Portfolio Optimization with Stochastic Dominance Constraints

Darinka Dentcheva* Andrzej Ruszczyński†

May 12, 2003

Abstract

We consider the problem of constructing a portfolio of finitely many assets whose returns are described by a discrete joint distribution. We propose a new portfolio optimization model involving stochastic dominance constraints on the portfolio return. We develop optimality and duality theory for these models. We construct equivalent optimization models with utility functions. Numerical illustration is provided.

KEYWORDS: Portfolio optimization, stochastic dominance, risk, utility functions.

1 Introduction

The problem of optimizing a portfolio of finitely many assets is a classical problem in theoretical and computational finance. Since the seminal work of Markowitz [15, 16, 17] it is generally agreed that portfolio performance should be measured in two distinct dimensions: the *mean* describing the expected return, and the *risk* which measures the uncertainty of the return. In the mean–risk approach, we select from the universe of all possible portfolios those that are *efficient*: for a given value of the mean they minimize the risk or, equivalently, for a given value of risk they maximize the mean. This approach allows one to formulate

*Stevens Institute of Technology, Department of Mathematics, Hoboken, NJ, e-mail: ddentche@stevens-tech.edu

†Rutgers University, Department of Management Science and Information Systems, Piscataway, NJ 08854, USA, e:mail: rusz@rutcor.rutgers.edu

the problem as a parametric optimization problem, and it facilitates the trade-off analysis between mean and risk.

Another theoretical approach to the portfolio selection problem is that of *stochastic dominance* (see [19, 30, 14]). The concept of *stochastic dominance* is related to models of risk-averse preferences [6]. It originated from the theory of majorization [9, 18] for the discrete case, was later extended to general distributions [23, 7, 8, 25], and is now widely used in economics and finance [14].

The usual (first order) definition of stochastic dominance gives a partial order in the space of real random variables. More important from the portfolio point of view is the notion of second-order dominance, which is also defined as a partial order. It is equivalent to this statement: a random variable R dominates the random variable Y if $\mathbb{E}\big[u(R)\big] \geq \mathbb{E}\big[u(Y)\big]$ for all nondecreasing concave functions $u(\cdot)$ for which these expected values are finite. Thus, no risk-averse decision maker will prefer a portfolio with return Y over a portfolio with return R.

In our earlier publications [2, 3, 4, 5] we have introduced a new stochastic optimization model with stochastic dominance constraints. In this paper we show how this theory can be used for risk-averse portfolio optimization. We add to the portfolio problem the condition that the portfolio return stochastically dominates a reference return, for example, the return of an index. We identify concave nondecreasing utility functions which correspond to dominance constraints. Maximizing the expected return modified by these utility functions, guarantees that the optimal portfolio return will dominate the given reference return.

2 The portfolio problem

Let $R_1, R_2, \ldots, R_n$ be random returns of assets $1, 2, \ldots, n$. We assume that $\mathbb{E}\big[|R_j|\big] < \infty$ for all $j = 1, \ldots, n$.

Our aim is to invest our capital in these assets in order to obtain some desirable characteristics of the total return on the investment. Denoting by $x_1, x_2, \ldots, x_n$ the fractions of the initial capital invested in assets $1, 2, \ldots, n$ we can easily derive the formula for the total return:

$$R(x) = R_1 x_1 + R_2 x_2 + \cdots + R_n x_n. \tag{1}$$

Clearly, the set of possible asset allocations can be defined as follows:

$$X = \{x \in \mathbb{R}^n : x_1 + x_2 + \cdots + x_n = 1,\ x_j \geq 0,\ j = 1, 2, \ldots, n\}.$$

In some applications one may introduce the possibility of *short positions*, i.e., allow some x_j's to become negative. Other restrictions may limit the exposure to particular assets or their groups, by imposing upper bounds on the x_j's or on their partial sums. One can also limit the absolute differences between the x_j's and some reference investments $\bar{x}_j$, which may represent the existing portfolio, etc. Our analysis will not depend on the detailed way this set is defined; we shall only use the fact that it is a convex polyhedron. All modifications discussed above define some convex polyhedral feasible sets, and are, therefore, covered by our approach.

The main difficulty in formulating a meaningful portfolio optimization problem is the definition of the preference structure among feasible portfolios. If we use only the mean return

$$\mu(x) = \mathbb{E}\big[R(x)\big],$$

then the resulting optimization problem has a trivial and meaningless solution: invest everything in assets that have the maximum expected return. For these reasons the practice of portfolio optimization resorts usually to two approaches.

In the first approach we associate with portfolio x some risk measure $\rho(x)$ representing the variability of the return $R(x)$. In the classical Markowitz model $\rho(x)$ is the variance of the return,

$$\rho(x) = \mathbb{V}\mathrm{ar}\big[R(x)\big],$$

but many other measures are possible here as well.

The mean–risk portfolio optimization problem is formulated as follows:

$$\max_{x \in X} \mu(x) - \lambda \rho(x). \tag{2}$$

Here, λ is a nonnegative parameter representing our desirable exchange rate of mean for risk. If $\lambda = 0$, the risk has no value and the problem reduces to the problem of maximizing the mean. If $\lambda > 0$ we look for a compromise between the mean and the risk. The general question of constructing mean–risk models which are in harmony with the stochastic dominance relations has been the subject of the analysis of the recent papers [20, 21, 22, 27]. We have identified there several primal risk measures, most notably central semideviations, and dual risk measures, based on the Lorenz curve, which are consistent with the stochastic dominance relations.

The second approach is to select a certain utility function $u : \mathbb{R} \to \mathbb{R}$ and to formulate the following optimization problem

$$\max_{x \in X} \mathbb{E}\big[u(R(x))\big]. \tag{3}$$

It is usually required that the function $u(\cdot)$ is concave and nondecreasing, thus representing preferences of a risk-averse decision maker. The challenge here is to select the appropriate utility function that represents well our preferences and whose application leads to non-trivial and meaningful solutions of (3). References ....

In this paper we propose an alternative approach, by introducing a comparison to a reference return into our optimization problem. The comparison is based on the stochastic dominance relation. More specifically, we shall consider only portfolios whose return stochastically dominates a certain reference return.

3 Stochastic dominance

In the stochastic dominance approach, random returns are compared by a point-wise comparison of some performance functions constructed from their distribution functions. For a real random variable V, its first performance function is defined as the right-continuous cumulative distribution function of V:

$$F(V;\eta) = \mathbb{P}\{V \le \eta\} \quad \text{for } \eta \in \mathbb{R}.$$

A random return V is said [13, 23] to *stochastically dominate* another random return S to the first order, denoted $V \succeq_{FSD} S$, if

$$F(V;\eta) \le F(S;\eta) \quad \text{for all } \eta \in \mathbb{R}.$$

The second performance function F_2 is given by areas below the distribution function F,

$$F_2(V;\eta) = \int_{-\infty}^{\eta} F(V;\xi)\, d\xi \quad \text{for } \eta \in \mathbb{R},$$

and defines the weak relation of the *second-order stochastic dominance* (SSD). That is, random return V stochastically dominates S to the second order, denoted $V \succeq_{SSD} S$, if

$$F_2(V;\eta) \le F_2(S;\eta) \quad \text{for all } \eta \in \mathbb{R}.$$

(see [7, 8, 25]). The corresponding strict dominance relations $\succ_{FSD}$ and $\succ_{SSD}$ are defined in the usual way: $V \succ S$ if and only if $V \succeq S$, $S \not\succeq V$.

By changing the order of integration we can express the function $F_2(V;\cdot)$ as the expected shortfall [20]: for each target value η we have

$$F_2(V;\eta) = \mathbb{E}\big[(\eta - V)_+\big], \tag{4}$$

where $(\eta - V)_+ = \max(\eta - V, 0)$. The function $F_{(2)}(V;\cdot)$ is continuous, convex, nonnegative and nondecreasing. It is well defined for all random variables V with finite expected value.

In the context of portfolio optimization, we shall consider stochastic dominance relations between random returns defined by (1). Thus, we say that portfolio x *dominates* portfolio y *under the FSD rules*, if

$$F(R(x);\eta) \le F(R(y);\eta) \quad \text{for all } \eta \in \mathbb{R},$$

where at least one strict inequality holds. Similarly, we say that x *dominates* y *under the SSD rules* ($R(x) \succ_{SSD} R(y)$), if

$$F_2(R(x);\eta) \leq F_2(R(y);\eta) \quad \text{for all } \eta \in \mathbb{R},$$

with at least one inequality strict. Recall that the individual returns R_j have finite expected values and thus the function $F_2(R(x);\cdot)$ is well defined.

Stochastic dominance relations are of crucial importance for decision theory. It is known that $R(x) \succeq_{FSD} R(y)$ if and only if

$$\mathbb{E}\big[u(R(x))\big] \geq \mathbb{E}\big[u(R(y))\big] \tag{5}$$

for any nondecreasing function $u(\cdot)$ for which these expected values are finite. Furthermore, $R(x) \succeq_{SSD} R(y)$ if and only if (5) holds true for every nondecreasing and concave $u(\cdot)$ for which these expected values are finite (see, e.g., [14]).

A portfolio x is called *SSD-efficient* (or *FSD-efficient*) in a set of portfolios X if there is no $y \in X$ such that $R(y) \succ_{SSD} R(x)$ (or $R(y) \succ_{FSD} R(x)$).

We shall focus our attention on the SSD relation, because of its consistency with risk-averse preferences: if $R(x) \succ_{SSD} R(y)$, then portfolio x is preferred to y by all risk-averse decision makers.

4 The dominance-constrained portfolio problem

The starting point for our model is the assumption that a reference random return Y having a finite expected value is available. It may have the form $Y = R(\bar{z})$, for some reference portfolio $\bar{z}$. It may be an index or our current portfolio. Our intention is to have the return of the new portfolio, $R(x)$, preferable over Y. Therefore, we introduce the following optimization problem:

$$\max f(x) \tag{6}$$

$$\text{subject to } R(x) \succeq_{(2)} Y, \tag{7}$$

$$x \in X. \tag{8}$$

Here $f : X \to \mathbb{R}$ is a concave continuous functional. In particular, we may use

$$f(x) = \mathbb{E}\big[R(x)\big]$$

and this will still lead to nontrivial solutions, due to the presence of the dominance constraint (7).

Proposition 1 *Assume that Y has a discrete distribution with realizations y_i, $i = 1, \ldots, m$. Then relation* (7) *is equivalent to*

$$\mathbb{E}\big[(y_i - R(x))_+\big] \le \mathbb{E}\big[(y_i - Y)_+\big], \quad i = 1, \ldots, m. \tag{9}$$

Proof. If relation (7) is true, then the equivalent representation (4) implies (9).

It is sufficient to prove that (9) imply that

$$F_2(R(x); \eta) \le F_2(Y; \eta) \quad \text{for all } \eta \in \mathbb{R}.$$

With no loss of generality we may assume that $y_1 < y_2 < \cdots < y_m$. The distribution function $F(Y; \cdot)$ is piecewise constant with jumps at y_i, $i = 1, \ldots, m$. Therefore, the function $F_2(Y; \cdot)$ is piecewise linear and has break points at y_i, $i = 1, \ldots, m$. Let us consider three cases, depending on the value of η.

Case 1: If $\eta \le y_1$ we have

$$0 \le F_2(R(x); \eta) \le F_2(R(x); y_1) \le F_2(Y; y_1) = 0.$$

Therefore the required relation holds as an equality.

Case 2: Let $\eta \in [y_i, y_{i+1}]$ for some i. Since, for any random return $R(x)$, the function $F_2(R(x);\cdot)$ is convex, inequalities (9) for i and $i+1$ imply that for all $\eta \in [y_i, y_{i+1}]$ one has

$$\begin{aligned} F_2(R(x);\eta) &\le \lambda F_2(R(x);y_i) + (1-\lambda)F_2(R(x);y_{i+1}) \\ &\le \lambda F_2(Y;y_i) + (1-\lambda)F_2(Y;y_{i+1}) = F_2(Y;\eta), \end{aligned}$$

where $\lambda = (y_{i+1}-\eta)/(y_{i+1}-y_i)$. The last equality follows from the linearity of $F_2(Y;\cdot)$ in the interval $[y_i, y_{i+1}]$.

Case 3: For $\eta > y_m$ the function $F_2(Y;\eta)$ is affine with slope 1, and therefore

$$\begin{aligned} F_2(Y;\eta) &= F_2(Y;y_m) + \eta - y_m \\ &\ge F_2(R(x);y_m) + \int_{y_m}^{\eta} F(R(x);\alpha)\,d\alpha = F_2(R(x);\eta), \end{aligned}$$

as required. □

Let us assume now that the returns have a discrete joint distribution with realizations r_{jt}, $t = 1,\dots,T$, $j = 1,\dots,n$, attained with probabilities p_t, $t = 1,2,\dots,T$. The the formulation of the stochastic dominance relation (7) resp. (9) simplifies even further. Introducing variables s_{it} representing shortfall of $R(x)$ below y_i in realization t, $i = 1,\dots,m$, $t = 1,\dots,T$, we obtain the following result.

Proposition 2 *Assume that R_j, $j = 1,\dots,n$, and Y have discrete distributions. Then problem* (6)–(8) *is equivalent to the problem*

$$\max\ f(x) \tag{10}$$

$$\text{subject to } \sum_{j=1}^{n} x_j r_{jt} + s_{it} \ge y_i, \quad i = 1,\dots,m, \quad t = 1,\dots,T, \tag{11}$$

$$\sum_{t=1}^{T} p_t s_{it} \le F_2(Y;y_i), \quad i = 1,\dots,m, \tag{12}$$

$$s_{it} \ge 0, \quad i = 1,\dots,m, \quad t = 1,\dots,T. \tag{13}$$

$$x \in X. \tag{14}$$

Proof. If $x \in \mathbb{R}^n$ is a feasible point of (6)–(8), then we can set

$$s_{it} = \max\Big(0, y_i - \sum_{j=1}^{n} x_j r_{jt}\Big), \quad i = 1, \dots, m, \quad t = 1, \dots, T.$$

The pair (x, s) is feasible for (11)–(14).

On the other hand, for any pair (x, s), which is feasible for (11)–(14), inequalities (11) and (13) imply that

$$s_{it} \geq \max\Big(0, y_i - \sum_{j=1}^{n} x_j r_{jt}\Big), \quad i = 1, \dots, m, \quad t = 1, \dots, T.$$

Taking the expected value of both sides and using (12) we obtain

$$F_2(R(x); y_i) \leq F_2(Y; y_i), \quad i = 1, \dots, m.$$

Proposition 1 implies that x is feasible for problem (6)–(8). $\square$

5 Optimality and Duality

From now on we shall assume that the probability distributions of the returns and of the reference outcome Y are discrete with finitely many realizations. We also assume that the realizations of Y are ordered: $y_1 < y_2 < \cdots < y_m$. The probabilities of the realizations are denoted by π_i, $i = 1, \dots, m$.

We define the set $\mathcal{U}$ of functions $u : \mathbb{R} \to \mathbb{R}$ satisfying the following conditions:

$u(\cdot)$ is concave and nondecreasing;

$u(\cdot)$ is piecewise linear with break points y_i, $i = 1, \dots, m$;

$u(t) = 0$ for all $t \geq y_m$.

It is evident that $\mathcal{U}$ is a convex cone.

Let us define the Lagrangian of (6)–(8), $L : \mathbb{R}^n \times \mathcal{U} \to \mathbb{R}$, as follows

$$L(x, u) = f(x) + \mathbb{E}\big[u(R(x))\big] - \mathbb{E}\big[u(Y)\big]. \tag{15}$$

It is well defined, because for every $u \in \mathcal{U}$ and every $x \in \mathbb{R}^n$ the expected value $\mathbb{E}\big[u(R(x))\big]$ exists and is finite.

Theorem 1 *If $\hat{x}$ is an optimal solution of* (6)–(8) *then there exists a function $\hat{u} \in \mathcal{U}$ such that*

$$L(\hat{x}, \hat{u}) = \max_{x \in X} L(x, \hat{u}) \tag{16}$$

and

$$\mathbb{E}\big[\hat{u}(R(\hat{x}))\big] = \mathbb{E}\big[\hat{u}(Y)\big]. \tag{17}$$

Conversely, if for some function $\hat{u} \in \mathcal{U}$ an optimal solution $\hat{x}$ of (16) *satisfies* (7) *and* (17), *then $\hat{x}$ is an optimal solution of* (6)–(8).

Proof. By Proposition 2 problem (6)–(8) is equivalent to problem (10)–(14). We associate Lagrange multipliers $\mu \in \mathbb{R}^m$ with constraints (12) and we formulate the Lagrangian $\Lambda : \mathbb{R}^n \times \mathbb{R}^{mT} \times \mathbb{R}^m \to \mathbb{R}$ as follows:

$$\Lambda(x, s, \mu) = f(x) + \sum_{i=1}^{m} \mu_i \Big(F_2(Y; y_i) - \sum_{t=1}^{T} p_t s_{it} \Big).$$

Let us define the set

$$Z = \Big\{ (x, s) \in X \times \mathbb{R}_+^{mT} : \sum_{j=1}^{n} x_j r_{jt} + s_{it} \geq y_i,\ i = 1, \dots, m,\ t = 1, \dots, T \Big\}.$$

Since Z is a convex closed polyhedral set, the constraints (12) are linear, and the objective function is concave, if the point $(\hat{x}, \hat{s})$ is an optimal solution of problem (6)–(8), then the following Karush-Kuhn-Tucker optimality conditions hold true. There exists a vector of multipliers $\hat{\mu} \geq 0$ such that:

$$\Lambda(\hat{x}, \hat{s}, \hat{\mu}) = \max_{(x,s) \in Z} \Lambda(x, s, \hat{\mu}) \tag{18}$$

and

$$\hat{\mu}_i \Big(F_2(Y; y_i) - \sum_{t=1}^{T} p_t \hat{s}_{it} \Big) = 0, \quad i = 1, \dots, m. \tag{19}$$

We can transform the Lagrangian Λ as follows:

$$\begin{aligned}\Lambda(x,s,\mu) &= f(x) + \sum_{i=1}^{m} \mu_i F_2(Y;y_i) - \sum_{i=1}^{m}\sum_{t=1}^{T} \mu_i p_t s_{it}\\ &= f(x) + \sum_{i=1}^{m} \mu_i F_2(Y;y_i) - \sum_{t=1}^{T} p_t \sum_{i=1}^{m} \mu_i s_{it}.\end{aligned}$$

For any fixed x the maximization with respect to s such that $(x,s)\in Z$ yields

$$s_{it} = \max\Big(0, y_i - \sum_{j=1}^{n} x_j r_{jt}\Big) = \max\big(0, y_i - [R(x)]_t\big), \quad i=1,\dots,m, \quad t=1,\dots,T,$$

where $[R(x)]_t$ is the t-th realization of the portfolio return. Define the functions $u_i:\mathbb{R}\to\mathbb{R}$, $i=1,\dots,m$ by

$$u_i(\eta) = -\max(0, y_i-\eta),$$

and let

$$u_\mu(\eta) = \sum_{i=1}^{m} \mu_i u_i(\eta).$$

Let us observe that $u_\mu \in \mathcal{U}$. We can rewrite the result of maximization of the Lagrangian Λ with respect to s as follows:

$$\begin{aligned}\max_s \Lambda(x,s,\mu) &= f(x) + \sum_{i=1}^{m} \mu_i F_2(Y;y_i) + \sum_{t=1}^{T} p_t \sum_{i=1}^{m} \mu_i u_i\big([R(x)]_t\big)\\ &= f(x) + \sum_{i=1}^{m} \mu_i F_2(Y;y_i) + \sum_{t=1}^{T} p_t u_\mu\big([R(x)]_t\big).\end{aligned} \tag{20}$$

Furthermore, we can obtain a similar expression for the sum involving Y:

$$\begin{aligned}\sum_{i=1}^{m} \mu_i F_2(Y;y_i) &= \sum_{i=1}^{m} \mu_i \sum_{k=1}^{m} \pi_k \max(0, y_i - y_k)\\ &= \sum_{k=1}^{m} \pi_k \sum_{i=1}^{m} \mu_i \max(0, y_i-y_k) = -\sum_{k=1}^{m} \pi_k u_\mu(y_k).\end{aligned}$$

Substituting into (20), we obtain

$$\max_s \Lambda(x,s,\mu) = f(x) + \mathbb{E}\big[u_\mu(R(x))\big] - \mathbb{E}\big[u_\mu(Y)\big] = L(x,u_\mu). \tag{21}$$

Setting $\hat{u} := u_{\hat{\mu}}$ we conclude that the conditions (18) imply (16), as required. Furthermore, adding the complementarity conditions (19) over $i = 1, \dots, m$, and using the same transformation we get (17).

To prove the converse, let us observe that for every $\hat{u} \in \mathcal{U}$ we can define

$$\hat{\mu}_i = \hat{u}'_-(y_i) - \hat{u}'_+(y_i), \quad i = 1, \dots, m,$$

with $\hat{u}'_-$ and $\hat{u}'_+$ denoting the left and right derivatives of $\hat{u}$:

$$\hat{u}'_-(\eta) = \lim_{t \uparrow \eta} \frac{\hat{u}(\eta) - \hat{u}(t)}{\eta - t}, \qquad \hat{u}'_+(\eta) = \lim_{t \downarrow \eta} \frac{\hat{u}(t) - \hat{u}(\eta)}{t - \eta}.$$

Since $\hat{u}$ is concave, $\hat{\mu} \geq 0$. Using the elementary functions $u_i(\eta) = -\max(0, y_i - \eta)$ we can represent $\hat{u}$ as follows:

$$\hat{u}(\eta) = \sum_{i=1}^{m} \hat{\mu}_i u_i(\eta).$$

Consequently, correspondence (21) holds true for $\hat{\mu}$ and $\hat{u}$. Therefore, if $\hat{x}$ is the maximizer of (16), then the pair $(\hat{x}, \hat{s})$, with

$$\hat{s}_{it} = \max\Big(0, y_i - \sum_{j=1}^{n} \hat{x}_j r_{jt}\Big), \quad i = 1, \dots, m, \quad t = 1, \dots, T,$$

is the maximizer of $\Lambda(x, s, \hat{\mu})$, over $(x, s) \in Z$. Our result follows then from standard sufficient conditions for problem (10)–(14) (see,e.g., [24, Thm. 28.1]). $\square$

We can also develop duality relations for our problem. With the Lagrangian (15) we can associate the dual function

$$D(u) = \max_{x \in X} L(x, u).$$

We are allowed to write the maximization operation here, because the set X is compact and $L(\cdot, u)$ is continuous.

The dual problem has the form

$$\min_{u \in \mathcal{U}} D(u). \tag{22}$$

The set $\mathcal{U}$ is a closed convex cone and $D(\cdot)$ is a convex functional, so (22) is a convex optimization problem.

Theorem 2 *Assume that* (6)–(8) *has an optimal solution. Then problem* (22) *has an optimal solution and the optimal values of both problems coincide. Furthermore, the set of optimal solutions of* (22) *is the set of functions* $\hat{u} \in \mathcal{U}$ *satisfying* (16)–(17) *for an optimal solution* $\hat{x}$ *of* (6)–(8).

Proof. The theorem is an easy consequence of Theorem 1 and general duality relations in convex nonlinear programming (see [1, Thm. 2.165]). Note that all constraints of our problem are linear or convex polyhedral, and therefore we do not need any constraint qualification conditions here.

6 Splitting

Let us now consider the special form of problem (6)–(8), with

$$f(x) = \mathbb{E}[R(x)].$$

Recall that the random returns R_j, $j = 1, \dots, n$, have discrete distributions with realizations r_{jt}, $t = 1, \dots, T$, attained with probabilities p_t.

In order to facilitate numerical solution of problem (6)–(8), it is convenient to consider its split-variable form:

$$\max \mathbb{E}[R(x)] \tag{23}$$

$$\text{subject to } R(x) \geq V, \quad a.s., \tag{24}$$

$$V \succeq_{(2)} Y, \tag{25}$$

$$x \in X. \tag{26}$$

In the above problem, V is a random variable having realizations v_t attained with probabilities p_t, $t = 1, \dots, T$, and relation (24) is understood almost surely. In the case of finitely many realizations it simply means that

$$\sum_{j=1}^{n} r_{jt} x_j \geq v_t, \quad t = 1, \dots, T. \tag{27}$$

We shall consider two groups of Lagrange multipliers: a utility function $u \in \mathcal{U}$, and a vector $\theta \in \mathbb{R}^T$, $\theta \geq 0$. The utility function $u(\cdot)$ will correspond to the dominance constraint (25), as in the preceding section. The multipliers $p_t\theta_t$, $t = 1, \ldots, T$, will correspond to the inequalities (27). The Lagrangian takes on the form

$$\begin{aligned} L(x, V, u, \theta) = &\sum_{t=1}^{T} p_t \sum_{j=1}^{n} r_{jt} x_j + \sum_{t=1}^{T} p_t \theta_t (\sum_{j=1}^{n} r_{jt} x_j - v_t) \\ &+ \sum_{t=1}^{T} p_t u(v_t) - \sum_{k=1}^{m} \pi_k u(y_k). \end{aligned} \tag{28}$$

The optimality conditions can be formulated as follows.

Theorem 3 *If $(\hat{x}, \hat{V})$ is an optimal solution of (23)–(26), then there exist $\hat{u} \in \mathcal{U}$ and a nonnegative vector $\hat{\theta} \in \mathbb{R}^T$, such that*

$$L(\hat{x}, \hat{V}, \hat{u}, \hat{\theta}) = \max_{(x,V) \in X \times \mathbb{R}^T} L(x, V, \hat{u}, \hat{\theta}), \tag{29}$$

$$\sum_{t=1}^{T} p_t \hat{u}(\hat{v}_t) - \sum_{k=1}^{m} \pi_k \hat{u}(y_k) = 0, \tag{30}$$

$$\hat{\theta}_t (\hat{v}_t - \sum_{j=1}^{n} r_{jt} \hat{x}_j) = 0, \quad t = 1, \ldots, T. \tag{31}$$

Conversely, if for some function $\hat{u} \in \mathcal{U}$ and nonnegative vector $\hat{\theta} \in \mathbb{R}^T$, an optimal solution $(\hat{x}, \hat{V})$ of (29) satisfies (24)–(25) and (30)–(31), then $(\hat{x}, \hat{V})$ is an optimal solution of (23)–(26).

Proof. By Proposition 1, the dominance constraint (25) is equivalent to finitely many inequalities

$$\mathbb{E}\big[(y_i - R(x))_+\big] \leq \mathbb{E}\big[(y_i - Y)_+\big], \quad i = 1, \ldots, m.$$

Problem (23)–(26) takes on the form:

$$\begin{aligned} &\max\ \mathbb{E}[R(x)] \\ &\text{subject to } \sum_{j=1}^{n} r_{jt}x_j \geq v_t, \quad t = 1, \ldots, T, \\ &\mathbb{E}\big[(y_i - R(x))_+\big] \leq \mathbb{E}\big[(y_i - Y)_+\big], \quad i = 1, \ldots, m, \\ &x \in X. \end{aligned}$$

Let us introduce Lagrange multipliers μ_i, $i = 1, \ldots, m$, associated with the dominance constraints. The standard Lagrangian takes on the form:

$$\begin{aligned} \Lambda(x, V, \mu, \theta) = &\sum_{t=1}^{T} p_t \sum_{j=1}^{n} r_{jt}x_j + \sum_{t=1}^{T} p_t\theta_t(\sum_{j=1}^{n} r_{jt}x_j - v_t) \\ &- \sum_{i=1}^{m} \mu_i \sum_{t=1}^{T} p_t[y_i - \sum_{j=1}^{n} r_{jt}x_j]_+ + \sum_{i=1}^{m} \mu_i \sum_{k=1}^{m} \pi_k[y_i - y_k]_+. \end{aligned}$$

Rearranging the last two sums, exactly as in the proof of Theorem 1, we obtain the following key relation. For every $\mu \geq 0$, setting

$$u_\mu(\eta) = -\sum_{i=1}^{m} \mu_i \max(0, y_i - \eta),$$

we have

$$\Lambda(x, V, \mu, \theta) = L(x, V, u_\mu, \theta).$$

The remaining part of the proof is the same as the proof of Theorem 1.

The dual function associated with the split-variable problem has the form

$$D(u, \theta) = \sup_{x \in X,\, V \in \mathbb{R}^T} L(x, V, u, \theta).$$

and the dual problem is, as usual,

$$\min_{u \in \mathcal{U}, \theta \geq 0} D(u, \theta). \tag{32}$$

The corresponding duality theorem is an immediate consequence of Theorem 3 and standard duality relations in convex programming. Note that all constraints of our problem (23)–(26) are linear or convex polyhedral, and therefore we do not need additional constraint qualification conditions here.

Theorem 4 *Assume that* (23)–(26) *has an optimal solution. Then the dual problem* (32) *has an optimal solution and the optimal values of both problems coincide. Furthermore, the set of optimal solutions of* (32) *is the set of functions* $\hat{u} \in \mathcal{U}$ *and vectors* $\hat{\theta} \geq 0$ *satisfying* (29)–(31) *for an optimal solution* $(\hat{x}, \hat{V})$ *of* (23)–(26).

Let us analyze in more detail the structure of the dual function:

$$
\begin{aligned}
D(u,\theta) &= \sup_{x\in X,\, V\in\mathbb{R}^T} \Big\{ \sum_{t=1}^T p_t \sum_{j=1}^n r_{jt}x_j + \sum_{t=1}^T p_t\theta_t\Big(\sum_{j=1}^n r_{jt}x_j - v_t\Big) + \sum_{t=1}^T p_t u(v_t) \Big\} - \sum_{k=1}^m \pi_k u(y_k) \\
&= \max_{x\in X} \sum_{j=1}^n \sum_{t=1}^T p_t(1+\theta_t) r_{jt} x_j + \sup_V \sum_{t=1}^T p_t \big[u(v_t) - \theta_t v_t\big] - \sum_{k=1}^m \pi_k u(y_k) \\
&= \max_{1\le j\le n} \sum_{t=1}^T p_t(1+\theta_t) r_{jt} + \sum_{t=1}^T p_t \sup_{v_t} \big[u(v_t) - \theta_t v_t\big] - \sum_{k=1}^m \pi_k u(y_k).
\end{aligned}
$$

In the last equation we have used the fact that X is a simplex and therefore the maximum of a linear form is attained at one of its vertices. It follows that the dual function can be expressed as the sum

$$
D(u,\theta) = D_0(\theta) + \sum_{t=1}^T p_t D_t(u,\theta_t) + D_{T+1}(u), \tag{33}
$$

with

$$
D_0(\theta) = \max_{1\le j\le n} \sum_{t=1}^T p_t(1+\theta_t) r_{jt}, \tag{34}
$$

$$
D_t(u,\theta_t) = \sup_{v_t} \big[u(v_t) - \theta_t v_t\big], \quad t = 1,\dots,T, \tag{35}
$$

and

$$
D_{T+1}(u) = -\sum_{k=1}^m \pi_k u(y_k). \tag{36}
$$

If the set X is a general convex polyhedron, the calculation of D_0 involves a linear programming problem with n variables.

To determine the domain of the dual function, observe that if $u'_-(y_1) < \theta_t$ then

$$\lim_{v_t \to \infty} \big[u(v_t) - \theta_t v_t \big] = +\infty,$$

and thus the supremum in (35) is equal to $+\infty$. On the other hand, if $u'_-(y_1) \geq \theta_t$, then the function $u(v_t) - \theta_t v_t$ has a nonnegative slope for $v_t \leq y_1$ and nonpositive slope $-\theta_t$ for $v_t \geq y_m$. It is piecewise linear and it achieves its maximum at one of the break points. Therefore

$$\mathrm{dom} D_t = \{ (u, \theta_t) \in \mathcal{U} \times \mathbb{R}_+ : u'_-(y_1) \geq \theta_t \}.$$

At any point of the domain,

$$D_t(u, \theta_t) = \max_{1 \leq k \leq m} \big[u(y_k) - \theta_t y_k \big]. \tag{37}$$

The domain of D_0 is the entire space $\mathbb{R}^T$.

7 Decomposition

It follows from our analysis that the dual functional can be expressed as a weighted sum of $T+2$ functions (34)–(36).

In order to analyze their properties and to develop a numerical method we need to find a proper representation of the utility function u. We represent the function u by its slopes between break points. Let us denote the values of u at its break points by

$$u_k = u(y_k), \quad k = 1, \ldots, m.$$

We introduce the slope variables

$$\beta_k = u'_-(y_k), \quad k = 1, \ldots, m.$$

The vector $\beta = (\beta_1, \ldots, \beta_m)$ is nonnegative, because u is nondecreasing. As u is concave, $\beta_k \geq \beta_{k+1}$, $k = 1, \ldots, m-1$. We can represent the values of u at break points as follows

$$u_k = -\sum_{\ell > k} \beta_\ell (y_\ell - y_{\ell-1}), \quad k = 1, \ldots, m.$$

The function (37) takes on the form

$$D_t(u, \theta_t) = \max_{1 \leq k \leq m} \big[u_k - \theta_t y_k \big] = \max_{1 \leq k \leq m} \Big[-\sum_{\ell > k} \beta_\ell (y_\ell - y_{\ell-1}) - \theta_t y_k \Big].$$

In this way we have expressed $D_t(u, \theta_t)$ as a function of the slope vector $\beta \in \mathbb{R}^m$ and of $\theta_t \in \mathbb{R}_+$. We denote

$$B_t(\beta, \theta_t) = \max_{1 \leq k \leq m} \Big[-\sum_{\ell > k} \beta_\ell (y_\ell - y_{\ell-1}) - \theta_t y_k \Big]. \tag{38}$$

Observe that B_t is the maximum of finitely many linear functions in its domain. The domain is a convex polyhedron defined by

$$0 \leq \theta_t \leq \beta_1.$$

Consequently, B_t is a convex polyhedral function. Therefore its subgradient at a point (β, θ_t) of the domain can be calculated as the gradient of the linear function at which the maximum in (38) is attained. Let k^* be the index of this linear function. Denoting by δ_ℓ the ℓth unit vector in $\mathbb{R}^m$ we obtain the following subgradient of $B_t(\beta, \theta_t)$:

$$\Big(-\sum_{\ell > k^*} \delta_\ell (y_\ell - y_{\ell-1}), -y_{k^*} \Big).$$

Similarly, function (36) can be represented as a function B_{T+1} of the slope vector β:

$$B_{T+1}(\beta) = \sum_{k=1}^{m} \pi_k \sum_{\ell > k} \beta_\ell (y_\ell - y_{\ell-1}).$$

It is linear in β and its gradient has the form

$$\sum_{\ell=1}^{n} \delta_\ell \sum_{k < \ell} \pi_k (y_\ell - y_{\ell-1}).$$

Finally, denoting by j^* the index at which the maximum in (34) is attained, we see that the vector with coordinates

$$p_t r_{j^*t}, \quad t = 1, \ldots, T, \tag{39}$$

is a subgradient of D_0.

Summing up, with our representation of the utility function by its slopes, the dual function is a sum of $T + 2$ convex polyhedral functions with known domains. Moreover, their subgradients are readily available. Therefore the dual problem can be solved by nonsmooth optimization methods (see [11, 10]). We have developed a specialized version of the regularized decomposition method described in [26]. This approach is particularly suitable, because the dual function is a sum of very many polyhedral functions.

After the dual problem is solved, we obtain not only the optimal dual solution $(\hat{\beta}, \hat{\theta})$, but also a collection of active cutting planes for each component of the dual function.

Let us denote by J_0 the collection of active cuts for D_0. Each cutting plane for D_0 provides a subgradient (39) at the optimal dual solution. A convex combination of these subgradients provides the subgradient of D_0 that enters the optimality conditions for the dual problem. The coefficients of this convex combination are also identified by the regularized decomposition method. Let g_0 denote this subgradient and let ν_j, $j \in J_0$ the corresponding coefficients. Then

$$g_0 = \sum_{t=1}^{T} \delta_t \sum_{j \in J_0} p_t r_{jt} \nu_j,$$

where

$$\nu_j \geq 0, \qquad \sum_{j \in J_0} \nu_j = 1.$$

For each t the subgradient of B_t with respect to θ_t entering the optimality conditions is

$$\hat{v}_t \in \operatorname{conv}\{y_{k^*} : k^* \text{ is a maximizer in (38)}\}.$$

Therefore

$$g_0 - \sum_{t=1}^{T} p_t \hat{v}_t = 0.$$

Using these relations we can verify that $\hat{v}$ is the vector of optimal portfolio returns in scenarios $t = 1, \ldots, T$. Thus the optimal portfolio has the weights

$$\begin{aligned} \hat{x}_j &= \nu_j, \qquad && j \in J_0, \\ \hat{x}_j &= 0, \qquad && j \notin J_0. \end{aligned}$$

We have tested our approach on a basket of 719 real-world assets, using 616 possible realizations of their joint returns [27]. Historical data on weekly returns in the 12 years from Spring 1990 to Spring 2002 were used as equally likely realizations.

We have used four reference returns Y. Each of them was constructed as return of a certain index composed of our assets. Since we actually know the past returns, for the purpose of comparison we have selected equally weighted indexes composed of the N assets having the highest average return in this period. Reference Portfolio 1 corresponds to $N = 26$, Reference Portfolio 2 corresponds to $N = 54$, Reference Portfolio 3 corresponds to $N = 82$, and Reference Portfolio 4 corresponds to $N = 200$. Our problem was to maximize the expected return, under the condition that the return of the reference portfolio is dominated. Since the reference point was a return of a portfolio composed from the same basket, we have $m = T = 616$ in this case.

The dual problem of minimizing (33) has 1335 decision variables: the utility function u, represented by the vector of slopes $\beta \in \mathbb{R}^{616}$, and the multiplier $\theta \in \mathbb{R}^{616}$. The number of functions in (33) equals 618.

Our method performed very well and converged to the optimal solution in 100–200 iterations, depending on the case, in *ca.* 20 min CPU time on a 1.6 GHz PC computer.

The utility functions, which play the role of the Lagrange multipliers for the dominance constraint are illustrated in Figure 1. We see that for Reference Portfolio 1, which contains only a small number of fast growing assets, the utility function is zero on almost the entire range of returns. Only very negative returns are penalized.

If the reference portfolio contains more assets, and is therefore more diversified and less risky, in order to dominate it, we have to use a utility function which introduces penalty

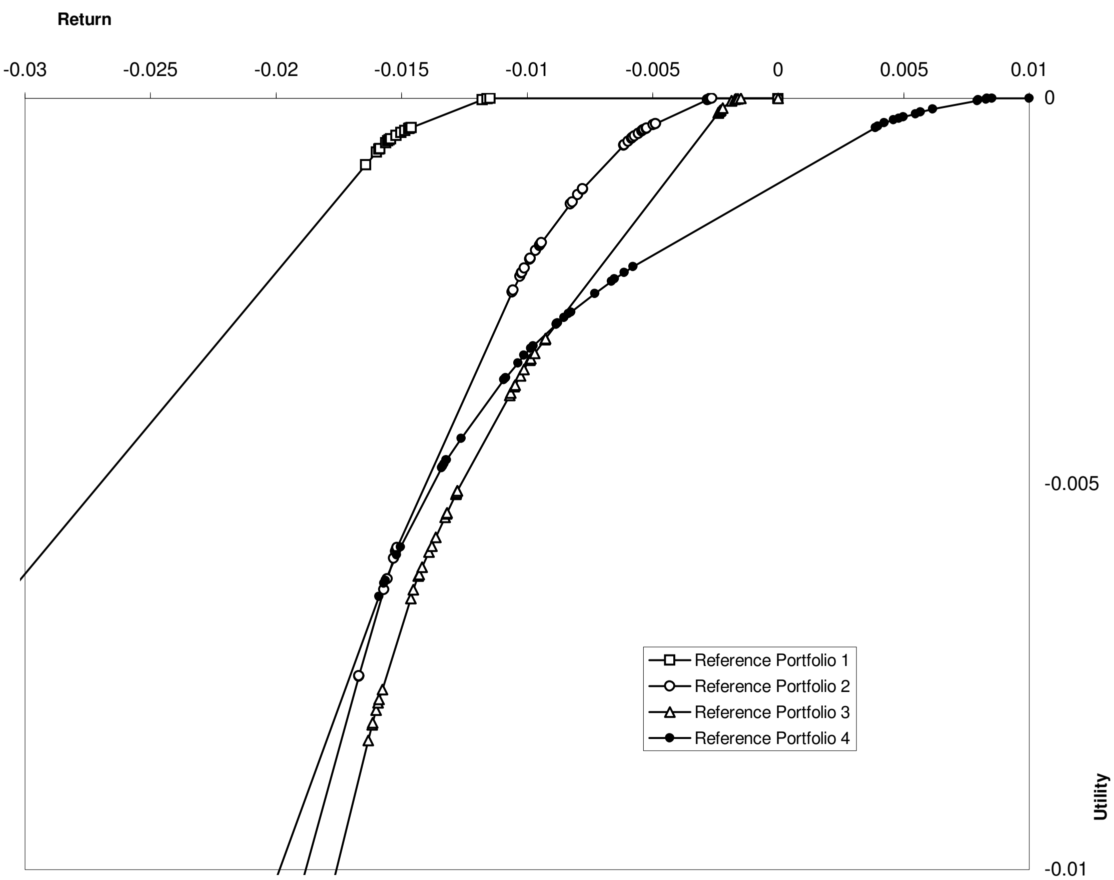

Figure 1: Utility functions corresponding to dominance constraints for four reference portfolios.

for a broader range of returns and is steeper. For the broadly based index in Reference Portfolio 4, the optimal utility function is more smooth and covers even positive returns.

It is worth mentioning that all these utility functions, although nondecreasing and concave, have rather complicated shapes. It would be a very hard task to guess the utility function that should be used to obtain a solution which dominates our reference portfolio.

References

[1] J.F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer-Verlag, New York, 2000.

[2] D. Dentcheva and A. Ruszczyński, Optimization under linear stochastic dominance, *Comptes Rendus de l'Academie Bulgare des Sciences* 56 (2003), No. 6, pp. 6–11.

[3] D. Dentcheva and A. Ruszczyński, Optimization under nonlinear stochastic dominance, *Comptes Rendus de l'Academie Bulgare des Sciences* 56 (2003), No. 7, pp. 19–25.

[4] D. Dentcheva and A. Ruszczyński, Optimization with stochastic dominance constraints, submitted for publication.

[5] D. Dentcheva and A. Ruszczyński, Optimality and duality theory for stochastic optimization problems with nonlinear dominance constraints, *Mathematical Programming*, accepted for publication.

[6] P. C. Fishburn, *Decision and Value Theory*, John Wiley & Sons, New York, 1964.

[7] J. Hadar and W. Russell, *Rules for ordering uncertain prospects*, The Amaerican Economic Review, 59 (1969), pp. 25–34.

[8] G. Hanoch and H. Levy, *The efficiency analysis of choices involving risk*, Review of Economic Studies, 36 (1969), pp. 335–346.

[9] G. H. Hardy, J. E. Littlewood and G. Pólya, *Inequalities*, Cambridge University Press, Cambridge, MA, 1934.

[10] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms*, Springer-Verlag, Berlin, 1993.

[11] K.C. Kiwiel, *Methods of Descent for Nondifferentiable Optimization*, Lecture Notes in Mathematics 1133, Springer-Verlag, Berlin, 1985.

[12] H. Konno and H. Yamazaki, *Mean–absolute deviation portfolio optimization model and its application to Tokyo stock market*, Management Science, 37 (1991), pp. 519–531.

[13] E. Lehmann, *Ordered families of distributions*, Annals of Mathematical Statistics, 26 (1955), pp. 399–419.

[14] H. Levy, *Stochastic dominance and expected utility: survey and analysis*, Management Science, 38 (1992), pp. 555–593.

[15] H. M. Markowitz, *Portfolio selection*, Journal of Finance, 7 (1952), pp. 77–91.

[16] H. M. Markowitz, *Portfolio Selection*, John Wiley & Sons, New York, 1959.

[17] H. M. Markowitz, *Mean–Variance Analysis in Portfolio Choice and Capital Markets*, Blackwell, Oxford, 1987.

[18] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*, Academic Press, San Diego, 1979.

[19] K. Mosler and M. Scarsini (Eds.), *Stochastic Orders and Decision Under Risk*, Institute of Mathematical Statistics, Hayward, California, 1991.

[20] W. Ogryczak and A. Ruszczyński, *From stochastic dominance to mean–risk models: semideviations as risk measures*, European Journal of Operational Research, 116 (1999), pp. 33–50.

[21] W. Ogryczak and A. Ruszczyński, *On consistency of stochastic dominance and mean–semideviation models*, Mathematical Programming, 89 (2001), pp. 217–232.

[22] W. Ogryczak and A. Ruszczyński, *Dual stochastic dominance and related mean–risk models*, SIAM Journal on Optimization, 13 (2002), pp. 60–78.

[23] J.P Quirk and R. Saposnik, *Admissibility and measurable utility functions*, Review of Economic Studies, 29 (1962), pp. 140–146.

[24] R. T. Rockafellar, *Convex Analysis*, Princeton Univ. Press, Princeton, NJ, 1970.

[25] M. Rothschild and J. E. Stiglitz, *Increasing risk: I. A definition*, Journal of Economic Theory, 2 (1969), pp. 225–243.

[26] A. Ruszczyński, A regularized decomposition method for minimizing a sum of polyhedral functions, *Mathematical Programming* 35 (1986) 309-333.

[27] A. Ruszczyński and R.J. Vanderbei, *Frontiers of stochastically nondominated portfolios*, Econometrica, to appear in 2003.

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

[29] R.J. Vanderbei, *Linear Programming: Foundations and Extensions.* Kluwer Academic Publishers, 2nd edition, 2001.

[30] G. A. Whitmore and M. C. Findlay, eds., *Stochastic Dominance: An Approach to Decision–Making Under Risk*, D.C.Heath, Lexington, MA, 1978.