Risk-Averse Two-Stage Stochastic Linear Programming: Modeling and Decomposition

Naomi Miller

RUTCOR, Rutgers University
640 Bartholomew Rd., Piscataway, NJ 08854
E-mail: nmiller@rutcor.rutgers.edu

Andrzej Ruszczyński

Department of Management Science and Information Systems
Rutgers University
94 Rockafeller Rd, Piscataway, NJ 80854
E-mail: rusz@business.rutgers.edu

July 29, 2009

Abstract

We formulate a risk-averse two-stage stochastic linear programming problem in which unresolved uncertainty remains after the second stage. The objective function is formulated as a composition of conditional risk measures. We analyze properties of the problem and derive necessary and sufficient optimality conditions. Next, we construct two decomposition methods for solving the problem. The first method is based on the generic cutting plane approach, while the second method exploits the composite structure of the objective function. We illustrate their performance on a portfolio optimization problem.

1 Introduction

The main objective of this paper is to introduce a risk-averse version of the two-stage model of stochastic programming, analyze its properties, and propose efficient numerical methods. In the modeling part we use methods of the modern theory of risk measures, initiated by Artzner, Delbaen, Eber, and Heath [2]. Particularly relevant for us are the duality theory [2, 8, 11, 12, 13, 18, 24, 30, 31], and the theory of conditional and dynamic risk measures [3, 6, 7, 9, 10, 14, 15, 22, 34, 23, 30, 32]. The optimization part generalizes and extends decomposition approaches to risk-neutral two-stage stochastic linear programming problems (see [5, 17, 27] and the references therein).

In section 2 we motivate our model by discussing general principles of dynamic risk measurement. Section 3 quickly reviews basic properties of conditional risk mappings. In section 4 we develop the nested form of a risk-averse two-stage problem. In section 5 we derive optimality conditions for the problem. In section 6 we discuss two decomposition methods for solving the problem: the basic method, and a new specialized multicut method, which exploits the composite structure of dynamic risk measures. Finally, section 7 contains results of numerical experiments on a portfolio example, which illustrate the efficiency of our approach.

Our model and results differ from earlier publications on risk-averse two-stage models [1, 35, 36] in several ways. We consider an extended two-stage model, in which there is still unresolved uncertainty after the second-stage decision is made. Because of that, we need to use a risk-averse version of the second-stage problem, while in earlier publications, similarly to the risk-neutral case, the second-stage problem was deterministic. Our approach allows for modeling important application problems, such as dynamic portfolio problems. Finally, our risk-averse multicut method of section 6 is a substantial improvement over the existing decomposition approaches, even for the simpler mean–risk models considered in [1, 35, 36].

2 The Model

Let $\Omega = \{1, \dots, N_2\}$ be a finite probability space with the σ-algebra $\mathscr{F}$ of all possible subsets of Ω, and with probabilities of elementary events $P[j] > 0$, $j = 1, \dots, N_2$. Let $\mathscr{F}_1 \subset \mathscr{F}$ by a σ-subalgebra given by disjoint events Ω_i, $i = 1, \dots, N_1$, constituting a partition of Ω and let $p_i = P[\Omega_i] > 0$, $i = 1, \dots, N_1$. For each elementary event $j \in \Omega_i$ we define the conditional probability $p_{ij} = P\big[j \mid \Omega_i\big] = P[j]/P[\Omega_i]$.

We have the following problem data: a deterministic matrix A of dimension $m_x \times n_x$, a deterministic vector $b \in \mathbb{R}^{m_x}$, random vectors $c \in \mathbb{R}^{n_x}$, $q \in \mathbb{R}^{n_y}$ and $h \in \mathbb{R}^{m_y}$, and random matrices T of dimension $m_y \times n_x$ and W of dimension

$m_y \times n_y$. We assume that c, h, T, and W are $\mathscr{F}_1$-measurable, while q is only $\mathscr{F}$-measurable. We denote the values of T, W, and h on Ω_i by T_i, W_i, and h_i, and the values of q on an elementary event $j \in \Omega_i$ by q_{ij}.

In a two-stage stochastic programming model two groups of decision variables can be distinguished. The first-stage decision vector $x \in \mathbb{R}^{n_x}$ has to be determined before any of the random problem data are observed. The second-stage decision vector $y \in \mathbb{R}^{n_y}$ is determined after an elementary event in $\mathscr{F}_1$ is observed. Therefore we can view the second-stage decision vector as a random vector Y, with realizations $y_i \in \mathbb{R}^{n_y}$ corresponding to the events Ω_i, $i = 1, \dots, N_1$.

A linear two-stage stochastic linear programming problem in its extended form is formulated as follows:

$$\begin{aligned} \min_{x,Y} \quad & \sum_{i=1}^{N_1} p_i \big[c_i^\mathsf{T} x + \sum_{j \in \Omega_i} p_{ij} q_{ij}^\mathsf{T} y_i \big] \\ \text{s.t.} \quad & Ax = b, \quad x \geq 0, \\ & T_i x + W_i y_i = h_i, \quad i = 1, \dots, N_1, \\ & y_i \geq 0, \quad i = 1, \dots, N_1. \end{aligned} \tag{1}$$

By employing interchangeability [25, Thm. 14.60] and conditioning, this problem can be rewritten in the nested form:

$$\begin{aligned} \min_{x} \quad & \bar{c}^\mathsf{T} x + \sum_{i=1}^{N_1} p_i V_i(x) \\ \text{s.t.} \quad & Ax = b, \quad x \geq 0. \end{aligned} \tag{2}$$

Here $\bar{c} = \mathbb{E}[c] = \sum_{i=1}^{N_1} p_i c_i$ and $V_i(x)$ is the realization of the second-stage cost in event Ω_i, which is defined as the optimal value of the second-stage problem

$$\begin{aligned} \min_{y} \quad & \bar{q}_i^\mathsf{T} y \\ \text{s.t.} \quad & T_i x + W_i y = h_i, \\ & y \geq 0, \end{aligned} \tag{3}$$

with $\bar{q}_i = \mathbb{E}[q|\Omega_i] = \sum_{j \in \Omega_i} p_{ij} q_{ij}$. All data of problem (3) are $\mathscr{F}_1$-measurable, and thus it can be solved separately for each $i = 1 \dots, N_1$. In fact, every solution of (1) corresponds to a solution of (2)–(3) and vice versa. The reader is referred to [37, Ch. 2] and the references therein for a detailed discussion of these issues.

Due to the fact that the objective function has the form of an expected value, the randomness of c and the randomness of q beyond $\mathscr{F}_1$ are irrelevant in the risk-neutral formulation. That is why risk-neutral two-stage models usually assume that c is deterministic and q becomes known after the first stage. However, in a risk-averse setting these simplifications are not justified; even for a single stage model a random cost $c^\top x$ is not indistinguishable from its expected value $\bar{c}^\top x$ and we may have strict preference among different random costs sharing the same expected value. A similar distinction has to be made at the second stage: a random cost $q_i^\top y$ is not equivalent to its conditional expectation $\bar{q}_i^\top y$.

Our objective is to formulate and analyze a risk-averse version of problem (1). Observe that our first-stage cost $c^\top x$ is an $\mathscr{F}_1$-measurable random variable, while the second-stage cost $q^\top Y$ is an $\mathscr{F}$-measurable random variable. Identifying them with vectors in $\mathbb{R}^{N_1}$ and $\mathbb{R}^{N_2}$, respectively, we can write an abstract risk-averse two-stage problem as follows:

$$\begin{aligned} \min_{x,Y} \quad & \varrho\big[c^\top x, q^\top Y\big] \\ \text{s.t.} \quad & Ax = b, \quad x \geq 0, \\ & T_i x + W_i y_i = h_i, \quad i = 1, \dots, N_1, \\ & y_i \geq 0, \quad i = 1, \dots, N_1. \end{aligned} \tag{4}$$

In this problem $\varrho : \mathbb{R}^{N_1} \times \mathbb{R}^{N_2} \to \mathbb{R}$ is a certain risk measure representing our preferences. We interpret the value of $\varrho[U, Z]$ as a fair one-time charge we would be willing to pay instead of random costs U and Z at stages 1 and 2. Many specific forms of ϱ can be employed here, following the general theory of dynamic measures of risk (see, *e.g.*, [7] and the references therein). The main feature that we want to preserve from the risk-neutral formulation is the possibility to define risk-averse first- and second-stage problems, and the equivalence of the extended formulation (4) and a nested formulation involving these problems, similarly to (2)–(3). In particular, for a solution $\hat{x}$, $\hat{Y}$ of problem (4), the decisions $\hat{y}_i$ should also be optimal for some second-stage problems, involving data available after the first stage. Moreover, for every $i = 1, \dots, N_1$, the second-stage problem should depend only on the elementary events $j \in \Omega_i$ that can actually happen after i.

To formalize these considerations, we assume that at the second stage we use another risk function $\rho_2(\cdot)$ to evaluate the risk of the second-stage cost $Z = q^\top Y$. As the first-stage event i is known at this time, we have $\rho_2 : \mathbb{R}^{N_2} \to \mathbb{R}^{N_1}$. Again, a useful interpretation of $\rho_2(Z)$ is the fair charge to be incurred after stage 1, instead of still uncertain second-stage cost Z: an $\mathscr{F}_1$-measurable equivalent of Z.

The following two concepts are fundamental for dynamic measures of risk (see, *e.g.*, [7]).

Local property. For every $i = 1, \dots, N_1$ there exists a function $\rho_{2i} : \mathbb{R}^{|\Omega_i|} \to \mathbb{R}$ such that for all $Z \in \mathbb{R}^{N_2}$ we have $\big[\rho_2(Z)\big]_i = \rho_{2i}(Z_i)$, where Z_i is the subvector of Z comprising the realizations Z_{ij}, $j \in \Omega_i$, that can actually be observed after the first-stage event i (this property is also called $\mathscr{F}_1$-regularity).

Time-consistency. For all $U, U' \in \mathbb{R}^{N_1}$ and $Z, Z' \in \mathbb{R}^{N_2}$ we have

$$\big\{U + \rho_2(Z) \leq U' + \rho_2(Z')\big\} \Rightarrow \big\{\varrho[U, Z] \leq \varrho[U', Z']\big\}.$$

In words, if for every first-stage event $i = 1, \dots, N_1$ the sum of the first-stage cost U_i and the second-stage risk $\rho_{2i}(Z_i)$ is smaller than $U_i' + \rho_{2i}(Z_i')$, then (U, Z) should be preferred over (U', Z') by the overall risk measure ϱ.

It follows from time-consistency that the value of $\varrho[U, Z]$ depends only on the sum $U + \rho_2(Z)$, and we can write it as a composition:

$$\varrho[U, Z] = \rho_1(U + \rho_2(Z)), \quad U \in \mathbb{R}^{N_1}, \quad Z \in \mathbb{R}^{N_2}. \tag{5}$$

In the formula above, $\rho_1 : \mathbb{R}^{N_1} \to \mathbb{R}$ is a real-valued nonincreasing function. It may be interesting to note that the structure (5) was introduced in [32] in a constructive way; here, we derive it from the property of time-consistency. For a more general discussion of this approach to time-consistency of dynamic risk measures, the reader is referred to [29].

3 Conditional Risk Measures

To proceed further, we impose on ρ_1 and ρ_2 more specific conditions. We follow the abstract construction proposed in [32], but with simplifications due to our assumption of a finite probability space. Uncertain outcomes are represented by vectors in a finite dimensional real space. We specify two vector spaces of uncertain outcomes: $\mathscr{Z} = \mathbb{R}^{N_2}$ and $\mathscr{U} = \mathbb{R}^{N_1}$. We also have a correspondence of the coordinates $i = 1, \dots, N_1$ in $\mathscr{U}$ to groups of coordinates $j \in \Omega_i$ in $\mathscr{Z}$, so that the sets $\Omega_1, \dots, \Omega_{N_1}$ form a partition of $\{1, \dots, N_2\}$. It is convenient to consider a matrix $\mathbb{I}$ of dimension $N_2 \times N_1$ with entries

$$I_{ji} = \begin{cases} 1 & \text{if } j \in \Omega_i, \\ 0 & \text{otherwise}, \end{cases} \qquad i = 1, \dots, N_1, \quad j = 1, \dots, N_2.$$

For a vector $U \in \mathscr{U}$ the vector $\mathbb{I}U \in \mathscr{Z}$ has groups of identical components, one for each Ω_i.

A *coherent conditional measure of risk* is a function $\rho : \mathscr{Z} \to \mathscr{U}$ satisfying the following axioms:

Convexity: $\rho(\alpha Z + (1 - \alpha)U) \leq \alpha\rho(Z) + (1 - \alpha)\rho(U)$, for all $U, Z \in \mathscr{Z}$ and all $\alpha \in [0, 1]$;

Monotonicity: If $U, Z \in \mathscr{Z}$ and $U \leq Z$, then $\rho(U) \leq \rho(Z)$;

Translation Equivariance: If $U \in \mathscr{U}$ and $Z \in \mathscr{Z}$, then $\rho(\mathbb{I}U + Z) = U + \rho(Z)$;

Positive Homogeneity: If $t > 0$ and $Z \in \mathscr{Z}$, then $\rho(tV) = t\rho(Z)$.

Inequalities are understood componentwise in all these conditions.

The axioms are formulated for the case when the uncertain outcomes represent cost; an equivalent set of axioms can be formulated when the outcomes represent profits. If $\mathscr{U} = \mathbb{R}$, a conditional risk measure becomes a coherent risk measure of [2].

We can also remark here that the conditions of monotonicity and translation equivariance imply the local property of a conditional measure of risk ρ, see [7, Prop. 3.3.].

Consider the setting of section 2, with $\mathscr{Z} = \mathbb{R}^{N_2}$ and $\mathscr{U} = \mathbb{R}^{N_1}$. Important examples of conditional measures of risk ρ_2 that can be used in (5) are obtained from conditional mean–risk models:

$$\rho_2(Z) = \mathbb{E}[Z|\mathscr{F}_1] + \varkappa r[Z|\mathscr{F}_1], \tag{6}$$

with a parameter $\varkappa > 0$ and with some risk functional $r : \mathscr{Z} \to \mathscr{U}$ representing the variability of the outcome. In (6) the conditional expectation $\mathbb{E}[Z|\mathscr{F}_1]$ is understood as a vector in $\mathbb{R}^{N_1}$ having components $\sum_{j\in\Omega_i} p_{ij} Z_j$, $i = 1, \dots, N_1$. The coefficient $\varkappa$ may be an $\mathscr{F}_1$-measurable random variable.

In particular, we may set $r[Z|\mathscr{F}_1]$ to be the *conditional semideviation* of order $p \geq 1$, having coordinates

$$\big(\sigma_p[Z|\mathscr{F}_1]\big)_i = \left[\sum_{j\in\Omega_i} p_{ij} \left(v_j - \sum_{k\in\Omega_i} p_{ik} v_k\right)_+^p\right]^{1/p}, \quad i = 1, \dots, N_1. \tag{7}$$

We can also use as $r[Z|\mathscr{F}_1]$ the *conditional weighted mean deviation from quantile*, with coordinates

$$\big(r_\alpha[Z|\mathscr{F}_1]\big)_i = \min_{\eta\in\mathbb{R}} \sum_{j\in\Omega_i} p_{ij} \max\left(\frac{1-\alpha}{\alpha}(v_j - \eta), \eta - v_j\right), \quad i = 1, \dots, N_1, \tag{8}$$

In both cases, for every $\mathscr{F}_1$-measurable $\varkappa \in [0, 1]$, the resulting function (6) satisfies the axioms of a conditional measure of risk.

If we substitute $\mathscr{Z} = \mathbb{R}^{N_1}$, $\mathscr{U} = \mathbb{R}$ and $\mathscr{F}_1 = \{\Omega, \emptyset\}$, we obtain examples of (unconditional) coherent measures of risk, which may be used as ρ_1 in (5), that is

$$\rho_1(U) = \mathbb{E}[U] + \varkappa r[U]. \tag{9}$$

We may set $r[U]$ to be the *semideviation* of order $p \geq 1$:

$$\sigma_p[U] = \left[\sum_{i=1}^{N_1} p_i \left(u_i - \sum_{k=1}^{N_1} p_k u_k\right)_+^p\right]^{1/p}, \tag{10}$$

or the *weighted mean deviation from quantile*:

$$r_\alpha[U] = \min_{\eta\in\mathbb{R}} \sum_{i=1}^{N_1} p_i \max\left(\frac{1-\alpha}{\alpha}(u_i-\eta), \eta - u_i\right). \tag{11}$$

The reader is referred to [19, 20, 21, 31, 32] for an extensive analysis of these risk measures.

The key property of conditional risk measures is their dual representation. The following is a special case of [32, Thm. 3.1]. Define for each $i = 1,\dots,N_1$ the set

$$\mathscr{P}_i = \left\{\mu \in \mathbb{R}_+^{|\Omega_i|} : \sum_{j\in\Omega_i} \mu_j = 1\right\}.$$

Theorem 1. *If $\rho : \mathscr{Z} \to \mathscr{U}$ is a conditional risk measure then there exist closed convex sets $\mathscr{A}_i \subset \mathscr{P}_i$, $i = 1,\dots,N_1$, such that for all $Z \in \mathscr{Z}$*

$$\big(\rho(Z)\big)_i = \max_{\mu\in\mathscr{A}_i} \sum_{j\in\Omega_i} \mu_j v_j, \quad i = 1,\dots,N_1. \tag{12}$$

Observe that due to (12) the value of $\rho(Z)$ in event i indeed depends only on the realizations v_j, $j \in \Omega_i$, that is, ρ has the local property.

4 The Nested Problem Formulation

Consider now the risk-averse two-stage problem (4) with ϱ defined as a composition (5) of a coherent measure of risk ρ_1 and a conditional measure of risk ρ_2. To simplify notation, it is convenient to introduce a matrix C of dimension $N_1 \times n_x$, whose rows are vectors c_i^T, $i = 1,\dots,N_1$. We obtain the formulation

$$\begin{aligned} \min_{x,Y} \quad & \rho_1\big(Cx + \rho_2(q^\mathsf{T}Y)\big) \\ \text{s.t.} \quad & Ax = b, \quad x \ge 0, \\ & T_i x + W_i y_i = h_i, \quad i = 1,\dots,N_1, \\ & y_i \ge 0, \quad i = 1,\dots,N_1. \end{aligned} \tag{13}$$

This formulation allows for a series of important simplifications. Define the set

$$X = \{x \in \mathbb{R}^{n_x} : Ax = b,\ x \ge 0\},$$

and for every $x \in X$ the sets

$$Y_i(x) = \left\{y \in \mathbb{R}_+^{n_y} : W_i y = h_i - T_i x\right\}, \quad i = 1,\dots,N_1,$$

Owing to the dual representation of ρ_2, an ith component of $\rho_2(q^\top Y)$ depends only on the realizations of $q^\top Y$ in elementary events $j \in \Omega_i$, that is, on y_i and on the realizations q_{ij} of q, $j \in \Omega_i$. It is convenient to define the matrix Q_i of dimension $|\Omega_i| \times n_y$ having as its rows the vectors $q_{ij}^\top$, $j \in \Omega_i$. With this notation, the ith component of $\rho_2(q^\top Y)$ can be written as

$$\big(\rho_2(q^\top Y)\big)_i = \rho_{2i}(Q_i y_i), \quad i = 1, \dots, N_1.$$

By the dual representation of $\rho_2(\cdot)$, each ρ_{2i} is a coherent risk measure defined by (12):

$$\rho_{2i}(Q_i y_i) = \max_{\nu \in \mathscr{A}_i} \sum_{j \in \Omega_i} \nu_j q_{ij}^\top y_i = \max_{\nu \in \mathscr{A}_i} \langle \nu, Q_i y_i \rangle, \quad i = 1, \dots, N_1. \tag{14}$$

Note that the sets $\mathscr{A}_i$ are compact, and thus the risk measures $\rho_{2i}(\cdot)$ are continuous.

Consider the first-stage *induced feasible sets*:

$$X^{\text{ind}} = \big\{ x \in \mathbb{R}^{n_x} : Y_i(x) \neq \emptyset \big\}, \quad i = 1, \dots, N_1,$$

and let

$$X^{\text{ind}} = \bigcap_{i=1}^{N_1} X_i^{\text{ind}}.$$

Exactly in the same way as in risk-neutral two-stage stochastic linear programming, we can show that the sets X_i^{ind}, $i = 1, \dots, N_1$, are closed convex polyhedra (see, *e.g.*, [37, Ch. 2]).

Define the second-stage optimal value functions:

$$V_i(x) = \inf_{y \in Y_i(x)} \rho_{2i}(Q_i y). \tag{15}$$

We assume that $V_i(x) > -\infty$ for all $x \in X_i^{\text{ind}}$ and for every $i = 1, \dots, N_1$. If $x \notin X_i^{\text{ind}}$, then $V_i(x) = +\infty$ by definition. Consequently, each $V_i(\cdot)$ is a proper convex function. The property that $V_i(x) > -\infty$ will be guaranteed by Assumption 2 to be formulated in the next section.

Due to the monotonicity of ρ_1, problem (13) can be equivalently transformed to a nested form

$$\min_{x \in X} \Big\{ \inf_{Y \in Y(x)} \rho_1\big(Cx + \rho_2(q^\top Y)\big) \Big\} = \min_{x \in X \cap X^{\text{ind}}} \rho_1\Big(Cx + \inf_{Y \in Y(x)} \rho_2(q^\top Y)\Big). \tag{16}$$

The innermost optimal value is understood as follows: for every $i = 1, \dots, N_1$ the value of $\inf_{Y \in Y(x)} \rho_2(q^\top Y)$ is given by (15). Equation (16) is the risk version of the *interchangeability principle* (see [32]), which is analogous to the transformation of (1) into (2)–(3). We can thus formulate the following *risk-averse first-stage problem*:

$$\min_{x \in X \cap X^{\text{ind}}} \rho_1\big(Cx + V(x)\big), \tag{17}$$

with the vector $V(x) \in \mathbb{R}^{N_1}$ having coordinates $V_i(x)$ defined by (15).

We have arrived at a formulation which is analogous to the risk-neutral model (2)–(3). The similarity becomes even more apparent, when we substitute for ρ_1 and ρ_{2i} the corresponding dual representations (12) and (14). We obtain the first-stage min-max problem:

$$\min_{x \in X \cap X^{\text{ind}}} \max_{\mu \in \mathscr{A}} \langle \mu, Cx + V(x) \rangle, \tag{18}$$

with each component $V_i(x)$ of $V(x)$ defined as the optimal value of the second-stage min-max problem

$$\min_{y \in Y_i(x)} \max_{\nu \in \mathscr{A}_i} \langle \nu, Q_i y \rangle, \quad i = 1, \ldots, N_1. \tag{19}$$

If $\mathscr{A}$ has only one element $(p_1, \ldots, p_{N_1})$, and if each $\mathscr{A}_i$, $i = 1, \ldots, N_1$, contains only one element $\big(p_{ij}, \ j \in \Omega_i\big)$, formulations (2)–(3) and (18)–(19) become identical.

5 Optimality Conditions

Let us consider the objective function of (17),

$$f(x) = \begin{cases} \rho_1\big(Cx + V(x)\big) & \text{if } x \in X^{\text{ind}}, \\ +\infty & \text{otherwise.} \end{cases} \tag{20}$$

The functions $V_i(\cdot)$, as optimal values of the second-stage problems (19) are convex and continuous on the sets X_i^{ind}. The function $\rho_1(\cdot)$ is convex and nondecreasing. Therefore, their composition (20) is convex. It is finite and continuous on the set X^{ind}.

Let us analyze the second-stage optimal value functions $V_i(\cdot)$. To this end we make the following assumption.

Assumption 2. *For every $i = 1, \ldots, N_1$ the set*

$$\mathscr{W}_i = \big\{\nu \in \mathscr{A}_i : (\exists\, \pi)\ W_i^{\mathsf{T}} \pi \le Q_i^{\mathsf{T}} \nu\big\} \tag{21}$$

is nonempty.

Theorem 3. *For all $x \in \mathbb{R}^{n_x}$ and every $i = 1, \ldots, N_1$ we have*

$$V_i(x) = \sup_{\pi \in \Pi_i} \langle \pi, h_i - T_i x \rangle, \tag{22}$$

where

$$\Pi_i = \big\{\pi \in \mathbb{R}^{m_y} : \exists\, (\nu \in \mathscr{A}_i)\ W_i^{\mathsf{T}} \pi \le Q_i^{\mathsf{T}} \nu\big\}. \tag{23}$$

Proof. From (19) we obtain

$$\begin{aligned} V_i(x) &= \inf_{y \in Y_i(x)} \max_{\nu \in \mathscr{A}_i} \langle \nu, Q_i y \rangle \\ &= \inf_{y \geq 0} \sup_{\pi \in \mathbb{R}^{m_x}} \max_{\nu \in \mathscr{A}_i} \big\{ \langle \nu, Q_i y \rangle + \langle \pi, h_i - T_i x - W_i y \rangle \big\} \\ &= \inf_{y \geq 0} \sup_{\nu \in \mathscr{A}_i} \left[\sup_{\pi \in \mathbb{R}^{m_x}} \big\{ \langle \nu, Q_i y \rangle + \langle \pi, h_i - T_i x - W_i y \rangle \big\} \right]. \end{aligned}$$

The function in brackets is convex and lower semicontinuous with respect to y and linear with respect to ν, and the set $\mathscr{A}_i$ is convex and compact. Therefore, we can exchange the outer "inf" and "sup" operations. We obtain

$$\begin{aligned} V_i(x) &= \sup_{\nu \in \mathscr{A}_i} \inf_{y \geq 0} \left[\sup_{\pi \in \mathbb{R}^{m_x}} \big\{ \langle \nu, Q_i y \rangle + \langle \pi, h_i - T_i x - W_i y \rangle \big\} \right] \\ &= \sup_{\nu \in \mathscr{A}_i} \left\{ \inf_{y \geq 0} \sup_{\pi \in \mathbb{R}^{m_x}} L_i(y, \pi, \nu) \right\}, \end{aligned} \tag{24}$$

where

$$L_i(y, \pi, \nu) = \langle \nu, Q_i y \rangle + \langle \pi, h_i - T_i x - W_i y \rangle.$$

Observe that $L_i(\cdot, \cdot, \nu)$ is the Lagrangian of the following linear programming problem having ν and x as its parameters:

$$\begin{aligned} \min_y \quad & \langle Q_i^{\mathsf{T}} \nu, y \rangle \\ \text{s.t.} \quad & W_i y = h_i - T_i x, \\ & y \geq 0. \end{aligned} \tag{25}$$

Its dual problem has the form

$$\begin{aligned} \max_\pi \quad & \langle \pi, h_i - T_i x \rangle \\ \text{s.t.} \quad & W_i^{\mathsf{T}} \pi \leq Q_i^{\mathsf{T}} \nu. \end{aligned} \tag{26}$$

From the duality theory of linear programming we know that the optimal values of both problems (finite or infinite) are equal, unless both problems have empty feasible sets.

Suppose $x \in X_i^{\text{ind}}$. Then the feasible set of the primal problem (25) is nonempty. As the "inf-sup" in (24) is its optimal value, we can exchange the "inf" and "sup" operations:

$$\inf_{y \geq 0} \sup_{\pi \in \mathbb{R}^{m_x}} L_i(y, \pi, \nu) = \sup_{\pi \in \mathbb{R}^{m_x}} \inf_{y \geq 0} L_i(y, \pi, \nu); \tag{27}$$

the right hand side of (27) is the optimal value of the dual problem. Applying this relation to (24), we get

$$V_i(x) = \sup_{\nu \in \mathscr{A}_i} \sup_{\pi \in \mathbb{R}^{m_x}} \inf_{y \geq 0} L_i(y, \pi, \nu). \tag{28}$$

Suppose $x \notin X_i^{\mathrm{ind}}$. Then the left hand side of (27) equals $+\infty$. If ν is such that the dual feasible set

$$\Pi_i(\nu) = \left\{ \pi \in \mathbb{R}^{m_x} : W_i^{\mathsf{T}} \pi \leq Q_i^{\mathsf{T}} \nu \right\}$$

is nonempty, then the duality relation (27) holds true. Hence, in this case we also have

$$\sup_{\pi \in \mathbb{R}^{m_x}} \inf_{y \geq 0} L_i(y, \pi, \nu) = +\infty.$$

By Assumption 2, the set of ν for which $\Pi_i(\nu) \neq \emptyset$ is nonempty, and thus it is sufficient in this case to consider $\nu \in \mathscr{W}_i \subset \mathscr{A}_i$ in formula (28) to conclude that $V_i(x) = \infty$. Therefore, formula (28) holds true in all cases.

Observe that the innermost "inf" in (28) is greater than $-\infty$ if and only if $\pi \in \Pi_i(\nu)$, in which case it is equal to $\langle \pi, h_i - T_i x \rangle$. Hence,

$$V_i(x) = \sup_{\nu \in \mathscr{A}_i} \sup_{\pi \in \Pi_i(\nu)} \langle \pi, h_i - T_i x \rangle.$$

As the function under the "sup-sup" operation does not depend on ν, we can write the last formula in the form (22), as required. □

Each set Π_i is convex and closed, because the set $\mathscr{A}_i$ is convex and compact. For $x \in X_i^{\mathrm{ind}}$ we define the solution set of the ith dual subproblem:

$$\mathscr{D}_i(x) = \left\{ \pi \in \Pi_i : \langle \pi, h_i - T_i x \rangle = V_i(x) \right\}.$$

Corollary 4. *At every $x \in X_i^{\mathrm{ind}}$ we have $\partial V_i(x) = -T_i^{\mathsf{T}} \mathscr{D}_i(x)$.*

Proof. Theorem 3 describes $V_i(\cdot)$ as the composition of the support function of the set Π_i,

$$\sigma_{\Pi_i}(\xi) = \sup_{\pi \in \Pi_i} \langle \pi, \xi \rangle,$$

with the affine function $x \mapsto h_i - T_i x$. Applying the formula for the subdifferential of the composition, we obtain

$$\partial V_i(x) = -T_i^{\mathsf{T}} \partial \sigma_{\Pi_i}(h_i - T_i x)$$

As Π_i is convex and closed, $\partial \sigma_{\Pi_i}(h_i - T_i x) = \mathscr{D}_i(x)$. □

We are now in the position to calculate the subdifferential of $f(\cdot)$ at $x \in X^{\mathrm{ind}}$. If both $\rho_1(\cdot)$ and $V(\cdot)$ were finite-valued, we could simply invoke standard results on subdifferentiating compositions of convex functions (see, *e.g.*, [16, Thm. 4.3.1]). The calculations below refine these results in our case.

We make an additional technical assumption.

Assumption 5. *For every $i = 1, \ldots, N_1$ there exists $\mu \in \mathscr{A}$ such that $\mu_i > 0$.*

We can always satisfy Assumption 5, by eliminating from considerations scenarios i for which $\mu_i = 0$ for all $\mu \in \mathscr{A}$.

Suppose $x \in X^{\text{ind}}$. Using the dual representation (12) of $\rho_1(\cdot)$ and Theorem 3, we get

$$f(x) = \sup_{\mu \in \mathscr{A}} \sum_{i=1}^{N_1} \mu_i \big(c_i^{\mathsf{T}} x + \sup_{\pi_i \in \Pi_i} \langle \pi_i, h_i - T_i x \rangle \big)$$
$$= \sup_{\mu \in \mathscr{A}} \sup_{\pi \in \Pi} \sum_{i=1}^{N_1} \mu_i \big(c_i^{\mathsf{T}} x + \langle \pi_i, h_i - T_i x \rangle \big), \tag{29}$$

where $\Pi = \Pi_1 \times \cdots \times \Pi_{N_1}$. If $x \notin X^{\text{ind}}$, then there exists j such that

$$V_j(x) = \sup_{\pi_j \in \Pi_j} \langle \pi_j, h_j - T_j x \rangle = +\infty.$$

By Assumption 5, we can find $\mu \in \mathscr{A}$ such that $\mu_j V_j(x) = +\infty$. As $V_i(x) > -\infty$ for all i, due to Theorem 3 (which uses Assumption 2), we conclude that the value of formula (29) is $+\infty$ in this case. Consequently, formula (29) is correct in all cases.

Define the function $\varphi : \mathbb{R}^{N_1} \times \mathbb{R}^{N_1 \times m_y} \to \overline{\mathbb{R}}$ as follows:

$$\varphi(\zeta, \xi_1, \ldots, \xi_{N_1}) = \sup_{\mu \in \mathscr{A}} \sup_{\pi \in \Pi} \sum_{i=1}^{N_1} \mu_i \big(\zeta_i + \langle \pi_i, \xi_i \rangle \big).$$

It is the support function of the set

$$\mathscr{S} = \big\{ s \in \mathbb{R}^{N_1} \times \mathbb{R}^{N_1 \times m_y} : s = (\mu, \mu_1 \pi_1, \ldots, \mu_{N_1} \pi_{N_1}),$$
$$\mu \in \mathscr{A},\ \pi_i \in \Pi_i,\ i = 1, \ldots, N_1 \big\}.$$

Lemma 6. *The set $\mathscr{S}$ is convex and closed.*

Proof. To prove convexity, consider $\mu^1, \mu^2 \in \mathscr{A}$, $\alpha \in (0, 1)$, and let

$$s^k = (\mu^k, \mu_1^k \pi_1^k, \ldots, \mu_{N_1}^k \pi_{N_1}^k), \quad k = 1, 2,$$
$$\pi_i^k \in \Pi_i, \quad i = 1, \ldots, N_1, \quad k = 1, 2,$$
$$s = \alpha s^1 + (1 - \alpha) s^2,$$
$$\mu = \alpha \mu^1 + (1 - \alpha) \mu^2.$$

By the convexity of $\mathscr{A}$, we have $\mu \in \mathscr{A}$. To show that $s \in \mathscr{S}$, it is sufficient to prove that for all $i = 1, \ldots, N_1$

$$\alpha \mu_i^1 \pi_i^1 + (1 - \alpha) \mu_i^2 \pi_i^2 \in \mu_i \Pi_i. \tag{30}$$

Observe that $\mu_i^1 \geq 0$ and $\mu_i^2 \geq 0$. Therefore $\mu_i \geq 0$. If $\mu_i = 0$, we must have $\mu_i^1 = \mu_i^2 = 0$ and (30) is trivial. It remains to consider the case of $\mu_i > 0$. Define

$$\beta_i = \frac{\alpha \mu_i^1}{\mu_i}.$$

The left hand side of (30) can be written as follows

$$\alpha \mu_i^1 \pi_i^1 + (1-\alpha)\mu_i^2 \pi_i^2 = \mu_i\big(\beta_i \pi_i^1 + (1-\beta_i)\pi_i^2\big).$$

Due to the convexity of Π_i, the right hand side is an element of $\mu_i \Pi_i$, which proves (30).

The closedness of $\mathscr{S}$ follows from the compactness of $\mathscr{A}$ and from the closedness of the sets Π_i. □

We can now describe the subdifferential of the objective function $f(\cdot)$.

Theorem 7. *At every $x \in X^{\text{ind}}$ we have*

$$\partial f(x) = \Bigg\{ g \in \mathbb{R}^{n_x} : g = C^{\mathsf{T}}\mu + \sum_{i=1}^{N_1} \mu_i T_i^{\mathsf{T}} \pi_i, \\ \mu \in \partial\rho_1\big(Cx + V(x)\big),\ \pi_i \in \mathscr{D}_i(x),\ i = 1, \dots, N_1 \Bigg\}. \tag{31}$$

Proof. The function $f(\cdot)$ is a composition of $\varphi(\cdot)$ with the affine function

$$x \mapsto (Cx, h_1 - T_1 x, \dots, h_{N_1} - T_{N_1} x).$$

Employing the formula for the subdifferential of a support function, and the formula for the subdifferential of the composition, we obtain

$$\partial f(x) = \Bigg\{ g \in \mathbb{R}^{n_x} : g = C^{\mathsf{T}}\mu + \sum_{i=1}^{N_1} \mu_i T_i^{\mathsf{T}} \pi_i, \\ (\mu, \mu_1\pi_1, \dots, \mu_{N_1}\pi_{N_1}) \in \hat{\mathscr{S}}(L(x)) \Bigg\},$$

where

$$\hat{\mathscr{S}}(\zeta, \xi_1, \dots, \xi_{N_1}) = \Bigg\{ (\mu, \mu_1\pi_1, \dots, \nu_{N_1}\pi_{N_1}) \in \mathscr{S} : \\ \langle \mu, \zeta \rangle + \sum_{i=1}^{N_1} \mu_i \langle \pi_i, \xi_i \rangle = \varphi(\zeta, \xi_1, \dots, \xi_{N_1}) \Bigg\}.$$

Observe that we have $\pi_i \in \mathscr{D}_i(x)$ whenever $\mu_i > 0$. Moreover $\mu \in \partial\rho_1\big(Cx + V(x)\big)$. Thus, the last expression is equivalent to (31). □

We complete this section by formulating optimality conditions for problem (17). The theorem below is an immediate consequence of standard optimality conditions in convex optimization (see, *e.g.*, [28, Thm. 3.33]), combined with the description of the subdifferential of $f(\cdot)$ provided in Theorem 7.

Theorem 8. *A point* $\hat{x} \in X \cap X^{\text{ind}}$ *is an optimal solution of problem* (17) *if and only if there exist* $\hat{\mu} \in \partial\rho_1\big(C\hat{x} + V(\hat{x})\big)$ *and* $\hat{\pi}_i \in \mathscr{D}_i(\hat{x})$, $i = 1, \dots, N_1$, *such that*

$$0 \in C^{\mathsf{T}}\hat{\mu} + \sum_{i=1}^{N_1} \hat{\mu}_i T_i^{\mathsf{T}}\hat{\pi}_i + \mathscr{N}_X(\hat{x}), \tag{32}$$

where $\mathscr{N}_X(\hat{x})$ *denotes the normal cone to* X *at* $\hat{x}$.

One obtain even more specific conditions, by employing a description of the normal cone $\mathscr{N}_X(\hat{x})$, but form (32) is sufficient for our purposes.

6 Decomposition Methods

Problem (17) can be solved by a cutting plane method, which is quite similar to the L-shaped decomposition of [38], and to the method discussed in [1].

Recall that feasibility of the second-stage problem (19) at $x = \bar{x}$ can be checked by the following Phase I problem:

$$\begin{aligned} \min_{y,s}\quad & \|s\|_1 \\ \text{s.t.}\quad & W_i y + s = h_i - T_i\bar{x}, \\ & y \geq 0. \end{aligned} \tag{33}$$

If the optimal value β_i of this problem is equal to 0, then $\bar{x} \in X_i^{\text{ind}}$. Otherwise, denoting by π_i the vector of Lagrange multipliers associated with the constraints, we can construct the following *feasibility cut*:

$$\beta_i - \big\langle T_i^{\mathsf{T}}\pi_i, x - \bar{x}\big\rangle \leq 0. \tag{34}$$

It is well-known that every point $x \in X_i^{\text{ind}}$ satisfies this inequality. If $\beta_i > 0$, the current point $\bar{x}$ is cut off.

In the algorithm below we assume that the set X is compact, and that we know a lower bound $\eta_{\min}$ for the optimal value of problem (17).

Basic Decomposition Method

Step 0: Set $k = 0$, $L^{\text{obj}} = \emptyset$, $L^{\text{feas}} = \emptyset$

Step 1: Solve the *master problem*

$$\begin{aligned} \min_{x,\eta} \quad & \eta \\ \text{s.t} \quad & \eta \geq \rho_1^\ell + \langle g^\ell, x - x^\ell \rangle, \quad \ell \in L^{\text{obj}}, \\ & \beta^\ell + \langle g^\ell, x - x^\ell \rangle \leq 0, \quad \ell \in L^{\text{feas}}, \\ & x \in X, \quad \eta \geq \eta_{\min}, \end{aligned}$$

and denote the solution by x^k and η^k.

Step 2a: Let $i = 1$.

Step 2b: Solve problem (33) and let β^k be its optimal value and π^k denote the vector of Lagrange multipliers. If $\beta^k > 0$ then let $g^k := -T_i^{\mathsf{T}}\pi^k$, $L^{\text{feas}} := L^{\text{feas}} \cup \{k\}$ and go to Step 6. Otherwise, continue.

Step 2c: Solve problem (19) and denote by V_i^k its optimal value and by π_i^k the vector of Lagrange multipliers.

Step 2d: If $i < N_1$ then increase i by 1 and go to Step 2b. Otherwise, go to Step 3.

Step 3: Solve the problem

$$\max_{\mu \in \mathscr{A}} \langle \mu, Cx^k + V^k \rangle$$

and denote its solution by μ^k and the optimal value by ρ_1^k.

Step 4: If $\rho_1^k = \eta^k$, then stop. Otherwise, continue.

Step 5: Let $L^{\text{obj}} := L^{\text{obj}} \cup \{k\}$ and let

$$g^k := \sum_{i=1}^{N} \mu_i^k \big(c_i - T_i^{\mathsf{T}}\pi_i^k\big).$$

Step 6: Increase k by 1 and go to Step 1.

Convergence of this method follows from general convergence properties of a cutting plane method (see, *e.g.*, [28, sec. 7.2]). In particular, when $\mathscr{A}$ and $\mathscr{A}(i)$, $i = 1, \dots, N_1$ are polyhedra, convergence is finite.

In the next algorithm we assume that the set X is compact, and that we know a lower bound $\eta_{\min}$ for the optimal value of problem (17), as well as lower bounds $[w_{\min}]_i$ on the optimal values of the second-stage problems (19). The method extends to the risk-averse case the idea of the multicut method for risk-neutral problems, developed in [4, 26].

Risk-Averse Multicut Method

Step 0: Set $k = 0$, $L_i^{\text{obj}} = \emptyset$, $L_i^{\text{feas}} = \emptyset$, $i = 1, \ldots, N_1$.

Step 1: Solve the *master problem*

$$\begin{aligned}
\min_{x,\eta,w} \quad & \eta \\
\text{s.t} \quad & \eta \geq \langle \mu^\ell, Cx + w \rangle, \quad \ell \in \bigcap_{i=1}^{N_1} L_i^{\text{obj}}, \\
& w_i \geq V_i^\ell + \langle g_i^\ell, x - x^\ell \rangle, \quad \ell \in L_i^{\text{obj}}, \quad i = 1, \ldots, N_1, \\
& \beta_i^\ell + \langle g_i^\ell, x - x^\ell \rangle \leq 0, \quad \ell \in L_i^{\text{feas}}, \quad i = 1, \ldots, N_1, \\
& x \in X, \quad \eta \geq \eta_{\min}, \quad w \geq w_{\min}.
\end{aligned}$$

and denote the solution by x^k, η^k, w^k.

Step 2a: Let $i = 1$.

Step 2b: Solve problem (33) and let β_i^k be its optimal value and π_i^k denote the vector of Lagrange multipliers. If $\beta_i^k > 0$ then let $g_i^k := -T_i^\mathsf{T} \pi_i^k$, $L_i^{\text{feas}} := L_i^{\text{feas}} \cup \{k\}$, and go to Step 2d; otherwise, continue.

Step 2c: Solve problem (19) and denote by V_i^k its optimal value and by π_i^k the vector of Lagrange multipliers.

Step 2d: If $i < N_1$ then increase i by 1 and go to Step 2b; otherwise, continue.

Step 2e: If $\beta_i^k > 0$ for at least one $i = 1, \ldots, N_1$, then go to Step 5; otherwise continue.

Step 3: Solve the problem

$$\max_{\mu \in \mathscr{A}} \langle \mu, Cx^k + V^k \rangle$$

and denote its solution by μ^k and the optimal value by ρ_1^k.

Step 4: If $\rho_1^k = \eta^k$, then stop. Otherwise, continue.

Step 5: Increase k by 1 and go to Step 1.

There are two differences between this method and the basic decomposition method. First, we do not aggregate cuts for $\rho_{2i}(\cdot)$, but rather maintain separate convex polyhedral models

$$\rho_{2i}(x) \geq V_i^\ell + \langle g_i^\ell, x - x^\ell \rangle, \quad \ell \in L_i^{\text{obj}}, \quad i = 1, \ldots, N_1.$$

This is similar to the idea employed in [4, 26] for risk-neutral models. Secondly, we memorize all previous measures μ^ℓ and we use them at Step 1, to construct a more accurate lower bound for $\rho_1(\cdot)$. This is specific for risk-averse models and leads to significant improvements, as we shall see in the next section.

Again, convergence of this method follows from general convergence properties of a cutting plane method (see, *e.g.*, [28, sec. 7.2]). In particular, when $\mathscr{A}$ and $\mathscr{A}(i)$, $i = 1, \ldots, N_1$ are polyhedra, the convergence is finite.

7 Application to Portfolio Optimization

In order to compare the methods presented in the previous section we consider the following two-stage portfolio optimization problem. There are n_x securities available, and we plan to invest in them in two stages. The amounts invested in the first stage are represented by the first-stage decision variables x_s, $s = 1, \ldots, n_x$. For simplicity, we assume that

$$x \in X = \left\{ x \in \mathbb{R}^{n_x} : \sum_{s=1}^{N_x} x_s = 1, \quad x_s \geq 0,\ s = 1, \ldots, n_x \right\}.$$

One of N_1 first-stage scenarios may occur, with probabilities p_i, $i = 1, \ldots, N_1$. The "cost vectors" c_i, $i = 1, \ldots, N_1$ are zero. Let R^1_{is} represent return of security s in scenario i in the first stage. After the first stage, the portfolio is rebalanced, and each second-stage decision y_{is} represents amount invested in security s after the first-stage scenario i. The second-stage "cost vectors" are defined as

$$q_{js} = -(1 + R^2_{js}), \quad j \in \Omega_i, \quad i = 1, \ldots, N_1, \quad s = 1, \ldots, n_x,$$

with R^2_{js} representing return of security s in scenario j in the second stage. As the rebalancing must be self-financing, the relation between the first and second-stage variables takes on the form

$$\sum_{s=1}^{n_x} y_{is} + \alpha \sum_{s=1}^{n_x} \left| y_{is} - (1 + R^1_{is}) x_s \right| \leq \sum_{s=1}^{n_x} (1 + R^1_{is}) x_s, \quad i = 1, \ldots, N_1,$$

where $\alpha \in (0, 1)$ is the coefficient of proportional transaction costs. It is clear that these conditions can be equivalently represented by linear inequalities. We also require that $y_{is} \geq 0$.

On both stages we used mean–semideviation measures of risk, given by (9) and (6), with (10) and (7), respectively. We set $p = 1$ and $\varkappa = 1$, and we considered $n_x = 500$.

Due to the polyhedral form of the mean–semideviation measures, our problem (18)–(19) could be also formulated as one large scale linear programming problem.

Scenario Tree	Direct Linear Programming				Decomposition Methods			
	Simplex		Barrier		Basic		Multicut	
$N_1 \times$ M	Iter.	CPU	Iter.	CPU	Iter.	CPU	Iter.	CPU
20 × 20	957	1.3	43	44.9	23	38.9	10	18.5
50 × 50	6168	36.4	33	1974.5	45	206.4	11	53.6
100 × 100	22797	291.3	-	-	45	473.1	14	158.1
200 × 200	-	-	-	-	56	1717	18	575.9

Table 1: Comparison of methods.

The table below compares performance of two linear programming methods, the simplex method and the interior point method, with our decomposition methods, for different numbers of first and second-stage scenarios. The sizes $N_1 \times M$ in the first column of Table 1 mean than we had N_1 first-stage scenarios, each of them followed by M second-stage scenarios. Thus the total number of elementary events is $N_2 = N_1 M$.

We can see from the results that the decomposition methods dramatically outperform direct linear programming methods. The risk-averse multicut method is uniformly more efficient than the basic decomposition method.

All calculations were carried out on a Toshiba notebook with 2.00 GHz Core(TM)2 CPU and 2 GB of RAM, by using version 9.1 of the CPLEX solver.

References

[1] Ahmed, S., Convexity and decomposition of mean-risk stochastic programs, *Mathematical Programming*, 106, 433-446, 2006.

[2] Artzner, P., F. Delbaen, J. M. Eber and D. Heath, Coherent measures of risk, *Mathematical Finance*, 9, 203–228, 1999.

[3] Artzner, P., F. Delbaen, J.-M. Eber, D. Heath, and H. Ku, Coherent multiperiod risk adjusted values and Bellman's principle, *Annals of Operations Research*, 152:5–22, 2007.

[4] Birge, J. R., and F. V. Louveaux, A multicut algorithm for two-stage stochastic linear programs, *European Journal of Operational Research*, 34, 384–392, 1988.

[5] Birge, J. R., and F. V. Louveaux, *Introduction to Stochastic Programming*, Springer-Verlag, New York, 1997.

[6] Cheridito, P., F. Delbaen, and M. Kupper, Coherent and convex risk measures for bounded càdlàg processes. *Stochastic Processes and their Applications*, 112:1–22, 2004.

[7] Cheridito, P., F. Delbaen, and M. Kupper, Dynamic monetary risk measures for bounded discrete-time processes. *Electronic Journal of Probability*, 11:57–106, 2006.

[8] Delbaen, F., Coherent risk measures on general probability spaces, In *Essays in Honour of Dieter Sondermann*. Springer-Verlag, Berlin, Germany, 2002.

[9] Eichhorn, A., and W. Römisch, Polyhedral risk measures in stochastic programming. *SIAM Journal on Optimization*, 16:69–95, 2005.

[10] Föllmer, and, Penner, I., Convex risk measures and the dynamics of their penalty functions, *Statistics and Decisions*, 24, 61–96, 2006.

[11] Föllmer, H., and A. Schied, *Convex measures of risk and trading constraints*, *Finance and Stochastics*, 6, 429–447, 2002.

[12] Föllmer, H., and A. Schied, *Stochastic Finance. An Introduction in Discrete Time*, Walter de Gruyter, Berlin, 2004.

[13] Fritelli, M, and E. Rosazza Gianin, Putting order in risk measures, *Journal of Banking and Finance*, 26, 1473–1486, 2002.

[14] Frittelli, M., and E. Rosazza Gianin. Dynamic convex risk measures. In G. Szegö, editor, *Risk Measures for the 21st Century*, pages 227–248. John Wiley & Sons, Chichester, 2005.

[15] Fritelli, M., and G. Scandolo, Risk measures and capital requirements for processes. *Mathematical Finance*, 16:589–612, 2006.

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

[17] Kall, P., and J. Mayer, *Stochastic Linear Programming*, Springer-Verlag, New York, 2005.

[18] Leitner, J., A short note on second-order stochastic dominance preserving coherent risk measures, *Mathematical Finance*, 15:649–651, 2005.

[19] Ogryczak, W. and A. Ruszczyński, From stochastic dominance to mean-risk models: Semideviations as risk measures, *European Journal of Operational Research*, 116, 33–50, 1999.

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

[21] Ogryczak, W. and A. Ruszczyński, Dual stochastic dominance and related mean-risk models, *SIAM Journal on Optimization*, 13(1), 60–78, 2002.

[22] Pflug, G. Ch., and W. Römisch. *Modeling, Measuring and Managing Risk*. World Scientific, Singapore, 2007.

[23] Riedel, F., Dynamic coherent risk measures, *Stochastic Processes and Their Applications*, 112, 185–200, 2004.

[24] Rockafellar, R. T., Uryasev, S., and M. Zabarankin, *Deviation measures in risk analysis and optimization*, Finance and Stochastics, 10, 51–74, 2006.

[25] Rockafellar, R. T., and R. J.-B. Wets, *Variational Analysis*, Springer-Verlag, Berlin, 1998.

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

[27] Ruszczyński, A., Decomposition Methods, in: (A. Ruszczyński and A. Shapiro (Eds.), *Stochastic Programming*, Elsevier, Amsterdam, 2003.

[28] Ruszczyński, A., *Nonlinear Optimization*, Princeton University Press, Princeton 2006.

[29] Ruszczyński, A., Risk-averse dynamic programming for Markov decision processes, *manuscript submitted for publication*.

[30] Ruszczyński, A. and A. Shapiro, Optimization of risk measures, In *Probabilistic and Randomized Methods for Design under Uncertainty*, G. Calafiore and F. Dabbene (Eds.), Springer, London, 2005.

[31] Ruszczyński, A. and A. Shapiro, Optimization of convex risk functions, *Mathematics of Operations Research*, 31, 433–452, 2006.

[32] Ruszczyński, A. and A. Shapiro, Conditional risk mappings, *Mathematics of Operations Research*, 31 544–561, 2006.

[33] Ruszczyński, A. and R. J. Vanderbei, Frontiers of stochastically nondominated portfolios, *Econometrica*, 71(4), 1287–1297, 2003.

[34] Scandolo, G., *Risk Measures in a Dynamic Setting*, PhD Thesis, Università degli Studi di Milano, Milan, 2003.

[35] Schultz, R., and S. Tiedemann, Risk aversion via excess probabilities in stochastic programs with mixedinteger recourse, *SIAM Journal on Optimization*, 14, 115—138, 2003.

[36] Schultz, R., and S. Tiedemann, Conditional value-at-risk in stochastic programs with mixed-integer recourse, *Mathematical Programming*, 105, 365-–386, 2006.

[37] Shapiro, A., D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory*, SIAM Publications, Philadelphia, 2009, *in press*.

[38] Van Slyke, R., and R. J.-B. Wets, L-shaped linear programs with applications to optimal control and stochastic programming, *SIAM Journal on Applied Mathematics* **17**, pp. 638–663, 1969.