Processing second-order stochastic dominance models using cutting-plane representations

Csaba I. Fábián · Gautam Mitra · Diana Roman

Abstract   Second-order stochastic dominance (SSD) is widely recognised as an important decision criterion in portfolio selection. Unfortunately, stochastic dominance models are known to be very demanding from a computational point of view. In this paper we consider two classes of models which use SSD as a choice criterion. The first, proposed by Dentcheva and Ruszczyński (J Bank Finance 30:433–451, 2006), uses a SSD constraint, which can be expressed as integrated chance constraints (ICCs). The second, proposed by Roman et al. (Math Program, Ser B 108:541–569, 2006) uses SSD through a multi-objective formulation with CVaR objectives. Cutting plane representations and algorithms were proposed by Klein Haneveld and Van der Vlerk (Comput Manage Sci 3:245–269, 2006) for ICCs, and by Künzi-Bay and Mayer (Comput Manage Sci 3:3–27, 2006) for CVaR minimization. These concepts are taken into consideration to propose representations and solution methods for the above class of SSD based models. We describe a cutting plane based solution algorithm and outline implementation details. A computational study is presented, which demonstrates the

C. I. Fábián
Institute of Informatics, Kecskemét College, 10 Izsáki út, 6000 Kecskemét, Hungary
e-mail: fabian.csaba@gamf.kefo.hu

C. I. Fábián
Department of OR, Loránd Eötvös University, Budapest, Hungary

G. Mitra (✉) · D. Roman
CARISMA: The Centre for the Analysis of Risk and Optimisation Modelling Applications,
School of Information Systems, Computing and Mathematics, Brunel University,
Uxbridge (Middlesex) UB8 3PH, UK
e-mail: gautam.mitra@brunel.ac.uk

D. Roman
e-mail: diana.roman@brunel.ac.uk

effectiveness and the scale-up properties of the solution algorithm, as applied to the SSD model of Roman et al. (Math Program, Ser B 108:541–569, 2006).

Mathematics Subject Classification (2000)   90C15 Stochastic programming · 90C25 Convex programming · 91G10 Portfolio theory · 91G60 Numerical methods · 91G70 Statistical methods

1 Introduction

The model of choice among random variables is a crucial issue in portfolio selection. Mean-risk models are convenient from a computational point of view and have an intuitive appeal. In their traditional form, however, they use only two (or a few) statistics to characterize the portfolio return distribution, and thus may ignore important information. Stochastic dominance, in contrast, takes into account the entire distribution of a random variable. It is considered to be a formal approach for comparing random variables; in particular, second-order stochastic dominance (SSD) is widely recognised as an important criterion of choice in portfolio selection.

New portfolio-optimization models which apply the concept of second-order stochastic dominance were proposed by [2], and by [36]. Dentcheva and Ruszczyński consider a second-order stochastic dominance constraint that can be expressed as a finite set of integrated chance constraints (ICC). Roman, Darby-Dowman, and Mitra introduce a uniform dominance concept, and formulate multi-objective Conditional Value-at-Risk (CVaR) minimization problems. These authors present their respective formulations as linear programming (LP) problems by introducing new variables which represent positive parts in the formulation of integrated chance constraints and CVaR, respectively. In this paper we introduce the term *lifting representations* for the linear programming problems that result from such formulations, as these transformations 'lift' the feasible sets of the problems into higher dimensional spaces. The resulting LP problems are very large, and their solution presents a computational challenge. Dentcheva and Ruszczyński exploit special structure of these problems: they adapt the Regularized Decomposition Method of [39] to the duals. The formulation proposed by Roman, Darby-Dowman, and Mitra, although powerful from a modeling point of view, could be solved in realistic time only for a limited number of scenarios. Dentcheva [1] has suggested a reformulation of the above problem as a two-stage stochastic program such that the computational scale up limitations can be addressed by applying a decomposition algorithm. Lim et al. [21] have proposed solution methods for CVaR minimization problems with large numbers of scenarios using non-differentiable optimization.

In this paper we propose cutting-plane methods to address the computational difficulty of processing this class of models based on second-order stochastic dominance. We consider methodological aspects of chance-constrained programming and the minimization of CVaR; we show important connections with the second-order stochastic dominance models. Klein Haneveld and Van der Vlerk [15] proposed a cutting-plane representation of integrated chance constraints, and [18] proposed a cutting-plane representation of CVaR. Based on their respective cutting-plane representations, these

authors developed cutting-plane methods for the solution of integrated-chance-constrained problems and CVaR minimization problems, respectively. They found that solution methods which exploit cutting-plane representations produce better results than direct methods based on lifting representations. For large test problems, the former approach was found to be *1–2 orders of magnitude faster* than the latter. Moreover, the former approach has much better scale-up properties for the processing of models of increasing size.

The interested reader may find further models and methods involving ICC and CVaR in [30] and [13].

In this paper we focus on adapting the cutting-plane representation of Künzi-Bay and Mayer to the model proposed by [36]. We present algorithmic description, implementation details, and a computational study that demonstrates the effectiveness of this approach: problems with thousands of scenarios were solved in seconds.

The paper is organised as follows. In Sect. 2, we review portfolio models which use SSD. In Sect. 3 we review cutting-plane based solution methods that are applicable to this class of problems. In Sect. 4 we describe a cutting-plane approach adapted to the model proposed by [36]. A computational study is presented in Sect. 5. Results are summarized, conclusions drawn, and further research directions are outlined in Sect. 6.

2 Portfolio selection using second-order stochastic dominance

Let n denote the number of assets available for investment at the beginning of a fixed time period. A portfolio $x = (x_1, \ldots, x_n)^T$ represents the fractions of the initial capital invested in the different assets ($x_j = w_j/w$ where w_j is the capital invested in asset j and w is the total amount of capital to be invested, $j = 1 \ldots n$). Let $X \subset \mathbb{R}^n$ denote the set of the feasible portfolios. Such a feasible set is often simply defined by the requirement that the weights must sum to 1 and short selling is not allowed: $X = \{(x_1, \ldots, x_n) / \sum_{j=1}^{n} x_j = 1, x_j \geq 0, \forall j \in \{1, \ldots, n\}\}$. We consider here a more general case and assume that X is a bounded convex polyhedron.

Let the random vector $\mathbf{R} = (R_1, \ldots, R_n)^T$ denote the returns of the different assets at the end of the investment period. The return of a portfolio x is the random variable $R_x = \mathbf{R}^T x$.

The problem of choosing between portfolios is thus a problem of choosing between random variables (with larger outcomes preferred). The criterion by which one random variable is considered "better" than another random variable is specified by models of choice, or models of preference. Specifying such a model of choice is a crucial issue in portfolio selection.

Mean-risk models have been widely used in portfolio selection, since their introduction in the early fifties [23]. In their traditional form, they describe random variables (and portfolios) by using two scalars: the expected value, or the "mean", and a "risk" value. Preference is then defined using a trade-off between the mean where a larger value is desirable, and risk, where a smaller value is desirable. Various risk measures have been proposed in the literature, see for example [9,23,27,28,34,35]. Mean-risk

models are convenient from a computational point of view and have an intuitive appeal. However, they use only two (or a few) statistics to characterize a distribution, and thus may ignore important information. One way of improvement is to apply mean-risk models in interactive decision-support frameworks: [16] and independently [22] observe that portfolio optimization with multiple risk-constraints for different time frames and at different confidence levels allows the shaping of distributions according to the decision maker's preferences. Another approach is to construct mean-risk that are consistent with expected utility maximization / stochastic dominance; this has been the research subject of several recent papers [27–29,44].

Stochastic dominance provides a completely different approach, since it takes into account the entire distribution of a random variable. It is closely connected to the expected utility theory [24] but it eliminates the need to explicitly specify a utility function. With stochastic dominance, random variables are ordered by pointwise comparison of functions constructed from their distribution functions.

The concept of stochastic dominance, that is, stochastic ordering of random variables, was inspired by earlier work in the theory of majorization [12]. In economics, stochastic dominance was introduced in the 1960s; [31] considered the first order stochastic dominance relation and demonstrated the connection to utility functions. Second order stochastic dominance was brought to economics by [11] and third order stochastic dominance by [41]. The interested reader may find a detailed discussion in [42] and a survey in [20].

Of particular importance in portfolio selection is the second-order stochastic dominance (SSD), as this relates to models of risk-averse preferences. We introduce these concepts below.

Consider two random variables R_x and R_y (they may represent the returns of two portfolios x and y) with cumulative distribution functions F_x and F_y respectively. It is known (see, for example [42]) that the three statements below are equivalent:

(a) $F_x^{(2)}(t) \leq F_y^{(2)}(t), \quad \forall t \in \mathbb{R}.$

   where: $F_x^{(2)}(t) = \int\limits_{-\infty}^{t} F_x(u)\, du, \quad \forall t \in \mathbb{R}.$

(b)

$$\mathrm{E}\left([t - R_x]_+\right) \leq \mathrm{E}\left([t - R_y]_+\right), \quad \forall t \in \mathbb{R} \tag{1}$$

   (where $[t - R_x]_+ = \max\{t - R_x, 0\}$)

(c) $\mathrm{E}\left(U(R_x)\right) \geq \mathrm{E}\left(U(R_y)\right)$, for any increasing and concave (integrable) utility function U.

If the relations above hold, R_x is said to (weakly) dominate R_y with respect to second order stochastic dominance (SSD). This dominance is denoted by $R_x \succeq_{\mathrm{SSD}} R_y$. The corresponding strict dominance relation $\succ_{\mathrm{SSD}}$ is defined in the usual way: $R_x \succ_{\mathrm{SSD}} R_y$ if and only if $R_x \succeq_{\mathrm{SSD}} R_y$ and $R_y \not\succeq_{\mathrm{SSD}} R_x$.

A portfolio x is said to dominate (or be preferred to) another portfolio y with respect to SSD if $R_x \succeq_{\mathrm{SSD}} R_y$, where R_x and R_y are the (random) returns of portfolios x and y respectively. A similar notation is used for the decision vectors x and y: $x \succeq_{\mathrm{SSD}} y$.

A portfolio $x^\star$ is said to be SSD-efficient if there is no feasible portfolio $x \in X$ such that $x \succ_{\text{SSD}} x^\star$.

It is known that increasing and concave utility functions express the preference of risk-averse investors [7,8]. This underlines the importance of second-order stochastic dominance, since risk aversion is the observed economic behavior; yet, practical application of SSD as a criterion of choice proves to be difficult. Generally, comparing two random variables with respect to SSD involves an infinite number of comparisons, as set out in (1). Few models that use SSD have been proposed in the literature; they use additional assumptions in order to be computationally tractable; even so, their computational difficulty is remarkable. In the rest of this section, we review two recently proposed portfolio selection models that use SSD.

In the practice of portfolio selection, it is usual to assume that the asset returns have a discrete joint distribution, described by realizations under S states of the world (these states of the world are obtained through scenario generation or finite sampling of historical data). This is also the assumption made in this paper. Let $r^{(1)}, \ldots, r^{(S)}$ denote the realizations of the return vector R, occurring with probabilities $p_1, \ldots, p_S$, respectively. Denote by r_{ij} be the return of asset j under scenario i, $i \in \{1, \ldots, S\}$, $j \in \{1, \ldots, n\}$. Thus, $r^{(i)} = (r_{i1}, \ldots, r_{in})$, $\forall i \in \{1, \ldots, S\}$. The random variable $R_x = R^T x$ representing the return of portfolio x is finitely distributed over $\{r_x^{(1)}, \ldots, r_x^{(S)}\}$, where $r_x^{(i)} = r^{(i)T} x = x_1 r_{i1} + \cdots + x_n r_{in}$ is the portfolio return under scenario i, $\forall i \in \{1, \ldots, S\}$.

2.1 Portfolio optimization with SSD constraints

Dentcheva and Ruszczyński [2] consider a reference return $\widehat{W}$; this may be, for example, the return of a reference portfolio, say an index. They propose the following portfolio-optimization model:

$$
\begin{aligned}
&\max \ f(x) \\
&\text{such that} \quad x \in X, \\
&\qquad\qquad R^T x \succeq_{\text{SSD}} \widehat{W},
\end{aligned} \tag{2}
$$

where f is a concave function. In particular, $f(x) := \mathrm{E}\left(R^T x\right) = \overline{R}^T x$ can be used. Thus, only portfolios whose return dominate the reference return with respect to SSD are considered.

In general, a SSD-constraint $R^T x \succeq_{\text{SSD}} \widehat{W}$ can be expressed by a continuum of constraints, in the form of (1). These authors observe that, in case $\widehat{W}$ has a finite discrete distribution with realizations $\widehat{w}^{(\ell)}$ ($\ell = 1, \ldots, L$), then $R^T x \succeq_{\text{SSD}} \widehat{W}$ is equivalent to the finite set of constraints

$$
\mathrm{E}\left([\widehat{w}^{(\ell)} - R^T x]_+\right) \ \leq \ \mathrm{E}\left([\widehat{w}^{(\ell)} - \widehat{W}]_+\right) \qquad (\ell = 1, \ldots, L). \tag{3}
$$

The relations (3) are then transformed into a set of linear constraints by introducing variables $y_{i\ell}$ to represent the shortfall of the portfolio return below $\widehat{w}^{(\ell)}$ in scenario

i: $y_{i\ell} = [\widehat{w}^{(\ell)} - r^{(i)T}x]_+$ $(i = 1, \ldots, S, \ \ell = 1, \ldots, L)$. They formulate the lifting representation of the SSD-constrained problem (2):

$$
\begin{aligned}
&\max \ f(x) \\
&\text{such that} \quad x \in X, \\
&y_{i\ell} \geq 0, \quad y_{i\ell} \geq \widehat{w}^{(\ell)} - r^{(i)T}x \quad (i = 1, \ldots, S, \ \ell = 1, \ldots, L), \\
&\sum_{i=1}^{S} p_i y_{i\ell} \ \leq \ \mathrm{E}\left([\widehat{w}^{(\ell)} - \widehat{W}]_+\right) \quad (\ell = 1, \ldots, L).
\end{aligned}
\tag{4}
$$

The above problem has a specific structure. For such specific problems, these authors have developed a duality theory in which the dual objects are utility functions. Based on this duality theory, they construct a dual problem for (4) that consists of the minimization of a weighted sum of polyhedral convex functions. Domains, function values, subgradients are computable. The authors adapted the Regularized Decomposition method of [39] to these special dual problems. The authors implemented this method, and report favorable performance. A problem with 719 real-world assets was solved using 616 possible realizations of their joint return rates.

2.2 Multi-objective formulation and SSD efficient portfolios

Roman et al. [36] propose a multi-objective model whose Pareto optimal solutions[1] are SSD-efficient portfolios. (The origin of this multi-objective formulation can be traced back to the work of [25,26].) A specific solution is chosen by considering a reference return $\widehat{W}$. The return distribution of this chosen portfolio comes close to, or emulates, the reference return $\widehat{W}$ in a *uniform* sense, as explained below.

Given a random wealth W, let $\mathrm{Tail}_\alpha(W)$ $(0 < \alpha \leq 1)$ denote the unconditional expectation of the least $\alpha * 100\%$ outcomes of W. This is a heuristic definition and further clarification is needed taking into consideration 'probability atoms' in the distribution of W. For a formal definition quantile functions can be used. The *first quantile function* $F_W^{(-1)}(\alpha) := \inf\{t | F_W(t) \geq \alpha\}$ is the generalised inverse of the cumulative distribution function F_W of W. The *second quantile function* is defined as $F_W^{(-2)}(\alpha) := \int_0^\alpha F_W^{(-1)}(\beta)d\beta$. (with $F_W^{(-2)}(0)$ as 0.)

Ogryczak and Ruszczyński [29] prove that the second quantile function $F_W^{(-2)}(\alpha)$ is the convex conjugate of the *second performance function* $F_W^{(2)}(t) := \int_{-\infty}^t F_W(u)du = \mathrm{E}\left([t - W]_+\right)$. Formally,

$$
F_W^{(-2)}(\alpha) = \sup_{t \in \mathbb{R}} \left\{\alpha t - F_W^{(2)}(t)\right\}
\tag{5}
$$

holds for $0 \leq \alpha \leq 1$. Theory of convex conjugacy can be found in [32]. Ogryczak and Ruszczyński [29] show that, due to the properties of convex conjugacy, $F_W^{(-2)}(\alpha) =$

---

[1] For a multi-objective model, a feasible solution is *Pareto optimal* if there is no other feasible solution that would improve on all objective functions, with at least one strict inequality.

$\alpha E(W|W \leq t)$ holds for $F_W(t) = \alpha$. Thus the second quantile function has the heuristic meaning of unconditional tail expectation, justifying the definition $\mathrm{Tail}_\alpha(W) := F_W^{(-2)}(\alpha)$.

Rockafellar and Uryasev [34,35] establish a risk measure whose optimization results in convex problems. Given a random loss M and a probability $0 \leq \beta < 1$, the quantity $\mathrm{CVaR}_\beta(M)$, i.e. β-CVaR, is the conditional mean value of the worst $(1 - \beta) \times 100\%$ losses. This definition requires further clarification respecting probability atoms. Rockafellar and Uryasev [34,35] give a definition that 'splits' probability atoms and results in a coherent risk measure - see also the summary in [33]. They establish the minimization rule

$$\mathrm{CVaR}_\beta(M) = \min_{u \in \mathbb{R}} \left\{ u + \frac{1}{1 - \beta} E\left([M - u]_+\right) \right\} \tag{6}$$

that facilitates the solution of optimization problems involving the CVaR risk measure.

Loss can be considered as shortage relative to a benchmark wealth. (This is the wealth that we intend to accumulate, a parameter to be set by the decision maker.) In this paper we set the benchmark wealth to 0, hence we have $M = -W$. It is easily seen that the definition of the second quantile function is compatible with the CVaR definition of Rockafellar and Uryasev, and from (5) and (6) we have

$$\mathrm{Tail}_\alpha(W) = -\alpha \, \mathrm{CVaR}_{1-\alpha}(-W) \qquad (0 < \alpha \leq 1). \tag{7}$$

Ogryczak and Ruszczyński [29] proved that the second-order dominance relation $W \succeq_{\mathrm{SSD}} W'$ between random variables W and W' is equivalent to

$$\mathrm{Tail}_\alpha(W) \geq \mathrm{Tail}_\alpha(W') \quad (0 < \alpha \leq 1). \tag{8}$$

Roman et al. [36] use the equally probable scenarios assumption ($p_1 = \cdots = p_S = 1/S$): the asset return vector $\boldsymbol{R}$ and the reference return $\widehat{W}$ have discrete finite distributions, each with S equally probable outcomes. This assumption is quite natural in scenario generation or in sampling of historical data. (Moreover, it is not restrictive as any discrete finite distribution can be approximated using equally probable scenarios.)

Under the equiprobability assumption, the characterization of second-order stochastic dominance in the form of (8) requires only a finite number of inequalities. Namely, let W and W' be the random variables to compare; $W \succeq_{\mathrm{SSD}} W'$ is equivalent to

$$\mathrm{Tail}_{\frac{i}{S}}(W) \geq \mathrm{Tail}_{\frac{i}{S}}(W') \quad \text{holding for} \quad i = 1, \ldots, S. \tag{9}$$

It is clear that, in the case of a random variable W with equally probable outcomes

$$\mathrm{Tail}_{\frac{i}{S}}(W) = \sum_{j=1}^{i} \frac{1}{S} w^{<j>} \qquad (i = 1, \ldots, S), \tag{10}$$

where $w^{<1>} \leq \cdots \leq w^{<S>}$ denote the ordered outcomes of W, each occurring with probability $1/S$.

Thus, under the equiprobability assumption, comparing random variables with respect to SSD reduces to comparing their cumulated outcomes.

The SSD efficient portfolios are Pareto optimal solutions of the following multi-objective model:

$$\max \quad \left(\mathrm{Tail}_{\frac{1}{S}} \left(R^T x \right), \ \ldots, \ \mathrm{Tail}_{\frac{i}{S}} \left(R^T x \right), \ \ldots, \ \mathrm{Tail}_{\frac{S}{S}} \left(R^T x \right) \right)$$
$$\text{such that} \quad x \in X. \tag{11}$$

The reference-point method of [43] is used in order to choose a particular SSD efficient solution: the portfolio whose return distribution comes close to (tracks) the reference return $\widehat{W}$. This is done by solving a single objective optimization problem, as described below.

A *reference point* is defined as

$$\widehat{\tau} = (\widehat{\tau}_1, \ldots, \widehat{\tau}_S) := \left(\mathrm{Tail}_{\frac{1}{S}} \left(\widehat{W} \right), \ldots, \mathrm{Tail}_{\frac{S}{S}} \left(\widehat{W} \right) \right).$$

The multi-objective problem (11) is reduced to a single objective maximization problem; the objective function is a concave achievement function $\Gamma_{\widehat{\tau}}$ whose arguments are the components of the objective in (11).

The simplest achievement function is

$$\Gamma_{\widehat{\tau}} (\tau_1, \ldots, \tau_S) := \min_{1 \leq i \leq S} (\tau_i - \widehat{\tau}_i). \tag{12}$$

(A more sophisticated approach is to replace the term $\tau_i - \widehat{\tau}_i$ in (12) with a *partial achievement function* for each component i. The ith partial achievement function is a concave monotone increasing function of τ_i that takes the value 0 for $\tau_i = \widehat{\tau}_i$. Moreover, a term $\varepsilon \sum_{i=1}^{S} (\tau_i - \widehat{\tau}_i)$ with a small positive ε is usually added to ensure Pareto-efficiency of the optimal solution.)

Using the above achievement function, the single-objective optimization problem takes the form

$$\max \quad \Gamma_{\widehat{\tau}} \left(\mathrm{Tail}_{\frac{1}{S}} \left(R^T x \right), \ldots, \mathrm{Tail}_{\frac{S}{S}} \left(R^T x \right) \right)$$
$$\text{such that} \quad x \in X. \tag{13}$$

Denoting by $\vartheta = \min_{1 \leq i \leq S} (\mathrm{Tail}_{\frac{i}{S}} \left(R^T x \right) - \widehat{\tau}_i)$ the worst partial achievement, the above problem is written as:

$$\begin{aligned} \max \quad & \vartheta \\ \text{such that} \quad & \vartheta \in \mathbb{R}, \quad x \in X \\ & \vartheta \leq \mathrm{Tail}_{\frac{i}{S}} \left(R^T x \right) - \widehat{\tau}_i \quad (i = 1, \ldots, S). \end{aligned} \tag{14}$$

To express the quantities $\text{Tail}_{\frac{i}{S}}(\boldsymbol{R}^T \boldsymbol{x})$, [36] use the CVaR-optimization formula (6). For the case when the random loss has a discrete finite distribution, [34,35] proposed transforming (6) into a linear programming problem by introducing new variables to represent positive parts. This is a lifting representation. Given a portfolio x, the random loss is the negative of the portfolio return $\boldsymbol{R}^T \boldsymbol{x}$. Hence the CVaR is the optimal value of the following problem (we substituted t for $-u$) :

$$\text{CVaR}_\beta\left(-\boldsymbol{R}^T \boldsymbol{x}\right) = \min -t + \frac{1}{1-\beta} \sum_{j=1}^{S} \frac{1}{S} d_j$$

$$\text{such that} \quad t \in \mathbb{R},$$
$$d_j \geq -\boldsymbol{r}^{(j)T} \boldsymbol{x} + t, \quad d_j \geq 0 \quad (j = 1, \ldots, S). \quad (15)$$

The lifting representation (15) and the formula (7) can then be used to compute the quantities $\text{Tail}_{\frac{i}{S}}(\boldsymbol{R}^T \boldsymbol{x})$ $(i = 1, \ldots, S)$ as optimal values of maximization problems. In the ith problem, the new variables are $t_i, d_{i1}, \ldots, d_{iS} \in \mathbb{R}$:

$$\text{Tail}_{\frac{i}{S}}\left(\boldsymbol{R}^T \boldsymbol{x}\right) = \frac{1}{S} \max i\, t_i - \sum_{j=1}^{S} d_{ij}$$

$$\text{such that} \quad t_i \in \mathbb{R},$$
$$-t_i + d_{ij} \geq -\boldsymbol{r}^{(j)T} \boldsymbol{x}, \quad d_{ij} \geq 0 \quad (j = 1, \ldots, S). \quad (16)$$

Using representation (16), [36] formulate the achievement-maximization problem (13) as a linear programming problem, with variables $\vartheta, \boldsymbol{x}, z_i$ $(i = 1, \ldots, S), t_i$ $(i = 1, \ldots, S), d_{ij}$ $(i, j = 1, \ldots, S)$:

$$
\begin{aligned}
&\max \vartheta \\
&\text{such that} \quad \boldsymbol{x} \in X \\
&z_i - \widehat{\tau}_i \geq \vartheta, \quad \text{for } i = 1 \ldots S \\
&z_i = \frac{1}{S}\left(i\, t_i - \sum_{j=1}^{S} d_{ij}\right), \quad \text{for } i = 1 \ldots S \\
&t_i - \sum_{j=1}^{n} x_j r_{ij} \leq d_{ij}, \quad \text{for } i, j = 1 \ldots S \\
&d_{ij} \geq 0, \quad \text{for } i, j = 1 \ldots S.
\end{aligned}
\quad (17)
$$

Due to lifting representation, a large number of new constraints and variables is introduced. (The number of new constraints and variables is in the order of S^2.)

These authors implemented the method outlined above, and made extensive testing on problems with 76 real-world assets using 132 possible realizations of their joint return rates. The modelling capability of their approach proved to be powerful;

however, the model could only be solved for a limited number of scenarios (less than 600), due to its size.

Remark 1 To be more precise, in [36] an additional term $\varepsilon \sum_{i=1}^{S}(z_i - \widehat{\tau_i})$, with ε an arbitrarily small positive number, is added to the objective function in (17). This term is added to ensure the SSD efficiency of the portfolio obtained, in case that the maximization of the worst partial achievement has non-unique optimal solutions.

The model described above is never infeasible, and its optimal solution is a SSD-efficient portfolio, irrespective of the reference return chosen by the decision maker.

The approach allows the decision maker to set and tune components of the reference return directly. (Hence the reference return may not be the return distribution of some feasible portfolio, but just composed of "desirable", "ideal" outcomes.) When solving the resulting LP, the sign of the optimal value distinguishes which type of reference has been used. There are three possible types of reference returns:

Type I :   a reference return $\widehat{W}$ that is not SSD efficient. In this case, the optimum is strictly positive, and the optimal portfolio has a return distribution that improves on $\widehat{W}$ (until SSD-efficiency is obtained).

Type II : a reference return that is SSD efficient. In this case, the optimum is zero, and the optimal solution is a portfolio whose return distribution is exactly the reference return.

Type III : a reference return $\widehat{W}$ that is unattainable (for example, when some of its components are too high and no feasible portfolio could have such a return distribution). In this case, the optimum is strictly negative, and the optimal solution is an SSD-efficient portfolio whose return comes closest, in a uniform sense, to $\widehat{W}$.

The model can thus be used for checking SSD efficiency of a given distribution.

Remark 2 Obviously, the reference return introduced by the user plays an important role in choosing the solution. The most common situation is type I, since usually the return of an index is used as a reference return; thus, there is a similarity with the model of Dentcheva and Ruszczyński.

There are also important differences between the models of [2] and [36]. The former model, while improving on a benchmark, is not guaranteed to improve until SSD efficiency; it could still result in dominated solutions. Moreover, there is (theoretically) the possibility that the constraint set has no feasible solution.

3 Cutting-plane based representations and solution methods

3.1 Integrated chance constraints

The constraints appearing in the SSD-formulation (3) have the form

$$\mathrm{E}\left(\left[h - \boldsymbol{R}^T \boldsymbol{x}\right]_{+}\right) \leq \gamma, \tag{18}$$

where h, γ are given numbers. Constraints of this type were introduced by [14], and are called *Integrated Chance Constraints* (ICCs). Klein Haneveld and Van der Vlerk [15] proposed a cutting-plane representation of ICCs. The constraint (18) is represented in the form

$$\sum_{j \in \mathcal{J}} p_j \left(h - r^{(j)T} x \right) \leq \gamma \quad \text{holding for each subset} \quad \mathcal{J} \subset \{1, \ldots, S\}. \quad (19)$$

Based on this representation, [15] also develop a cutting-plane method. We outline this method as applied to the expectation-maximization problem

$$\begin{aligned} \max \ & \overline{R}^T x \\ \text{such that} \ & x \in X, \\ & \mathrm{E}\left(\left[h - R^T x \right]_+ \right) \leq \gamma, \end{aligned} \quad (20)$$

where X is a convex bounded polyhedron.

Using the cutting-plane representation, problem (20) takes the form:

$$\begin{aligned} \max \ & \overline{R}^T x \\ \text{such that} \ & x \in X, \\ & \psi(x) \leq 0, \end{aligned} \quad (21)$$

with the constraint function

$$\psi(x) := \max_{\mathcal{J} \subset \{1, \ldots, S\}} \sum_{j \in \mathcal{J}} p_j \left(h - r^{(j)T} x \right) - \gamma. \quad (22)$$

A sequence of iterates $x_1, \ldots, x_\kappa \in X$ is computed; at each iterate x_ι, a linear function $l_\iota^\sharp$ is constructed:

$$l_\iota^\sharp(x) := \sum_{j \in \mathcal{J}_\iota} p_j \left(h - r^{(j)T} x \right) - \gamma \quad \text{with} \quad \mathcal{J}_\iota := \left\{ 1 \leq j \leq S \ \middle|\ h - r^{(j)T} x_\iota > 0 \right\}.$$

It is easily seen that this function satisfies

$$l_\iota^\sharp(x) \leq \psi(x) \quad (x \in \mathbb{R}^n) \quad \text{and} \quad l_\iota^\sharp(x_\iota) = \psi(x_\iota).$$

In words, the graph of $l_\iota^\sharp$ is a supporting hyperplane to the epigraph of ψ at the point $(x_\iota, \psi(x_\iota))$.

Remark 3 Restricting the above graph and epigraph to the space of the x-vectors, we can say: if the level set $\{x \mid \psi(x) \leq 0\}$ is not empty, and $\psi(x_\iota) > 0$, then the hyperplane $\left\{ x \mid l_\iota^\sharp(x) = 0 \right\}$ separates the level set from x_ι. Hence the above cutting-plane method can be viewed as constructing a cutting-plane model of the level set. In the

forthcoming overview, however, we prefer viewing the method as constructing a cutting-plane model of the function ψ, because such view will allow the application of constrained convex programming methods. In "Appendix A." we consider applying the Constrained Level Method of [19] to the problem (21).

A cutting-plane model of ψ is constructed as

$$\psi_\kappa(\boldsymbol{x}) := \max_{1 \leq \iota \leq \kappa} l_\iota^\sharp(\boldsymbol{x}). \tag{23}$$

The next iterate $\boldsymbol{x}_{\kappa+1}$ will be an optimal solution of the model problem

$$
\begin{aligned}
&\max \ \overline{\boldsymbol{R}}^T \boldsymbol{x} \\
&\text{such that} \quad \boldsymbol{x} \in X, \\
&\psi_\kappa(\boldsymbol{x}) \leq 0.
\end{aligned}
\tag{24}
$$

The method stops if the model problem becomes infeasible, or if the current iterate becomes a feasible solution of the original problem (i.e., $\psi(\boldsymbol{x}_{\kappa+1}) \leq 0$ holds).

Klein Haneveld and Van der Vlerk [15] observe that the procedure stops in finitely many steps. They implemented the method and their test results show that only a small fraction of the possible 2^S cuts needs to be generated. They found that for a fixed dimension n, the average number of cuts (slowly) increases with the number S of scenarios.

These authors also formulated the lifting representation of problem (20) by introducing new variables y_j to represent $\left[h - \boldsymbol{r}^{(j)\,T} \boldsymbol{x}\right]_+ \ (j = 1, \ldots, S)$:

$$
\begin{aligned}
&\max \ \overline{\boldsymbol{R}}^T \boldsymbol{x} \\
&\text{such that} \quad \boldsymbol{x} \in X, \\
&y_j \geq 0, \quad y_j \geq h - \boldsymbol{r}^{(j)\,T} \boldsymbol{x} \quad (j = 1, \ldots, S), \\
&\sum_{j=1}^{S} p_j y_j \ \leq \gamma.
\end{aligned}
\tag{25}
$$

They solved this problem with a benchmark interior-point solver, and compared results with their cutting-plane representation approach. On smaller problem instances, the cutting-plane algorithm could not beat the interior-point solver. However, the cutting-plane representation approach proved much faster on larger instances.

Using this cutting-plane representation of ICCs, the SSD-constrained portfolio-optimization model (2) can be formulated as

$$
\begin{aligned}
&\max \ f(\boldsymbol{x}) \\
&\text{such that} \quad \boldsymbol{x} \in X, \\
&\sum_{j \in \mathcal{J}_\ell} p_j \left(h_\ell - \boldsymbol{r}^{(j)\,T} \boldsymbol{x}\right) \leq \gamma_\ell \quad (\mathcal{J}_\ell \subset \{1, \ldots, S\}, \quad \ell = 1, \ldots, L),
\end{aligned}
\tag{26}
$$

where $h_\ell = \widehat{w}^{(\ell)}$ and $\gamma_\ell = \mathrm{E}\left([\widehat{w}^{(\ell)} - \widehat{W}]_+\right) (\ell = 1, \ldots, L)$.

If $f(x) = \mathrm{E}\left(R^T x\right) = \overline{R}^T x$, then the cutting-plane method of [15] can be easily adapted to problem (26). (The feasibility cuts can be applied independently for $\ell = 1, \ldots, L$).

3.2 Conditional value-at-risk

Künzi-Bay and Mayer [18] proposed the following cutting-plane representation of CVaR_β $(0 \le \beta < 1)$:

$$\mathrm{CVaR}_\beta\left(-R^T x\right) = \min \; z + \frac{1}{1-\beta} v$$
$$\text{such that} \;\; z, v \in \mathbb{R},$$
$$\sum_{j \in \mathcal{J}} p_j \left(-r^{(j)T} x - z\right) \le v \quad (\mathcal{J} \subset \{1, \ldots, S\}). \quad (27)$$

The cut belonging to the empty set $\mathcal{J} = \emptyset$ just prescribes the non-negativity of v. (This representation even allows a random benchmark wealth that we assume to be 0, hence the loss is expressed by $-R^T x$.)

The representation (27) is the CVaR-analogue of the representation (19) of [15]. The two approaches employ the same idea, originally developed by Klein Haneveld and Van der Vlerk, as indicated in the References. Künzi-Bay and Mayer [18] obtained the CVaR representation (27) independently, by considering the (single-stage) CVaR-minimization problem as a two-stage stochastic programming problem. They found that the second-stage problems have a special structure and that their duals can be solved explicitly. These authors also did extensive testing, and their results confirm the effectiveness of the cutting-plane approach. (The two computational studies actually complement each other: [15] employ the cutting-plane model of a constraint function, while [18] employ the cutting-plane model of the objective function.)

Using the above cutting-plane representation, a CVaR-minimization problem takes the form:

$$\min \; z + \frac{1}{1-\beta} \varphi(x, z) \;\; \text{such that} \;\; x \in X, \; z \in \mathbb{R}, \quad (28)$$

where X is a convex bounded polyhedron, and the objective function is

$$\varphi(x, z) := \max_{\mathcal{J} \subset \{1, \ldots, S\}} \sum_{j \in \mathcal{J}} p_j \left(-r^{(j)T} x - z\right). \quad (29)$$

For the minimization of CVaR in one-stage stochastic problems, [18] developed a special cutting-plane method. They construct a sequence of iterates $(x_1, z_1), \ldots, (x_K, z_K) \in X \times \mathbb{R}$. At each iterate (x_ι, z_ι), they construct a linear function l_ι as

$$l_\iota(\boldsymbol{x}, z) := \sum_{j \in \mathcal{J}_\iota} p_j \left(-\boldsymbol{r}^{(j)\,T} \boldsymbol{x} - z \right) \quad \text{with} \quad \mathcal{J}_\iota := \left\{ 1 \le j \le S \,\middle|\, -\boldsymbol{r}^{(j)\,T} \boldsymbol{x}_\iota - z_\iota > 0 \right\}.$$

It is easily seen that this function satisfies

$$l_\iota(\boldsymbol{x}, z) \le \varphi(\boldsymbol{x}, z) \quad (\boldsymbol{x} \in \mathbb{R}^n, z \in \mathbb{R}) \quad \text{and} \quad l_\iota(\boldsymbol{x}_\iota, z_\iota) = \varphi(\boldsymbol{x}_\iota, z_\iota).$$

In words, the graph of l_ι is a supporting hyperplane to the epigraph of φ at the point $((\boldsymbol{x}_\iota, z_\iota), \varphi(\boldsymbol{x}_\iota, z_\iota))$. A cutting-plane model of φ is constructed as

$$\varphi_\kappa(\boldsymbol{x}, z) := \max_{1 \le \iota \le \kappa} l_\iota(\boldsymbol{x}, z). \tag{30}$$

The next iterate $(\boldsymbol{x}_{\kappa+1}, z_{\kappa+1})$ is an optimal solution of the problem

$$\min \; z + \frac{1}{1-\beta} \, \varphi_\kappa(\boldsymbol{x}, z) \quad \text{such that} \quad \boldsymbol{x} \in X, \; z \in \mathbb{R}. \tag{31}$$

The method stops if $\varphi_\kappa(\boldsymbol{x}_{\kappa+1}, z_{\kappa+1}) = \varphi(\boldsymbol{x}_{\kappa+1}, z_{\kappa+1})$ holds, i.e., in the current iterate, the model function value coincides with the original objective value.

The above procedure finds an optimal solution in a finite number of steps.

Künzi-Bay and Mayer [18] implemented their method and solved several CVaR-minimization test problems with their experimental solver called CVaRMin. They also solved the test problems with general-purpose LP solvers. These solvers were used to solve LP-equivalent problems constructed using the lifting representation (15). (The lifting representation contains the constraints describing feasible portfolios, and $2S$ additional constraints. These are *individual* constraints for each scenario, as opposed to the *aggregate* cuts that appear in the cutting-plane representation (27).) Additionally, they also solved the test problems in two-stage recourse forms, by employing a benchmark stochastic solver. Their experimental results show the clear superiority of the solver CVaRMin in case of CVaR-problems. For the largest test problems, CVaR-Min was by at least one order of magnitude faster than either of the other solvers involved. They also propose a version of the method that finds an ϵ-optimal solution with a prescribed stopping tolerance $\epsilon > 0$.

Remark 4 A cutting-plane scheme for optimization of (SSD-compatible) coherent risk measures that involve p-order moments of loss distributions is discussed in [17].

4 The proposed algorithm

In this section we adapt the cutting-plane representation of CVaR (27) to the model proposed by [36], described in Sect. 2.2.

In the present application we assume $p_1 = \cdots = p_S = 1/S$ and only consider confidence levels $\beta = \frac{S-i}{S}$ $(i = 1, \ldots, S)$. Let us substitute these into (27). Moreover,

let us introduce a new variable $\nu := z + \frac{1}{1-\beta}\nu = z + \frac{S}{i}\nu$ instead of ν. We get:

$$
\begin{aligned}
\text{CVaR}_{\frac{S-i}{S}}\left(-\boldsymbol{R}^T\boldsymbol{x}\right) = \min \ & \nu \\
\text{such that} \ \ & z, \nu \in \mathbb{R}, \\
& \frac{1}{i}\sum_{j\in\mathcal{J}}\left(-\boldsymbol{r}^{(j)\,T}\boldsymbol{x}\right) - \left(\frac{|\mathcal{J}|}{i} - 1\right)z \leq \nu \quad (\mathcal{J} \subset \{1,\ldots,S\}),
\end{aligned}
$$
(32)

where $|\mathcal{J}|$ denotes the cardinality of the set $\mathcal{J}$.

Fábián [4] proposed a simplified version of the cutting-plane representation (27) for single-stage models. Fábián observes that under the present conditions, only sets $\mathcal{J}$ of cardinality $|\mathcal{J}| = i$ need to be considered in the ith problem ($i = 1, \ldots, S$). Hence the coefficient of z is 0 in the constraints of (32). It follows that

$$
\begin{aligned}
\text{CVaR}_{\frac{S-i}{S}}\left(-\boldsymbol{R}^T\boldsymbol{x}\right) = \max \ & \frac{1}{i}\sum_{j\in\mathcal{J}}-\boldsymbol{r}^{(j)\,T}\boldsymbol{x} \\
\text{such that} \ \ & \mathcal{J} \subset \{1,\ldots,S\}, \ \ |\mathcal{J}| = i.
\end{aligned}
$$
(33)

Using (7) and (33) we get the following cutting-plane representation for $\text{Tail}_{\frac{i}{S}}$ ($i = 1, \ldots, S$):

$$
\begin{aligned}
\text{Tail}_{\frac{i}{S}}\left(\boldsymbol{R}^T\boldsymbol{x}\right) = \frac{1}{S} \min \ & \sum_{j\in\mathcal{J}}\boldsymbol{r}^{(j)\,T}\boldsymbol{x} \\
\text{such that} \ \ & \mathcal{J} \subset \{1,\ldots,S\}, \ \ |\mathcal{J}| = i.
\end{aligned}
$$
(34)

Remark 5 Equality (34) clearly follows from (10). We went through the steps (32) - (33) in order to show that the cutting-plane representation (34) is a special form of representation (27). Hence a cutting-plane method using (34) inherits the computational properties of the method proposed by [18].

In contrast to the lifting representation (16), the cutting-plane representation (34) introduces no additional variables. Using (34), the achievement-maximization problem (13) can be re-formulated to:

$$
\begin{aligned}
\max \ & \vartheta \\
\text{such that} \ \ & \vartheta \in \mathbb{R}, \quad \boldsymbol{x} \in X, \\
& \vartheta + \widehat{\tau}_i \ \leq \ \frac{1}{S}\sum_{j\in\mathcal{J}_i}\boldsymbol{r}^{(j)\,T}\boldsymbol{x} \ \ \text{for each} \ \ \mathcal{J}_i \subset \{1,\ldots,S\}, \ \ |\mathcal{J}_i| = i,
\end{aligned}
$$
(35)

where i = 1, …, S.

No additional variables are introduced in the above formulation. Theoretically an astronomical number of cuts are required, but in practice only a few of them are needed. The problem can be solved by the following cutting-plane method:

0. *Initialize.*
Set the stopping tolerance $\epsilon > 0$.
The initial cutting-plane model contains the constraints $x \in X$, and a single cut that we select arbitrarily as, e.g.,

$$\vartheta + \widehat{\tau}_1 \ \leq \ \frac{1}{S} r^{(1)\,T} x.$$

(This cut belongs to the set $\mathcal{J}_1 = \{1\}$. It is added to make the objective function of the model problem finite.)
1. *Solve the model.*
Let $(\vartheta^\star, x^\star)$ be an optimal solution of the current cutting-plane model.
Let $w^{\star < j_1^\star >} \leq \cdots \leq w^{\star < j_S^\star >}$ denote the ordered outcomes of the current portfolio return $W^\star = R^T x^\star$.
Let $\mathcal{J}_i^\star := \{j_1^\star, \ldots, j_i^\star\}$ $(i = 1, \ldots, S)$.
2. *Check for optimality*
If

$$\vartheta^\star + \widehat{\tau}_i \ \leq \ \frac{1}{S} \sum_{j \in \mathcal{J}_i^\star} r^{(j)\,T} x^\star + \epsilon \quad \text{holds for each} \quad i = 1, \ldots, S, \qquad (36)$$

then $x^\star$ is an ϵ-optimal solution; stop.
If some of the above inequalities are not satisfied, then consider the violations

$$\vartheta^\star + \widehat{\tau}_i \ - \ \frac{1}{S} \sum_{j \in \mathcal{J}_i^\star} r^{(j)\,T} x^\star \quad (i = 1, \ldots, S).$$

Let $\hat{\imath}$ $(1 \leq \hat{\imath} \leq S)$ denote the index that maximizes violation.
3. *Append cuts*
Append the following cut to the model:

$$\vartheta + \widehat{\tau}_{\hat{\imath}} \ \leq \ \frac{1}{S} \sum_{j \in \mathcal{J}_{\hat{\imath}}^\star} r^{(j)\,T} x.$$

Repeat from *step 1.*

Remark 6 The cutting plane representation (35) slightly deviates from the model proposed by [36], where a regularization term is added to the objective function, as described in Remark 1. Thus, solving the cutting plane representation (35) may still, theoretically, result in a SSD dominated solution. (If (35) has a unique optimal solution $(\vartheta^\star, x^\star)$, then $x^\star$ is a SSD efficient portfolio. However, in the case of multiple optimal solutions, some of them may represent portfolios that are dominated with respect to SSD.) In its present form, the representation and the algorithm proposed only guarantees to improve on the reference return with respect to SSD (if possible). However, our numerical experiments show that, in practice, the representation (35)

usually leads to the same solution, as the representation proposed by [36]. (For small datasets, the model has been solved using both the cutting-plane representation and the lifting representation with regularization. In all instances, both representations led to the same solutions, that is, SSD efficient portfolios).

In the "Appendix" we propose enhancements and extensions to the approach described above, both from the solution method and the modeling point of view. The Level Method of [19] is suggested as an alternative to the cutting plane method for solving (35). A possible application to multi-objective models with variance taken into account is also suggested.

5 Computational study

5.1 Scope and implementation issues

In [36], the model described in Sect. 2.2 was tested on a data set of 132 historical monthly returns of 76 stocks (all the stocks that belonged to the FTSE 100 index during the period January 1993–December 2003). The returns of the 132 time periods were considered as equally probable scenarios for the next time period (January 2004). Tests were made with different types of reference returns. Powerful modelling capabilities were demonstrated by in-sample and out-of-sample analysis of the return distributions of the optimal portfolios. Unfortunately, using the lifting representation, only models of relatively small scenario sizes (less than 500 scenarios) could be solved in a realistic time.

Here, we solve this model using the cutting plane representation (35) instead of the lifting representation. We implement the algorithm described in Sect. 4. However, in the optimality check of step *2*, we use a relative tolerance instead of the absolute tolerance of (36). Hence we check whether

$$\vartheta^{\star} + \widehat{\tau}_i \ \leq \ (1 + \epsilon) \, \frac{1}{S} \sum_{j \in \mathcal{J}_i^{\star}} r^{(j)\,T} x^{\star} \quad (i = 1, \ldots, S)$$

holds. Accordingly, we only add cuts where the relative violations are significant. The use of relative tolerances is justified from a decision maker's point of view. Moreover we found this approach more effective. We test the scale-up properties of this approach.

The method was implemented using the AMPL modelling system [10] and the AMPL COM Component Library [40], integrated with C functions. Under AMPL we use the FortMP solver. FortMP was developed at Brunel University and NAG Ltd by [3], the project being co-ordinated by E.F.D. Ellison.

In our cutting-plane system, cut generation is implemented in C, and cutting-plane model problem data are forwarded to AMPL in each iteration. Hence the bulk of the arithmetic computations is done in C, since the number of scenarios is typically large as compared to the number of assets. Moreover, our test results imply that acceptable accuracy can be achieved by a relatively small number of cuts in the master problem.

Table 1 *Type I reference return*; numbers of iterations

Column headers show the stopping tolerances. *Row headers* show the numbers of the scenarios

	1e–04	1e–05	1e–06	1e–07
5,000	26	26	19	19
7,000	22	23	23	18
10,000	25	28	28	28

Table 2 *Type I reference*; optimal objective values obtained with different stopping tolerances

	1e–04	1e–05	1e–06	1e–07
5,000	0.005506	0.005734	0.005734	0.005734
7,000	0.005508	0.005506	0.005506	0.005506
10,000	0.005719	0.005715	0.005715	0.005715

Table 3 *Type II reference*; numbers of iterations

	1e–04	1e–05	1e–06	1e–07
5,000	9	6	6	6
7,000	9	9	9	9
10,000	6	6	6	6

Hence the sizes of the master problems do not directly depend on the number of scenarios.

Our tests were run on a Windows XP machine with Intel CPU T2250 @ 1.73 GHz and 2 GB RAM.

5.2 Test problems and test results

We generated large scenario sets using the Geometric Brownian Motion, which is well accepted in finance for modelling asset prices, see e.g., [38]. The parameters for scenario generation were derived from the historical dataset used by [36], described in Sect. 5.1. We generated scenario sets containing 5,000, 7,000 and 10,000 scenarios, respectively. (A single scenario describes returns for the 76 component stocks and also for the reference return.)

We solved the achievement-maximization problem with the 3 types of reference return described in [36] and summarized in Remark 1, Sect. 2.2 of this paper.

The problems were solved with different values for the relative stopping tolerance (1e-04, 1e-05, 1e-06, and 1e-07). In all these cases, the models were solved in less than 10 s. Thus, the superiority and effectiveness of the cutting plane approach (as compared to the lifting representation) is obvious.

Our test results are summarized in Tables 1, 2, 3, 4, 5. In these tables, column headers display the stopping tolerances, and row headers display the numbers of scenarios.

Type I is a reference return distribution that is not SSD efficient. To illustrate this case, we chose as reference the return distribution of the FTSE100 index. (Indeed, this benchmark distribution is not SSD efficient, since a strictly positive optimum is obtained; the optimal portfolio has a return distribution that dominates the reference

Table 4 *Type III reference*;
numbers of iterations

	1e–04	1e–05	1e–06	1e–07
5,000	12	14	19	20
7,000	12	14	16	18
10,000	14	16	19	21

Table 5 *Type III reference*; optimal objective values

	1e–04	1e–05	1e–06	1e–07
5,000	−6.166302035	−6.210836095	−6.211805798	−6.211895451
7,000	−7.911551621	−7.968123422	−7.976466103	−7.977045489
10,000	−8.897442933	−9.068657844	−9.072698606	−9.073370813

with respect to SSD.) The number of cuts required to reach near-optimality are presented in Table 1. The optimal objectives obtained are presented in Table 2. In this case, stopping tolerances of 1e-05, 1e-06 and 1e-07 resulted exactly the same solution. The stopping tolerance 1e-04 resulted a fractionally different solution.

Type II is a reference return distribution that is efficient with respect to SSD. In this case, the optimal objective value is zero and the model finds a portfolio whose return distribution is exactly the reference. To illustrate this case, we chose as reference the return distribution of the stock with the highest expected return. (This distribution is chosen only for demonstrative purposes. The optimal portfolio consists of the single stock used for creating the reference. It is obviously not desirable, due to lack of diversification.) The number of cuts required to reach near-optimality are presented in Table 3. In all cases, the optimal objective values are zero, and the optimal solutions are the same.

Type III is a reference return that is not attainable, in the sense that there is no feasible portfolio whose return distribution is this reference, or dominates this reference with respect to SSD. In this case, the optimal objective is negative and the model finds a portfolio whose return distribution comes uniformly close to the reference. For illustrating this case, we chose the reference vector $\widehat{\tau}$ composed of the best individual tails. More precisely, in the case of S scenarios, we solved S optimization problems; in the ith problem we maximize $\text{Tail}_{\frac{i}{S}}$ and set $\widehat{\tau}_i$ equal to the optimum obtained, $(i = 1, \ldots, S)$. The reference $\widehat{\tau} = (\widehat{\tau}_1, \ldots, \widehat{\tau}_S)$ is obviously unattainable. The number of cuts required to reach near-optimality are presented in Table 4. The optimal objective values are presented in Table 5.

Running times largely depend on the type of the reference distribution used. For the 10, 000-scenario problems, running times were as follows. Type-I problems: 9–10 s, Type-II problems: 4–5 s, Type-III problems: 6–8 s. We found that considering a certain reference distribution type, the time of single iteration is a slightly increasing function of the number of the scenarios.

In the present computational setup, processing times of the cutting-plane model problems make the dominant part of the total running time. (In an earlier implementation, cut generation was also done in AMPL, and it took the dominant part of the

Table 6 *Type I reference*; in-sample performances of P1 and the FTSE 100 index: statistics of their return distribution

	P1	FTSE 100 Index
Mean	0.0105	0.0034
SD	0.0033	0.0018
Minimum	−0.0001	−0.0034
Maximum	0.0224	0.0102
Kurtosis	−0.0523	−0.0263
Skewness	0.0119	0.0100
Range	0.0225	0.0136

P1 is the portfolio obtained with our model, 10,000 scenarios; reference = return distribution of FTSE 100 index

total time. The present C implementation of cut generation is more efficient.) We note that the processing time of a cutting-plane model problem includes the time used by AMPL to read model and data, generate the problem, and run the solver (FortMP). Also there is some overhead introduced by invoking AMPL through COM.

Remark 7 As it can be seen from Tables 1 and 3, the number of iterations required to solve a problem sometimes decreases with smaller tolerance. This curious behavior is due to our use of relative tolerances: when a tighter tolerance is prescribed, we may in an early iteration add a certain cut that greatly helps in the solution process. Applying a slacker tolerance, on the other hand, we may not add this certain cut, because it is not violated significantly with respect to the slack tolerance.

We also solved the hardest, Type I-reference distribution problems with the tight absolute stopping tolerance of 1e-8. (In this experiment, we used the method as described in Sect. 4; and did not apply relative tolerances as described in Sect. 5.1.) For each scenario set, the optimal objective values coincided with the respective values given in the last columns of Table 2 (up to the first 7 digits given in Table 2). However, the number of iterations and the running times were about 3 times larger with the tight absolute tolerance, justifying our approach of using relative stopping tolerances as described in Sect. 5.1.

Remark 8 In our type-III test problems, the reference tails are composed of the best individual tails of the different assets. Hence the reference vector $\hat{\tau}$ gets 'stronger' as the sample size increases. This is the explanation of the large decreases inside columns of Table 5.

Although the present paper concerns cutting plane representations and an effective solution algorithm, we wish to highlight the power of the SSD model under consideration. A full treatment (considering a limited number of scenarios) is given in [36]. We consider the type I benchmark (in our example, the return distribution of the FTSE100 index), 10,000 scenarios and a tolerance level of 1e-07. The portfolio obtained as a solution of our model is denoted by P1. Its return distribution strictly dominates the index with respect to SSD. The statistics of the return distributions of P1 and of FTSE100 are presented in Table 6. The statistics are clearly better in the case of P1: higher expected return, higher minimum, higher maximum. The standard deviation and the range are also higher, but, in this case, it is not a drawback, since it involves 'spread' at higher return levels: the range for the possible returns of P1 is

	P1	FTSE 100 Index
Mean	0.0189	0.0057
SD	0.0270	0.0225
Minimum	−0.0044	−0.0278
Maximum	0.0801	0.0355

Table 7 *Type I reference*; out-of-sample performances of P1 and the FTSE 100 index: statistics of their series of return values for the 10 months following the date of selection (January–October 2004)

'shifted upwards'. This underlines again the difference between our approach and a classical approach like the mean-variance model.

The out-of-sample performance of P1 and FSTE100 index is analyzed over the 10 historical time periods following the date of selection (January–October 2004). The statistics of the out-of-sample returns are presented in Table 7. The out-of-sample performance of P1 is obviously better.

6 Discussion

In this paper we have proposed cutting-plane representations and solution methods which dramatically improve the computational performance of decision models based on second-order stochastic dominance. The methods are particularly suitable for two recently proposed portfolio optimization models. The first model due to [2] has a SSD constraint that can be expressed as a set of integrated chance constraints. The second model proposed by [36] uses minimization of CVaR at multiple confidence levels. In their original form, both models were reduced to LPs by introducing additional variables. This approach has severe limitations in respect of the computational scale up properties. Thus, even problems with a reasonable number of assets are difficult to solve when the number of scenarios is large.

Cutting plane representations of ICCs and CVaR were recently proposed by [15] and [18], respectively. In our proposed method we have adapted the representations and solution techniques proposed by these authors to the portfolio optimization models based on SSD.

We have focused on adapting the cutting-plane representation of [18] to the model proposed by [36]; we have presented algorithmic description, implementation details, and a computational study that demonstrates effectiveness of this approach. Problems with 10,000 scenarios were solved within 30 s.

In "Appendix A" we suggest possible enhancements and extensions of the work presented in this paper.

Appendix A: Extensions and other applications

A.1 The level methods

In Sects. 3.1 and 3.2, we described specialized cutting-plane methods for solving the cutting-plane representation problems (21) (of ICC) and (28) (of CVaR).

As an alternative to these cutting-plane methods, we propose solving the cutting-plane representation problems (21) and (28) by the Constrained Level Method of [19]. This is a bundle-type method.

In the case of ICC models, it can solve problems more general than (21): it can maximize a concave objective function instead of the present linear objective. Hence the Constrained Level Method can handle the SSD-constrained portfolio-optimization model in the general form (2) proposed by Dentcheva and Ruszczyński.

Lemaréchal, Nemirovskii, and Nesterov prove the following efficiency estimate: To obtain an ϵ-*optimal* solution, the Constrained Level Method performs no more than

$$c \left(\frac{D\Lambda}{\epsilon}\right)^2 \ln \left(\frac{D\Lambda}{\epsilon}\right) \tag{37}$$

iterations, where D is the diameter of the feasible polyhedron X, Λ is a common Lipschitz constant of the objective and constraint functions, and c is a constant that depends only on the parameters of the method.

In the case of unconstrained minimization models, Lemaréchal, Nemirovskii, and Nesterov prove the following efficiency estimate: to obtain an ϵ-*optimal* solution, the (unconstrained) Level Method performs no more than

$$c' \left(\frac{D\Lambda}{\epsilon}\right)^2 \tag{38}$$

iterations, where D is the diameter of the feasible polyhedron X, Λ is a Lipschitz constant of the objective function, and c' is a constant that depends only on the parameters of the method.

Lemaréchal, Nemirovskii, and Nesterov report on successful application of the above methods to a variety of problems. Their experimental results suggest much better practical behavior than the above estimates.

Fábián and Szőke [5] experimented with variants of the Level Methods adapted for the solution of stochastic programming problems. Their test results are favorable. They found that for a fixed stopping tolerance ϵ, the required number of steps does not depend on the number S of the scenarios.

In the unconstrained case, [5] solved problems with increasing accuracy, and found that the number of steps required to find an ϵ-optimal solution, grew in proportion with $\log 1/\epsilon$. This suggest a much better practical behavior than the theoretical estimate (38).

Remark 9 If we use the Level Method for the portfolio selection model proposed by [36], then we can handle general concave partial achievement functions instead of the simplest achievement function (12).

A.2 Multi-objective models with variance

Roman et al. [37] proposed and studied portfolio optimization models which take into account both variance and CVaR. This approach represents a compromise between regulators' requirements for short tails and classical fund managers' requirements for small variance. A portfolio is characterized with 3 items of data: expected yield, variance, and CVaR. An approximation of the 3-dimensional efficient frontier is constructed. (In making the final choice, the decision maker plays a key role.) The authors tested the model on real-life data. Several levels of expected yield were considered; and for each level, five portfolios were selected from the mean/variance/CVaR efficient frontier: the minimum variance portfolio, the minimum CVaR portfolio, and three intermediate ones. The selected portfolios were then tested. Both in-sample and out-of-sample analysis shows that the performance of intermediate portfolios are superior.

The multi-objective model described in Sect. 2.2 can be extended to take into account variance as well. Given a portfolio x, let us consider the achievement function value

$$\Gamma_{\widehat{\tau}} \left(\operatorname{Tail}_{\frac{1}{S}} \left(r^T x \right), \ldots, \operatorname{Tail}_{\frac{S}{S}} \left(r^T x \right) \right)$$

as an efficiency measure of x. (The achievement function $\Gamma_{\widehat{\tau}}$ was defined in (12)). We can characterize a portfolio with 2 items of data: variance and the above efficiency measure. Applying the cutting-plane approach of Sect. 3.2, we can construct an approximation of the 2-dimensional efficient frontier.

A.3 Two-stage SP recourse models

We can easily formulate two-stage extensions of the SSD-models described in Sect. 2: Suppose we have two time-periods, and we can rebalance our portfolio at the beginning of each period. Let us compare benchmark yield and portfolio yield at the end of the second time period.

For the 2-stage extension of the SSD-constrained problem described in Sect. 2.1, [6] proposed a dual decomposition scheme and a solution method. Variance terms are included in the objective function following the ideas of [37]. Besides the economic advantages observed by Roman et al., the variance terms yield a technical advantage: They make the objective function strictly convex, and hence enable solution of the primal problem through a dual approach.

We plan to develop a solution method for the 2-stage extension of the multi-objective problem described in Sect. 2.2.

References

1. Dentcheva, D.: Private communication (2006)
2. Dentcheva, D., Ruszczyński, A.: Portfolio optimization with stochastic dominance constraints. J. Bank. Finance **30**, 433–451 (2006)
3. Ellison, E.F.D., Hajian, M., Levkovitz, R., Maros, I., Mitra, G.: A Fortran based mathematical programming system FortMP. Brunel University/Uxbridge, NAG Ltd/Oxford (1999)
4. Fábián, C.I.: Handling CVaR objectives and constraints in two-stage stochastic models. Eur. J. Oper. Res. **191**, 888–911 (special issue on continuous optimization in industry. Illés, T., Lopez, M., Vörös, J., Terlaky, T., Weber, G-W. eds.) (2008)
5. Fábián, C.I., Szőke, Z.: Solving two-stage stochastic programming problems with level decomposition. Comput. Manage. Sci. **4**, 313–353 (2007)
6. Fábián, C.I., Veszprémi, A.: Algorithms for handling CVaR-constraints in dynamic stochastic programming models with applications to finance. J. Risk **10**, 111–131 (2008)
7. Fishburn, P.C.: Decision and Value Theory. Wiley, New York (1964)
8. Fishburn, P.C.: Utility Theory for Decision Making. Wiley, New York (1970)
9. Fishburn, P.C.: Mean-risk analysis with risk associated with below-target returns. Am. Econ. Rev. **67**, 116–126 (1977)
10. Fourer, R., Gay, D.M., Kernighan, B.: AMPL: a mathematical programming language (1989)
11. Hadar, J., Russell, W.: Rules for ordering uncertain prospects. Am. Econ. Rev. **59**, 25–34 (1969)
12. Hardy, G.H., Littlewood, J.E., Polya, G.: Inequalities. Cambridge University Press, Cambridge (1934)
13. Kall, P., Mayer, J.: Stochastic linear programming: models, theory, and computation. Springer, Int. Ser. Oper. Res. Manage. Sci. (2005)
14. Klein Haneveld, W.K.: Duality in Stochastic Linear and Dynamic Programming. Lecture Notes in Economics and Math. Systems, vol. 274. Springer, New York (1986)
15. Klein Haneveld, W.K., van der Vlerk, M.H.: Integrated chance constraints: reduced forms and an algorithm. Comput. Manage. Sci. **3**, 245–269 (2006), First published as SOM Research Report 02A33, University of Groningen (2002)
16. Krokhmal, P., Palmquist, J., Uryasev, S.: Portfolio optimization with conditional value-at-risk objective and constraints. J. Risk **4**, 11–27 (2002)
17. Krokhmal, P., Soberanis, P.: Risk optimization with p-order conic constraints: a linear programming approach. Working paper, Department of Mechanical and Industrial Engineering, The University of Iowa, IA 52242 (2008)
18. Künzi-Bay, A., Mayer, J.: Computational aspects of minimizing conditional value-at-risk. Comput. Manage. Sci. **3**, 3–27 (2006)
19. Lemaréchal, C., Nemirovskii, A., Nesterov, Yu.: New variants of bundle methods. Math. Program. **69**, 111–147 (1995)
20. Levy, H.: Stochastic dominance and expected utility: survety and analysis. Manage. Sci. **38**, 555–593 (1992)
21. Lim, C., Sherali, H.D., Uryasev, S.: Portfolio optimization by minimizing conditional value-at-risk via nondifferentiable optimization. Comput. Optim. Appl. (2008) doi:10.1007/s10589-008-9196-3
22. Mansini, R., Ogryczak, W., Speranza, M.G.: Conditional value-at-risk and related linear programming models for portfolio optimization. Ann. Oper. Res. **152**, 227–256 (2007)
23. Markowitz, H.: Portfolio selection. J. Finance **7**, 77–91 (1952)
24. von Neumann, J., Morgenstern, O.: Theory of Games and Economic Behavior. Princeton University Press, Princeton (1947)
25. Ogryczak, W.: Multiple criteria linear programming model for portfolio selection. Ann. Oper. Res. **97**, 143–162 (2000)
26. Ogryczak, W.: Multiple criteria optimization and decisions under risk. Control Cybern. **31**, 975–1003 (2002)
27. Ogryczak, W., Ruszczyński, A.: From stochastic dominance to mean-risk models: semideviations as risk measures. Eur. J. Oper. Res. **116**, 33–50 (1999)
28. Ogryczak, W., Ruszczyński, A.: On consistency of stochastic dominance and mean-semideviations models. Math. Program. **89**, 217–232 (2001)
29. Ogryczak, W., Ruszczyński, A.: Dual stochastic dominance and related mean-risk models. SIAM J. Optim. **13**, 60–78 (2002)

30. Prékopa, A.: Probabilistic programming. In: Ruszczyński, A., Shapiro, A. (eds.) Stochastic Programming, Handbooks in Operations Research and Management Science, vol. 10, pp. 267–351. Elsevier, Amsterdam (2003)
31. Quirk, J.P., Saposnik, R.: Admissibility and measurable utility functions. Rev. Econ. Stud. **29**, 140–146 (1962)
32. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)
33. Rockafellar, R.T.: Coherent approaches to risk in optimization under uncertainty. Tutorials in Oper. Res. INFORMS, 38–61 (2007)
34. Rockafellar, R.T., Uryasev, S.: Optimization of conditional value-at-risk. J. Risk **2**, 21–41 (2000)
35. Rockafellar, R.T., Uryasev, S.: Conditional value-at-risk for general loss distributions. J. Bank. Finance **26**, 1443–1471 (2002)
36. Roman, D., Darby-Dowman, K., Mitra, G.: Portfolio construction based on stochastic dominance and target return distributions. Math. Program. Ser. B **108**, 541–569 (2006)
37. Roman, D., Mitra, G., Darby-Dowman, K.: Mean-risk models using two risk measures: a multi-objective approach. Quant. Finance **7**, 443–458 (2007)
38. Ross, S.M.: An Elementary Introduction to Mathematical Finance. Cambridge University Press, Cambridge (2002)
39. Ruszczyński, A.: A regularized decomposition method for minimizing the sum of polyhedral functions. Math. Program. **35**, 309–333 (1986)
40. Sadki, A.M.: AMPL COM Component Library, User's Guide Version 1.6. Internal T. Report (2005). See also http://www.optirik-systems.com/products/AMPLCOM
41. Whitmore, G.A.: Third-degree stochastic dominance. Am. Econ. Rev. **60**, 457–459 (1970)
42. Whitmore, G.A., Findlay, M.C.: Stochastic Dominance: An Approach to Decision-Making Under Risk. D.C.Heath, Lexington, MA (1978)
43. Wierzbicki, A.: A mathematical basis for satisficing decision making. Math. Model. **3**, 391–405 (1982)
44. Yamai, Y., Yoshiba, T.: Comparative analyses of expected shortfall and value-at-risk: expected utility maximization and tail risk. Monet. Econ. Stud. **20**, 57–86 (2002)