DOCUMENTOS DE TRABAJO

BILTOKI

D.T. 2015.01

Some experiments on solving
multistage stochastic mixed 0-1 programs
with time stochastic dominance constraints

Laureano. F. Escudero, María Araceli Garín, María Merino, Gloria Pérez

eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

Facultad de Ciencias Económicas.
Avda. Lehendakari Aguirre, 83
48015 BILBAO.

Documento de Trabajo BILTOKI DT2015.01
Editado por el Departamento de Economía Aplicada III (Econometría y Estadística)
de la Universidad del País Vasco.

ISSN: 1134-8984

Some experiments on solving multistage stochastic mixed 0-1 programs with time stochastic dominance constraints

Laureano F. Escudero[1], **María Araceli Garín**[2], **María Merino**[3], **Gloria Pérez**[3]

[1]Dpto. Estadística e Investigación Operativa
Universidad Rey Juan Carlos, Móstoles (Madrid), Spain
laureano.escudero@urjc.es

[2]Dpto. de Economía Aplicada III
Universidad del País Vasco, Bilbao (Vizcaya), Spain
mariaaraceli.garin@ehu.es

[3]Dpto. de Matemática Aplicada, Estadística e Investigación Operativa
Universidad del País Vasco, Leioa (Vizcaya), Spain
maria.merino@ehu.es, gloria.perez@ehu.es

March 4, 2015

Abstract

In this work we extend to the multistage case two recent risk averse measures for two-stage stochastic programs based on first- and second-order stochastic dominance constraints induced by mixed-integer-linear recourse. Additionally, we consider Time Stochastic Dominance (TSD) along a given horizon. Given the dimensions of medium-sized problems augmented by the new variables and constraints required by those risk measures, it is unrealistic to solve the problem up to optimality by plain use of MIP solvers in a reasonable computing time, at least. Instead of it, decomposition algorithms of some type should be used. We present an extension of our Branch-and-Fix Coordination algorithm, so named BFC-TSD, where a special treatment is given to cross scenario group constraints that link variables from different scenario groups. A broad computational experience is presented by comparing the risk neutral approach and the tested risk averse strategies. The performance of the new version of the BFC algorithm versus the plain use of a state-of-the-art MIP solver is also reported.

Keywords: Multistage stochastic mixed 0-1 optimization, scenario clustering, mixed 0-1 deterministic equivalent model, risk averse measures, stochastic dominance constraints.

1 Introduction

A multistage stochastic optimization model has a more complex scenario information structuring than its related sometimes approximating two-stage model. For the general formulation of a multistage model, where decisions on each stage have to be made stage-wise, let Ω denote the set of scenarios that are considered to be representative of the uncertainty in the problem and $\mathcal{T}$ is the set of stages in the given time horizon. At every state, i.e., node (also so-named scenario group), say g in set $\mathcal{G}^t$, where it is the set of nodes in stage $t \in \mathcal{T}$ in the related scenario tree, there is information about the past that is known with certainty. That information consists

of the realization of the uncertain parameters in the node and in its ancestor ones, say set $\mathcal{A}^g$. On the other hand, it is also known at that node g what scenarios will not happen in the future, i.e., the set of scenarios in set Ω/Ω^g where Ω^g denotes the set of scenarios that belong to group g, since the realizations of their uncertain parameters are identical up to stage t, for $t \in \mathcal{T} : g \in \mathcal{G}^t$. So, the decision variables at each node g should be based on the known information without anticipating future information although using as much information as possible (being represented in set Ω^g), i.e., the nonanticipativity (for short, NA) principle introduced in [41] for two-stage problems should be satisfied. Additionally, at a first glance it seems appropriate to consider that the optimal policy (i.e., the decision variables) at node g in stage t should not depend on scenarios which do not below to set Ω^g, i.e., 'cannot happen in the future' as stated in [39]. It is so-named time consistency principle, see [37, 39]; for slightly different contexts, see [30, 36].

Notice that the NA principle paved the way for a dual block-angular representation of the problem, introduced in [42] and known as Deterministic Equivalent Model (DEM, for short). Traditionally, special attention has been given to the DEM by optimizing (in our case, maximizing) the objective function expected value over the set of scenarios, along the time horizon, subject to the satisfaction of all the problem constraints in the defined scenarios, i.e., the so named risk neutral (RN) strategy. Currently, we are able to solve RN-based medium-sized mixed 0-1 DEMs by using different types of decomposition approaches, e.g., see in [17, 18] a description of our Branch-and-Fix Coordination (BFC) algorithm. However, the optimization of the RN-based objective function expected value has the inconvenience of providing a solution that ignores the variability of the objective function value over the scenarios. So, it does not hedge against the occurrence of low-probability high-consequence events (i,e., the so named black swan events). For multistage optimization the RN measure, obviously, satisfies the time consistency principle, since it is only based on averages (expectations) in the objective function value for feasible (implementable) decision variables (i.e., the ones satisfying the NA principle).

Alternatively, there are some multistage approaches that also deal with the type of risk management that provides hedging solutions against the occurrence of some non-desired scenarios, see some surveys in [2, 35], by considering semi-deviations, scenario optimization, value-at-risk (VaR), conditional value-at-risk (CVaR), mean-expected shortfall and mean-risk, among other risk averse measures. Our BFC approach can be easily adapted to those strategies. Additionally, there are some other risk averse strategies as the VaR measure, among others, that require the extension of our BFC procedure due to the requirement of considering cross scenario group constraints, a difficult issue (see below) for multistage problems, being the subject of this work. Note: A cross scenario constraint is a constraint where there are variables with nonzero elements that belong to different scenarios. It also applies to scenario group that belong to the same stage.

Some risk averse measures satisfy the time consistency principle and some others not, see [39], although all of them are implementable. Some risk averse CVaR-based approaches with time consistency have been proposed in the literature, see [6, 9, 12, 21, 34, 36], among others.

Recently, new risk averse measures have appeared in the literature, in particular, the so named first- and second-order Stochastic Dominance Constraints (SDC) strategies for a set of profiles, each one included by a

threshold value on a given function and some types of shortfall related bounds on reaching it. See [14, 15] for the case of continuous variables where the problem is considered as a semi-infinite one, [13] for the general continuous case, and [24] for first-order and [25] for second-order SDC induced by mixed integer-linear recourse on finite scenarios. More recently, an application in mining is presented in [2] (by plain use of a MIP solver for a multistage problem), some applications in energy are described in [11, 16, 25] (by using Lagrangean heuristics and cutting plane approaches for two-stage problems) and some applications in finance are introduced in [5, 22, 31] (the first one for a multistage problem, the second one for a single stage problem and the last one for a two stage environment, particularly for the second-order measure), among others.

By construction, SDC-based measures as well as any other measure that requires cross scenario (or scenario group) constraints (as VaR [23], mean-expected shortfall [20, 26] and mean-risk [38], among others) do not satisfy the time consistency principle as it has been enunciated above. The reason is that the optimal values of the decision variables at any node also require to use information from scenarios that at that time is known that will not happen but, for hedging purposes, that 'outside' information is required. So, they do not satisfy the time consistency principle.

Given the broad applicability of those risk averse measures we consider the time stochatic dominance (for short, TSD) policy to cover those situations where risk reduction should be performed for hedging values of given functions (the objective one, in our case) along the time horizon. The risk neutral strategy satisfies the time consistency principle. However, in cases where its 'optimal' solution results in high variability in the objective function value over the scenarios, the TSD policy has the advantage of better risk reduction of non-wanted scenarios. The decision on how far the objective function expected value is from a scenario solution depends on the problem's structure and the modeler-driven risk management safeguards.

In this work we consider a threshold value on the objective function at the end of the time horizon, as well as at a modeler-driven subset of stages. So, our approach extends a mixture of the two recent two-stage SDC strategies introduced in [24, 25] to the multistage risk averse environment with TSD. The new measure requires additional variables and constraints. The new variables are binary ones for the first-order stochastic dominance constraints strategy (FSD), and they are continuous ones for the second-order one (SSD). The difference between FSD and SSD strategies is that the former considers bounds on the probability of failure that the objective function value reaches the thresholds, and the latter considers bounds on the expected shortfall on reaching them. Given the dimensions of medium-sized instances augmented by the new variables and constraints required by the risk measures to consider, it is unrealistic to solve the problem up to optimality by plain use of MIP solvers in a reasonable computing time. Instead of it, one should use decomposition algorithms of some type. However, since both strategies (FSD and SSD) require cross scenario group constraints, the nice decomposable scenario group-based structure of the problem is partially destroyed. So, traditional Benders-based [7] decomposition and splitting variable schemes such as scenario group-based decomposition [18], cluster Lagrangean decomposition [19] and stage-based decomposition [33] cannot be used, among other types of decomposition algorithms. The key idea of those algorithms is that the original scenario tree-base model can be easily decomposed in independent scenario-related or stage-related submodels and, then,

models with cross scenario group constraints along the stages can not be solved by those schemes without an appropriate modification.

We propose in this work the new multistage risk averse strategy so named TSD as a mixture of the soft FSD and SSD measures, presenting its advantages as well as the computational price to be paid for using plain MIP solvers, instead of appropriate decomposition approaches. Additionally, an extension of the BFC algorithm is also proposed for dealing with the cross scenario group constraints that are required particularly by the risk measures to consider in this work (besides some important refinements for branch cutting purposes). The rationale behind considering soft versions of the measures is that it could be possible that both FSD and SSD may not be simultaneously satisfied. Then, since they are modeler-driven risk management policies (i.e., they are not physical requirements) and their potential violation is to be heavily penalized, the resulting solution may still be useful for decision making, see below. So, let BFC-TSD be the name for the new proposed algorithm. A broad computational experience is presented by comparing the risk neutral approach and the tested risk averse strategies as well as the impact of including the TSD policy and the advantage of considering soft SDC bounds instead of the hard ones. The performance of the new version of the BFC algorithm versus the plain use of a state-of-the-art MIP solver is also reported.

The remainder of the paper is organized as follows. In Section 2 the multistage mixed 0-1 optimization problem with the risk neutral (RN) strategy is presented as well as the model for the TSD strategy. Section 3 presents the corresponding original RN DEM as a mixture of the splitting variable and compact representations, and the related scenario cluster submodels on which the original model can be decomposed. Section 4 presents the BFC-TSD algorithm for problem solving. Section 5 reports computational results of using the new algorithm for solving a test bed of medium-sized instances. A comparison between the RN and SDC strategies (i.e., FSD, SSD and the new one, TSD) is performed as mentioned above. Finally, some conclusions are withdrawn in Section 6. The appendices present additional information on model dimensions, some illustrative graphics and the meaning of abbreviations used in the work.

2 TSD measure in multistage stochastic mixed 0-1 problems

Without loss of generality, let us consider the compact representation of the multistage mixed 0-1 model for maximizing the objective function expected value over a set of scenarios, say Ω, in the scenario tree that is used for representing the random parameters and decision variables. The risk neutral (RN) strategy is used.

$$\begin{aligned} z_{RN} = \quad & \max \sum_{g \in \mathcal{G}} w^g (a^g x^g + b^g y^g) \\ \text{s.t.} \quad & \sum_{q \in \mathcal{A}^g} (A^q_g x^q + B^q_g y^q) = h^g \qquad \forall g \in \mathcal{G} \\ & x^g \in \{0,1\}^{nx(g)}, y^g \in I\!R^{ny(g)} \qquad \forall g \in \mathcal{G}, \end{aligned} \tag{1}$$

where w^g is the likelihood or probability of scenario group g to be computed as $\sum_{\omega \in \Omega^g} w^\omega$; ω is a scenario in set $\Omega^g \subseteq \Omega$ of group g; w^ω is the modeler-driven weight assigned to scenario $\omega \in \Omega$; $\mathcal{G}$ is the set of scenario groups (i.e., nodes in the scenario tree); $\mathcal{A}^g$ is the set of indexes for the ancestor scenario groups to group g

(including itself) whose decisions have direct influence (i.e., have non zero elements) on the constraints for scenario group g, where $\mathcal{A}^1 = \{1\}$; x^g and y^g are the vectors of the 0-1 and continuous variables for scenario group g, respectively; a^g and b^g are the vectors of the objective function coefficients for the 0-1 and continuous variables, respectively; A_g^q and B_g^q are the constraint matrices of ancestor scenario group q of group g for the vectors x^q and y^q with direct influence in the constraints of group g, respectively; h^g is the right-hand-side vector (rhs) for scenario group g; and $nx(g)$ and $ny(g)$ are the numbers of 0-1 and continuous variables, for $g \in \mathcal{G}$, respectively.

Notice that the constraint system in model (1) is a generalization of the traditional one, where if $\mathcal{A}^g/\{g\}$ is singleton then q for $q \in \mathcal{A}^g/\{g\}$ is usually the index of the immediate ancestor scenario group of group g. Examples of a non-singleton set $\mathcal{A}^g$ are frequent in transportation models where the trip starting period is different from its arrival one, supply chain management (where the decision on suppliers' selection and raw material supplying commitment at the beginning of the time horizon impacts on the volume to be supplied in later periods), site location in a dynamic setting, power unit commitment lasting for more than one period, investment planning where the period at which the investment is made in factories, power plants, energy transmission lines, warehouses, distributions centers, etc. is different from the period at which the investment is made available for being used. In those situations the constraints of a given scenario group have nonzero elements for the variables of ancestor scenario groups.

Additionally, let $\mathcal{G}^t \subseteq \mathcal{G}$ denote the set of scenario groups in stage $t \in \mathcal{T}$, where $\mathcal{T}$ is the set of stages in the considered horizon; $t(g)$ is the stage from set $\mathcal{T}$ to which group g belongs to; $T = |\mathcal{T}|$; and observe that Ω^g is a singleton for $g \in \mathcal{G}^T$. See e.g., [8] for the main concepts on stochastic optimization via scenario tree analysis.

The RN model (1) aims to maximize the objective function expected value alone. The main criticism that can be made to this very popular strategy, as stated above, is that it ignores the variability of the objective function value over the scenarios and, in particular in our case, the "left" tail of the non-wanted scenarios. However, there are some risk averse approaches that additionally deal with risk management; among them, the TSD strategy below reduces the risk of wrong solutions in a better way than the other ones under some hypotheses, see e.g., [2]. That strategy also aims to maximize the objective function expected value as RN (1). Additionally, a set of given thresholds on a function value for each scenario should be satisfied with a bound *target* on the probability of failure due to a shortfall on reaching each threshold as well as a bound *target* on the expected shortfall on reaching it. Particularly, TSD strategy requires a set of modeler-driven profiles, say $\mathcal{P}^t$, given by the n-tupla $(\phi^p, e^p, \beta^p, S^p) \forall p \in \mathcal{P}^t, t \in \tilde{\mathcal{T}}$, where ϕ^p is the objective function value (for short, the profit) threshold to be satisfied up to node (i.e., scenario group) g in the scenario tree, e^p is the upper bound *target* on the expected shortfall on reaching profit threshold ϕ^p with a *target* on the shortfall probability bound β^p, and S^p is the maximum shortfall that is allowed on reaching the threshold for scenario group g, for $g \in \mathcal{G}^t, t \in \tilde{\mathcal{T}}$, where $\tilde{\mathcal{T}}$ denotes the set of stages in $\mathcal{T}$ for which the TSD policy is considered. The model can

be expressed as follows,

$$
\begin{array}{lll}
z_{TSD} = & \max \sum_{g \in \mathcal{G}} w^g (a^g x^g + b^g y^g) - \sum_{t \in \tilde{\mathcal{T}}} \sum_{p \in \mathcal{P}^t} (M_e^p \varepsilon_e^p + M_\beta^p \varepsilon_\beta^p) & \\
\text{s.t.} & \sum_{q \in \mathcal{A}^g} A_g^q x^q + B_g^q y^q = h^g & \forall g \in \mathcal{G} \\
& \sum_{q \in \tilde{\mathcal{A}}^g} (a^q x^q + b^q y^q) + s^{gp} \geq \phi^p & \forall g \in \mathcal{G}^t,\ p \in \mathcal{P}^t,\ t \in \tilde{\mathcal{T}} \\
& 0 \leq s^{gp} \leq S^p \nu^{gp} & \forall g \in \mathcal{G}^t,\ p \in \mathcal{P}^t,\ t \in \tilde{\mathcal{T}} \\
& \sum_{g \in \mathcal{G}^t} w^g s^{gp} \leq e^p + \varepsilon_e^p & \forall p \in \mathcal{P}^t,\ t \in \tilde{\mathcal{T}} \\
& \sum_{g \in \mathcal{G}^t} w^g \nu^{gp} \leq \beta^p + \varepsilon_\beta^p & \forall p \in \mathcal{P}^t,\ t \in \tilde{\mathcal{T}} \\
& x^g \in \{0,1\}^{nx(g)}, y^g \in I\!R^{ny(g)} & \forall g \in \mathcal{G} \\
& \nu^{gp} \in \{0,1\} & \forall g \in \mathcal{G}^t,\ p \in \mathcal{P}^t,\ t \in \tilde{\mathcal{T}} \\
& 0 \leq \varepsilon_e^p \leq S^p - e^p,\ 0 \leq \varepsilon_\beta^p \leq 1 - \beta^p & \forall p \in \mathcal{P}^t,\ t \in \tilde{\mathcal{T}},
\end{array}
\tag{2}
$$

where $\tilde{\mathcal{A}}^g$ is the set of indices for the ancestor scenario groups of group g and, so, $\mathcal{A}^g \subseteq \tilde{\mathcal{A}}^g$; s^{gp} is the shortfall (continuous) variable that, obviously, is equal to the difference (if it is positive) between threshold ϕ^p and the profit for scenario group g, ν^{gp} is a 0-1 variable such that its value is 1 if the profit for scenario group g has a shortfall, and otherwise, 0; and ε_e^p and ε_β^p are the slack variables that take the violation of the e- and β-bounds, respectively, being M_e^p and M_β^p the related big enough M-parameters for penalizing those variables in the objective function. Observe that the constraints with the rhs e^p and β^p are precisely the cross scenario group constraints for the last stage T assuming that $T \in \tilde{\mathcal{T}}$ and, any case, they are the cross scenario group constraints for any stage in set $\tilde{\mathcal{T}}$, since the weighted variables s^{gp} and ν^{gp} are summed up for all scenario groups in set $\mathcal{G}^t$ for $p \in \mathcal{P}^t$, $t \in \tilde{\mathcal{T}}$. Notice that the TSD policy consists of controlling the objective function value at modeler-driven stages, instead of only performing it at the end of the time horizon. It is very useful for applications with long horizons. The concept of the expected shortfall on reaching a given profit threshold has its roots in the Integrated Chance Constraints concept introduced in [28], see also [29]; in different contexts, see [32, 40].

3 Scenario clustering in risk averse modeling

In [18] we propose a decomposition of the nonsymmetric scenario tree into a set of scenario cluster subtrees. Based on this cluster decomposition concept a splitting-compact representation of the original RN multistage stochastic mixed 0-1 model is presented. The reason for such decomposition is based on the way in which our BFC decomposition algorithm works. It explicitly considers the nonanticipativity constraints (NAC) of the variables of scenario groups in different cluster subtrees. By construction, those scenario groups belong to stages up to a given so-named break stage, t^* (see below). On the other hand, the NAC of the variables of scenario groups that belong from the next stage to the break one until the last stage are implicitly considered while solving the scenario cluster submodels. Let us consider the following definitions taken from [18].

Definition 1 *A **break stage** t^* is a stage such that the number of scenario clusters is $C = |\mathcal{G}^{t^*+1}|$, where $t^*+1 \in \mathcal{T}$. In this case, any cluster $c \in \mathcal{C}$ is induced by a group $g \in \mathcal{G}^{t^*+1}$ and contains all the scenarios belonging to that group.*

Definition 2 *The **scenario cluster** submodels are those that result from the relaxation of the NAC until break stage t^*.*

For clarification purposes, see Figure 1, where the NAC of the variables for node 1 (res. nodes 1, 2 and 3) in the scenario tree shown in the left part of the figure have been relaxed for break stage $t^* = 1$ (res. $t^* = 2$) in the center (res. right) part of the figure.

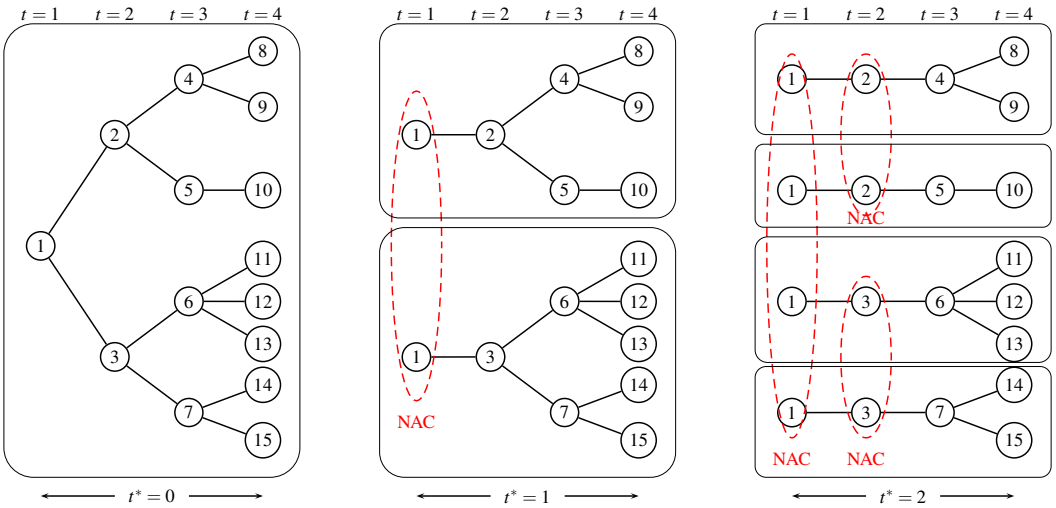

Figure 1: Scenario clustering for $t^* = 0$ (one cluster), $t^* = 1$ (two clusters) and $t^* = 2$ (four clusters)

Once decided the break stage t^*, the corresponding t^*-cluster partition is given and then, the number of scenario clusters C is fixed to $|\mathcal{G}^{t^*+1}|$, i.e., where each group in $\mathcal{G}^{t^*+1}$ belongs to just one scenario cluster in set $\mathcal{C} = \{1,...,C\}$, with $C = |\mathcal{C}|$. Let $\Omega_c = \Omega^g$ for $g \in \mathcal{G}^{t^*+1}$ denote the set of scenarios and $\mathcal{G}_c \subseteq \mathcal{G}$ the set of scenario groups in cluster $c \in \mathcal{C}$, respectively, where $g \in \mathcal{G}_c$ provided that $\Omega^g \cap \Omega_c \neq \emptyset$; x_c^g and y_c^g denote the replicas of the variables vectors x^g and y^g for scenario group $g \in \mathcal{G}_c$ in scenario cluster $c \in \mathcal{C}$, respectively; and x_c and y_c are the vectors that include the set of variables in the vectors x_c^g and y_c^g for all scenario groups $g \in \mathcal{G}_c$ in cluster $c \in \mathcal{C}$, respectively. Note: $\mathcal{G}^t \cap \mathcal{G}_c$ is singleton for $t \in \mathcal{T} : t \leq t^*+1$.

Now, we can formulate the cluster submodels in compact representation, such that the original RN model (1) can be formulated via a mixture of the splitting variable and compact representations, so that the cluster submodels are linked by the explicit NAC up to break stage t^*. So, the strategy RN via compact representation

of each cluster submodel c for $c \in \mathcal{C}$ can be formulated as follows,

$$\begin{aligned} z_c = \quad & \max \sum_{g \in \mathcal{G}_c} w_c^g (a^g x_c^g + b^g y_c^g) \\ \text{s.t.} \quad & \sum_{q \in \mathcal{A}^g} A_g^q x_c^q + B_g^q y_c^q = h^g \qquad \forall g \in \mathcal{G}_c \\ & x_c^g \in \{0,1\}^{nx(g)}, y_c^g \in I\!R^{ny(g)} \quad \forall g \in \mathcal{G}_c, \end{aligned} \tag{3}$$

where $w_c^g = \sum_{\omega \in \Omega^g \cap \Omega_c} w^\omega$, $w_c^g = w^q$ being $q \in \mathcal{G}^{t^*+1} \cap \mathcal{G}_c$ for $g : t(g) \in \mathcal{T}_1$, and $w_c^g = w^g$ for $g : t(g) \in \mathcal{T}_2$.

Let us first split the set of stages $\mathcal{T}$ in two subsets, such that $\mathcal{T} = \mathcal{T}_1 \bigcup \mathcal{T}_2$, where $\mathcal{T}_1 = \{1, \ldots, t^*\}$, and $\mathcal{T}_2 = \{t^*+1, \ldots, T\}$. Now, in the splitting-compact representation of RN (1), the nonanticipativity principle is implicitly taken into account for the stages $t \in \mathcal{T}_2$, since the submodel for each cluster is formulated via a compact representation. On the other hand, the (explicit) NAC of the scenario clusters for the stages in set $\mathcal{T}_1$ can be formulated by observing that the clusters c and c' have the scenario group g in common if $g \in \mathcal{G}_c \cap \mathcal{G}_{c'}$, and it could only happen for $g \in \mathcal{G}^t : t \in \mathcal{T}_1$. So, the cluster submodels (3) are linked by the NAC to be formulated as follows,

$$x_c^g - x_{c'}^g = 0 \quad \forall c, c' \in \mathcal{C} : c \neq c',\ g \in \mathcal{G}_c \cap \mathcal{G}_{c'} \tag{4}$$

$$y_c^g - y_{c'}^g = 0 \quad \forall c, c' \in \mathcal{C} : c \neq c',\ g \in \mathcal{G}_c \cap \mathcal{G}_{c'}. \tag{5}$$

RN model (1) can be represented by a mixture of the splitting variable representation (for explicitly satisfying the NAC between the cluster submodels) and the compact representation (for implicitly satisfying the NAC of each cluster submodel, besides the other constraints in the submodel). So, the cluster splitting-compact representation can be expressed

$$\begin{aligned} z_{RN} = \quad & \max \sum_{c \in \mathcal{C}} \sum_{g \in \mathcal{G}_c} w_c^g (a^g x_c^g + b^g y_c^g) \\ \text{s.t.} \quad & \sum_{q \in \mathcal{A}^g} A_g^q x_c^q + B_g^q y_c^q = h^g \qquad \forall g \in \mathcal{G}_c, c \in \mathcal{C} \\ & \text{NAC } (4) - (5) \\ & x_c^g \in \{0,1\}^{nx(g)}, y_c^g \in I\!R^{ny(g)} \quad \forall g \in \mathcal{G}_c, c \in \mathcal{C}. \end{aligned} \tag{6}$$

4 BFC-TSD algorithm

Plain use of state-of-the-art MIP optimization engines for solving the stochastic version of medium-sized instances of a multistage mixed 0-1 problem may require unaffordable computing effort (in memory and elapsed time). So, a decomposition algorithm for optimizing this type of problems is required. This requirement is strongly influenced by the risk averse measure TSD whose modeling needs additional continuous and 0-1 variables and scenario group- and stage-based constraints as well as cross scenario group constraints.

4.1 General methodology

The general methodology consists of a branching approach through the scenario tree to be presented below. It requires the following additional notation: $\overline{t} \in \mathcal{T}_1$, current branching stage; $\overline{g} \in \mathcal{G}^{\overline{t}}$, branching scenario group;

I^g, set of indices of vector x^g; and $\bar{i} \in I^{\bar{g}}$, index of the current branching variable in vector $x^{\bar{g}}$.

The main concepts of the BFC methodology have been introduced in [3, 4] and lately refined mainly in [17]. They are particularized here in the following elements:

1. Scenario cluster based Branch-and-Fix (BF) tree, i.e., Branch-and-Bound tree for any cluster, such that the optimization of the submodel (3) for any scenario cluster $c \in \mathcal{C}$ is performed in a coordinated way with the submodels for the other clusters, in order to solve up to optimality the original RN model (6).

2. Common variables, i.e., variables, say, $(x_c^g)_i$ and $(x_{c'}^g)_i$ for scenario clusters c and c' where $c, c' \in \mathcal{C}$: $c \neq c', i \in I^g, g \in \mathcal{G}_c \cap \mathcal{G}_{c'}$. Remember that $g \in \mathcal{G}^t, t \in \mathcal{T}_1$.

3. Twin nodes, i.e., nodes from different BF trees whose paths from their root nodes are such that their x-common variables $(x_c^g)_i$ and $(x_{c'}^g)_i$, if any, have been branched on or fixed to the same 0-1 value, for $c, c' \in \mathcal{C} : c \neq c', i \in I^g, g \in \mathcal{G}_c \cap \mathcal{G}_{c'}$. The branched or fixed common variable $(x_c^g)_i$ may imply fixing other common variables, say, $(x_c^{g'})_j$ for $j \in I^{g'}, g' \in \mathcal{G}_c$, such that themselves, by construction, force the replicas of those variables, say, $(x_{c'}^{g'})_j$ in any other scenario cluster $c' \in \mathcal{C} : g' \in \mathcal{G}_c \cap \mathcal{G}_{c'}$ to be fixed to the same 0-1 values, and the implications could go further.

4. Twin Node Family (TNF), i.e., set of nodes such that any node is a twin node to all the other node members in the family (from different BF trees).

5. Candidate TNF, i.e., a TNF such that there is one x-common variable in the node members, at least, that has not been yet branched on, nor fixed to 0-1 values.

6. Integer TNF, i.e., a TNF such that all x-common variables have already been branched on or fixed to 0-1 values.

The branching procedure BFC-TSD, Step 5, see Section 4.2, reaches a given candidate TNF for the branching triplet $(\bar{t}, \bar{g}, \bar{i})$. Let $\overline{I}_1$ denote the set of indices of the already branched on or fixed (common) variables at the candidate TNF, i.e., $\overline{I}_1 = \cup_{g \in \mathcal{A}^{\bar{g}} \setminus \{\bar{g}\}} I^g \cup \{i \in I^{\bar{g}} : i \leq \bar{i}\}$. Let $\hat{x}^g$ denote the 0-1 value vector for vector x^g, where $\hat{x}^g = \hat{x}_c^g$ for any scenario cluster $c \in \mathcal{C} : g \in \mathcal{G}_c$.

The pruning scheme at a candidate TNF is done by comparing the incumbent value, say, z_{TSD} for TSD model (2) and $z = \sum_{c \in \mathcal{C}} z_c$, where z_c is the solution value of submodel (3) for scenario cluster c, such that

- it has been obtained for that scenario cluster c while branching on a x-variable of scenario group $g \in \mathcal{G}_c \cap \mathcal{G}^{\bar{t}} : g > \bar{g}$ at the previous stage $\bar{t} - 1$,
- it has already been obtained when branching on a x-variable of group $g \in \mathcal{G}_c \cap \mathcal{G}^{\bar{t}} : g < \bar{g}$, or
- it has just been obtained for group $\bar{g} \in \mathcal{C}_c$.

The x-solution for an integer TNF can be obtained in procedure BFC-TSD, Step 6 as the (partial) solution of the scenario cluster submodels (3) for all $c \in \mathcal{C}$ and, in the related candidate TNF defined by the branching triplet $(\bar{t}, \bar{g}, \bar{i})$. Notice that each NAC (4) of the x-variables in model (2):

- either it has been algorithmically satisfied (case of already branched on or fixed variables $(x)_i \forall i \in \bar{I}_1$) or
- it has been implicitly satisfied while solving submodels (3) for $(x^g)_i$ where $i \in \{I^g, g \in \mathcal{G}^t, t \in \mathcal{T}_2\}$.
- Additionally, if its testing for the 0-1 $(\hat{x}^g)_i$-values of $(x^g)_i$ for $i \in \{I^g, g \in \mathcal{G}^t, t \in \mathcal{T}_1\} \setminus I_1$, where those values come from the solution of the cluster submodels, is positive (i.e., the NAC are satisfied) then an integer TNF has been obtained.

At each integer TNF a feasible solution for the original model (2) could be obtained by fixing the x-variables to their current 0-1 values, such that the model can be expressed

$$\begin{aligned} z_T^{TNF} = \quad & \max \sum_{g \in \mathcal{G}} w^g (a^g x^g + b^g y^g) - \sum_{t \in \tilde{\mathcal{T}}} \sum_{p \in \mathcal{P}^t} (M_e^p \varepsilon_e^p + M_\beta^p \varepsilon_\beta^p) \\ \text{s.t.} \quad & \text{Constraint system in TSD model (2)} \\ & x^g = \hat{x}^g \quad \forall g \in \mathcal{G}. \end{aligned} \tag{7}$$

Notice that the TSD 0-1 ν-variables are the only ones that are integer in the model. If it is feasible then the new incumbent solution value is $z_{TSD} := \max\{z_T^{TNF},\ z_{TSD}\}$.

In order to guarantee the optimality of the solution provided by the procedure, model (8) should be solved for any integer TNF where the variables up to break stage t^* have been fully branched on or fixed.

$$\begin{aligned} z_{OPT}^{TNF} = \quad & \max \sum_{g \in \mathcal{G}} w^g (a^g x^g + b^g y^g) - \sum_{t \in \tilde{\mathcal{T}}} \sum_{p \in \mathcal{P}^t} (M_e^p \varepsilon_e^p + M_\beta^p \varepsilon_\beta^p) & \\ \text{s.t.} \quad & \text{Constraint system in TSD model (2)} & \\ & x^g = \hat{x}^g & \forall g \in \mathcal{G}^t, t \in \mathcal{T}_1 \\ & x^g \in \{0,1\}^{nx(g)} & \forall g \in \mathcal{G}^t, t \in \mathcal{T}_2. \end{aligned} \tag{8}$$

Model (8) guarantees that there is not any area in the feasible region that explicitly or implicitly has not been explored. The rationale behind it has similarities with our works [1, 17, 18], but with some main differences: (a) The model forces the satisfaction of the cross scenario group constraints; and (b) The TSD policy requirements force the replacement of the paradigm given by the pair stage - scenario cluster with the paradigm given by the pair scenario group - scenario cluster.

4.2 BFC-TSD procedure for chosen t^* and label OPT being either $'Y'$ or $'N'$

The main steps of the procedure are as follows:

Step 1: (Root node)

Solve the independent scenario cluster MIP submodels (3) to obtain $z_c \forall c \in \mathcal{C}$.

Compute $z = \sum_{c \in \mathcal{C}} z_c$.

If all variables in x_c and y_c satisfy the NAC (4) and (5), respectively, and the linking cross scenario group constraints in (2) are satisfied for the ε-variables equal to zero in the TSD strategy then its optimal

solution is found, so, set $z_{TSD} := z$ and **STOP**.

Set $\overline{t} := 0$, $z_{TSD} := -\infty$.

Step 2: (Next stage)

Reset $\overline{t} := \overline{t} + 1$ and $\overline{i} := 0$. If $\overline{t} > t^*$ then go to Step 9.

Step 3: (Next scenario group)

Select $\overline{g} \in \mathcal{G}^{\overline{t}}$, unless all the variables in its groups have already been branched, in this case go to Step 2.

Reset $\overline{i} := 0$.

Step 4: (Next node)

Reset $\overline{i} := \overline{i} + 1$. If $\overline{i} > nx(\overline{g})$ then go to Step 3.

Step 5: (Branching. Candidate TNF building)

Set $(x_c^{\overline{g}})_{\overline{i}} := 1 \, \forall c \in \mathcal{C} : \overline{g} \in \mathcal{G}_c$.

Step 6: (Integer TNF building)

(a) If the current branching value $(\hat{x}^{\overline{g}})_{\overline{i}}$ has already been obtained for variable $(x^{\overline{g}})_{\overline{i}}$ while solving the appropriate submodel (3) at the immediate ancestor branching variable $(x^g)_i$, where either $g = \overline{g}$, $i = nx(g)$ such that g is the immediate ancestor scenario group in $\mathcal{G}^{\overline{t}}$ or $g \in \mathcal{G}^{\overline{t}-1}$ then go to Step 4. (Notice that, in this case, the solution vector $\hat{x}^g \, \forall g \in \mathcal{G}$ and its value z (see below) have already been obtained and then Steps 6 and 7 do not need to be executed).

(b) Solve the independent scenario cluster MIP submodels (3) for obtaining solution value $z_c \, \forall c \in \mathcal{C} : \overline{g} \in \mathcal{G}_c$, where the already branched variables in the TNF (i.e., $\overline{I}_1$) are fixed to the related 0-1 values and all the other x-variables (then, until the groups in $\mathcal{G}^T$, including themselves) keep their integrality.

(c) Compute $z = \sum_{c \in \mathcal{C}} z_c$, where z_c has already been obtained for that scenario cluster c while branching on scenario group $g \in \mathcal{G}_c \cap \mathcal{G}^{\overline{t}} : g > \overline{g}$ at the previous stage $\overline{t} - 1$, when branching on $g \in \mathcal{G}_c \cap \mathcal{G}^{\overline{t}} : g < \overline{g}$ or it has just been obtained for $\overline{g} \in \mathcal{C}_c$.

(d) If $z \leq z_{TSD}$ then go to Step 8.

(e) If any variable in $x_c \, \forall c \in \mathcal{C}$ does not satisfy NAC (4) then go to Step 4.

(f) If all variables in $y_c \, \forall c \in \mathcal{C}$ do satisfy NAC (5) and the linking cross scenario group constraints in (2) are also satisfied for the ε-variables equal to zero in the TSD strategy then update $z_{TSD} := z$ and go to Step 8.

Step 7: (Satisfying the cross scenario group constraints for integer TNF)

Solve MIP model (7) for obtaining a feasible solution to the original model (2).

If it is feasible then update $z_{TSD} := \max\{z_T^{TNF}, \, z_{TSD}\}$.

If $OPT =' N'$ or $\overline{t} < t^*$ or $\overline{g}$ is not the last group in set $\mathcal{G}^{\overline{g}}$ or $\overline{i} < nx(\overline{g})$ then go to Step 4.

Solve MIP model (8) for obtaining another feasible solution to model (2) and guaranteeing optimality. If it is feasible then update $z_{TSD} := \max\{z^{TNF}_{OPT},\ z_{TSD}\}$.

Step 8: (Branch pruning)

If variable $(x^{\overline{g}}_c)_{\overline{i}}$ has been branched on or fixed to 1 for any cluster $c \in \mathcal{C} : \overline{g} \in \mathcal{G}_c$ then go to Step 11.
(Note: The branched on or fixed to 0-1 value for variable $(x^{\overline{g}}_c)_{\overline{i}}$ is the same one for all clusters $c \in \mathcal{C} : \overline{g} \in \mathcal{G}_c$).

Step 9: (Backward to previous node)

Reset $\overline{i} := \overline{i} - 1$.
If $\overline{i} = 0$ and $\overline{t} \leq 1$ then, the incumbent solution, if any, and its value z_{TSD} has been found, **STOP**.
If $\overline{i} = 0$ and the x-common variables of all the groups in $\mathcal{G}^{\overline{t}}$ have already been branched on or fixed then $\overline{t} := \overline{t} - 1$ and select the last group $\overline{g} \in \mathcal{G}^{\overline{t}}$.
If $\overline{i} = 0$ and not all the x-common variables of all the groups in $\mathcal{G}^{\overline{t}}$ have already been branched on or fixed then select the previous group $\overline{g} \in \mathcal{G}^{\overline{t}}$.

Step 10: (Prune checking)

If $(x^{\overline{g}}_c)_{\overline{i}} = 0$ for any $c \in \mathcal{C} : \overline{g} \in \mathcal{G}_c$ then go to Step 9.

Step 11: (Opposite branching)

Reset $(x^{\overline{g}}_c)_{\overline{i}} := 0 \forall c \in \mathcal{C} : \overline{g} \in \mathcal{G}_c$ and go to Step 6.

It is worth to point out that procedure BFC-TSD for the modeler-driven label $OPT =' N'$ may not obtain the optimal solution to the original model (2) and, in the worst case, it may not even find any feasible solution. Notice that Step 6 only performs branching until break stage t^* that could be much smaller than the last stage of the time horizon, and model (7) to be solved in Step 7 may not be feasible for the values of the x-variables as obtained by scenario cluster submodels (3) in Step 6. However, the elapsed time required by the procedure for obtaining the incumbent solution could be affordable for medium-sized instances. In fact, the computational experiment reported in Section 5 for a broad test bed of instances shows that the solutions that have been obtained for all instances considering $t^* = 1$ for label $OPT =' N'$ require very small computing time; and their incumbent values coincide with the ones obtained by plain use of CPLEX V12.5 (even for the optimal ones) or they are very close for the instances where CPLEX is not running out of memory. On the other hand, the optimal solution of the original TSD model (2) could have 0-1 values for some x-variables that have not been obtained by solving those submodels. Then, there is a guarantee for label $OPT =' Y'$ that those values are to be found, if any, since all candidate TNFs will be implicit or explicitly explored.

5 Computational experience

The BFC-TSD algorithm has been implemented in a C++ code. It uses one of the state-of-the-art commercial optimization engines, in particular CPLEX V12.5, see [27], within the open source engine COIN-OR v1.6.0

[10]. This optimizer is used by the algorithm for solving the LP relaxation and the mixed 0-1 submodels; all the tolerances have been set up to the default values.

The computational experiments were conducted in the SW/HW platform given by a workstation Dell Precision T7600 under Linux operating system (version Debian2.6.32-48) with 64 bits, processor Intel(R) Xeon(R) CPU E5-2630 @ 2.3 GHz, 12 Gb of RAM and 8 cores.

The algorithm has also been tested on a variety of randomly generated instances by playing with different sets of profiles for the TSD risk averse measures. The dimensions of the compact representation (1) of RN DEM of the stochastic problems are given in Table 1. Its headings are as follows: *Instance*, member in the test bed we have experimented with; m, number of constraints; nx, number of 0-1 variables; ny, number of continuous variables; nel, number of nonzero coefficients in the constraint matrix; $dens$, constraint matrix density (in %); $|\Omega|$, number of scenarios; $|\mathcal{G}|$, number of scenario groups; and T, number of stages.

Table 1: RN (1) dimensions

Instance	m	nx	ny	nel	$dens$	$\|\Omega\|$	$\|\mathcal{G}\|$	T
P1	2114	453	1359	44040	1.15	113	151	4
P2	2002	429	1287	41861	1.22	80	143	5
P3	7072	1632	4896	295577	0.64	217	272	4
P4	9248	2176	6528	515768	0.64	217	272	4
P5	7072	1632	4896	295577	0.64	217	272	4
P6	12766	2946	8838	534563	0.35	340	491	5
P7	14400	3456	10368	1206875	0.60	182	288	5
P8	17380	3950	11850	606173	0.22	574	790	5
P9	24552	5580	16740	856121	0.16	844	1116	5

Let FSD (for soft first-order stochastic dominance constraints strategy) denote the particular case (9) of TSD strategy where the constraints related to the expected shortfall e-bound target in model (2) are deleted. On the other hand, let SSD (for soft second-order stochastic dominance constraints strategy) denote the particular case (10) of TSD strategy where the constraints related to the shortfall probability β-bound target in model (2) are deleted. Observe that FSD may have the drawback that it does not control the expected shortfall over the scenario groups, although the maximum shortfall S^p is imposed up to each scenario group g at stage t on reaching threshold ϕ^p for $p \in \mathcal{P}^t, t \in \tilde{\mathcal{T}}$. On the other hand, SSD may have the drawback that it does not control the fraction of scenario groups with shortfall on reaching the thresholds, although the maximum shortfall S^p is also imposed. So, TSD (2) avoids those potential drawbacks, but at a price of a higher computing effort (in time and memory). Since it could be possible that the FSD β bound and the SSD e bounds could not be simultaneously satisfied in TSD strategy, and since the bounds are only modeler-driven risk management policies (i.e., they are not physical requirements), they are also considered as soft constraints, heavily penalized in the objective function.

$$
\begin{aligned}
z_{FSD} = \quad & \max \sum_{g\in\mathcal{G}} w^g(a^g x^g + b^g y^g) - \sum_{t\in\tilde{\mathcal{T}}}\sum_{p\in\mathcal{P}^t} M^p_\beta \varepsilon^p_\beta \\
\text{s.t.} \quad & \sum_{q\in\mathcal{A}^g} A^q_g x^q + B^q_g y^q = h^g && \forall g\in\mathcal{G} \\
& \sum_{q\in\tilde{\mathcal{A}}^g} (a^q x^q + b^q y^q) + S^p \nu^{gp} \geq \phi^p && \forall g\in\mathcal{G}^t,\ p\in\mathcal{P}^t,\ t\in\tilde{\mathcal{T}} \\
& \sum_{g\in\mathcal{G}^t} w^g \nu^{gp} \leq \beta^p + \varepsilon^p_\beta && \forall p\in\mathcal{P}^t,\ t\in\tilde{\mathcal{T}} \\
& x^g \in \{0,1\}^{nx(g)}, y^g \in I\!R^{ny(g)} && \forall g\in\mathcal{G} \\
& \nu^{gp} \in \{0,1\} && \forall g\in\mathcal{G}^t,\ p\in\mathcal{P}^t,\ t\in\tilde{\mathcal{T}} \\
& 0 \leq \varepsilon^p_\beta \leq 1-\beta^p && \forall p\in\mathcal{P}^t,\ t\in\tilde{\mathcal{T}}.
\end{aligned}
\tag{9}
$$

$$
\begin{aligned}
z_{SSD} = \quad & \max \sum_{g\in\mathcal{G}} w^g(a^g x^g + b^g y^g) - \sum_{t\in\tilde{\mathcal{T}}}\sum_{p\in\mathcal{P}^t} M^p_e \varepsilon^p_e \\
\text{s.t.} \quad & \sum_{q\in\mathcal{A}^g} A^q_g x^q + B^q_g y^q = h^g && \forall g\in\mathcal{G} \\
& \sum_{q\in\tilde{\mathcal{A}}^g} (a^q x^q + b^q y^q) + s^{gp} \geq \phi^p && \forall g\in\mathcal{G}^t,\ p\in\mathcal{P}^t,\ t\in\tilde{\mathcal{T}} \\
& \sum_{g\in\mathcal{G}^t} w^g s^{gp} \leq e^p + \varepsilon^p_e && \forall p\in\mathcal{P}^t,\ t\in\tilde{\mathcal{T}} \\
& x^g \in \{0,1\}^{nx(g)}, y^g \in I\!R^{ny(g)} && \forall g\in\mathcal{G} \\
& 0 \leq s^{gp} \leq S^p && \forall g\in\mathcal{G}^t,\ p\in\mathcal{P}^t,\ t\in\tilde{\mathcal{T}} \\
& 0 \leq \varepsilon^p_e \leq S^p - e^p && \forall p\in\mathcal{P}^t,\ t\in\tilde{\mathcal{T}}.
\end{aligned}
\tag{10}
$$

The strategies to consider in the section are FSD (9), SSD (10), TSD (2) with a singleton $\tilde{\mathcal{T}} = \{T\}$ (risk averse for the last stage only) and TSD with a non-singleton $\tilde{\mathcal{T}}$ (that is, a subset of stages in set $\mathcal{T}$) besides the RN strategy.

The so-named "Constraint system in TSD model (2)" of models (7) and (8) to be solved at Step 7 of the BFC-TSD procedure should be accordingly replaced with the constraint system of models (9) and (10) whenever the particular cases FSD and SSD are considered, respectively.

Appendix A shows in Table 9 the dimensions of the original risk averse model (2) for the strategies FSD, SSD and TSD with a singleton $\tilde{\mathcal{T}}$ for a set of $|\mathcal{P}|$=4 profiles, and TSD with a non-singleton $\tilde{\mathcal{T}}$ for a set of $|\mathcal{P}|$=6 profiles, two profiles for stage $T-1$ and four profiles for stage T.

Section 5.1 reports the main results for the Risk Neutral (RN) strategy. Sections 5.2, 5.3 and 5.4 present the main results for the FSD, SSD and TSD strategies, respectively.

5.1 Computational results for the Risk Neutral strategy

The main computational results are presented in Table 2, which shows the LP relaxation and integer optimal solution values as well as the elapsed time for the BFC-TSD algorithm versus the plain use of CPLEX. Its headings are as follows: T, number of stages; C, number of scenario clusters in which the scenario tree has been divided into; z^{RN}_{LP}, solution value of the LP relaxation of RN model (1); z_0, expected solution value

obtained at the BF root node while solving independently the scenario cluster MIP submodels (3); n^n, number of explored twin nodes up to optimality while solving the original model (1); n^{TNF}, number of explored TNF integer sets; z^{RN}_{BFC} and z^{RN}_{CPX}, solution values of the RN model (1) obtained by using BFC-TSD algorithm and plain CPLEX, respectively; $OG\%$, CPLEX optimality gap, defined as $100\frac{z^{bn}-z^{RN}_{CPX}}{z^{RN}_{CPX}}$, where z^{bn} is the solution value of the active branch-and-bound node with the greatest solution value at the time instant at which CPLEX execution was interrupted; $GG\%$, goodness gap of BFC-TSD algorithm versus CPLEX, defined as $100\frac{z^{RN}_{CPX}-z^{RN}_{BFC}}{z^{RN}_{BFC}}$; and t^{RN}_{BFC} and t^{RN}_{CPX}, elapsed times (in secs) to obtain the related solutions of RN model by using BFC-TSD algorithm and plain CPLEX, respectively.

Table 2: BFC-TSD performance of RN (1)

Instance	T	C	z^{RN}_{LP}	z_0	n^n	n^{TNF}	z^{RN}_{BFC}	$z^{RN}_{CPX}(OG\%)$	$GG\%$	t^{RN}_{BFC}	t^{RN}_{CPX}
P1	4	8	157109	156324	1	0	156324	156324(*)	*	0	3
P2	5	6	6310	6068	1	0	6068	6068(*)	*	1	2
P3	4	10	293677	292118	8	3	292108	292108(0.07)	*	39	–
P4	4	10	285706	284151	17	7	283938	283938(0.27)	*	1372	–
P5	4	6	79957	78004	10	1	78004	78004(*)	*	4	8
P6	5	8	36303	35960	3	1	35960	35960(*)	*	30	1037
P7	5	7	270222	269447	3	1	269441	269441(*)	*	63	345
P8	5	9	155077	154813	1	0	154813	154795(0.04)	-0.01	16	–
P9	5	10	226377	225752	1	0	225752	225752(0.09)	*	48	–

(*): Optimality gap achieved ($< 0.01\%$)

*: Goodness gap achieved ($< 0.01\%$)

–: Out of memory (12Gb) or time limit (6h) exceeded

The break-stage that has been chosen is $t^* = 1$ and, so, according to the scenario tree structure, $C = |\mathcal{G}^2|$ scenario clusters $\Omega^c,\ \forall c \in \mathcal{C}$ are considered in the experiment. They are constructed by favoring the approach that shows higher scenario clustering for greater number of scenario groups in common. Notice the very strong upper bound z_0 and the very small elapsed time to obtain the optimal RN solution by using the decomposition BFC-TSD algorithm. Observe that the plain use of CPLEX cannot solve the instances P3, P4, P8 and P9 given the computer platform and the allowed elapsed time.

Risk results for RN and Wait-and-See (WS) solutions

A modeler familiar with the considered multistage stochastic TSD model (2) may not have high difficulty for providing the set of profiles for FSD, SSD and TSD risk averse strategies. However it is useful to firstly obtaining the RN solution value by solving model (1) and its negative semi-deviation (in case of a maximization) for confirming and, in case, reassessing the chosen set of profiles to realizing the potentiality of the feasible region of the TSD model (2). Moreover, based on our computational experience on dealing with this type of risk averse strategies for the multistage case, it is very advisable that the non-familiar modeler with the TSD strategies to work first with the solution values of the scenarios in RN (1) and WS models. In this way a

justified set of profiles could be chosen for the TSD strategy. Notice that a RN solution satisfies the NAC but the TSD constraints have been relaxed, and a WS solution value is the average of the solution values obtained independently for each scenario and therefore that solution does not necessarily satisfy the NAC, nor the TSD constraints.

Let $\hat{x}^g, \hat{y}^g$ denote the x, y vectors related to scenario group $g \in \mathcal{G}$ in the solution of RN model (1). Given a profile $p \in \mathcal{P}^T$ and a threshold on the objective function to be satisfied, ϕ^p, the set of scenarios, say, Ω^p_{RN} whose objective function value in that solution has a shortfall on reaching threshold ϕ^p can be expressed

$$\Omega^p_{RN} = \{\omega \in \Omega^g, g \in \mathcal{G}^T : \sum_{q \in \mathcal{A}^g} a^q \hat{x}^q + b^q \hat{y}^q < \phi^p\}.$$

So, let β^p_{RN} denote the probability of the set Ω^p_{RN}, i.e., $\beta^p_{RN} = \sum_{\omega \in \Omega^p_{RN}} w^\omega$, for $p \in \mathcal{P}^T$.

Similarly, let $\tilde{x}^g, \tilde{y}^g$ denote the x, y vectors related to scenario group $g \in \mathcal{G}$ in the solution for the WS approach. So, for profile $p \in \mathcal{P}^T$ the set of scenarios, say, Ω^p_{WS} whose objective function value in that solution has a shortfall on reaching threshold ϕ^p can be expressed

$$\Omega^p_{WS} = \{\omega \in \Omega^g, g \in \mathcal{G}^T : \sum_{q \in \mathcal{A}^g} a^q \tilde{x}^q + b^q \tilde{y}^q < \phi^p\}.$$

And, let β^p_{WS} denote the probability of set Ω^p_{WS}, i.e., $\beta^p_{WS} = \sum_{\omega \in \Omega^p_{WS}} w^\omega$, for $p \in \mathcal{P}^T$.

On the other hand, let e^p_{WS} and e^p_{RN} denote the expected shortfall on reaching threshold ϕ^p in the WS and RN solutions, for profile $p \in \mathcal{P}^T$, respectively.

The risk values β^p_{RN}, β^p_{WS}, and e^p_{RN}, e^p_{WS} reported from the solution of RN model (1) (where, by construction, the NAC (4)-(5) are fully satisfied) and WS separable models (i.e., model (1) where those NAC are fully relaxed) are shown in Table 3 and 5, respectively. They can be compared with the β^p- and e^p-upper bounds imposed in risk averse strategies FSD and SSD.

5.2 Computational results for risk averse strategy FSD

5.2.1 Singleton profile sets

The RN optimal solution obviously satisfies FSD model (9) for any choice of the upper bound β^p such that $\beta^p \geq \beta^p_{RN}$. Then, we are interested on values of $\beta^p < \beta^p_{RN}$ for $p \in \mathcal{P}^T$.

Remark 1. For FSD maximization problems, the following inequality is satisfied: $\beta^p_{WS} \leq \beta^p_{RN}$, $p \in \mathcal{P}$. So, it results that $\beta^p_{WS} \leq \beta^p \leq \beta^p_{RN}$ for each $p \in \mathcal{P}^T$.

From Remark 1, it can be easily shown that model (9) is infeasible for $\beta^p < \beta^p_{WS}$ for the pair (ϕ^p, β^p). Then, let $\underline{\beta}^p$ denote the smallest value of β^p for which model is feasible for the pair (ϕ^p, β^p).

Remark 2. For each profile $p \in \mathcal{P}^T$ and given a threshold ϕ^p, the interval of feasible values for the upper bound β^p of the fraction of scenarios with shortfall is given by $[\underline{\beta}^p, 1]$, where $\beta^p_{WS} \leq \underline{\beta}^p \leq \beta^p_{RN}$.

So, the scheme consists of solving model (9) for each $p \in \mathcal{P}^T$, with $\beta^p = \beta^p_{WS}$, and obtain the value of ε^p_β such that $\underline{\beta}^p$ will be equal to $\beta^p_{WS} + \varepsilon^p_\beta$.

For RN (1) and WS models, Table 3 shows the $\underline{\beta}^p-$ computed values, i.e., the fraction of scenarios whose optimal objective function values do not reach the modeler driven threshold ϕ^p for the last stage T in the time horizon. The threshold is computed as $\phi^p = \delta^p z_{RN}$, where δ^p is a modeler-driven multiplicative factor of the solution value z_{RN} reported in Table 2 for RN model (1).

Table 3: FSD risk averse strategy with $\tilde{\mathcal{T}} = \{T\}$. Feasible bounds for singleton profile sets

Instance	p	δ^p	ϕ^p	$\beta^p_{WS}(\lvert\Omega^p_{WS}\rvert)$	$\underline{\beta}^p(\lvert\Omega^p\rvert)$	$\beta^p_{RN}(\lvert\Omega^p_{RN}\rvert)$
P1	1	1.1	171956.4	0.708(80)	0.708(80)	0.726(82)
$\lvert\Omega\rvert = 113$	2	1.0	156324.0	0.434(49)	0.443(50)	0.487(55)
	3	0.9	140691.6	0.231(26)	0.231(26)	0.248(28)
	4	0.8	125059.2	0.116(13)	0.116(13)	0.124(14)
P2	1	1.1	6674.26	0.538(43)	0.563(45)	0.575(46)
$\lvert\Omega\rvert = 80$	2	1.0	6067.51	0.313(25)	0.4(31)	0.4(32)
	3	0.9	5460.75	0.213(17)	0.25(20)	0.3(24)
	4	0.8	4854.00	0.163(13)	0.175(14)	0.188(15)
P3	1	0.95	277502.6	0.309(67)	0.323(70)	0.328(71)
$\lvert\Omega\rvert = 217$	2	0.9	262897.2	0.176(38)	0.185(40)	0.194(42)
	3	0.85	248292.8	0.056(12)	0.06(13)	0.065(14)
	4	0.8	233686.4	0.01(2)	0.019(4)	0.019(4)
P4	1	1.0	283938.0	0.48(104)	0.498(108)	0.535(116)
$\lvert\Omega\rvert = 217$	2	0.95	269741.1	0.249(54)	0.272(59)	0.282(61)
	3	0.9	255544.2	0.116(25)	0.143(31)	0.148(32)
	4	0.8	227150.4	0(0)	0.005(1)	0.01(2)
P5	1	0.95	74104	0.194(42)	0.365(79)	0.420(91)
$\lvert\Omega\rvert = 217$	2	0.9	70203.8	0.116(25)	0.189(41)	0.231(50)
	3	0.85	66303.6	0.047(10)	0.07(15)	0.093(20)
	4	0.8	62403.4	0.01(2)	0.014(3)	0.042(9)
P6	1	0.95	34161.8	0.656(223)	0.656(223)	0.674(229)
$\lvert\Omega\rvert = 340$	2	0.9	32363.8	0.618(210)	0.618(210)	0.633(215)
	3	0.85	30565.8	0.583(198)	0.583(198)	0.592(201)
	4	0.8	28767.8	0.571(194)	0.571(194)	0.574(195)
P7	1	0.95	255969	0.204(37)	0.215(39)	0.275(50)
$\lvert\Omega\rvert = 182$	2	0.9	242497	0.066(12)	0.072(13)	0.121(22)
	3	0.85	229025	0(0)	0(0)	0.033(6)
	4	0.8	215553	0(0)	0(0)	0(0)
P8	1	0.95	147072	0.345(198)	0.356(204)	0.384(220)
$\lvert\Omega\rvert = 574$	2	0.9	139332	0.223(128)	0.234(134)	0.248(142)
	3	0.85	131591	0.128(73)	0.131(75)	0.152(87)
	4	0.8	123850	0.068(39)	0.072(41)	0.082(47)
P9	1	0.95	214464	0.368(310)	0.376(317)	0.387(326)
$\lvert\Omega\rvert = 844$	2	0.9	203177	0.266(224)	0.269(227)	0.280(236)
	3	0.85	191889	0.149(125)	0.157(132)	0.170(143)
	4	0.8	180602	0.067(56)	0.07(59)	0.086(72)

The headings of Table 3 are as follows: p, profile; δ^p, multiplicative factor; threshold ϕ^p; $\beta^p_{WS}(\lvert\Omega^p_{WS}\rvert)$ and $\beta^p_{RN}(\lvert\Omega^p_{RN}\rvert)$), probabilities of the sets Ω^p_{WS} and Ω^p_{RN}, where those sets have been defined in Section 5.1, such that they denote the subset of scenarios whose objective function values in the WS and RN optimal solutions have a shortfall on reaching threshold ϕ^p, respectively; and $\underline{\beta}^p(\lvert\Omega^p\rvert)$, the smallest bound for which model

FSD is feasible for profile p.

Observe in Table 3 the above β-bounds and the risk averse results for $\tilde{\mathcal{T}} = \{T\}$, i.e., last stage, in the test bed that we have experimented with. Notice in Table 3 the goodness quality of the solutions obtained for singleton profile sets by FSD strategy, since the values of $\underline{\beta}^p$ are strongly decreasing while the values of δ^p are decreasing at a substantial smaller path through the profiles in $\mathcal{P}^T$. For example, it can be guaranteed that there is not more than 23.1% ($\underline{\beta}^3$) and 11.6% ($\underline{\beta}^4$) of the scenarios in instance P1 whose solution values are smaller than the thresholds whose values are 90% and 80% of the RN solution value z_{RN} (see Table 3), respectively.

5.2.2 Non-Singleton profile sets

It is up to the modeler the decision on risk reduction with one or more profiles and, then, the modeler after analizing the variability of the objective function values for the scenarios of the RN solution, may decide on the aceptable risk for one or more thresholds unsatisfaction quantified by the upper bounds on the scenarios with failure on reaching the threshold and the related expected shortfall.

The BFC-TSD algorithm considers the availability of different profiles without incrementing the computational effort due to the replacement of cross scenario constraints (singleton $\tilde{\mathcal{T}}$) with cross scenario group constraints (non-singleton $\tilde{\mathcal{T}}$).

Table 4 reports our computational experience for a non-singleton profile set in the FSD risk averse strategy taken from the singleton profiles presented in Table 3, but considering them together in the same FSD model (9). The new headings of the table are as follows: $\beta^p(|\Omega^p|)$, upper bound on the probability of the scenarios with shortfall (and, then, at the last stage T); t^{FSD}_{BFC} and t^{FSD}_{CPX}, elapsed time (secs) to obtain the solution value z^{FSD}_{BFC} by the BFC-TSD algorithm and z^{FSD}_{CPX} by plain use of CPLEX, respectively; $OG\%$, CPLEX optimality gap, defined as $100\frac{z^{bn}-z^{FSD}_{CPX}}{z^{FSD}_{CPX}}$, where z^{bn} is as above; and $GG\%$, goodness gap of BFC-TSD algorithm versus CPLEX, defined as $100\frac{z^{FSD}_{CPX}-z^{FSD}_{BFC}}{z^{FSD}_{BFC}}$.

Observe in Table 4 that the feasibility of the model may require that the values of β^p change slightly, $\forall p \in \mathcal{P}^T$. Notice the increase in elapsed time when considering non-singleton profiles versus the RN scheme (see Table 2), due to the increase in model dimensions and its higher tightening. On the other side, we can notice that the number of scenarios $|\Omega^p_{RN}|$ with a shortfall on reaching threshold ϕ^p (and, then, the probability β^p_{RN} of the set of scenarios in Ω whose profit do not reach the threshold) reported in Table 3 for RN strategy has been reduced in FSD strategy (see $|\Omega^p|$ in Table 4). Notice that β^p is smaller than β^p_{RN}. It is is obtained at a price of a reduction of the estimated profit (i.e., solution value) and the increase on the elapsed time. It is worth pointing out that the estimated profit shown in Table 4 have not a high reduction when comparing it with the profit shown in Table 2 and, otherwise, the elapsed time required by BFC-TSD algorithm shown in Table 4 is still affordable even for the difficult instances P3, P4, P8 and P9 (while the plain use of CPLEX cannot still provide a solution). As an example, let us consider instance P4 whose elapsed time for the FSD model (9) is the same one (1372 secs) as for the RN model (1) (see Table 2), and the number of scenarios for RN whose profit does not reach e.g. threshold $\phi^1 = 283938$ is $|\Omega^1_{RN}|$=116 (see Table 3) while this number is

Table 4: FSD risk averse strategy with $\tilde{\mathcal{T}} = \{T\}$. Computational results for a non-singleton profile set

Inst.	$\{\ \beta^p(\lvert\Omega^p\rvert),\quad p=\ 1,\ 2,\ 3,\ 4\ \}$	n^n	n^{TNF}	z^{FSD}_{BFC}	$z^{FSD}_{CPX}(OG\%)$	$GG\%$	t^{FSD}_{BFC}	t^{FSD}_{CPX}
P1	{0.726(82), 0.461(52), 0.248(28), 0.116(13)}	3	2	156210	156280(*)	0.04	4	15
P2	{0.575(46), 0.4(32), 0.3(22), 0.175(14)}	3	2	6049	6049(*)	*	5	37
P3	{0.328(71), 0.185(40), 0.06(13), 0.019(4)}	12	4	292002	292034(0.12)	0.01	77	–
P4	{0.507(112), 0.277(60), 0.143(31), 0.005(1)}	17	7	283902	283902(0.37)	*	1372	–
P5	{0.406(88), 0.203(44), 0.079(17), 0.037(8)}	3	1	77961	77965(*)	*	18	373
P6	{0.662(225), 0.627(213), 0.586(199), 0.571(194)}	3	2	35951	35953(*)	*	292	12798
P7	{0.248(45), 0.094(17), 0.017(3), 0(0)}	3	1	269396	269406(*)	*	109	2229
P8	{0.366(210), 0.241(138), 0.136(78), 0.077(44)}	3	2	154655	154690(0.13)	0.02	357	–
P9	{0.381(321), 0.274(231), 0.163(137), 0.079(66)}	3	2	225746	225764(0.09)	*	531	–

(*): Optimality gap achieved ($< 0.01\%$)

*: Goodness gap achieved ($< 0.01\%$)

–: Out of memory (12Gb) or time limit (6h) exceeded

$|\Omega^1|$=112 for FSD satisfying the β^1-bound 50.7% (while β^1_{RN} is 53.5%). Finally, observe that the goodness gap $GG\%$ of the FSD solution given by BFC-TSD algorithm versus CPLEX is very small, in practically 50% of the instances CPLEX is running out of memory or the time limit (6h) is exceeded, and on the contrary BFC-TSD always obtains a solution value in all instances that is similar to the CPLEX one, being the elapsed time very small.

5.3 Computational results for risk averse strategy SSD

5.3.1 Singleton profile sets

The Remarks 1 and 2 for FSD maximization (Section 5.2.1) also apply in a similar way for SSD maximization. Table 5 shows the $e-$ computed values for RN (1) and WS models, i.e., the expected shortfalls of the optimal objective function values on reaching the modeler driven threshold ϕ^p for the last stage T in the time horizon. The threshold is computed as $\phi^p = \delta^p z_{RN}$, where δ^p is a modeler-driven multiplicative factor of the solution value z_{RN} reported in Table 2 for RN model (1).

Table 5 shows the expected shortfall bounds in SSD model (10) for $\tilde{\mathcal{T}} = \{T\}$ for the singleton profile set in the test bed we have experimented with. The headings of Table 5 are as follows: p, profile; δ^p, multiplicative factor; threshold ϕ^p; e^p_{WS} and e^p_{RN}, expected shortfalls for RN and WS models; and $\underline{e}^p$, i.e., the smallest expected shortfall upper bound for which SSD model is feasible for profile p.

Observe in Table 5 the goodness quality of the solution obtained for the SSD strategy, since the value of $\underline{e}^p$ is strongly decreasing while the value of δ^p is decreasing at a smaller path through the profiles in set $\mathcal{P}^T$. As in the case of the β-bounds, for obtaining the value $\underline{e}^p$ for each $p \in \mathcal{P}^T$, SSD model (10) has been solved with $e^p = e^p_{WS}$, and the value $\underline{e}^p$ is the sum $e^p_{WS} + \varepsilon^p_e$.

Table 5: SSD risk averse strategy with $\tilde{\mathcal{T}} = \{T\}$. Feasible bounds for singleton profile sets

Instance	p	δ^p	ϕ^p	e^p_{WS}	$\underline{e}^p$	e^p_{RN}
P1	1	1.1	171956.4	18454.5	19552	19573.5
$\|\Omega\| = 113$	2	1.0	156324.0	9345.9	10082	10191.3
	3	0.9	140691.6	4201.1	4568	4742.8
	4	0.8	125059.2	1653.9	1697	1904.9
P2	1	1.1	6674.3	1948.5	2094	2099.6
$\|\Omega\| = 80$	2	1.0	6067.5	1682.7	1788	1801.1
	3	0.9	5460.7	1523.1	1573	1581.6
	4	0.8	4854.0	1408.5	1432	1434.1
P3	1	0.95	277502.6	5594.2	6178	6277.6
$\|\Omega\| = 217$	2	0.9	262897.2	2122.4	2399	2545.3
	3	0.85	248292.8	487.1	535	598.2
	4	0.8	233686.4	69.9	112	114.5
P4	1	1.0	283938	8630.8	9930	9979.1
$\|\Omega\| = 217$	2	0.95	269741.1	3586.3	4300	4457.1
	3	0.9	255544.2	1041.3	1415	1574.1
	4	0.8	227150.4	0	0	21.9
P5	1	0.95	74104	1023.4	1973	2237.8
$\|\Omega\| = 217$	2	0.9	70203.8	427.4	757	973.8
	3	0.85	66303.6	111.3	218	408
	4	0.8	62403.4	26.4	36	185.2
P6	1	0.95	34161.8	7842.4	8184	8201.1
$\|\Omega\| = 340$	2	0.9	32363.8	6696.8	7004	7029.9
	3	0.85	30565.8	5622.8	5891	5927.6
	4	0.8	28767.8	4588.7	4832	4877.1
P7	1	0.95	255969	2159.1	3005	3498.9
$\|\Omega\| = 182$	2	0.9	242497	348.6	503	886.9
	3	0.85	229025	0	0	56.6
	4	0.8	215553	0	0	0
P8	1	0.95	147072	4894.5	5394	5597.4
$\|\Omega\| = 574$	2	0.9	139332	2741.4	3007	3235.1
	3	0.85	131591	1409.5	1148	1753.4
	4	0.8	123850	684.3	750	912.6
P9	1	0.95	214464	8225.7	8978	9177.8
$\|\Omega\| = 844$	2	0.9	203177	4706.5	5242	5434.7
	3	0.85	191889	2352.8	2656	2868.7
	4	0.8	180602	1107.1	1258	1404.2

5.3.2 Non-Singleton profile sets

Table 6 reports our computational experience for a non-singleton profile set in the SSD strategy taken from the singleton profiles presented in Table 5, but considering them simultaneously in the same SSD model. The new headings of the table are as follows: e^p, upper bound on the expected shortfall on reaching threshold ϕ^p over the scenarios (and, then, at last stage $t = T$); t^{SSD}_{BFC} and t^{SSD}_{CPX}, elapsed time (secs) to obtain the solution value z^{SSD}_{BFC} by BFC-TSD algorithm and z^{SSD}_{CPX} by plain use of CPLEX, respectively; $OG\%$, CPLEX optimality gap, defined as $100\frac{z^{bn} - z^{SSD}_{CPX}}{z^{SSD}_{CPX}}$, where z^{bn} is as above; and $GG\%$, goodness gap of BFC-TSD algorithm versus

CPLEX, defined as $100\frac{z^{SSD}_{CPX}-z^{SSD}_{BFC}}{z^{SSD}_{BFC}}$.

Table 6: SSD risk averse strategy with $\tilde{\mathcal{T}}=\{T\}$. Computational results for a non-singleton profile set

Inst.	$\{\ e^p,\quad p=\ 1,\ 2,\ 3,\ 4\ \}$	n^n	n^{TNF}	z^{SSD}_{BFC}	$z^{SSD}_{CPX}(OG\%)$	$GG\%$	t^{SSD}_{BFC}	t^{SSD}_{CPX}
P1	{19560, 10170, 4720, 1900}	3	2	156304	156304(*)	*	2	5
P2	{2097, 1790, 1576, 1434}	3	2	6059	6059(*)	*	3	3
P3	{6250, 2520, 560, 113}	25	6	291790	291850(0.09)	0.02	118	–
P4	{9955, 4400, 1490, 5}	17	8	283802	283867(0.34)	0.02	1372	–
P5	{2137, 873, 330, 134}	3	1	77945	77968(*)	0.03	9	10
P6	{8193, 7018, 5914, 4860}	3	2	35957	35957(*)	*	32	1320
P7	{3240, 650, 35, 0}	3	1	269275	269291(*)	*	63	1010
P8	{5476, 3110, 1626, 885}	3	2	154549	154752(0.04)	0.13	49	–
P9	{9110, 5340, 2758, 1320}	3	2	225769	225774(0.07)	*	162	–

(*): Optimality gap achieved ($< 0.01\%$)

*: Goodness gap achieved ($< 0.01\%$)

(-): Incumbent solution value at the time limit (6h)

Observe in Table 6 that while considering a set of profiles simultaneously, the feasibility of the whole model may require that the values of $e^p\,\forall p \in \mathcal{P}^T$ should be increased, as it happens with β^p for the FSD strategy (see Table 4), although for SSD the changes are higher. On the other hand, the performance of BFC-TSD algorithm versus CPLEX in solution goodness as well as in elapsed time has the same pattern as for FSD strategy such that *GG%* is very small, CPLEX requires an unaffordable elapsed time and, on the contrary, the elapsed time required by BFC-TSD is very small.

5.4 Computational results for risk averse strategy TSD. Non-singleton profile set

Tables 7 and 8 report our computational experience for the TSD model (2) for $\tilde{\mathcal{T}}=\{T\}$ and $\tilde{\mathcal{T}}=\{T-1,T\}$, respectively, in a non-singleton profile environment set for the test bed considered in the previous sections. The new headings of the tables are as follows: z^{TSD}_{BFC} and t^{TSD}_{CPX}, solution value and elapsed time (secs) by using BFC-TSD algorithm; z^{TSD}_{CPX} and t^{TSD}_{CPX}, solution value and elapsed time (secs) by plain use of CPLEX, respectively; *OG%*, CPLEX optimality gap, defined as $100\frac{z^{bn}-z^{TSD}_{CPX}}{z^{TSD}_{CPX}}$, where z^{bn} is as above; and *GG%*, goodness gap of BFC-TSD algorithm versus CPLEX, defined as $100\frac{z^{TSD}_{CPX}-z^{TSD}_{BFC}}{z^{TSD}_{BFC}}$.

The β^p- and e^p-upper bounds shown in the tables are not the bounds shown in Tables 4 and 6 for FSD and SSD, respectively, since TSD (2) is infeasible taken them together. So, while simultaneously considering a set of β- and e-upper bounds, the feasibility of the whole model may require that the values of $e^p\,\forall p \in \mathcal{P}^T$ are higher than the values for SSD strategy (see Table 6), but the values of $\beta^p\,\forall p \in \mathcal{P}^T$ only change slightly with respect to the values for FSD (see Table 4). In any case, BFC-TSD algorithm provides the solution in very small elapsed time for all instances even for the very difficult ones P3, P4, P8 and P9. On the contrary, CPLEX is running out of memory (12Gb) or time limit (6h) exceeded in all instances but P1, P5 and P7 in Table 7 and P1, P2 and P5 in Table 8.

The profile set used for the results reported in Table 7 continues only considering the last stage, so $\tilde{\mathcal{T}} = \{T\}$. However the table also shows, as an illustrative example, a modeler-driven non-wanted profit threshold ϕ^p at stage $T-1$ for $p = 1,2$ in case that the TSD solution is accepted. The table also shows (in boldface symbols) the columns $\boldsymbol{\beta}^{\mathbf{p}}_{\mathbf{TSD}}(|\boldsymbol{\Omega}^{\mathbf{p}}_{\mathbf{TSD}}|)$ and $\mathbf{e}^{\mathbf{p}}_{\mathbf{TSD}}$ for the probability of scenarios (number of scenarios) in set Ω with shortfall on reaching profit threshold ϕ^p $\forall p \in \mathcal{P}^{T-1} = \{1,2\}$ and the expected shortfall, respectively. Those values *have been computed from the TSD solution* with $\tilde{\mathcal{T}} = \{T\}$ for the profit obtained up to stage $T-1$, say $z_{T-1} = \sum_{g \in \mathcal{G}^{T-1}} w^g \sum_{q \in \tilde{\mathcal{A}}^g} (a^q x^q + b^q y^q)$.

Table 8 reports the main results for TSD model (2) with $\tilde{\mathcal{T}} = \{T-1, T\}$, where $|\mathcal{P}^T| = 4$ and $|\mathcal{P}^{T-1}| = 2$. The β^p- and e^p-upper bounds are shown in the first two columns for set $\mathcal{P}^T$ and the third and fourth columns for set $\mathcal{P}^{T-1}$.

Observe that the profile set $\mathcal{P}^T$ in Table 8 is as in Table 7. However, profile set $\mathcal{P}^{T-1}$ reduces the risk of non-wanted scenarios, so that TSD is considered, since $\beta^p < \boldsymbol{\beta}^{\mathbf{p}}$ and $e^p < \mathbf{e}^{\mathbf{p}}$ for the profit threshold ϕ^p shown in Table 7 for all $p \in \mathcal{P}^{T-1}$.

The values of the slack variables ε^p_e and ε^p_β for all $p \in \mathcal{P}^t, t \in \tilde{\mathcal{T}}$, where $\tilde{\mathcal{T}} = \{T\}$ in Table 7 and $\tilde{\mathcal{T}} = \{T-1, T\}$ in Table 8, have been zero in the reported solutions. Additionally, the solution value (i.e., expected profit) z^{TSD}_{BFC} for TSD when $\tilde{\mathcal{T}} = \{T, T-1\}$ (Table 8) is slightly worse than the profit for TSD for $\tilde{\mathcal{T}} = \{T\}$ (Table 7). On the other hand, the results shown in both tables have the same pattern observed in the other tables (for the FSD and SSD strategies). Notice that the goodness gap *GG%* of BFC-TSD algorithm versus CPLEX is very small, CPLEX has difficulties on solving the instances in the test bed and, on the contrary, BFC-TSD algorithm requires a very small elapsed time.

Finally, Appendix B presents the risk reduction trajectory from WS, through RN, FSD, SSD and TSD, with singleton and non-singleton $\tilde{\mathcal{T}}$, by using instance P7 as an illustrative example. In what follows the same instance P7 is used for clarifying the results shown in Tables 7 and 8. Observe that the optimal solution of RN model (1) gives the expected profit z_{RN}=269441 whose number of scenarios with shortfall on reaching the modeler-driven thresholds $\phi^p = \delta^p \times z_{RN}$, for $p \in \mathcal{P}^T$, where ϕ^1=255969, ϕ^2=242497, ϕ^3=229025 and ϕ^4=215553 (in decreasing order) are $|\Omega^1_{RN}|$=50, $|\Omega^2_{RN}|$=22, $|\Omega^3_{RN}|$=6 and $|\Omega^4_{RN}|$=0 (in decreasing order as well) out of 128 scenarios, respectively (see Table 3), and the related expected shortfall is e^1_{RN}=3498.9, e^2_{RN}=886.9, e^3_{RN}=56.6 and e^4_{RN}=0.0, respectively (see Table 5).

For illustrative purposes, let me assume that the risk neutral picture is carrying out an excessive profit risk, and consider that it has been decided that the upper bound on the number of scenarios with shortfall (i.e., at the last stage $T = 5$) has been reduced from 50 to $|\Omega^1|$=45, from 22 to $|\Omega^2|$=17, from 6 to $|\Omega^3|$=3 and the zero scenario policy is kept for profile 4 (i.e., $|\Omega^4|$=0). Additionally, assume that the upper bound on the expected shortfall has been reduced from 3498.9 to e^1=3240.0, from 886.9 to e^2=650.0, from 56.6 to e^3=35.0 and the zero shortfall policy is obviously kept for profile 4 (i.e., e^4=0.0). The optimal solution of TSD model (2) is shown in Table 7. Let us consider the expected profit z^{TSD}_{CPX}=269273 obtained by plain use of CPLEX (observe that the BFC-TSD algorithm gives a close value, z^{TSD}_{BFC}=269247 whose goodness gap is smaller than 0.01%). Observe that the expected profit of RN model (1), z_{RN}=269441 has only been slightly reduced by

Table 7: TSD risk averse strategy with $\tilde{\mathcal{T}} = \{T\}$. Computational results for a non-singleton profile set

Inst.	$p \in \mathcal{P}^T = \{1,2,3,4\}$		$p \in \mathcal{P}^{T-1} = \{1,2\}$		n^n	n^{TNF}	z^{TSD}_{BFC}	$z^{TSD}_{CPX}(OG\%)$	GG%	t^{TSD}_{BFC}	t^{TSD}_{CPX}
	$\beta^p(\lvert\Omega^p\rvert)$	e^p	$\boldsymbol{\beta}^{\mathbf{p}}(\lvert\boldsymbol{\Omega}^{\mathbf{p}}\rvert)$	$\mathbf{e}^{\mathbf{p}}/\phi^p$							
P1	0.726(82)	19570	**0.646(73)**	**5601**	3	2	156269	156269 (*)	*	3	52
	0.478(54)	10170	**0.407 (46)**	**1363**							
	0.248(28)	4710		ϕ^1 =15632.4							
	0.115(13)	1870		ϕ^2 =7816.2							
P2	0.575(46)	2097	**0.500(40)**	**1747**	3	2	6052	6052 (0.31%)	*	5	–
	0.400(32)	1797	**0.375(30)**	**1474**							
	0.275(22)	1579		ϕ^1 = 4247.26							
	0.188(15)	1434		ϕ^2 =3640.51							
P3	0.327(71)	6270	**0.382(83)**	**1770**	12	3	292048	292049 (0.18%)	*	67	–
	0.189(41)	2540	**0.083(18)**	**110**							
	0.060(13)	590		ϕ^1 =11684.3							
	0.018(4)	113		ϕ^2 =5842.16							
P4	0.526(114)	9957	**0.903(196)**	**13201**	17	8	283875	283878 (0.31%)	*	1432	–
	0.277(60)	4430	**0.516(112)**	**2438**							
	0.148(32)	1560		ϕ^1 = 28393.8							
	0.005(1)	21		ϕ^2 =14196.9							
P5	0.410(89)	2100	**0.737(160)**	**3847**	3	1	77882	77956 (*)	0.09	13	108
	0.212(46)	900	**0.424(92)**	**1659**							
	0.083(18)	390		ϕ^1 =50702.7							
	0.032(7)	160		ϕ^2 =46802.5							
P6	0.665(226)	8193	**0.621(211)**	**3185**	3	2	35938	35938 (0.11%)	*	178	–
	0.626(213)	7018	**0.409(139)**	**1248**							
	0.591(201)	5914		ϕ^1 = 14383.9							
	0.571(194)	4860		ϕ^2 = 10787.9							
P7	0.248(45)	3240	**0.495(90)**	**5839**	3	1	269247	269273 (*)	*	77	3135
	0.094(17)	650	**0.181(33)**	**1063**							
	0.017(3)	35		ϕ^1 = 67360.2							
	0(0)	0		ϕ^2 = 53888.2							
P8	0.378(217)	5479	**0.131(75)**	**215**	3	2	154557	154748 (0.06%)	0.12	272	–
	0.240(138)	3110	**0.016(9)**	**31**							
	0.146(84)	1627		ϕ^1 = 6192.52							
	0.077(44)	885		ϕ^2 = 3096.26							
P9	0.381(321)	9110	**0.254(214)**	**862**	3	2	225741	225712 (0.12%)	-0.01	630	–
	0.274(231)	5340	**0.070(59)**	**147**							
	0.163(137)	2758		ϕ^1 = 9030.08							
	0.079(66)	1320		ϕ^2 = 4515.04							

(*): Optimality gap achieved ($< 0.01\%$)

*: Goodness gap achieved ($< 0.01\%$)

–: Out of memory (12Gb) or time limit (6h) exceeded

the model TSD (2) and, on the other hand, the profit risk of non-wanted scenarios has been reduced to the modeler-driven bounds.

However, let us assume that the above risk averse TSD picture is also carrying out an excessive profit risk on the number of scenarios with shortfall and expected shortfall on reaching modeler-driven thresholds on an intermediate stage, say $T-1=4$, being the thresholds ϕ^1 =67360.2 for its profile $p=1$ and ϕ^2 =53888.2 for

Table 8: TSD risk averse strategy with $\tilde{\mathcal{T}} = \{T-1, T\}$. Computational results for a non-singleton profile set

Inst.	$p \in \mathcal{P}^T = \{1,2,3,4\}$		$p \in \mathcal{P}^{T-1} = \{1,2\}$		n^n	n^{TNF}	z^{TSD}_{BFC}	$z^{TSD}_{CPX}(OG\%)$	GG%	t^{TSD}_{BFC}	t^{TSD}_{CPX}
	$\beta^p(\lvert\Omega^p\rvert)$	e^p	$\beta^p(\lvert\Omega^p\rvert)$	e^p							
P1	0.726(82)	19570	0.619(70)	5550	3	2	156240	156242 (*)	*	4	21
	0.478(54)	10170	0.363(41)	1300							
	0.248(28)	4710									
	0.115(13)	1870									
P2	0.575(46)	2097	0.463(37)	1690	3	2	6020	6020 (*)	*	5	362
	0.400(32)	1797	0.350(28)	1450							
	0.275(22)	1579									
	0.188(15)	1434									
P3	0.327(71)	6270	0.359(78)	1586	12	5	291881	291969 (0.20%)	0.03	102	–
	0.189(41)	2540	0.046(10)	73							
	0.060(13)	590									
	0.018(4)	113									
P4	0.526(114)	9957	0.894(194)	12600	17	9	283253	283619 (0.32%)	0.13	1466	–
	0.277(60)	4430	0.475(103)	1889							
	0.148(32)	1560									
	0.005(1)	21									
P5	0.410(89)	2100	0.737(160)	3846	3	2	77799	77955 (*)	0.20	29	75
	0.212(46)	900	0.424(92)	1658							
	0.083(18)	390									
	0.032(7)	160									
P6	0.665(226)	8193	0.591(201)	2900	3	2	35882	35891 (0.08%)	0.03	338	–
	0.626(213)	7018	0.379(129)	1025							
	0.591(201)	5914									
	0.571(194)	4860									
P7	0.248(45)	3240	0.440(80)	5400	3	1	269204	269222 (0.01%)	*	225	–
	0.094(17)	650	0.143(26)	800							
	0.017(3)	35									
	0(0)	0									
P8	0.378(217)	5479	0.111(64)	181	3	2	154527	154729 (0.07%)	0.13	291	–
	0.240(138)	3110	0.009(5)	11							
	0.146(84)	1627									
	0.077(44)	885									
P9	0.381(321)	9110	0.243(205)	755	3	2	225712	225733 (0.11%)	*	1208	–
	0.274(231)	5340	0.058(49)	98							
	0.163(137)	2758									
	0.079(66)	1320									

(*): Optimality gap achieved ($< 0.01\%$)

*: Goodness gap achieved ($< 0.01\%$)

–: Out of memory (12Gb) or time limit (6h) exceeded

its $p = 2$, the number of scenarios with shortfall are $|\mathbf{\Omega^1}| = \mathbf{90}$ and $|\mathbf{\Omega^2}| = \mathbf{33}$, and the expected shortfall is $\mathbf{e^1}$ =**5839** and $\mathbf{e^2}$ =**1063** (see Table 7). In order to reduce the non-admissible profit risk up to stage $T-1=4$ within the feasible region restricted by the risk bounds already imposed for stage $T=5$, let us also assume that it has been decided that the upper bound on the number of scenarios with shortfall for stage $T-1=4$ has been reduced from 90 to $|\Omega^1|$=80 and from 33 to $|\Omega^2|$=26, and the upper bound on the expected shortfall

has been reduced from 5839 to e^1=5400 and from 1063 to e^2=800. The optimal solution of TSD model (2) is obtained where the profiles above are set up in the sets $\mathcal{P}^t$ for $t \in \tilde{\mathcal{T}} = \{4,5\}$, see the results in Table 8. Let us consider the solution value z_{CPX}^{TSD}=269222 obtained by plain use of CPLEX (observe that the BFC-TSD algorithm gives a close value, z_{BFC}^{TSD}=269204 whose goodness gap is also smaller than 0.01%). Observe also that the expected profit of TSD model (2) with $\tilde{\mathcal{T}} = \{5\}$, 269273, has only been slightly reduced by TSD model (2) with $\tilde{\mathcal{T}} = \{4,5\}$ and, on the other hand, the profit risk of non-wanted scenarios up to stage $T-1=4$ has been reduced to the modeler-driven bounds without modifying the also modeler-driven profit risk bounds for last stage $T=5$.

6 Conclusions

In this work we have presented an extension of the Branch-and-Fix Coordination BFC algorithm that we have proposed elsewhere [18] for solving multistage mixed 0-1 models where the uncertainty appears anywhere and is represented via nonsymmetric scenario trees. The extension allows to consider some risk averse strategies, namely the first- and second-order stochastic dominance constraints (SDC) measures induced by mixed integer-linear recourse, say FSD and SSD, respectively. Those strategies have been recently proposed in [24, 25] for dealing with the two-stage problem and their implementation requires cross scenario constraints. Additionally, our multistage model (2) considers the Time Stochastic Dominance policy requirements for those measures, so that a new SDC strategy has been introduced, so named TSD, by simultaneously considering bound targets on the probability of failure and the expected shortfall over the scenarios on reaching the modeler-driven thresholds.

The treatment of the related FSD and SSD constraint systems has been performed in a similar way as it is done for the nonanticipativity constraints (NAC) (4)-(5) for the stages up to the so-named break stage t^* introduced in [18], where they are algoritmically satisfied. That is, relaxing those risk averse constraint systems in the original model (2) and solving the scenario cluster submodels (3) in a coordinated way for the so-named candidate Twin Node Families (TNF) along the execution of the algorithm. As a result, a so-named integer TNF is obtained such that the 0-1 problem variables are fixed to their 0-1 values. At that point, model (7) is solved for satisfying the NAC (5) for the continuous variables as well as satisfying the cross scenario group constraint systems for the 0-1 x-solution from submodels (3). Finally, for guaranteeing optimality a modeler-driven label allows to solve model (8). As a result, the extension of the BFC algorithm, so named BFC-TSD, paves the way for considering other risk averse strategies, see [2]. In any case, the TSD strategies are much more computationally complicated than the other ones.

The TSD strategy (2) for multistage stochastic mixed integer recourse problems (considering singleton as well as non-singleton sets $\tilde{\mathcal{T}}$) has been tested against the risk neutral strategy RN (1), i.e., maximization of the objective function expected value over the scenarios along a time horizon, such that a risk reduction profile set is modeled for given functions (in our case, the objective function) up to selected stages along the time horizon. Additionally, since the risk reduction constraints system considered in the TSD strategy is a

managerial type of system (as opposed to physical constraints, balance equations, etc.), the model considers those constraints as soft ones. So, the fraction of scenarios with shortfall and the expected shortfall related bounds are not considered as hard bounds, but as targets to reach, so that appropriate slack variables (highly penalized in the objective function) avoid potential policy infeasibility and, in case, help to identify feasible bounds.

A definitive conclusion that can be drawn from the analysis of our computational experience is that solving the original TSD model (2) by plain use of a MIP solver may require so much computing effort for medium-sized instances that a type of decomposition scheme should be used. The algorithm that is proposed in this work allows to decompose the original model (2) by a scheme based on scenario cluster submodels, whose main ingredients are the Twin Node Family [17] and break stage [18] concepts. For that purpose we represent the original model as a mixture of the splitting variable and compact representations. A BFC scheme [3, 4] has been used for handling the splitting variable representation to obtain the solution value of the original stochastic MIP problem. The compact representation of the independent MIP submodels related to the scenario clusters for the stages from the break one until the last stage has been optimized by plain use of CPLEX (where 8 threads have been used).

From our computational experiment we can conclude that the risk averse TSD strategy with $|\tilde{\mathcal{T}}| > 1$ has higher risk reduction than the risk neutral (RN) strategy in case of high variability in the objective function value over the scenarios for the RN solution. It is not easy to draw conclusions from the computational comparison of the risk averse strategies FSD, SSD and TSD, but the computational requirements by the TSD strategy seems not to be an inconvenience (by using appropriate decomposition approaches) on performing better risk management on the problem. However, more computational experience is required to draw a definitive conclusion about the issue.

Appendix A. Dimensions of FSD, SSD and TSD models

Table 9 shows the dimensions of the models for the risk averse strategies that are treated in the work, such that models FSD (9), SSD (10) and TSD (2) for $|\mathcal{P}^T| = 4$ profiles where $\tilde{\mathcal{T}} = \{T\}$ (TSD1). The table also shows the dimensions of model TSD (2) for $|\mathcal{P}^T| = 4$ and $|\mathcal{P}^{T-1}| = 2$ profiles where $\tilde{\mathcal{T}} = \{T-1, T\}$ (TSD2). Its new headings are as follows: $n\nu$ and ns, additional number of 0-1 and continuous variables, respectively; and n_{ε_β} and n_{ε_e}, number of slack ε_β- and ε_e-variables, respectively.

Appendix B. Performance of the models WS, RN, FSD, SSD, TSD1 and TSD2. Illustrative instance

Let us consider instance P7 ($|\Omega| = 128$ scenarios), and z_M be the objective function expected value and z_M^ω be the related value for scenario ω from the solution of model M, where it could be the reference WS or RN (1) models; the risk averse models FSD (9), SSD (10), TSD (2) with $\tilde{\mathcal{T}} = \{T\}$ (TSD1) or TSD (2) with $\tilde{\mathcal{T}} = \{T-1, T\}$ (TSD2). Figure 2 depicts the left tail of the probability density and cumulative distribution

Table 9: DEM (2) dimensions for FSD, SSD and TSD1 (with $\tilde{\mathcal{T}} = \{T\}$) and TSD2 (with $\tilde{\mathcal{T}} = \{T-1, T\}$)

Inst.	Risk Measure	m	nx	ny	$n\nu$	n_{ε_β}	ns	n_{ε_e}	nel	$dens$	$\|\Omega\|$	$\|\mathcal{G}\|$	T
P1	FSD	2570	453	1359	452	4	0	0	66640	1.145	113	151	4
	SSD	2570	453	1359	0	0	452	4	66640	1.145	113	151	4
	TSD1	3026	453	1359	452	4	452	4	68004	0.825	113	151	4
	TSD2	3146	453	1359	510	6	510	6	70386	0.787	113	151	4
P2	FSD	2326	429	1287	320	4	0	0	61701	1.303	80	143	5
	SSD	2326	429	1287	0	0	320	4	61701	1.303	80	143	5
	TSD1	2650	429	1287	320	4	320	4	62669	1.000	80	143	5
	TSD2	2814	429	1287	400	6	400	6	66913	0.941	80	143	5
P3	FSD	7944	1632	4896	868	4	0	0	380641	0.648	217	272	4
	SSD	7944	1632	4896	0	0	868	4	380641	0.648	217	272	4
	TSD1	8816	1632	4896	868	4	868	4	383253	0.526	217	272	4
	TSD2	8996	1632	4896	956	6	956	6	390033	0.513	217	272	4
P4	FSD	10120	2176	6528	868	4	0	0	628608	0.649	217	272	4
	SSD	10120	2176	6528	0	0	868	4	628608	0.649	217	272	4
	TSD1	10992	2176	6528	868	4	868	4	631220	0.550	217	272	4
	TSD2	11172	2176	6528	956	6	956	6	640112	0.539	217	272	4
P5	FSD	7944	1632	4896	868	4	0	0	359808	0.612	217	272	4
	SSD	7944	1632	4896	0	0	868	4	359808	0.612	217	272	4
	TSD1	8816	1632	4896	868	4	868	4	362420	0.497	217	272	4
	TSD2	8996	1632	4896	956	6	956	6	367616	0.483	217	272	4
P6	FSD	14130	2946	8838	1360	4	0	0	700451	0.377	340	491	5
	SSD	14130	2946	8838	0	0	1360	4	700451	0.377	340	491	5
	TSD1	15494	2946	8838	1360	4	1360	4	704539	0.313	340	491	5
	TSD2	15950	2946	8838	1586	6	1586	6	727369	0.305	340	491	5
P7	FSD	15132	3456	10368	728	4	0	0	1383051	0.628	182	288	5
	SSD	15132	3456	10368	0	0	728	4	1383051	0.628	182	288	5
	TSD1	15864	3456	10368	728	4	728	4	1385243	0.571	182	288	5
	TSD2	16164	3456	10368	876	6	876	6	1414403	0.561	182	288	5
P8	FSD	19680	3950	11850	2296	4	0	0	840249	0.236	574	790	5
	SSD	19680	3950	11850	0	0	2296	4	840249	0.236	574	790	5
	TSD1	21980	3950	11850	2296	4	2296	4	847145	0.189	574	790	5
	TSD2	22652	3950	11850	2630	6	2630	6	875539	0.183	574	790	5
P9	FSD	27932	5580	16740	3376	4	0	0	1200301	0.167	844	1116	5
	SSD	27932	5580	16740	0	0	3376	4	1200301	0.167	844	1116	5
	TSD1	31312	5580	16740	3376	4	3376	4	1210437	0.133	844	1116	5
	TSD2	32184	5580	16740	3810	6	3810	6	1247331	0.129	844	1116	5

curves of the objective function of models M. Figure 3 shows the comparison between the objective function values of the WS and RN strategies and each risk averse strategy in particular.

We can observe in Figure 2 that the objective function of the risk averse strategies have more scenarios with value close to the expected one than the RN strategy has and, on the other hand, they have fewer scenarios with a value in the "left tail" of the probability density function. It means a smaller probability of having a non-wanted scenario (for a maximization of the objective function, as it is our case) and, then, higher risk reduction. It can also be observed that the TSD1 solution has higher risk reduction than the FSD or SSD solutions, and the TSD2 solution has higher risk reduction than the TSD1 solution. The kurtosis vector is as

follows: $(WS, RN, FSD, SSD, TSD1, TSD2) = (-0.2638, -0.2552, -0.2447, -0.2485, -0.2362, -0.2176)$. On the other hand, observe that the cumulative distribution curves for the risk averse strategies are below the other curves.

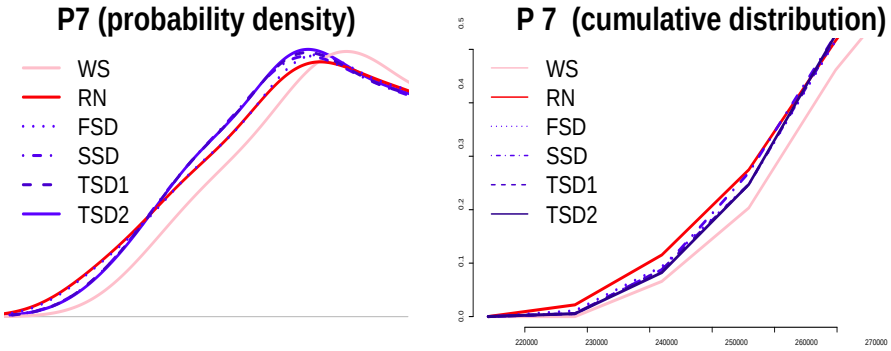

Figure 2: Probability density and cumulative distribution curves of WS, RN, FSD, SSD, TSD1 and TSD2

Given the threshold ϕ^p as the multiplicative factor δ^p times the solution value z_{RN} for $p = 1,2,3,4$, i.e., $(\phi^1, \phi^2, \phi^3, \phi^4) = (0.95, 0.90, 0.85, 0.80) \cdot z_{RN}$, the abscissa of Figure 3 gives the objective function value for the scenarios up to $1.0 \cdot z_{RN}$, the thresholds are depicted in vertical and the ordinate gives the cumulative fraction of the scenarios, say β, whose objective function values are smaller than the value in abscissa, in particular $\beta^p = P(z_M < \phi^p)$ for $M \in \{FSD, SSD, TSD1, TSD2\}$, i.e., fraction of scenarios with shortfall on reaching threshold ϕ^p. Notice that

$$P(z_{WS} < \phi^p) \leq P(z_M < \phi^p) \leq P(z_{RN} < \phi^p) \ \forall M \in \{FSD, SSD, TSD1, TSD2\},\ p = 1,2,3,4,$$

that is, the fraction of scenarios with shortfall on reaching the thresholds has significatively been reduced in the risk averse solutions versus the RN solution.

Appendix C. Meaning of abbreviations

BF, Branch-and-Fix.

BFC, Branch-and-Fix Coordination.

BFC-TSD, Branch-and-Fix Coordination procedure with Time Stochastic Dominance strategy.

CVaR, Conditional Value-at-Risk.

DEM, Deterministic Equivalent Model

FSD, First-order Stochastic Dominance.

MIP, Mixed Integer Programming.

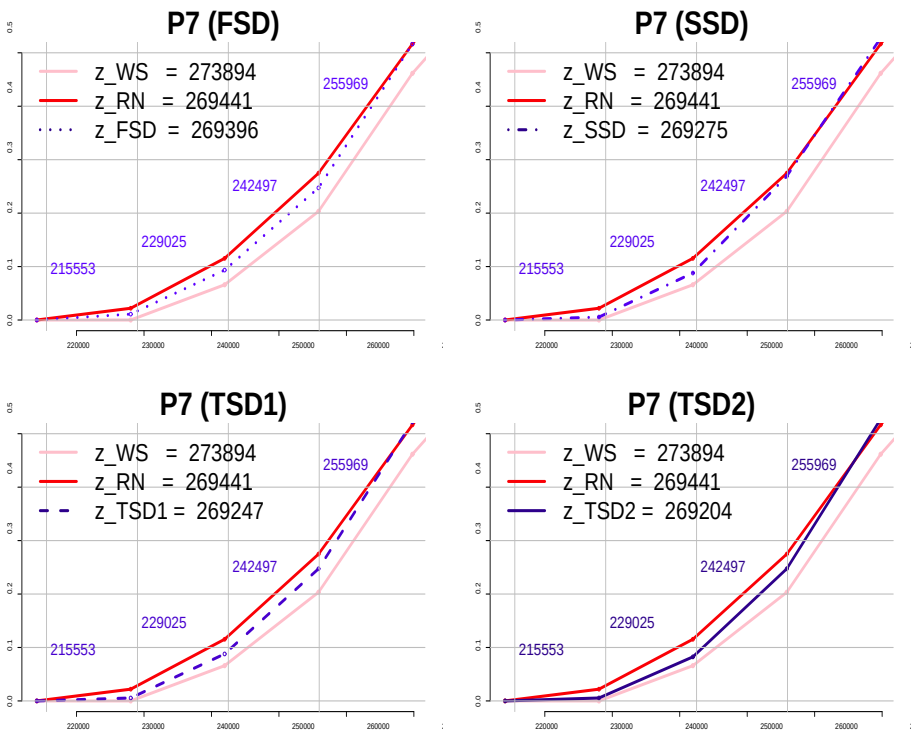

Figure 3: Cumulative distributions curves of FSD, SSD, TSD1 and TSD2 versus WS and RN

NAC, Nonanticipativity Constraints.

RN, Risk Neutral.

SDC, Stochastic Dominance Constraints.

SSD, Second-order Stochastic Dominance.

TNF, Twin Node Family.

TSD, Time Stochastic Dominance.

TSD1, Time Stochastic Dominance for $|\tilde{\mathcal{T}}| = 1$.

TSD2, Time Stochastic Dominance for $|\tilde{\mathcal{T}}| = 2$.

WS, Wait-and-See.

Acknowledgements

This research has been partially supported by the projects Grupo de Investigación EOPT (IT-567-13) from the Basque Country Government, BETS UPV/EHU Research and Training Unit (UFI), Project P711RT0278 in

the Programa Iberoamericano de Ciencia y Tecnología para el Desarrollo (CYTED), and MTM2012-31514 from the Spanish Ministry of Economy and Competitiveness. The authors would like to thank to the two anonymous reviewers for their help on clarifying some concepts presented in the manuscript and strongly improving its presentation.

References

[1] U. Aldasoro, L.F. Escudero, M. Merino and G. Pérez. An algorithmic framework for solving large scale multistage stochastic mixed 0-1 problems with nonsymmetric scenario trees. Part II: Parallelization. *Computers & Operations Research*, 40:2950-2960, 2013.

[2] A. Alonso-Ayuso, F. Carvallo, L.F. Escudero, M. Guignard-Spielberg, J. Pi, R. Puranmalka and A. Weinteraub. Medium range optimization of copper extraction planning under uncertainty in future copper prices. *European Journal of Operational Research*, 233:711-726, 2014.

[3] A. Alonso-Ayuso, L.F. Escudero and M.T. Ortuño. BFC, a Branch-and-Fix Coordination algorithmic framework for solving some types of stochastic pure and mixed 0–1 programs. *European Journal of Operational Research*, 151:503-519, 2003.

[4] A. Alonso-Ayuso, L.F. Escudero, A. Garín, M.T. Ortuño and G. Pérez. A Stochastic 0-1 Program based approach for Strategic Supply Chain Planning under Uncertainty. *Journal of Global Optimization*, 26:97-124, 2003.

[5] L. Aranburu, L.F. Escudero, A Garín and G. Pérez. Stochastic models for optimizing immunization strategies in fixed-income security portfolios under some sources of uncertainty. In H. Gassmann, S.W. Wallace and W.T. Ziemba (eds.), *Applications in Finance, Energy, Planning and Logistics*. World Scientific Publishers, 2012, pp. 173-220.

[6] N. Bauerle and A. Mundt. Dynamic mean-risk optimization in a binomial model. *Mathematical Methods of Operations Research*, 70219-239, 2009.

[7] J.F. Benders. Partitioning procedures for solving mixed variables programming problems. *Numerische Mathematik*, 4:238–252, 1962.

[8] J.R. Birge and F.V. Louveaux. *Introduction to Stochastic Programming*. Springer, 2nd edition 2011.

[9] K. Boda and J. Filar. Time consistent dynamic risk measures. *Mathematics of Operations Research*, 63:169-186, 2006.

[10] COIN-OR Foundation. COIN-OR: COmputational INfrastructure for Operations Research. `http://www.coin-or.org`, 2010.

[11] M. Carrión, U. Gotzes and R. Schultz. Risk aversion for an electric retailer with second-order stochastic dominance constraints. *Computational Management Science*, 6:233-250, 2009.

[12] D. Cuoco, H. He, S. Issaenko. Optimal dynamic trading strategies with risk limits. *Operations Research*, 56:358-368, 2008.

[13] D. Dentcheva and G. Martinez. Two-stage stochastic optimization problems with stochastic ordering constraints on the recourse. *European Journal of Operational Research*, 219:1-8, 2012.

[14] D. Dentcheva and A. Ruszczynski. Optimization with stochastic dominance constraints. *SIAM Journal on Optimization*, 14:548-566, 2003.

[15] D. Dentcheva and A. Ruszczynski. Robust optimization dominance and its application to risk-averse optimization. *Mathematical Programming, Ser. B*, 123:85-100, 2010.

[16] D. Drapkin, R. Gollmer, U.Gotzes, F. Neise and R. Schultz. Risk management with Stochastic Dominance models with disperse generation. In M. Bertochi, G. Consigli and M.A.H. Dempster (eds.), *Stochastic Optimization methods in Finance and Energy*. Springer, 2011, pp. 253-271.

[17] L.F. Escudero, A. Garín, M. Merino and G. Pérez. On BFC-MSMIP strategies for scenario cluster partitioning and Twin Nodes Families branching selection and bounding for multistage stochastic mixed integer programming. *Computers & Operations Research*, 37:738-753, 2010.

[18] L.F. Escudero, A. Garín, M. Merino and G. Pérez. An algorithmic framework for solving large scale multistage stochastic mixed 0-1 problems with nonsymmetric scenario trees. *Computers & Operations Research*, 39:1133-1144, 2012.

[19] L.F. Escudero, A. Garín and A. Unzueta. Cluster Lagrangean decomposition in multistage stochastic optimization. *Submitted for publication*.

[20] G.D. Eppen, R.K. Martin, L. Schrage. Scenario approach to capacity planning. *Operations Research* 34:517-527, 1989.

[21] C. Fabian and A. Veszpremi. Algorithms for handling Cvar constraints in dynamic stochastic models with applications in Finance. *Journal of Risk*, 10:111-131, 2008.

[22] C. Fabian, G. Mitra, D. Roman and V. Zverovich. An enhanced model for portfolio choice with SSD criteria: a constructive approach *Mathematical Programming, Ser. B*, 108:541-569, 2010.

[23] A.A. Gaivoronski, G. Pflug. Finding optimal portfolios with constraints on value-at-risk. In B. Green, editor, *Proceedings of the Third International Stockholm Seminar on Risk Behaviour and Risk Management*. Stockholm University, 1999.

[24] R. Gollmer, F. Neise and R. Schultz. Stochastic programs with first-order stochastic dominance constraints induced by mixed-integer linear recourse. *SIAM Journal on Optimization*, 19:552-571, 2008.

[25] R. Gollmer, U. Gotzes and R. Schultz. A note on second-order stochastic dominance constraints induced by mixed-integer linear recourse. *Mathematical Programming, Ser. A*, 126:179-190, 2011.

[26] V. Guigues. SDDP for some interstage dependent risk-averse problems and application to hydro-thermal planning. *Computational Optimization and Applications*, 57:167-203, 2014.

[27] IBM ILOG. CPLEX 12.5. *http://www.ilog.com/products/cplex*, 2011.

[28] W.K. Klein Haneveld. *Duality in stochastic linear and dynamic programming*. Springer-Verlag, 1986.

[29] W.K. Klein Haneveld and M.H. van der Vlerk. Integrated chance constraints: reduced forms and an algorithm. *Computational Management Science*, 3:245-269, 2006.

[30] R. Kovacevic and G. Pflug. Time consistency and information monotonicity of multiperiod acceptability functionals. *Radon Series on Computational and Applied Mathematics*, 8:1-24, 2008.

[31] A. Lizyayev. Stochastic dominance efficiency analysis of diversified portfolios: classification, comparison and refinements. *Annals of Operations Research*, 196:391-410, 2012.

[32] J. Luedtke. A branch-and-cut decomposition algorithm for solving chance-constrained mathematical programs with finite support. *Mathematical Programming, Ser. A*, 146:219-244, 2014.

[33] A. Martin, D. Morgenstern and A. Zelmer. A scenario tree-based decomposition for solving multistage stochastic programs with applications. *Submitted for publication*

[34] A. Philpott and V.L. de Matos. Dynamic sampling algorithms for multi-stage stochastic programs with risk avesion. *European Journal of Operational Research*, 218:470-483, 2012.

[35] G.Ch. Pflug and W. Römisch. *Modeling, measuring and managing risk*. World Scientific, 2007.

[36] B. Rudloff, A. Street and D.M. Valladäo. Time consistency and risk averse dynamic decision models: Definition, interpretation and practical consequences. *European Journal of Operational Research*, 234:743-750, 2014.

[37] A. Ruszczynski. Risk-averse dynamic programming for Markov decision processes. *Mathematical Programming, Ser. B*, 125:235-261, 2010.

[38] R. Schultz, S. Tiedemann. Risk Averse via excess probabilities in stochastic programs with mixed-integer recourse. *SIAM Journal on Optimization* 14:115-138, 2003.

[39] A. Shapiro. On a time consistency concept in risk averse multistage programming. *Operations Research Letters*, 37:143-147, 2009.

[40] S. Shen, J. Smith and S. Ahmed. Expectation and chance-constrained models and algorithms for insuring critical paths. *Management Science*, 56:1794-1814, 2010.

[41] R. J-B. Wets. Stochastic programs with fixed recourse: The equivalent deterministic program. *SIAM Review*, 16:309-339, 1974.

[42] R. J-B. Wets. On the relation between stochastic and deterministic optimization. In A. Bensoussan and J.L. Lions (eds), *Control Theory, Numerical Methods and Computer Systems Modelling*. Springer-Verlag, 1975, pp. 350-361.