing and concave. This is generally accepted in investment theory. Further modes of economic behaviour (e.g. decreasing risk aversion, "ruin-aversion") that would impose additional conditions on utility functions (e.g. reducing concavity with wealth level) are less observed and arguable.

Once a utility function is decided, the non-dominated choices (random variables) are obtained by solving an optimisation problem in which the expected utility is maximised.

In the portfolio selection problem, we solve an optimisation problem with decision variables $x_1, \ldots, x_n$:

$$\begin{array}{rl} \max & E[U(R_x)] \\ \text{Subject to:} & (x_1, \ldots, x_n) \in X \end{array}$$

A major difficulty with the expected utility maximisation is that the specification of a utility function is a subjective task. Moreover, two utility functions belonging to the same class (e.g. nondecreasing and concave) could lead to a different ranking of random variables. Stochastic dominance (or stochastic ordering of random variables) overcomes this difficulty.

5 Stochastic Dominance (SD)

Stochastic dominance ranks random variables under assumptions about general characteristics of utility functions that follow from prevalent modes of economic behaviour. Stochastic dominance ensures that all individuals, whose utility functions are in the same class, rank choices in the same way.
In the stochastic dominance approach two random variables are compared by pointwise comparison of some performance functions constructed from their distribution functions.

For a random variable R_x with a distribution function F_x, we define recursively:

$$F_x^{(1)} = F_x;$$
$$F_x^{(k)}(r) = \int_{-\infty}^{r} F_x^{(k-1)}(t)\, dt, \quad \forall r \in \mathbb{R}, \quad \text{for } k \geq 2.$$

.

Definition 8 *Let F_x and F_y be the cumulative distribution functions of R_x and R_y respectively and $k \in \mathbb{N}$, $k \geq 1$.* R_x is preferred to R_y with respect to k-th order stochastic dominance *(denoted: $R_x \succ_{(k)} R_y$) if and only if: $F_x^{(k)}(r) \leq F_y^{(k)}(r)$, $\forall r \in \mathbb{R}$, with at least one strict inequality.*

Alternatively, we may say that choice x is preferred to choice y with respect to k-th order stochastic dominance.

In portfolio selection, of particular importance is the second order stochastic dominance (SSD), due to its relation to models of risk-averse behaviour, as explained below.
There are progressively stronger assumptions about investors' behaviour (as described in section 4) that are used in expected utility theory, leading to first order stochastic dominance (FSD), second order stochastic dominance (SSD), and third order stochastic dominance (TSD). It is known that:

- $R_x \succ_{(1)} R_y$ if and only if $E[U(R_x)] \geq E[U(R_y)]$, for every *non-decreasing* utility function U, with at least one strict inequality. In this case, all rational investors prefer R_x to R_y.
- $R_x \succ_{(2)} R_y$ if and only if $E[U(R_x)] \geq E[U(R_y)]$, for every *non-decreasing and concave* utility function U, with at least one strict inequality. In this case, all rational and risk-averse investors prefer R_x to R_y.
- $R_x \succ_{(3)} R_y$ if and only if $E[U(R_x)] \geq E[U(R_y)]$, for every *non-decreasing and concave* utility function U whose *first derivative is convex* , with at least one strict inequality. In this case, all rational, risk-averse and ruin-averse investors prefer R_x to R_y.

(for a proof see [51])

Thus, second order stochastic dominance (SSD) describes the preference of rational and risk averse investors. Important results regarding SSD may be found in [31], [32].

Another stochastic dominance relation, which is related to decreasing absolute risk aversion (DARA), is discussed in the literature. R_x dominates R_y with respect to DARA stochastic dominance if and only if $E[U(R_x)] \geq E[U(R_y)]$, for every *non-decreasing and concave* utility function U whose *Arrow-Pratt absolute risk aversion coefficient is a decreasing function of* w - see Section 4. The interested reader may find more in [51]).

Remark 9 *SSD is stronger than FSD in the sense that it is able to order more pairs of random variables. We could have indifference between R_x and R_y with respect to FSD but prefer R_x or R_y with respect to SSD.*
The set of efficient solutions with respect to SSD is included in the set of efficient solutions with respect to FSD.

An interesting aspect is the connection of SD with lower partial moments (LPM). In [31], [32], Ogryczak and Ruszczynski pointed out that dominance with respect to k-th order SD relations ($k \geq 2$) is equivalent to dominance with respect to LPM of order $k-1$ for all possible targets. To be precise, they proved that, for a random variable R_x with performance functions $F_x^{(k)}$,

$$F_x^{(k)}(r) = \frac{1}{(k-1)!} LPM_{k-1}(r, R_x), \quad \forall r \in \mathbb{R}, \forall k \geq 2. \tag{13}$$

This means that $F_x^{(k)}$ measures the under-achievement of R_x with respect to all possible targets. Thus, finding the efficient solutions with respect to a SD relation is a multi-objective model with a continuum of objectives.
Relation 13 above leads to a well-known and useful characterisation of SSD:

$$R_x \succ_{(2)} R_y \Leftrightarrow E\left([t - R_x]_+\right) \ \leq \ \mathrm{E}\left([t - R_y]_+\right) \quad \forall t \in \mathbb{R} \tag{14}$$

where $[t - R_x]_+$ is $t - R_x$, if $R_x \leq t$, and 0 otherwise.
Thus, the SSD dominance may be expressed by comparing expected shortfalls with respect to *every* target t:

As a conclusion, theoretically, stochastic dominance relations are a sound choice criterion for making investment decisions, corresponding to observed economic behaviour. Unfortunately, they are difficult to apply in practice. The theory requires the decision-maker to take the complete distribution of outcomes into consideration. Only the comparison of two random variables with respect to SSD (and in general, to stochastic dominance relations) involves an infinite number of inequalities. Until recently, portfolio models based on SSD were considered intractable. Recently there have been some computational breakthroughs in respect of tractable portfolio models that use SSD as a choice criterion. Several authors proposed formulations and solution methods for such models ([7], [8], [42], [10]).

6 Risk measures consistent with stochastic dominance

An important research effort has been put into combining the practicality of mean-risk models with the theoretical soundness of SD relations. This effort mostly resulted in proposal of risk measures that are consistent with SD relations, in particular with SSD. This consistency guarantees that any efficient solution chosen by the corresponding mean-risk model is efficient for the whole class of risk-averse investors.

Definition 9 *A risk measure ρ is* consistent with the k-th order stochastic dominance *($k = 1, 2, 3$) if: $R_x \succ_{(k)} R_y \Rightarrow \rho(R_x) \leq \rho(R_y)$, for every pair of random variables R_x and R_y.*

We are mainly interested in consistency with SSD (thus, when $k = 2$).
It is well-known that: $R_x \succ_{(k)} R_y \quad (k = 1, 2) \Rightarrow E(R_x) \geq E(R_y)$ - see for example [51] for a proof. (In the case of FSD, i.e. $k = 1$, the inequality is strict.)
Thus, in the case of a risk measure ρ consistent with SSD, $R_x \succ_{(2)} R_y$ implies that R_x is preferred to R_x in the mean-risk model with risk defined by ρ. This implies that the efficient solutions of the mean-risk model with risk defined by ρ are also efficient with respect to SSD, thus efficient for the class of risk-averse decision makers. This aspect underlines the importance of risk measures' consistency with SSD. (Obviously, consistency with SSD implies consistency with FSD).

- In general, variance is not consistent with stochastic dominance relations, in particular with SSD (see [51]);
- In general, MAD is not consistent with stochastic dominance relations, in particular with SSD ([51]) - which means that a mean-MAD efficient solutions is not guaranteed to be efficient with respect to SSD. However, an important part of the mean-MAD efficient frontier (the upper part, corresponding to high mean - high risk portfolios) represents portfolios that are efficient to SSD (see [32]);
- Lower partial moments are consistent with stochastic dominance relations. In particular, lower partial moments of order 2 (with target semi-variance as a special case) are consistent with SSD. (Lower partial moments of order 1 are consistent with FSD, and lower partial moments of order 3 are consistent with TSD); (see [12]);
- Central semi-deviations (in particular, semi-variance) are not generally consistent with SSD. However, like in the case of MAD, a large portion of the efficient frontier of the mean-semivariance model is consistent with SSD;
- VaR (at any confidence level $\alpha \in (0, 1)$) is consistent with FSD but not with SSD ([35])
- CVaR (at any confidence level $\alpha \in (0, 1)$) is consistent with SSD ([35], [52]).

In table 1 the properties of the risk measures discussed so far are summarised. Columns 3 refers to axiomatic properties of risk measures. As specified in section 3.6, different conditions are required from deviation-type risk measures (variance, MAD, lower partial moments and central semideviations) and from tail risk measures (VaR and CVaR). For the first type of risk measures we checked if they satisfy the conditions of a "deviation measure" (described in section 3.6). We considered the normalised versions of variance and semivariance (meaning, standard deviation and semi standard deviation) since only the normalised versions could satisfy the positive homogeneity condition (required for all types of risk measures). For the second type of risk measures we checked if they are coherent.

7 Conclusions and new directions

In portfolio selection, a crucial issue is the model used for choosing between payoff (return) distributions. By far the most accepted and frequently applied approach is given by the mean-risk models, in which return distributions are characterised by two scalars: the expected value ("mean") and a risk value. The risk associated to a return distribution can be quantified in a variety of ways. A common

Risk measure	FSD Consistency	SSD Consistency	Axiomatic conditions
standard deviation	no	no	✓(deviation measure)
MAD	no	no	✓(deviation measure)
semi st deviation	no	no	✓(deviation measure)
target semi st deviation	✓	✓	no
VaR	✓	no	no
CVaR	✓	✓	✓(coherent)

Table 1: Risk measures and their properties.

approach (and the first one, chronologically speaking) is to use variance as a measure of risk. The mean-variance model is still the most used by practitioners.
However, variance as a risk measure has been criticised as inappropriate, (due to its symmetric nature, lack of consistency with stochastic dominance relations). Alternative risk measures were firstly proposed in order to capture risk more intuitively (as an adverse event). Lower partial moments quantify risk considering only outcomes less than a predefined target. Central deviations are similar, but consider as target the expected value. All these risk measures quantify risk in terms of deviation from a target (fixed or distribution dependent). A different category of risk measures are the so called "tail" risk measures, which consider a small percentage of worst-case scenarios and evaluate the magnitude of losses under these conditions. Notable examples are VaR and CVaR.
From a computational point of view, the minimisation of most of the risk measures above lead to tractable optimisation models. For below-target risk measures, mostly order 1 (expected downside risk) and order 2 (semi-variance) are used. Minimisation of variance and of (target) semivariance are QP models. Minimisation of MAD, expected downside risk and CVaR are LP models. VaR is the only risk measure whose minimisation leads to a non-convex program, difficult to solve; thus, although largely used for evaluating portfolios' risk (it is a standard in regulation), VaR has not been used as a parameter in a model for portfolio construction.
The risk measures consistent with axiomatic models for choice under risk (second order stochastic dominance) are lower partial moments and CVaR. They in general are risk measures with good theoretical and practical properties.

In spite of these risk measures proposed in academia (some of them with good properties and leading to tractable models), variance is "the" risk measure used in the practice of portfolio selection.

A first possible motivation is a somewhat general belief that assets' returns are normally distributed. It is known that, if this was the case, then minimising variance leads to exactly the same solutions as in the case of minimising other risk measures (MAD, central semideviations, CVaR). Thus, there is no need to replace the (traditional and easy-to-interpret) mean-variance approach[5]
However, empirical results show that assets' returns are not normally (or elliptically) distributed ([11]); particularly, they exhibit longer and fatter tails than in the case of normal distribution.
A second possible motivation could be that, using a different risk measure could lead to results drastically different from the current and traditional approach. More precisely, a portfolio obtained by minimising an alternative risk measure could have a return distribution with an unacceptably large variance.
A possible way of overcoming this situation was proposed by us in [43]; the return distributions are characterised by three statistics: mean, variance and CVaR at a specified confidence level. In this approach, a portfolio dominates another if its return distribution has greater expected value, smaller variance and smaller CVaR. The solutions are then found by solving optimisation problems in which

[5]In fact, the minimisation of these risk measures leads to the same results in a more general framework than the multivariate normal distribution: precisely, when the distribution of the assets' returns belong to the more general class of elliptic distributions - [9].

variance is minimised and constraints on the expected return and on CVaR are imposed. Using this approach, the mean-variance efficient portfolios are not excluded (nor are the mean-CVaR efficient portfolios); instead, it can be avoided a situation where there is low variance, but high CVaR (or viceversa: low CVaR, high variance).
Stochastic dominance is another criterion for choice among random returns. As opposed to mean-risk models, it takes into account the whole distribution of returns. In particular, the second order stochastic dominance (SSD) is important in portfolio selection, because it describes the preference of risk-averse investors. The implementation of SSD (and in general, of any stochastic dominance relation) proved to be a difficult task; however, models for portfolio selection that use SSD as a choice criterion were recently proposed.

References

[1] C. ACERBI and D. TASCHE, *On the coherence of expected shortfall*, Journal of Banking and Finance **26** (2002), 1487–1503.

[2] P. ALBRECHT, *Risk measures*, John Wiley and Sons, New York, 2004, In: Encyclopaedia of Actuarial Science, Eds: Teugels, J. and Sundt, B.

[3] F. D. ARDITTI, *Risk and the required return on equity*, The Journal of Finance **22** (1967), 19–36.

[4] P. ARTZNER, F. DELBAEN, J.M. EBER, and D. HEATH, *Coherent measures of risk*, Mathematical Finance **9** (1999), 203–228.

[5] V. BAWA, *Safety-first, stochastic dominance and optimal portfolio choice*, Journal of Financial and Quantitative Analysis **13** (1978), 255–271.

[6] D. BERNOULLI, *Exposition of a new theory on the measurement of risk*, Econometrica **22** (1954), 23–36, Translation of a paper originally published in Latin in St. Petersburg in 1738.

[7] D. DENTCHEVA and A. RUSZCZYNSKI, *Optimization with stochastic dominance constraints*, SIAM Journal on Optimization **14** (2003), no. 2, 548–566.

[8] ______, *Portfolio optimization with stochastic dominance constraints*, Journal of Banking and Finance **30** (2006), no. 2, 433451.

[9] P. EMBRECHTS, A. MCNEIL, and D. STRAUMAN, *Correlation and dependence in risk management: Properties and pitfalls*, Risk Management: Value-at-Risk and Beyond (M. Dempster, ed.), Cambridge University Press, Cambridge, 1999.

[10] C. FABIAN, G. MITRA, and D. ROMAN, *Processing second-order stochastic dominance models using cutting-plane representations*, 2008.

[11] E. F. FAMA, *The behaviour of stock market prices*, Journal of Business **38** (1965), 34–105.

[12] P.C. FISHBURN, *Mean-risk analysis with risk associated with below-target returns*, American Economic Review **67** (1977), no. 2, 116–126.

[13] J. HADAR and W. RUSSELL, *Rules for ordering uncertain prospects*, The American Economic Review **59** (1969), 25–34.

[14] C. HARVEY, R. LIECHTY, M. W. LIECHTY, and P. MUELLER, *Portfolio selection with higher moments*, 2003, Working Paper, Duke University, available from: http://ssrn.com/abstract=634141.

[15] P. JORION, *Portfolio optimisation with tracking-error constraints*, Financial Analysts Journal **59** (2003), 70–82.

[16] M. KAUT and S. WALLACE, *Evaluation of scenario-generation methods for stochastic programming*, 2003, Stochastic Programming E-Print Series (SPEPS), vol. 14.

[17] H. KONNO, H. SHIRAKAWA, and H. YAMAZAKI, *A mean-absolute deviation-skewness portfolio optimisation model*, Annals of Operations Research **45** (1993), 205–220.

[18] H. KONNO and H. YAMAZAKI, *Mean-absolute deviation portfolio optimization model and its application to Tokyo stock market*, Management Science **37** (1991), 519–531.

[19] Y. KROLL and H. LEVY, *Stochastic dominance: A review and some new evidence*, Research in Finance **2** (1980), 163–227.

[20] N. LARSEN, H. MAUSSER, and S. URYASEV, *Algorithms for optimisation of Value-at-Risk*, Financial Engineering, E-Commerce and Supply Chain (P. Pardalos and V. K. Tsitsiringos, eds.), Kluwer Academic Publishers, Norwell, 2002, pp. 129–157.

[21] J. P. LAURENT, *Sensitivity analysis of risk measures for discrete distributions*, 2003, Available from: http://laurent.jeanpaul.free.fr.

[22] E. LEHMANN, *Ordered families of distributions*, Ann. Math. Statistics **26** (1955), 388–419.

[23] S. LEYFFER and J.S. PANG, *On the global minimization of the Value-at-Risk*, 2005, Available from: http://www.citebase.org/cgi-bin/citations?id=oai:arXiv.org:math/0401063.

[24] R. MANSINI, W. OGRYCZAK, and M.G. SPERANZA, *Conditional value at risk and related linear programming models for portfolio optimization*, Research Report 03-02, Warsaw University of Technology, Institute of Control and Computation Engineering, 2002.

[25] H. MARKOWITZ, *Portfolio selection*, Journal of Finance **7** (1952), 77–91.

[26] ______, *Portfolio selection*, John Wiley and Sons, New York, 1959.

[27] A.W. MARSHALL and I. OLKIN, *Inequalities: Theory of majorization and its applications*, Academic Press, San Diego, 1979.

[28] J. P. MORGAN, *Riskmetrics: Technical report*, J. P. MORGAN, New York, 1996.

[29] J. von NEUMANN and O. MORGENSTERN, *Theory of games and economic behaviour*, Princeton, N. J., Princeton University Press, 1947.

[30] Group of Thirty, *Derivatives: Practices and principles*, Group of Thirty, Washington, D. C., 1994.

[31] W. OGRYCZAK and A. RUSZCZYNSKI, *From stochastic dominance to mean- risk models: Semideviations as risk measures*, European Journal of Operational Research **116** (1999), 33–50.

[32] ______, *On consistency of stochastic dominance and mean- semideviations models*, Mathematical Programming **89** (2001), 217–232.

[33] S. ORTOBELLI, S. RACHEV, S. STOYANOV, F. FABOZZI, and BIGLOVA A., *The proper use of the risk measures in the portfolio theory*, International Journal of Theo- retical and Applied Finance **8** (2005), 1107–1134.

[34] A. F. PEROLD, *Large scale portfolio optimisation*, Management Science **30** (1984), 1143–1160.

[35] G. C. PFLUG, *Some remarks on the Value-at-Risk and the Conditional Value-at-Risk*, Probabilistic Constrained Optimization: Methodology and Applications (S. Uryasev, ed.), Kluwer Academic Publishers, 2000.

[36] J. W. PRATT, *Risk aversion in the small and in the large*, Econometrica **32** (1964), 122–136.

[37] J. P. QUIRK and R. SAPOSNIK, *Admissibility and measurable utility functions*, Review of Economic Studies **29** (1962), 140–146.

[38] R. T. ROCKAFELLAR, *Convex analysis*, Princeton University Press, Princeton, N. J., 1970.

[39] R.T. ROCKAFELLER and S. URYASEV, *Optimization of Conditional Value-at-Risk*, Journal of Risk **2** (2000), 21–42.

[40] ______, *Conditional Value-at-Risk for general loss distributions*, Journal of Banking and Finance **26** (2002), 1443–1471.

[41] R.T. ROCKAFELLER, S. URYASEV, and M. ZABARANKIN, *Deviation measures in risk analysis and optimization*, 2002, Research Report 2002-7, Risk Management and Engineering Lab/Center for Applied Optimization, University of Florida.

[42] D. ROMAN, K. DARBY-DOWMAN, and G. MITRA, *Portfolio construction based on stochastic dominance and target return distributions*, Mathematical Programming series B **108** (2006), 541–569.

[43] ______, *Mean-risk models using two risk measures: A multi-objective approach*, Quantitative Finance **7** (2007), 443–458.

[44] A. D. ROY, *Safety first and the holding of assets*, Econometrica **20** (1952), 431–449.

[45] R. C. SCOTT and P. A. HORVATH, *On the direction of preference for moments of higher order than the variance*, The Journal of Finance **35** (1980), 915–919.

[46] W. F. SHARPE, *A linear programming approximation for the general portfolio analysis problem*, Journal of Financial and Quantitative Analysis **6** (1971), 1263–1275.

[47] M. G. SPERANZA, *Linear programming models for portfolio optimization*, Finance **14** (1993), 107–123.

[48] D. TASCHE, *Expected shortfall and beyond*, Journal of Banking and Finance **26** (2002), 1519–1533.

[49] M. UNSER, *Lower Partial Moments as perceived measures of risk: an experimental study*, Journal of Economic Psychology **21** (2000), 253–280.

[50] G.A. WHITMORE, *Third-degree stochastic dominance*, The American Economic Review **60** (1970), 457–459.

[51] G.A. WHITMORE and M.C. FINDLAY (eds.), *Stochastic dominance: An approach to decision-making under risk*, D.C.Heath, Lexington, MA., 1978.

[52] Y YAMAI and T. YOSHIBA, *Comparative analyses of Expected Shortfall and Value-at-Risk: Expected utility maximization and tail risk*, Monetary and Economic Studies **20** (2002), 57–86.

[53] S. YITZHAKI, *Stochastic dominance, mean variance and Gini's mean difference*, American Economic Review **72** (1977), no. 1, 178–185.