Annals of Operations Research manuscript No.
(will be inserted by the editor)

Reduced Cost-Based Variable Fixing in Two-Stage Stochastic Programming

**Teodor G. Crainic · Francesca Maggioni ·
Guido Perboli · Walter Rei**

Received: date / Accepted: date

Abstract The explicit consideration of uncertainty is essential in addressing most planning and operation issues encountered in the management of complex systems. Unfortunately, the resulting stochastic programming formulations, integer ones in particular, are generally hard to solve when applied to realistically-sized instances. A common approach is to consider the simpler deterministic version of the formulation, even if it is well known that the

Partial funding for this project was provided by the Italian University and Research Ministry under the UrbeLOG project-Smart Cities and Communities. The second author acknowledges financial support from the grants "Fondi di Ricerca di Ateneo di Bergamo 2014", managed by Francesca Maggioni. Partial funding for this project has also been provided by the Natural Sciences and Engineering Council of Canada (NSERC) through its Discovery Grants program. We also gratefully acknowledge the support of the Fonds de recherche du Québec through their infrastructure grants. While working on this project, the first author was Adjunct Professor with the Department of Computer Science and Operations Research, Université de Montréal.

T.G. Crainic
Dept. de management et technologie, ESG, U.Q.A.M and CIRRELT - Centre Interuniversitaire de Recherche
sur les Réseaux d'Entreprise, la Logistique et le Transport
Montreal, Canada
E-mail: TeodorGabriel.Crainic@cirrelt.net

F. Maggioni
Department of Management, Economics and Quantitative Methods
University of Bergamo, Bergamo, Italy
E-mail: francesca.maggioni@unibg.it

G. Peboli
DAUIN, Politecnico di Torino, Turin, Italy and CIRRELT, Montreal, Canada
E-mail: guido.perboli@polito.it

W. Rei
Dept. de management et technologie, ESG, U.Q.A.M and CIRRELT, Montreal, Canada
E-mail: rei.walter@uqam.ca

solution quality could be arbitrarily bad. In this paper, we aim to identify meaningful information, which can be extracted from the solution of the deterministic problem, in order to reduce the size of the stochastic one. Focusing on two-stage formulations, we show how and under which conditions the reduced costs associated to the variables in the deterministic formulation can be used as an indicator for excluding/retaining decision variables in the stochastic model. We introduce a new measure, the *Loss of Reduced Costs-based Variable Fixing* (*LRCVF*), computed as the difference between the optimal values of the stochastic problem and its reduced version obtained by fixing a certain number of variables. We relate the *LRCVF* with existing measures and show how to select the set of variables to fix. We then illustrate the interest of the proposed *LRCVF* and related heuristic procedure, in terms of computational time reduction and accuracy in finding the optimal solution, by applying them to a wide range of problems from the literature.

Keywords Stochastic programming · value of stochastic solution · skeleton solution · reduced costs based variable fixing solution · expected value solution.

1 Introduction

The explicit consideration of uncertainty is essential in addressing most management problems, particularly for the planning and operations of complex systems in transportation, logistics, finance, marketing, energy, health care, production, to name but a few important areas (Prékopa, 1995; Kall and Wallace, 1994; Gaivoronski, 2005; Birge and Louveaux, 2011; King and Wallace, 2012).

Two-stage stochastic programs offer a classical modelling framework for those problems, strategic and tactical planning formulations, see in particular Birge and Louveaux (2011). In such programs, the first stage groups all decisions to be implemented before the realization of the random variables representing the stochastic parameters of the problem. In the second stage, all random information becomes known and a set recourse actions are taken to adjust the decisions made in the previous stage. The two-stage stochastic model then optimizes (without loss of generality, we use minimization in the following) a total system cost combining the cost of the first stage decisions, plus the expected cost of the recourse over all possible realizations of the random variables (the developments in this paper may be extended to the multistage case, but for simplicity of presentation we focus on the two-stage case).

Stochastic programs, in particular stochastic integer ones, are known to be generally very difficult, if not close to impossible, to address for realistically-sized instances. A formulation approximating the original stochastic model is then often used. This approximation generally takes the form of a deterministic formulation, such as the *expected-value problem*, obtained by replacing the random parameters with their expected values (or with other single-point forecast) or the extensive form of the equivalent deterministic problem obtained through sampling of a finite number of scenarios (Birge and Louveaux, 2011). Due to its generally very large dimensions produced by the scenario approximation, the latter is generally not much easier to address than the stochastic one, particularly for formulations involving integer-valued decision variables. As for the former, it is known that the use of single-point forecasts can lead to finding arbitrarily bad solutions when compared to the optimal solution of the stochastic program, e.g., (Lium et al., 2009).

But what insights can be derived from an optimal expected-value solution even if its single-point forecast defines an inaccurate estimator of the stochastic parameters of the considered problem? Specifically, two important questions then arise: 1) what can be inferred about the optimal stochastic solution from this optimal deterministic solution even when it is not of high quality and 2) can we use this information to reduce the computational effort of the stochastic program without affecting the stochastic solution quality?

We proceed in two steps: the first aims to achieve a deeper understanding of the relation between the expected-value and the stochastic solutions. What could be identified as "inherited" from the former to the latter? Can we identify a subset of variables with zero value in the deterministic solution to fix at zero in the stochastic formulation in order to guide the search toward the optimal stochastic solution? In the affirmative, are the *reduced costs* of the optimal solution of the (continuous relaxation of, in the case of integer formulations) deterministic problem a good estimation of bad/good variables to include into the stochastic solution? Can we infer a general trend from the several cases considered or is the behavior of the deterministic solution problem dependent?

To achieve these goals, we introduce the *Loss of Reduced Costs-based Variable Fixing LRCVF*, a measure of the badness/goodness of deterministic solutions based on the information offered by the reduced costs of the solution of the (continuous relaxation of the) deterministic formulation. We relate *LRCVF* to other measures present in the literature, the *Value of the Stochastic Solution* (*VSS*) (Birge, 1982) and the *Loss Using the Skeleton Solution (LUSS)* (Maggioni and Wallace, 2012). We then show experimentally that the *LRCVF* helps to identify the "good" variables that the stochastic solution should inherit from the expected-value deterministic solution and thus, provides better insights into what defines the structure of the solution to the stochastic programming model than *VSS* and *LUSS*.

We analyze, in the second step, the general trends observed during the Step 1 of the experimental campaign. This analysis aims to identify how the reduced costs associated to the non basic variables in the expected-value deterministic solution can be used to guide the selection of the variables to exclude from the stochastic formulation, making it solvable for larger instances, while preserving the quality of the final solution. The skeleton of a first heuristic procedure implementing the hints provided by this analysis is then experimentally evaluated on a large set of problems from the literature, including some large-sized stochastic Traveling Salesman Problem (TSP) instances (Ahmed et al., 2015). The results illustrate the performance and interest of *LRCVF* and the heuristic idea.

To sum up, the main contributions of this paper are to:

1. Provide a more comprehensive understanding of the structure of the optimal solution of two-stage stochastic problems and its links to the optimal solution of the expected-value corresponding deterministic version (its linear relaxation for integer formulations);
2. Define *LRCVF*, a new measure of goodness/badness of the deterministic solution with respect to the stochastic formulation;
3. Show, using *LRCVF*, how the reduced costs in the deterministic solution lead, under certain conditions, to the identification of the variables to retain/exclude in the stochastic solution;
4. Show, by means of an extensive experimental campaign, the interest of the proposed *LRCVF*, and how the reduced-costs rules may yield a heuristic effective in terms of computational time reduction and accurate in the approximation of the optimal solution.
5. Define new and more realistic standard benchmark for Stochastic Programming. It should be noted that our experimental campaign was conducted using the instances available in the *SIPLIB* library. In addition, numerical tests were also conducted using a set of larger stochastic programming problems that represent more realistic settings. These additional instances have been added to the *SIPLIB* library to complement the overall benchmark set available to the stochastic programming community.

The paper is organized as follows. The problem statement and literature review are presented in Section 2, while Section 3 defines the *LRCVF* measure. The experimental plan is described in Section 4, including how we use *LRCVF* and the problems and formulations considered in the experimentation. Numerical results are presented and analyzed in the same section. We sum up the highlights and general trends observed from this experiments in Section 5. Given the trends identified, we derive and algorithmic procedure based on *LRCVF* and we test it on a wide set of highly combinatorial instances taken from the literature. We conclude in Section 6.

2 Literature Review and Problem Statement

We focus our brief literature review on the characterization of the solutions of deterministic versions of stochastic formulations in relation to the solutions to the latter. A main concern is the identification of structures that might migrate from the deterministic solution to the stochastic one.

As already mentioned, stochastic programs, in particular integer ones, are generally very difficult to address for realistically-sized instances. Bounding techniques are therefore quite useful in practice, and several approaches and bounds on the optimal objective-function value have thus been proposed.

The standard measure of the expected gain from solving a stochastic model rather than its deterministic counterpart is given by the *Value of the Stochastic*

Solution, (*VSS*) (Birge (1982); Maggioni and Wallace (2012); Escudero et al. (2007)), computed by comparing the solution values of the stochastic and expected-value deterministic variants of the problem. A high *VSS* indicates that stochastic programming models are necessary despite the computational efforts involved.

Other approaches (e.g., Frauendorfer (1988); Hausch and Ziemba (1983); Huang et al. (1977a,b)) generalize the Edmundson-Madansky inequality (Madansky (1960)) for upper bounding and Jensen's inequality (Jensen (1906)) for lower bounding. Bounds have been proposed in Birge (1985) and Rosa and Takriti (1999) by aggregating constraints and variables in the extensive-form, while bounds based on the barycentric approximation scheme are investigated in Kuhn (2005). Bounds for convex multistage stochastic programs have been extensively elaborated in Kuhn (2008) by means of an integrated stage-aggregation and space-discretization. Other bounds for multistage linear programs have been analyzed in Maggioni et al. (2014a) by means of measures of information, measures of quality of the expected value solution, and rolling horizon measures. Maggioni and Pflug (2016) also provides bounds and approximations for multistage convex problems with concave risk functionals as objective. Maggioni et al. (2016) proposed a bounding approach, extending that of Birge (1982); Maggioni et al. (2014a) and Sandıkçı et al. (2013), which works for multistage stochastic mixed integer linear programs. The latter considers an alternative way of forming sub-problems and merging their results, with the significant advantage of dividing a given problem into independent sub-problems, which may take advantage of parallel-machine architectures. Worst-case analysis of approximated solutions in a stochastic setting has been performed in Bertazzi and Maggioni (2015) for a capacitated traveling salesmen location problem and in Bertazzi and Maggioni (2017) for a fixed charge transportation problem.

The main drawback of all these methodologies is that they measure, in different ways, the quality of the approximating solution in terms of objective-function values, but they do not provide any information on the structure of the stochastic solution. An open research question is then the following: can we learn from an approximating formulation solution, irrespective of its quality, measured in terms of objective function value?

It is well known that, in general, the expected-value solution can behave very badly in a stochastic environment. The structural differences between the two solutions within the context of particular combinatorial optimization problems have been studied in Lium et al. (2009); Thapalia et al. (2011, 2012a,b); Wang and Wallace (2016), observing both the general bad behavior of the expected value solution and hinting that some structures from the deterministic solution find their way into the stochastic one. An approach proposed in the literature to assess the value of a given solution is to approximate its relative gap to the optimum value of the stochastic problem. For example, a Monte Carlo sampling-based procedure was proposed in Mak et al. (1999) and Bayraksan and Morton (2006). Escudero et al. (2007) proposed to use the expected value solution in a multistage setting by solving subsets of scenarios and testing the obtained solution in a dynamic way.

However, from all these experiments, it is still generally not clear where the badness of the expected value solution comes from: is it because the *wrong variables* are fixed at non-zero levels or because they have been assigned *wrong values*?

An attempt to answer this question has been proposed in Maggioni and Wallace (2012). Starting from the solution of the expected value problem, it assesses whether 1) the deterministic model produced the right non-zero variables, but possibly was off on the values of the basic variables; and 2) the deterministic solution is upgradable to become good (if not optimal) in the stochastic setting. The resulting measures, called *Loss Using the Skeleton Solution (LUSS)* and the *Loss of Upgrading the Deterministic Solution (LUDS)* in Maggioni and Wallace (2012) (see Maggioni et al., 2014a, for the extension to the multistage setting), are obtained by restricting the values of the first stage variables based on the solution of the expected-value problem. *LUSS* is obtained by fixing at zero (or at the lower bound) the first stage variables which are at zero (or at the lower bound) in the expected value solution (i.e., for linear programs, the non basic variables), solving the stochastic program, and contrasting it to the solution of the original stochastic model. *LUDS* is measured by first solving a restricted stochastic model obtained by fixing the lower bound of all variables to their corresponding values in the expected value solution, and contrasting it to the solution of the original stochastic model. Unfortunately, this approach leads to suboptimal solutions, in particular when large combinatorial stochastic problems must be solved. We compare in our experimental-results section the performance of *LRCVF*, the new measure we propose, to that of *LUSS* and *LUDS*.

Notice also that, approaches were proposed in the literature on deterministic combinatorial optimization to fix to zero the largest part of the non basic variables in the continuous relaxation of the problem in order to reduce the computational time (Angelelli et al., 2010; Perboli et al., 2011). Then, to identify the appropriate core set of non basic variables to be included in the restricted problem, the search is performed starting from the ones with the smallest reduced cost Perboli et al. (2011).

One may conclude from this brief review of previous work that a systematic way to identify the structure of the stochastic solution out of the expected-value deterministic one is still missing. The goal of this paper is to fill this gap providing a tool to analyze and compare the expected value solution with respect to the stochastic one. In the next section, we introduce the concepts and a procedural way to compute the *Reduced Costs-based Variable Fixing* (*RCVF*) and the *Loss of Reduced Costs-based Variable Fixing* (*LRCVF*). *LRCVF* will provide the means to investigate, even in the case of a large *VSS*, what can be inherited from the structure of the expected value solution in its stochastic counterpart, by taking into account the information on reduced costs associated to the variables at zero (or lower bound) in the expected value solution.

3 The Value of Variable Fixing

We first define the standard notation used in this paper, and then move to introduce *RCVF* and *LRCVF*.

3.1 Notation and definitions

The following mathematical model represents a general formulation of a stochastic program in which a decision maker needs to determine x in order to minimize (expected) costs or outcomes (Kall and Wallace, 1994; Birge and Louveaux, 2011):

$$\min_{x \in X} E_{\boldsymbol{\xi}} z(x, \boldsymbol{\xi}) = \min_{x \in X} \left\{ f_1(x) + E_{\boldsymbol{\xi}} \left[h_2(x, \boldsymbol{\xi}) \right] \right\}, \tag{1}$$

where x is a first-stage decision vector restricted to the set $X \subseteq \mathbb{R}^n_+$, with $\mathbb{R}^n_+$ is the set of non negative real vectors of dimension n, and $E_{\boldsymbol{\xi}}$ stands for the expectation with respect to a random vector $\boldsymbol{\xi}$, defined on some probability space $(\Omega, \mathscr{A}, p)$ with support Ω and given probability distribution p on the $\sigma-$algebra $\mathscr{A}$. The function h_2 is the value function of another optimization problem defined as

$$h_2(x, \xi) = \min_{y \in Y(x,\xi)} f_2(y; x, \xi), \tag{2}$$

which is used to reflect the costs associated with adapting to information revealed through a realization ξ of the random vector $\boldsymbol{\xi}$. The term $E_{\boldsymbol{\xi}} \left[h_2(x, \boldsymbol{\xi}) \right]$ in (1) is referred to as the recourse function. We make the assumption in this paper that functions f_1 and f_2 are linear in their unknowns. The solution x^* obtained by solving problem (1), is called the *here and now solution* and

$$RP = E_{\boldsymbol{\xi}} z(x^*, \boldsymbol{\xi}), \tag{3}$$

is the optimal value of the associated objective function.

A simpler approach is to consider the *Expected Value Problem*, where the decision maker replaces all random variables by their expected values and solves a deterministic program:

$$EV = \min_{x \in X} z(x, \bar{\xi}), \tag{4}$$

where $\bar{\xi} = E(\boldsymbol{\xi})$. Let $\bar{x}(\bar{\xi})$ be an optimal solution to (4), called the *Expected Value Solution* and let EEV be the expected cost when using the solution $\bar{x}(\bar{\xi})$:

$$EEV = E_{\boldsymbol{\xi}} \left(z \left(\bar{x}(\bar{\xi}), \boldsymbol{\xi} \right) \right). \tag{5}$$

The *Value of the Stochastic Solution* is then defined as

$$VSS = EEV - RP, \tag{6}$$

measuring the expected increase in value when solving the simpler deterministic model rather than its stochastic version. Relations and bounds on EV, EEV and RP can be found for instance in Birge (1982) and Birge and Louveaux (2011).

Let $\mathcal{J} = \{1, \ldots, J\}$ be the set of indices for which the components of the expected value solution $\bar{x}(\bar{\xi})$ are at zero or at their lower bound (non basic variables). Then let $\hat{x}$ be the solution of:

$$\begin{aligned} \min_{x \in X} \; & E_{\boldsymbol{\xi}} z\left(x, \boldsymbol{\xi}\right) \\ \text{s.t.} \quad & x_j = \bar{x}_j(\bar{\xi}), \; j \in \mathcal{J}. \end{aligned} \tag{7}$$

We then compute the *Expected Skeleton Solution Value*

$$ESSV = E_{\boldsymbol{\xi}}\left(z\left(\hat{x}, \boldsymbol{\xi}\right)\right), \tag{8}$$

and we compare it with RP by means of the *Loss Using Skeleton Solution*

$$LUSS = ESSV - RP. \tag{9}$$

A $LUSS$ close to zero means that the variables chosen by the expected value solution are the correct ones but their values may be off. We have:

$$RP \leq ESSV \leq EEV, \tag{10}$$

and consequently,

$$VSS \geq LUSS \geq 0. \tag{11}$$

Notice that the case $LUSS = 0$ corresponds to the *perfect skeleton solution* in which the condition $x_j = \bar{x}_j(\bar{\xi}),\ j \in \mathcal{J}$, is satisfied by the stochastic solution x^* even without being enforced by a constraint (i.e., $\hat{x} = x^*$); on the other hand, if there exists $j \in \mathcal{J}$ such that $x_j^* \neq \bar{x}_j(\bar{\xi})$ in any optimal stochastic solutions x^*, then $0 < LUSS < VSS$. Finally, one observes $LUSS = VSS$, if the $\hat{x} = \bar{x}(\bar{\xi})$.

3.2 Defining the *LRCVF*

We now define *RCVF* and *LRCVF*, together with a procedural way to compute them.

Let $\mathscr{R} = \{r_1, \ldots, r_j, \ldots, r_J\}$ be the set of *reduced costs*, with respect to the recourse function, of the components $\bar{x}_j(\bar{\xi}),\ j \in \mathcal{J}$, of the expected-value solution $\bar{x}(\bar{\xi})$ at zero or at their lower bound (i.e., non basic variables). We recall that a reduced cost is the amount by which an objective function coefficient would have to improve (increase, for maximization problems and decrease for minimization ones) before it would be possible for the corresponding variable to assume a positive value in the optimal solution and become a basis variable. Since the reduced costs of all basis variables (also the ones at the related upper bounds) are zero, they will be not fixed. In the following, we make the assumption that in the case of a problem with first stage integer variables, we compute the reduced costs on the continuous relaxation.

Let $r^{max} = \max_{j\in\mathcal{J}}\{r_j : r_j \in \mathscr{R}\}$ and $r^{min} = \min_{j\in\mathcal{J}}\{r_j : r_j \in \mathscr{R}\}$ be respectively the maximum and the minimum of the reduced costs of the variables $\bar{x}_j(\bar{\xi}),\ j \in \mathcal{J}$. We divide the difference $r^{max} - r^{min}$ into N classes $\mathscr{R}_1, \ldots, \mathscr{R}_N$ of constant width $\frac{r^{max}-r^{min}}{N}$ such that the $p-$class is defined as follows

$$\mathscr{R}_p = \left\{ r_j : r^{min} + (p-1) \cdot \frac{(r^{max} - r^{min})}{N} \le r_j \le r^{min} + p \cdot \frac{(r^{max} - r^{min})}{N} \right\}, \quad (12)$$

with $p = 1, \ldots, N$. Let $\mathcal{J}_p$ be the set of indices associated to the variables $\bar{x}_j(\bar{\xi})$ with reduced costs $r_j \in \mathscr{R}_p$. Then let $\tilde{x}_p$ be the solution of

$$\begin{aligned} &\min_{x\in X}\ E_{\boldsymbol{\xi}} z\,(x, \boldsymbol{\xi}) \\ &\qquad \text{s.t.} \quad x_j = \bar{x}_j(\bar{\xi}),\ j \in \mathcal{J}_p, \ldots, \mathcal{J}_N, \end{aligned} \quad (13)$$

where we fix at zero or lower bounds only the variables with indices belonging to the last p classes $\mathcal{J}_p, \ldots, \mathcal{J}_N$, i.e., with the highest reduced costs.

We then compute the *Reduced Costs-based Variables Fixing*

$$RCVF(p, N) = E_{\boldsymbol{\xi}}\,(z\,(\tilde{x}_p, \boldsymbol{\xi}))\ , \quad p = 1, \ldots, N, \quad (14)$$

and we compare it with RP by means of the *Loss of Reduced Costs-based Variable Fixing*

$$LRCVF(p, N) = RCVF(p, N) - RP\ , \quad p = 1, \ldots, N. \quad (15)$$

Notice that $RCVF(1, N) = ESSV$ and consequently $LRCVF(1, N) = LUSS$.

Furthermore, considering that both $RCVF$ and $LRCVF$ are defined on the basis of restricting only a subset of the N classes that partition the non basic variables according to their respective values, these bounds $RCVF(p, N)$, $p = 1, \ldots, N$ can be improved (as is clearly stated in the two propositions that will follow). Also, as will be described in the subsequent section of this paper, by varying the values of parameters p and N, a systematic search can be performed to both assess the quality of the obtained bounds and inferring what the actual restriction to be applied on the overall stochastic model should be. We now prove that the following inequalities hold true:

Proposition 3.1 *For a fixed* $N \in \mathbb{N}\setminus\{0, 1\}$ *(where* $\mathbb{N}$ *is the set of natural numbers),*

$$LRCVF(p, N) \ge LRCVF(p+1, N)\ , \quad p = 1, \ldots, N-1. \quad (16)$$

Proof
Any feasible solution of problem $RCVF(p, N)$ is also a solution of problem $RCVF(p{+}1, N)$, since the former is more restricted than the latter, and so, the relation (16) holds true. If $LRCVF(p, N) = \infty$, the inequality is automatically satisfied. □

Proposition 3.2 *For a given* $N \in \mathbb{N} \setminus \{0\}$ *and a fixed* $p \in \mathbb{N} \setminus \{0\}$ *such that* $p = 1, \ldots, N$,

$$LRCVF(p, N+1) \geq LRCVF(p, N). \tag{17}$$

Proof
If $p = 1$ then $LRCVF(p, N+1) = LRCVF(p, N) = LUSS$. Furthermore, any feasible solution of problem $RCVF(p, N+1)$ is also a solution of problem $RCVF(p, N)$, since the former is more restricted than the latter, and so, the relation (17) holds true. If $LRCVF(p, N+1) = \infty$, the inequality is automatically satisfied. □

The two previous properties can be generalized in the following corollary:

Corollary 3.1 *For given* $N_1, N_2 \in \mathbb{N} \setminus \{0\}$ *and* $p_1, p_2 \in \mathbb{N} \setminus \{0\}$, *with* $p_1 = 1, \ldots, N_1$, $p_2 = 1, \ldots, N_2$ *and such that* $\frac{p_1}{N_1} \leq \frac{p_2}{N_2}$

$$LRCVF(p_1, N_1) \geq LRCVF(p_2, N_2). \tag{18}$$

Proof
If $p_1 = p_2 = 1$ then $LRCVF(p_1, N_1) = LRCVF(p_2, N_2) = LUSS$. Furthermore, if $\frac{p_1}{N_1} \leq \frac{p_2}{N_2}$ then the number of variables at zero with highest reduced cost to be fixed is respectively $\frac{N_1 - p_1}{N_1}|\mathscr{R}| \geq \frac{N_2 - p_2}{N_2}|\mathscr{R}|$. Consequently $RCVF(p_1, N_1)$ is more restricted than $RCVF(p_2, N_2)$, and the relation (18) holds true. □

Notice that, variables are unbounded in the minimization problem setting considered (1). One might, however, consider problem settings where the variables have limited upper bounds. In these cases, non basic variables might be at zero (or at their lower bound values) with positive reduced cost or at their upper bounds with negative reduced costs (Ahuja et al., 1993). The variable fixing procedure we propose implicitly considers this case, as non basic variables at their upper bounds correspond, due to their negative reduced costs, to the sets $\mathscr{R}_p$ with the lowest reduced cost values. Therefore, such variables are unlikely to be fixed to 0 by the procedure.

LRCVF measures how much we lose in terms of solution quality when we consider the reduced costs-based variable fixing solution. But how can one use it in order to analyze and derive the structure of the stochastic solution? How should we choose the number of classes N and p? We answer these questions in the following sections, by presenting a procedure using *LRCVF* and applying it to a wide set of problems from the literature.

4 Experimental Plan and Results

This section describes the experimental plan and the instance sets considered. Our goal is to assess the validity of *LRCVF* for extracting information about the skeleton of the stochastic solution from the reduced costs of the expected-value solution (or its linear relaxation for integer formulations). We therefore performed an experimental analysis to explore the behavior of *RCVF* and

LRCVF, compared to *LUSS*, while varying the values of p and N, according to three axes:

- *Computational effort.* What number of variables can we fix in order to drastically reduce the effort of the stochastic solution computation?
- *Feasibility.* What are the effects of fixing a subset of the variables from the expected-value solution with regards to the feasibility of the stochastic model?
- *Optimality.* How to use the *LRCVF* to find an optimal or near optimal stochastic solution?

We used instances corresponding to stochastic optimization models related to three real-case applications: a single-sink transportation problem, a power generation scheduling case, and a supply transportation problem. All numerical experiments were conducted on a 64-bit machine with 12 GB of RAM and a Intel Core i7-3520M CPU 2.90 GHz processor, using CPLEX 12.5 as MIP solver.

Section 4.1 presents our methodology for computing the two measures, including a proposed approach to set up the number of classes N and the class parameter p of $LRCVF(p, N)$. Section 4.2 gives a short description of the test instances, while computational results are discussed in Section 4.3.

4.1 Computing RCVF and LRCVF

We computed VSS, $LUSS$ and $LRCVF(p, N)$ for each instance set. The optimal solutions of the stochastic formulations were either taken from the literature, when available, or computed, otherwise. We now briefly describe the procedure we developed, which can be applied and extended to any stochastic programming problem.

Recall that parameter N defines the number of classes, or sets, in which the non basic variables of $\bar{x}(\bar{\xi})$ are grouped, and that these sets provide a characterization of the variables with respect to their reduced costs. Thus, the higher the value of N, the closer the reduced-cost values of the variables included in each set. We therefore start by considering a rough characterization given by three classes, $N = 3$, where the non basic variables of $\bar{x}(\bar{\xi})$ are included in a high, low or medium-range reduced-cost set. Finally, the size of the "supply transportation" problem allowed us to test other values of N ($N = 3, 10, 50, 100$) and to analyze the sensitivity of the results when N increases.

For a given value N, our objective while generating sets $\mathscr{R}_1, \ldots, \mathscr{R}_N$ and the partition of the variables $\mathcal{J}_1, \ldots, \mathcal{J}_N$, is to identify which non basic variables of $\bar{x}(\bar{\xi})$ should be fixed in the stochastic model to produce an optimal, or near-optimal, solution. To do so, the parameter p is first fixed to its upper limit (i.e., $p = N$) to compute $LRCVF(N, N)$. Parameter p is then iteratively decreased by a value of one as long as the following condition is verified: $LRCVF(p, N) = LRCVF(p-1, N)$. In fact, from Property 3.1, we have that, for a fixed N, $LRCVF(p, N)$ can only increase when p decreases.

4.2 Test instances

The instances used in this experimental phase are taken from the literature:

- *Power generation scheduling* based on an economic scheduling model formulated in Williams (2013) and Garver (1962) as a deterministic mixed integer program and extended in Maggioni and Wallace (2012) as a stochastic optimization problem; Power generation scheduling involves the selection of generating units to be put into operation and the allocation of the stochastic power demand among the units over a set of time periods;
- *Supply transportation problem* inspired by a real case of *gypsum* replenishment in Italy, provided by the primary Italian cement producer. The logistics system is organized as follows: 24 suppliers, each of them having several plants located all around Italy, are used to satisfy the demand for gypsum of 15 cement factories belonging to the same company; the demands for gypsum at the 15 cement factories are considered stochastic; See Maggioni et al. (2017) for more details.

In order to ensure the fluidity of the paper, the problem descriptions and the two-stage models as reported in the literature are included in Annex A, while their corresponding numerical data are summarized in Annex B. Notice that annexes A and B, include also the description and numerical results of a *Single-sink transportation problem*, inspired by a real case of *clinker* replenishment, provided by the largest Italian cement producer located in Sicily (Maggioni et al., 2009).

4.3 Numerical results

We now present and analyze the results obtained by applying the $LRCVF(p, N)$ measure to the problems described above. We followed the procedure described in 4.1, computing each time VSS, $LUSS$ and $LRCVF(p, N)$, $p = 1, \ldots, N$. Detailed solutions of the different instances for the first three test problems may be found at: http://www.francescamaggioni.it/index.php?id=lrcvf.

4.3.1 The power generation problem

The power generation problem (PGP) (Annex A.2) selects power units of type 1 or 2 to operate and allocates the power demand among the selected units. We run the model for 10 different instances with demand randomly generated in the interval $\left[d^{min}, d^{max}\right]$, where $d^{min} = 33$ and $d^{max} = 687$ are respectively the minimum and maximum demand observed in the historical data. The number of scenarios is 20. Summary statistics of the adjusted problem derived for our test case are reported in Table 1. Columns 3-4-5-6 display the total number of variables and the total number of integer variables, respectively. Notice that presolve eliminates 68 constraints and 2 variables.

Results are reported in Tables 2 and 3. The former reports the deviations (in %) with respect to RP for VSS, $LRCVF(p, 3)$, $p = 1, \ldots, 3$, and

Table 1 Summary statistics for the PGP

CPU time (ss)	# simplex iterations	# variables				# constraints	
		1-stage v.	2-stage v.	Int. 1-stage v.	Int. 2-stage v.	ineq.	eq.
0.015625	46	8	120	4	80	152	0

$LRCVF(p,4)$. The latter illustrates the discussion that follows with the results obtained for the first instance, displaying the first stage solutions of generating units u_i^2, the number of started up generators s_i^2, total output rate x_i^2 $(i \in \mathscr{I})$ and the total cost.

Table 2 Results for the PGP (% deviation from RP)

Instance	VSS	$LRCVF(p,3)$			$LRCVF(p,4)$			
		1	2	3	1	2	3	4
1	10.17	10.17	0	0	10.17	0	0	0
2	10.03	10.03	0	0	10.03	0	0	0
3	7.05	0	0	0	0	0	0	0
4	10.66	10.66	0	0	10.66	0	0	0
5	10.07	10.07	0	0	10.07	0	0	0
6	6.78	0	0	0	0	0	0	0
7	6.14	0	0	0	0	0	0	0
8	5.93	0	0	0	0	0	0	0
9	6.84	6.84	0	0	6.84	0	0	0
10	7.44	7.44	0	0	7.44	0	0	0
Mean	8.11	5.44	0	0	5.44	0	0	0

We evaluated the expected value solution under the mean scenario $\bar{D}$ in the stochastic model. As illustrated in the case of instance 1, Table 3, the deterministic model closes down as many units as possible to simply cover the considered demand, ending up with only four units of type 1. Because the deterministic solution only keeps 4 units running, instead of the 8 (4 units of type 1 and 4 of type 2) included in the stochastic solution, the associated total cost reduces to 104 285 € compared to 117 927.5 € for the stochastic counterpart. However, the 4 units working in the deterministic solution are not enough to satisfy the high demand scenarios, yielding

$$VSS = 129\,927.5 - 117\,927.5 = 12\,000, \tag{19}$$

causing a loss of 10.17% given the need to restart some units at the second stage. We now investigate why the deterministic solution is bad by means of $LUSS$ and *LRCVF*.

According to the $LUSS$, we follow the skeleton solution from the deterministic model and close the units of type 2 that are not required to satisfy the deterministic demand. The stochastic model reacts by opening units of type 2 at the second stage at higher cost. As a consequence, the associated *Expected Skeleton Solution Value* $ESSV$ is the same as EEV and $LUSS$ is VSS. It

Table 3 Optimal solutions for different problem types for PGP instance 1

Problem type	u_1^2	u_2^2	s_1^2	s_2^2	x_1^2	x_2^2	Objective value (€)
EV	4	0	0	0	300	0	104 285
RP	4	4	0	0	180	120	117 927.5
EEV	4	0	0	0	300	0	129 927.5
$ESSV$	4	0	0	0	300	0	127 877.5
$RCVF(p,N) = RP$, $p = 2, \ldots, N$	4	4	0	0	180	120	117 927.5
$RCVF(1,N) = ESSV$	4	0	0	0	300	0	129 927.5

confirms that deterministic solution has a bad structure (required units for the stochastic environment that are closed in the deterministic case).

Applying $LRCVF(p,N)$ idea, the reduced costs of the decision variables at zero in the skeleton solution from the deterministic model are computed. It closes the units of type 2, $u_2^2 = 0$, yielding $x_2^2 = 0$, and do not start up any generator, $s_i^2 = 0$; thus $r_{u_2^2} = 500$ and $r_{s_1^2} = 14000$, $r_{s_2^2} = 16000$ and $r_{x_2^2} = 50$. Let define $r^{max} = r_{s_2^2} = 16000$ and $r^{min} = r_{x_2^2} = 50$.

Notice that, since the number of variables at zero in the deterministic solution is 4, the maximum number of classes to consider is $N = 4$. We computed two measures, dividing the difference $r^{max} - r^{min}$ into $N = 3$ and $N = 4$ classes of constant width, respectively. With the two values of N we have that $LRCVF(p,N) = 0$, with $p \in \{2, \ldots, N\}$, while the percentage gap between $LRCVF$ and RP becomes 10.17 % when $p = 1$. It shows how the wrong choice from the deterministic solution is in the selection of variable u_2^2.

The results obtained on instances 3, 6, 7 and 8 show a different behavior since $LRCVF(1,N) = LUSS = 0$ while the VSS is around 6.5% (Table 2). This means that, in these instances, the EV problem is able to identify the appropriate structure in terms of zero and non-zero variables, but fails in providing the correct first-stage non zero values.

In conclusion, the deterministic solution is bad because it tends to follow in every period the market profile, thus closing units that could be needed in the following time periods. However, we again obtain the optimal stochastic solution by applying the new measure and procedure, that is, by following the skeleton of the deterministic solution with highest reduced costs (i.e., do not starting up any generator, $s_i^2 = 0$).

4.3.2 The supply transportation problem

VSS, $LUSS$ and $LRCVF(p,N)$, $p = 1, \ldots, N$ were computed for the supply transportation problem (STP) (Annex A.3), which identifies the number of vehicles to book for each plant of each supplier, for the replenishment of gypsum at minimum total cost. Data represents the first week of March 2014. We run the model for 10 different instances with demand randomly generated in the interval $\left[d_j^{min}, d_j^{max}\right]$, where d_j^{min} and d_j^{max} are the minimum and maximum demand observed in the historical data, respectively in destination $j \in \mathscr{D}$ (see Table 5).

The number of scenarios is 48. Summary statistics of the adjusted problem derived for our test case are reported in Table 4. Columns 3-4-5-6 display the total number integer variables. Notice that in this problem all the decision variables are integer, and presolve eliminates 1261 constraints.

Table 4 Summary statistics for the STP

CPU time (ss)	# simplex iterations	# variables		# constraints	
		Int. 1-stage v.	Int. 2-stage v.	inequality	equality
1.8125	9156	480	23760	23954	0

Table 5 Minimum and maximum observed demand (second and third columns) over all the destinations $j \in \mathscr{D}$. Fourth and fifth columns report the total number of booked vehicles at each destination respectively in deterministic and stochastic solution. These values are averaged and rounded over 10 instances

	Demand		Solution	
Destination $j \in \mathscr{D}$	Minimum d_j^{min}	Maximum d_j^{max}	Deterministic $\sum_{k\in\mathscr{K}}\sum_{i\in\mathscr{O}_k}\bar{x}_{ij}$	Stochastic $\sum_{k\in\mathscr{K}}\sum_{i\in\mathscr{O}_k}x_{ij}$
1	27.45	298.43	6	9
2	202.01	1479.89	29	26
3	171.78	680.16	14	21
4	0	216.96	4	7
5	0	101.26	2	3
6	0	196.93	4	6
7	0	216.20	4	7
8	0	200.43	4	6
9	0	545.19	10	15
10	0	234.37	4	7
11	0	318.89	6	9
12	0	430.36	7	11
13	0	199.42	4	6
14	0	223.50	4	7
15	0	723.46	12	20

The cost values associated to the solutions of the deterministic, the EV model (4), and stochastic formulations are reported in Table 6 for the 10 instances. The deterministic model will always book the exact number of vehicles $\bar{x}_{ij}$ needed for the next period for each plant $i \in \mathscr{O}_k$ of supplier $k \in \mathscr{K}$, to destination $j \in \mathscr{D}$; it sorts the suppliers and their plants according to the transportation costs and books a full production capacity from the cheapest one, followed by the next-cheapest, and so on. As long as there is sufficient transportation capacity, the model will never purchase extra gypsum from external sources, i.e. $y_j = 0, \forall j \in \mathscr{D}$. The total cost then reduces to the booking cost at the first stage.

The last two columns of Table 5 show the total number of booked vehicles at each cement factory averaged and rounded over the 10 instances, for the expected value solution and the optimal solution of the stochastic problem, respectively.

The headings of Table 6 are as follows: Instance code in column 1; EV_1 and RP_1 give in columns 2 and 3 the objective function terms (i.e., cost) related to the first stage in the deterministic model and stochastic model, respectively; EV and RP give in columns 4 and 5 the total cost of the solution of the deterministic and stochastic models, respectively.

The deterministic model books much fewer vehicles than the stochastic one, resulting in a solution costing only 83% of the stochastic counterpart (Table 6). The EEV is infeasible, however, resulting in $VSS = \infty$, which shows that the expected value solution is not appropriate in a stochastic setting.

Table 6 Optimal solution values for STP

Instance	EV_1 (€)	RP_1 (€)	EV (€)	RP (€)
1	79 709.30	123 673.00	79 709.30	95 738.53
2	76 768.80	119 626.00	76 768.79	93 468.32
3	77 386.50	121 147.00	77 386.50	93 297.18
4	74 332.00	122 054.00	74 332.00	90 734.68
5	76 101.30	119 638.00	76 101.30	93 014.14
6	75 066.10	118 657.00	75 066.10	91 661.99
7	76 992.00	119 004.00	76 992.00	94 066.06
8	78 859.30	123 980.00	78 859.30	94 306.65
9	75 111.80	119 033.00	75 111.80	91 966.36
10	80 344.30	123 781.00	80 344.30	96 296.00
Mean	77 067.14	121 059.35	77 067.14	93 454.99

Why is the deterministic solution bad? Is it because of an overly optimistic guess on the randomness, leading to too *few booked vehicles* from the plants $i \in \mathscr{O}_k$ of suppliers $\mathscr{K}$, or is it because of wrong choices being made regarding the *suppliers and plants*? $LUSS$ and $LRCVF$ are used to answer these questions.

To compute the $LUSS$, we follow the skeleton solution from the deterministic model, not allowing to book vehicles from the plants $i \in \mathscr{O}_k$ (for all suppliers $k \in \mathscr{K}$), such that $\bar{x}_{ij}(\bar{\xi}) = 0, \;\; j \in \mathscr{D}$, in the expected value solution. The *Expected Skeleton Solution Value* $ESSV$ is still infeasible and then, the associated *Loss Using the Skeleton Solution* $LUSS = \infty$. Therefore, the chosen suppliers and associated plants, derived from the solution to the deterministic model, are unsuited for the stochastic case. We can thus conclude that the deterministic solution is inappropriate because a wrong number of vehicles are booked from the wrong suppliers and plants.

We then turn to $LRCVF(p, N)$ and analyze the reduced costs of the variables at zero in the deterministic solution, illustrated in Figure 1 for the first instance. The range of reduced costs, from $r^{min} = 131$ to $r^{max} = 683$, is

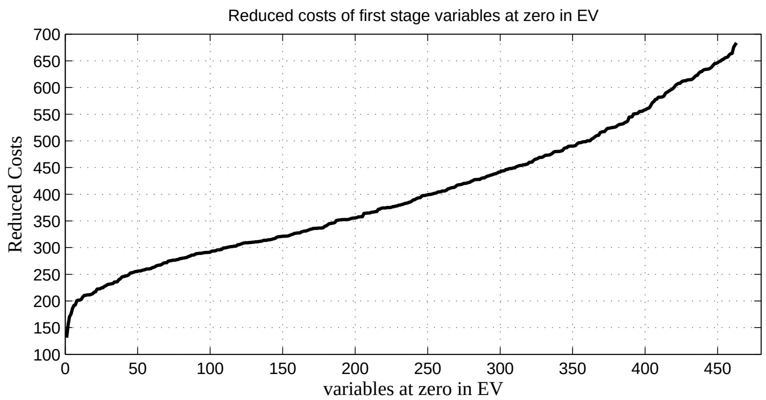

Fig. 1 Reduced costs of the variables at zero in the EV solution of STP instance 1

sufficiently broad to allow testing the sensitivity of the results with a large number of classes N. We therefore divide the difference $r^{max} - r^{min} = 552$ into $N = 3, 10, 50, 100,$ classes $\mathscr{R}_1, \ldots, \mathscr{R}_N$ of constant width, respectively. Results are reported in Tables 8, 9 and 10, respectively.

Contrary to the VSS and $LUSS$, $LRCVF(p, N)$ is able to find optimal results when a limited subset of variables are fixed. In the case of $LRCVF(p, 3)$, for $p = 1, 2, 3$, the appropriate variables from the deterministic solution are identified as the ones included in the last two classes (i.e., $\left[r^{min} + \frac{r^{max} - r^{min}}{3}, r_{max}\right]$), considering that $LRCVF(2, 3) = LRCVF(3, 3) = 0$. On the other hand, fixing at zero in the stochastic model all the variables at zero in the deterministic solution yielding $LRCVF(1, 3) = \infty$. It should also be noticed that $LRCVF(p, 3)$, $p = 2, 3$ is able to replicate the optimal values of the stochastic problem while reducing the computational effort by 50% when $p = 3$ and by 75% when $p = 2$; see Table 7.

More refined information on the wrong variables from the deterministic solution is obtained by increasing the number of classes to $N = 10$, and identifying the good variables to fix as the ones belonging to the classes in the interval $\left[r^{min} + \frac{3(r_{max} - r_{min})}{10}, r^{max}\right]$. These results are displayed in Table 8. Furthermore, by also fixing the variables belonging to the interval $\left[\frac{2(r_{max} - r_{min})}{10}, \frac{3(r_{max} - r_{min})}{10}\right]$, a nearly optimal solution can be obtained. Adding class $p = 2$, results in $LRCVF(2, 10) = \infty$ and consequently $LRCVF(1, 10) = \infty$. As previously observed, $LRCVF(4, 10)$ is able to replicate the optimal values of the stochastic problem while reducing the computational effort by a significant margin (i.e., 80%), see Table 7.

Increasing the number of classes to $N = 50$, see Table 9, further refines the information deduced from the deterministic-model solution regarding the good variables to fix, as $LRCVF(p, 50) = 0$, with $p = 15, \ldots, 50$. In terms of the computational effort, the observed gains increase to 84% with $p = 15$. By

Table 7 CPU time (seconds) and optimal objective values for the computation of EV, RP, $LRCVF(p,3)$, $p = 1,\ldots,3$ and $LRCVF(p,10)$, $p = 1,\ldots,10$, for the STP instance 1

Measure	CPU time (ss)	Objective value
EV	0.07	79 709.27
RP	3.07	95 738.53
$LRCVF(3,3)$	1.49	95 738.53
$LRCVF(2,3)$	0.74	95 738.53
$LRCVF(1,3)$	0.14	∞
Total $LRCVF(p,3)$	2.24	–
$LRCVF(10,10)$	1.79	95 738.53
$LRCVF(9,10)$	1.80	95 738.53
$LRCVF(8,10)$	1.48	95 738.53
$LRCVF(7,10)$	1.35	95 738.53
$LRCVF(6,10)$	1.29	95 738.53
$LRCVF(5,10)$	0.92	95 738.53
$LRCVF(4,10)$	0.60	95 738.53
$LRCVF(3,10)$	0.39	95 744.87
$LRCVF(2,10)$	0.14	∞
$LRCVF(1,10)$	0.15	∞
Total $LRCVF(p,10)$	9.95	–

Table 8 Results of $LRCVF(p,3)$ and $LRCVF(p,10)$ for STP as % from RP

Instance	VSS	$LRCVF(p,3)$			$LRCVF(p,10)$									
		1	2	3	1	2	3	4	5	6	7	8	9	10
1	∞	∞	0	0	∞	∞	0.006	0	0	0	0	0	0	0
2	∞	∞	0	0	∞	∞	0.006	0	0	0	0	0	0	0
3	∞	∞	0	0	∞	∞	0.008	0	0	0	0	0	0	0
4	∞	∞	0	0	∞	∞	0.084	0	0	0	0	0	0	0
5	∞	∞	0	0	∞	∞	0.001	0	0	0	0	0	0	0
6	∞	∞	0	0	∞	∞	0	0	0	0	0	0	0	0
7	∞	∞	0	0	∞	∞	0.01	0	0	0	0	0	0	0
8	∞	∞	0	0	∞	∞	0.006	0	0	0	0	0	0	0
9	∞	∞	0	0	∞	∞	0.002	0	0	0	0	0	0	0
10	∞	∞	0	0	∞	∞	0.002	0	0	0	0	0	0	0
Mean	∞	∞	0	0	∞	∞	0.012	0	0	0	0	0	0	0

setting $N = 100$, two extra variables at zero from the deterministic solution can be detected, $LRCVF(p,100) = 0$, with $p = 28,\ldots,100$, see Table 10. In this case, the gain in computational effort is 81% with $p = 28$.

Regarding the distribution of the reduced costs in the expected-value deterministic solution, one idea is to compute them, and plot or pass them through a statistical package, to see if one can observe a trend referable to a known probability distribution. Unfortunately, the answer appears to be "no", even though the distribution seems to have a certain regularity for low values of the number of classes N. For larger numbers of classes, this regularity is less evident. We illustrate this phenomenon with the results obtained for instance 1.

Table 9 Results of $LRCVF(p, 50)$ for STP as % from RP

Instance	VSS	$LRCVF(p, 50)$			
1	∞	$p \le 10 : \infty$	$11 \le p \le 14 : 0.006$		$p \ge 15 : 0$
2	∞	$p \le 8 : \infty$	$9 \le p \le 10 : 0.164$	$11 \le p \le 14 : 0.006$	$p \ge 15 : 0$
3	∞	$p \le 10 : \infty$	$11 \le p \le 14 : 0.008$		$p \ge 15 : 0$
4	∞	$p \le 10 : \infty$	$11 \le p \le 14 : 0.084$		$p \ge 15 : 0$
5	∞	$p \le 10 : \infty$	$11 \le p \le 14 : 0.001$		$p \ge 15 : 0$
6	∞	$p \le 10 : \infty$			$p \ge 11 : 0$
7	∞	$p \le 10 : \infty$	$11 \le p \le 14 : 0.010$		$p \ge 15 : 0$
8	∞	$p \le 10 : \infty$	$11 \le p \le 14 : 0.006$		$p \ge 15 : 0$
9	∞	$p \le 10 : \infty$	$11 \le p \le 14 : 0.002$		$p \ge 15 : 0$
10	∞	$p \le 10 : \infty$	$11 \le p \le 14 : 0.002$		$p \ge 15 : 0$

Table 10 Results of $LRCVF(p, 100)$ for STP as % from RP

Instance	VSS	$LRCVF(p, 100)$			
1	∞	$p \le 19 : \infty$	$p = 20 : 0.035$	$21 \le p \le 27 : 0.006$	$p \ge 28 : 0$
2	∞	$p \le 15 : \infty$	$p = 16 : 0.173$	$17 \le p \le 19 : 0.164$	$p = 20 : 0.039$
				$21 \le p \le 27 : 0.006$	$p \ge 28 : 0$
3	∞	$p \le 19 : \infty$	$p = 20 : 0.021$	$21 \le p \le 27 : 0.008$	$p \ge 28 : 0$
4	∞	$p \le 19 : \infty$	$p = 20 : 0.188$	$21 \le p \le 27 : 0.084$	$p \ge 28 : 0$
5	∞	$p \le 19 : \infty$	$p = 20 : 0.017$	$21 \le p \le 27 : 0.001$	$p \ge 28 : 0$
6	∞	$p \le 19 : \infty$	$p = 20 : 0.027$		$p \ge 21 : 0$
7	∞	$p \le 19 : \infty$	$p = 20 : 0.042$	$21 \le p \le 27 : 0.010$	$p \ge 28 : 0$
8	∞	$p \le 19 : \infty$	$p = 20 : 0.049$	$21 \le p \le 27 : 0.006$	$p \ge 28 : 0$
9	∞	$p \le 19 : \infty$	$p = 20 : 0.020$	$21 \le p \le 27 : 0.002$	$p \ge 28 : 0$
10	∞	$p \le 19 : \infty$	$p = 20 : 0.027$	$21 \le p \le 27 : 0.002$	$p \ge 28 : 0$

Figure 2 displays the histograms of the distribution of the reduced costs. The graphs show how, up to N=10, the distribution has almost a Gumbel shape. Its behavior becomes very irregular when increasing the number of classes to 50 and then to 100, and a probability distribution is difficult to be identified. Similar observations were made for other instances and problem classes.

We observed feasibility issues when fixing subsets of variables from the deterministic solution, following the computation of $LRCVF(1, 3)$, $LRCVF(p, 10)$, with $p = 1, 2$, $LRCVF(p, 50)$ with $p = 1, \ldots, 10$ and $LRCVF(p, 100)$ with $p = 1, \ldots, 19$. We therefore performed a sensitivity analysis on the values of a number of parameters, the stochastic demand d_j^s and the minimum capacity requirement capacity a_k of supplier $k \in \mathscr{K}$, aiming to obtain the largest set of variables from the deterministic solution that causes infeasibility in the stochastic one.

The results of this analysis show that the infeasibility comes out in classes $LRCVF(p, 100)$, with $p = 1, \ldots, 19$, for $a_k < 16.13\%\ v_k$, $k = 4, 6, 10, 11, 15, 16$ and $a_1 < 2000$ (Table 24), since too large a number of variables were fixed not allowing to satisfy the constraint on the minimum capacity requirement. For $a_k = 16.13\%\ v_k$, $k = 4, 6, 10, 11, 15, 16$ and $a_1 = 2000$, the stochastic problem itself becomes infeasible and, consequently, also all $LRCVF(p, 100)$,

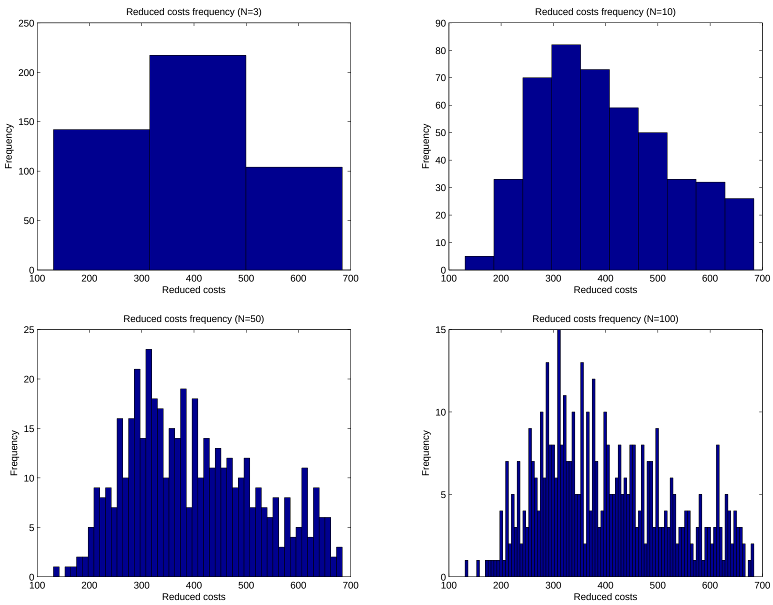

Fig. 2 Absolute frequency of reduced costs of out of basis variables in the EV solution for the STP instance 1, for $N = 3, 10, 50, 100$

$p = 1, \ldots, 100$. The reason of the infeasibility is that when the value of the minimum capacity requirement a_k is increased, the model decides to transport at least for the required quantity. Consequently, for a scenario with low demand, the constraint limiting the maximum storage capacity at the customers is no longer satisfied, generating the infeasibility. On the other hand, high demand scenarios will not bring infeasibility since the model includes the possibility to acquire extra product from external sources at a higher price.

Histograms of the distributions of the reduced costs are plotted in Figure 3 for $N = 3, 10, 50, 100$. First, one may notice how, when considering higher values of N, the classes containing the largest number of variables become the ones in the middle and the left tail, i.e., the classes characterized by the lowest reduced costs. Moreover, the results show empirical evidence that the $LRCVF$ is stable also from the point of view of feasibility, if one does not try to fix p close to 1, i.e., one fixes to 0 the largest part of non basic variables.

5 General Trends and Skeleton of a Heuristic Procedure

The detailed results presented in Section 4.3 show how the $LRCVF$ can be used to derive the structure of the stochastic solution (or a good part of it, at least) starting from data extracted from the continuous relaxation of the expected-value solution. We summarize in this section the lessons learned from

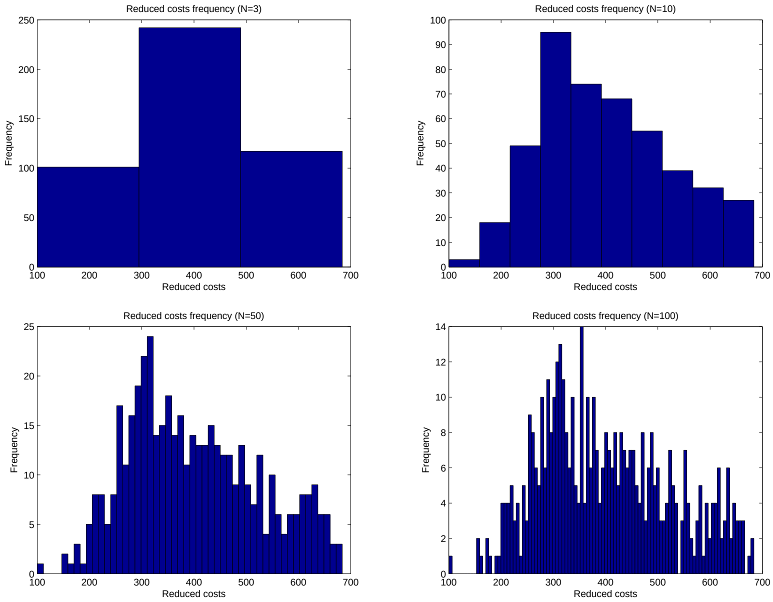

Fig. 3 Absolute frequency of reduced costs of out of basis variables in the EV solution for the STP instance 1, for $N = 3, 10, 50, 100$, with $a_1 = 2000$ and $a_k = 16.13\%\ v_k$, $k = 4, 6, 10, 11, 15, 16$.

our experiments applying $LRCVF$ to different problems, considering the issues of computational effort, feasibility and optimality. We then sketch in Section 5.1 the skeleton of a heuristic method to use $LRCVF$ in an iterative way and an algorithmic procedure. The method is applied to a wide set of additional instances normally used in the literature to test stochastic programming solvers (Sections 5.2 and 5.3). Trends and perspectives are also highlighted (Section 5.4).

5.1 Toward an algorithmic procedure for stochastic programming

We derive a hint from the cases studied above on how to proceed when we want to apply $LRCVF$ to a new problem. The core heuristic idea is the following:

- Solve the (continuous relaxation of the) deterministic version of the original problem;
- Divide the resulting reduced costs in N intervals and fix in the stochastic formulation, the first stage variables belonging to the last class only, i.e., the non basic variables with highest reduced costs;
- If feasibility issues appear, split the interval again into N sub-intervals; then fix in the stochastic formulation, to zero only the first stage variables belonging to the new N class and so on.

When feasibility issues do not appear, the process to obtain the variables to fix in the stochastic problem is summarized in Algorithm 1. The procedure begins by solving the expected value problem EV, finding its first stage solution $\bar{x}(\bar{\xi})$, with minimum and maximum reduced costs $r^{\min}$ and r^{max} associated to the non basic variables. It initializes parameters $p = N = N_0$ (line: 1) and computes the corresponding $LRCVF(p, N)$ (line: 2). In the main loop (lines: 3 to 12), $LRCVF(p, N)$ is updated until it is equal to 0 or, parameter p reaches the value 1. The algorithm provides the variables to fix to zero in the stochastic solution, i.e. the ones with indices belonging to $\mathcal{J}_p, \ldots, \mathcal{J}_N$. It is important to realize that the value to which parameter N is fixed greatly influences the overall numerical effort involved.

Algorithm 1 Using $LRCVF(p, N)$

Require: RP, $\bar{x}(\bar{\xi})$, $N_0 \in \mathbb{N}$, $r^{\min}$, $r^{\max}$,
1: $p = N = N_0$,
2: $LRCVF(p, N) := RCVF(p, N) - RP$,
3: **while** $1 < p \leq N$ **do**
4:     **if** $LRCVF(p, N) = 0$ **then**
5:         $p = p - 1$,
6:         $LRCVF(p, N) = RCVF(p, N) - RP$,
7:     **else**
8:         $N = N \cdot N_0$, $p = p \cdot N_0$,
9:         $LRCVF(p, N) = RCVF(p, N) - RP$,
10:     **end if**
11: **end while**
12: **return** $\mathcal{J}_p, \ldots, \mathcal{J}_N$.

We illustrate this heuristic idea on a wide set of SIPLIB problems (Ahmed et al., 2015).

5.2 SIPLIB instances

The widely-available SIPLIB library is a collection of test problems used to facilitate computational and algorithmic research in stochastic integer programming. We use the problems characterized by a two-stage formulation and the presence of integer, or binary, variables in the first stage:

$DCAP$ test set is a collection of stochastic integer programs arising in dynamic capacity acquisition and allocation under uncertainty. All problem instances have complete recourse, mixed-integer first-stage variables, pure binary second-stage variables, and discrete distributions.

$SSLP$ test set consists of two-stage stochastic mixed-integer programs arising in server location under uncertainty. The problems have pure binary first-stage variables, mixed-binary second-stage variables, and discrete distributions.

$SEMI$ test set consists of instances of a two-stage multi-period stochastic integer problem arising in the planning of semiconductor tool purchases. The instances have mixed-integer first-stage variables and continuous second-stage variables.

$mpTSP_s$ test set are instances of the multi-path Traveling Salesman Problem with stochastic travel times ($mpTSP_s$), a variant of the deterministic TSP, where each pair of nodes is connected by several paths and each path entails a stochastic travel time. The problem, arising in the domain of City Logistics, aims to find an expected minimum Hamiltonian tour connecting all nodes (Maggioni et al., 2014b; Tadei et al., 2014). These problems are large and highly combinatorial, reaching easily more than 1000 binary variables (fixing the 1st-stage variables still leaves binary 2nd-stage variables). Moreover, the continuous relaxation is highly degenerated.

Table 11 details the instance sizes, where the *Type* indicates the range in terms of number of integer/binary variables: S, between 0 and 100; M, between 100 and 1000; and L, greater than 1000. Columns 4-5 and 6-7 display for the first and second stages the total number of variables and the total number of integer variables, respectively. The last column gives the number of scenarios.

Table 11 SIPLIB instance set description

Problem	Inst #	Type	1-stage v.	2-stage v.	Int. 1-stage v.	Int. 2-stage v.	$\|S\|$
$DCAP$	12	S	12	[25,35]	6	[25,35]	[200,500]
$SSLP$	10	M	[5,15]	[100,700]	[5,15]	[100,700]	[5,2000]
$SEMI$	3	M	614	9800	612	0	[2,4]
$mpTSP_s$	5	L	[2500,10000]	[7500, 30000]	[2500,10000]	[7500, 30000]	100

Table 12 Results of SIPLIB *DCAP* instances

Instance	% VSS	% *LRCVF*	Time RP (ss)	Time *LRCVF* (ss)	p
dcap233.200	12.6	0	18.15	5.56	3
dcap233.500	44.2	0	15.89	29.85	3
dcap243.200	32.9	1.28	51.68	3.03	3
dcap243.300	1.6	0.01	23.48	46.38	3
dcap243.500	17.4	0.86	83.84	24.68	3
dcap332.200	57.3	7.16	133.34	5.4	3
dcap332.300	28.9	3.84	141.32	14.39	3
dcap332.500	24.7	8.82	199.67	48.2	3
dcap342.200	35.4	9.41	131.94	8.15	3
dcap342.300	48.6	6.96	493.15	15.68	3
dcap342.500	61.5	6.5	349.98	25.1	3
Mean	33.17	4.08	149.31	20.58	

5.3 SIPLIB computational results

We discuss in the following the results of $LRCVF(p, N)$ with respect to $LUSS$ and VSS for the set of SIPLIB library instances. The results obtained on the SIPLIB problems have been integrated to the SIPLIB library.

Results were obtained using the best known solutions of the RP, i.e., the proven optima for all the instances. Table 12 summarizes the results obtained for the $DCAP$ instances, where Column 1 gives the instance name, Columns 2-3 show the gaps (in %) relative to the optimal values of the stochastic formulation (the RP) for the VSS and the $LRCVF$ at the end of the heuristic idea presented in Section 5.1, while Columns 3-4 display the corresponding computational times in CPU seconds. Finally, Column 5 gives the value of the class $p = 1, \ldots, N$ at the end of the process. Notice that a value of $p = N_0 = 3$ means that the heuristic idea in Section 5.1 stopped at the first iteration without feasibility issues. The reason of our choice relies in the tests performed on the other problems.

The results illustrate how the first-stage solution obtained by solving the expected-value problem fails to provide a good solution in the stochastic case, with a VSS mean error of 31%. With the proposed measure and procedure, we obtain a deviation from the proven optima of 4% and $p = 3$. We also obtain a large reduction of the computational effort (about 7 times on average). These reductions reach one order of magnitude on the largest instances.

The $SSLP$ results are reported in Table 13 (same column definitions as the previous table). The VSS is very high (47% on average). This means that, the EV problem preserves the structure in terms of basic and non basic variables, but fails in providing the correct first-stage basic values. Our procedure is able to find the same RP solution while reducing the computational time by a factor of 3. This is far from marginal considering that, when the size of the instances increases, solving the full stochastic formulation reaches 10 000 seconds, while we find the optimal solution within a computational time that, in the instances with the largest RP CPU times, is reduced 5 times.

Table 13 Results for SIPLIB *SSLP* instances

Instance	% VSS	% *LRCVF*	Time RP (ss)	Time *LRCVF* (ss)	*p*
sslp5.25.50	43.36	0	0.8	0.6	3
sslp5.25.100	42.83	0	1.5	1.5	3
sslp10.50.50	30.43	0	920.8	498.6	3
sslp10.50.100	31.64	0	3608.9	2711.8	3
sslp10.50.500	32.24	0	3620	2702.2	3
sslp10.50.1000	32.16	0	10936	2705.2	3
sslp10.50.2000	32.93	0	40683	4618.4	3
sslp15.45.5	78.68	0	3.8	9.8	3
sslp15.45.10	74.24	0	8.5	6.3	3
sslp15.45.15	73.48	0	262.2	83.1	3
Mean	47.19	0	6004.6	1333.7	

A similar behavior is observed for the $SEMI$ instances, as displayed in Table 14 (same organization as the previous table). The VSS is providing a relatively good gap (close to 5% on average). Again, we find the optimal results at the first iteration. In this case, the gain in terms of computational effort is somewhat limited, being reduced by a factor of 2 only.

Table 14 Results for SIPLIB *SEMI* instances

Instance	% VSS	% *LRCVF*	Time RP (ss)	Time *LRCVF* (ss)	*p*
semi2	3.59	0	2164	1701.1	3
semi3	4.38	0	8914	5568	3
semi4	5.68	0	27519	15923	3
Mean	4.55	0	12865.67	7730.62	3

Up to now, we examined the effect of our algorithm on small and medium-sized instances. What about larger-sized instances? Are we able to replicate optimal or near optimal values while reducing the computational effort? The answer is "yes" to both questions, as can be seen in Table 15 for the SIPLIB $mpTSP_s$ instances. Once again, we replicate the optimal values. In this case, the computational effort is reduced by a full order of magnitude. This means that problems usually not solvable by a MIP solver in a reasonable computational time can be solved giving optimal or near optimal solutions.

Table 15 Results for SIPLIB $mpTSP_s$ instances

Instance	% VSS	% *LRCVF*	Time RP (ss)	Time *LRCVF* (ss)	*p*
D0.50	4.22	0	473.3	265.05	3
D1.50	4.88	0	137.4	127	3
D2.50	2.05	0	655.5	411.1	3
D3.50	3.75	0	2069.1	567.8	3
D1.100	4.22	0	12376	256.1	3
Mean	3.82	0	3142.26	325.41	3

To conclude, it is clear how the $LRCVF$ can be effectively used to find high quality solutions to stochastic problems by starting from the EV solutions. Furthermore, when compared to the effort needed to find the optimal solution to the full stochastic formulation, $LRCVF$ considerably reduces the computational times. Finally, it might be incorporated in an iterative heuristic algorithm providing high quality solutions. As a further heuristic insight, when one desires a greater precision and the number of non basic variables appearing in the continuous relaxation of the deterministic approximation is sufficiently large, the set of non basic variables may be split into 10 bids and the values $p = 2, 3$, and 4 seem appropriate.

5.4 General trends

One of the main issues that emerges when using $LRCVF$ is how to choose the number N of classes dividing the reduced costs of non-basic variables. On the one hand, it would be preferable to fix the largest possible number of variables in order to reduce the problem size, and, on the other hand, fixing too large a number may result in errors in terms of feasibility and optimality. The general trend emerging from the empirical observations is that fixing to 0 about a third of the non-basic variables with the highest reduced costs is a good compromise. Indeed, applying this policy, we reached the optimal stochastic solutions without feasibility issues and reducing the computational time up to two orders of magnitude for the largest instance (Table 15).

From the point of view of problem optimality, it seems that, as already noted for deterministic combinatorial models (Perboli et al. (2011)), the reduced costs of the deterministic solution give an hint on the variables to make inactive in the stochastic program. Moreover, the results show that, even when the VSS is high and the objective function of the expected-value deterministic model far from the one of the stochastic problem, the expected-value deterministic solution provides correct information about the optimal stochastic solution. On the other hand, problems with just a few variables with positive reduced costs in their deterministic solution (e.g., the $DCAP$ instances in SIPLIB, Table 12) highlight the need to extend the $LRCVF$ approach by defining a measure for ranking also the basic variables associated to the continuous relaxation of the expected value deterministic problem.

It is interesting to note that the $LRCVF$ idea of fixing to 0 the non basic variables with the highest reduced costs in the solution to the deterministic version, which appears to perform so well, is the complete opposite of what may be seen in a number of approaches for deterministic combinatorial optimization, where the search for the variables to fix starts with the smallest reduced costs. A possible explanation is that one has to remove a lot of variables in order to obtain a substantial reduction in computational effort in the deterministic combinatorial case, while removing just a small subset of non basicvariables from a stochastic program, pays a lot in terms of computational effort (see the results in Table 15).

6 Conclusions and Future Directions

In this paper, we analyzed the *quality* of the expected value solution with respect to the stochastic one, in particular the part of its structure that could be relevant for the solution to the original stochastic formulation.

We introduced the *Loss of Reduced Costs-based Variable Fixing*, *LRCVF*, a new measure of goodness/badness of the deterministic solution that goes beyond the standard measures. *LRCVF* takes into account the information on the reduced costs of non basic variables in the deterministic solution. These costs are sorted, grouped into homogeneous classes, and, then, those in classes with the highest reduced costs are fixed in the associated stochastic formulation, significantly reducing its size and, thus, the associated computational effort. We examined the relations between the new measure and traditional ones, and provided the procedure to compute it, as well as the skeleton of a heuristic method using it to address two-stage combinatorial stochastic optimization problems.

We performed a wide range of experiments on instances drawn both from the Stochastic Programming literature and from real cases. The experiments provided the opportunity of a deeper understanding of the relations between the deterministic and stochastic solutions, of the main causes of goodness in the variables with highest reduced cost in the deterministic solution, measured by $0 = LRCVF(p, N) \leq LUSS \leq VSS$. The results also showed that the *LRCVF* can help identify the good and bad variables from the deterministic solution to fix in the stochastic formulation. In all the cases considered, fixing the variables with *high* reduced costs when solving the *RP* allowed us to reach exactly the stochastic solution.

The proposed *LRCVF* measure and algorithmic procedure can hence be effectively used both for problems actually solvable but that must be run very often, and for intractable real-world problems, to reduce the computational time of solving the stochastic problem, without loosing in terms of solution quality.

The proposed methodology and the results obtained also point to how a smart usage of the information coming from the linear programming theory can be effectively incorporated in a Stochastic Programming resolution approach in order to build accurate solutions. The introduction of the *LRCVF* thus opens a number of interesting future research directions, including how to extend and incorporate this idea into various algorithmic frameworks, such as progressive hedging, diving procedures, etc.

The computational analysis performed in this paper points to a second avenue: when classifying the non basic variables according to their reduced costs, the ones with the *highest* values can be hard-fixed to zero without affecting the stochastic solution. On the other side, the variables with the *lowest* reduced costs should be present in the stochastic model. But what about the variables which are in between these two extreme classes? Can we define a way to identify those hedging variables and to incorporate this? What is the appropriate number N of classes needed and their usage (which ones to be fixed and which ones not)? A related, but different research avenue concerns the case, studied within the branch-and-bound literature for deterministic formulations, of identifying a measure of the willingness to fix a basic variable and how to fix it. We expect to report on some of these issues in the near future.

Acknowledgments

Mircea Simionica and Daniele Magnaldi are acknowledged for their collaboration.

References

S. Ahmed, R. Garcia, N. Kong, L. Ntaimo, G. Parija, F. Qiu, and S. S. Siplib: A stochastic integer programming test problem library, 2015. http://www.isye.gatech.edu/ sahmed/siplib.

Ravindra K. Ahuja, Thomas L. Magnanti, and James B. Orlin. *Network Flows: Theory, Algorithms, and Applications*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.

E. Angelelli, R. Mansini, and M. Speranza. Kernel search: A general heuristic for the multi-dimensional knapsack problem. *Computers & Operations Research*, 37(11):2017 – 2026, 2010. ISSN 0305-0548. doi: http://dx.doi.org/10.1016/j.cor.2010.02.002. URL `http://www.sciencedirect.com/science/article/pii/S0305054810000328`.

G. Bayraksan and D. P. Morton. Assessing solution quality in stochastic programs. *Mathematical Programming*, 108(2-3), 2006.

L. Bertazzi and F. Maggioni. Solution approaches for the stochastic capacitated traveling salesmen location problem with recourse. *Journal of Optimization Theory and Applications*, 166(1):321–342, Jul 2015. ISSN 1573-2878. doi: 10.1007/s10957-014-0638-z. URL `https://doi.org/10.1007/s10957-014-0638-z`.

L. Bertazzi and F. Maggioni. A stochastic multi-stage fixed charge transportation problem: Worst-case analysis of the rolling horizon approach. *European Journal of Operational Research*, pages –, 2017. ISSN 0377-2217. doi: https://doi.org/10.1016/j.ejor.2017.12.004. URL `https://www.sciencedirect.com/science/article/pii/S037722171731086X`.

J. Birge. The value of the stochastic solution in stochastic linear programs with fixed recourse. *Mathematical Programming*, 24(1):314–325, 1982. ISSN 0025-5610. doi: 10.1007/BF01585113. URL `http://dx.doi.org/10.1007/BF01585113`.

J. R. Birge. Aggregation bounds in stochastic linear programming. *Mathematical Programming*, 31(1):25–41, 1985. ISSN 1436-4646. doi: 10.1007/BF02591859. URL `http://dx.doi.org/10.1007/BF02591859`.

J. R. Birge and F. Louveaux. *Introduction to stochastic programming, 2nd edition*. Springer, Princeton, NJ, USA, 2011.

L. F. Escudero, A. Garín, M. Merino, and G. Pérez. The value of the stochastic solution in multistage problems. *TOP*, 15(1):48–64, 2007. ISSN 1863-8279. doi: 10.1007/s11750-007-0005-4. URL http://dx.doi.org/10.1007/s11750-007-0005-4.

K. Frauendorfer. Solving slp recourse problems with binary multivariate distributionsï¿½the dependent case. *Math Oper Res*, 13(3):377–394, 1988.

A. A. Gaivoronski. Stochastic optimization problems in telecommunications. In S. W. Wallace and W. T. Ziemba, editors, *Applications of Stochastic Programming*, volume 5 of *MPS/SIAM Series on Optimization*, pages 669–704. SIAM, Philadelphia, PA, 2005.

L. Garver. Power generation scheduling by integer programming-development of theory. *Power Apparatus and Systems, Part III. Transactions of the American Institute of Electrical Engineers*, 81(3):730–734, April 1962. ISSN 0097-2460. doi: 10.1109/AIEEPAS.1962.4501405.

D. Hausch and W. Ziemba. Bounds on the value of information in uncertain decision problems ii. *Stochastics*, 10:181–217, 1983.

C. Huang, I. Vertinsky, and W. Ziemba. Sharp bounds on the value of perfect information. *Oper Res*, 25(1):128–139, 1977a.

C. Huang, W. Ziemba, and B.-T. A. Bounds on the expectation of a convex function of a random variable: with applications to stochastic programming. *Oper Res*, 25(2):315–325, 1977b.

J. Jensen. Sur les fonctions convexes et les inï¿½galitï¿½s entre les valeurs moyennes. *Acta Mathematica*, 30(1):175–193, 1906.

P. Kall and S. Wallace. *Stochastic Programming*. John Wiley & Sons, 1994.

A.J. King and S.W.. Wallace. *Modeling with Stochastic Programming*. Springer, New York, 2012.

D. Kuhn. *Generalized Bounds for Convex Multistage Stochastic Programs*, volume 548 of *Lecture Notes in Economics and Mathematical Systems*. Spinger Verlag Berlin Heidelberg, 2005.

D. Kuhn. Aggregation and discretization in multistage stochastic programming. *Mathematical Programming*, 113(1):61–94, 2008. ISSN 1436-4646. doi: 10.1007/s10107-006-0048-6. URL http://dx.doi.org/10.1007/s10107-006-0048-6.

A. G. Lium, T. G. Crainic, and S. W. Wallace. A study of demand stochasticity in service network design. *Transportation Science*, 43(2):144–157, 2009. doi: 10.1287/trsc.1090.0265. URL http://dx.doi.org/10.1287/trsc.1090.0265.

A. Madansky. Inequalities for stochastic linear programming problems. *Manag Sci*, 6:197–204, 1960.

F. Maggioni and G. C. Pflug. Bounds and approximations for multistage stochastic programs. *SIAM Journal on Optimization*, 26(1):831–855, 2016. doi: 10.1137/140971889. URL http://dx.doi.org/10.1137/140971889.

F. Maggioni and S. Wallace. Analyzing the quality of the expected value solution in stochastic programming. *Annals of Operations Research*, 200:37–54, 2012.

F. Maggioni, M. Kaut, and L. Bertazzi. Stochastic optimization models for a single-sink transportation problem. *Computational Management Science*, 6: 251–267, 2009.

F. Maggioni, E. Allevi, and M. Bertocchi. Bounds in multistage linear stochastic programming. *Journal of Optimization Theory and Applications*, 163 (1):200–229, 2014a. ISSN 0022-3239. doi: 10.1007/s10957-013-0450-1. URL `http://dx.doi.org/10.1007/s10957-013-0450-1`.

F. Maggioni, G. Perboli, and R. Tadei. The multi-path traveling salesman problem with stochastic travel costs: Building realistic instances for city logistics applications. *Transportation Research Procedia*, 3:528–536, 2014b.

F. Maggioni, E. Allevi, and M. Bertocchi. Monotonic bounds in multistage mixed-integer stochastic programming. *Computational Management Science*, 13(3):423–457, Jul 2016. ISSN 1619-6988. doi: 10.1007/s10287-016-0254-5. URL `https://doi.org/10.1007/s10287-016-0254-5`.

F. Maggioni, F. A. Potra, and M. Bertocchi. A scenario-based framework for supply planning under uncertainty: stochastic programming versus robust optimization approaches. *Computational Management Science*, 14(1):5–44, Jan 2017. ISSN 1619-6988. doi: 10.1007/s10287-016-0272-3. URL `https://doi.org/10.1007/s10287-016-0272-3`.

W.-K. Mak, D. P. Morton, and R. Wood. Monte carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters*, 24(1â€“2):47 – 56, 1999. ISSN 0167-6377. doi: http://dx.doi.org/10.1016/S0167-6377(98)00054-6. URL `http://www.sciencedirect.com/science/article/pii/S0167637798000546`.

G. Perboli, R. Tadei, and D. Vigo. The two-echelon capacitated vehicle routing problem: Models and math-based heuristics. *Transportation Science*, 45: 364–380, 2011.

A. Prékopa. *Stochastic Programming.* Springer, 1995.

C. H. Rosa and S. Takriti. Improving aggregation bounds for two-stage stochastic programs. *Operations Research Letters*, 24(3):127 – 137, 1999. ISSN 0167-6377. doi: http://dx.doi.org/10.1016/S0167-6377(99)00019-X. URL `http://www.sciencedirect.com/science/article/pii/S016763779900019X`.

B. Sandıkçı, N. Kong, and A. J. Schaefer. A hierarchy of bounds for stochastic mixed-integer programs. *Mathematical Programming*, 138(1):253–272, 2013. ISSN 1436-4646. doi: 10.1007/s10107-012-0526-y. URL `http://dx.doi.org/10.1007/s10107-012-0526-y`.

R. Tadei, G. Perboli, and F. Perfetti. The multi-path traveling salesman problem with stochastic travel cost. *EURO Journal on Transportation and Logistics*, forthcoming:doi:10.1007/s13676–014–0056–2, 2014. doi: 10.1007/s13676-014-0056-2.

B. Thapalia, T. Crainic, M. Kaut, and S. Wallace. Single-commodity stochastic network design with multiple sources and sinks. *INFOR*, 49(3):195–214, 2011.

B. Thapalia, T. Crainic, M. Kaut, and S. Wallace. Single-commodity stochastic network design with random edge capacities. *European Journal of Operational Research*, 9:139–160, 2012a.

B. Thapalia, S. Wallace, M. Kaut, and T. Crainic. Single source single-commodity stochastic network design. *Computational Management Science*, 9(1):139–160, 2012b. ISSN 1619-697X. doi: 10.1007/s10287-010-0129-0. URL http://dx.doi.org/10.1007/s10287-010-0129-0.

C. T. Wang, X. and S. Wallace. Stochastic scheduled service network design: The value of deterministic solutions. Technical report, Publication CIRRELT-2016-14, Centre interuniversitaire de recherche sur les r´eseaux d'entreprise, la logistique et le transport, Universit´e de Montr´eal, Montr´eal, QC, Canada., 2016.

H. Williams. *Model buiding in mathematical programming*. Wiley & Sons, 2013.

Annex

A Test Problem Description

A.1 A single-sink transportation problem

This problem is inspired by a real case of *clinker* replenishment, provided by the largest Italian cement producer located in Sicily Maggioni et al. (2009). The logistics system is organized as follows: *clinker* is produced by four plants located in Palermo (PA), Agrigento (AG), Cosenza (CS) and Vibo Valentia (VV) and the warehouse to be replenished is in Catania. The production capacities of the four plants, as well as the demand for clinker at Catania, are considered stochastic.

All the vehicles are leased from an external transportation company, which we assume to have an unlimited fleet. The vehicles must be booked in advance, before the demand and production capacities are revealed. Only full-load shipments are allowed. When the demand and the production capacity become known, there is an option to cancel some of the bookings against a cancellation fee α. If the quantity delivered from the four suppliers using the booked vehicles is not enough to satisfy the demand in Catania, the residual quantity is purchased from an external company at a higher price b. The problem is to determine, for each supplier, the number of vehicles to book in order to minimize the total costs, given by the sum of the transportation costs (including the cancellation fee for vehicles booked but not used) and the costs of the product purchased from the external company. The notation adopted is:

$$\mathscr{I} = \{i : i = 1, \ldots, I\} : \text{set of suppliers (AG, CS, PA, VV)} ;$$
$$\mathscr{S} = \{s : s = 1, \ldots, S\} : \text{set of scenarios.}$$

t_i : unit transportation costs of supplier $i \in \mathscr{I}$;
c_i : unit production costs of supplier $i \in \mathscr{I}$;
b : buying cost from an external source (we assume that $b > \max_i(t_i + c_i)$) ;
q : vehicle capacity ;
g : maximum capacity that can be booked ;
l_0 : initial inventory level at the customer ;
$l_{\max}$: storage capacity at the customer ;
p^s : probability of scenario $k \in \mathscr{S}$;
a_i^s : production capacity of supplier $i \in \mathscr{I}$in scenario $s \in \mathscr{S}$;
d^s : customer demand at scenario $s \in \mathscr{S}$;
α : cancellation fee ;

with the decision variables

$x_i \in \mathbb{N}$: number of vehicles booked from supplier $i \in \mathscr{I}$;
$z_i^s \in \mathbb{N}$: number of vehicles actually used from $i \in \mathscr{I}$ in $s \in \mathscr{S}$;
y^s : product to purchase from an external source in scenario $s \in \mathscr{S}$;

In the two-stage (one-period) case, we get the following mixed-integer stochastic programming model with recourse:

$$\min \; q \sum_{i=1}^{I} t_i x_i + \sum_{s=1}^{S} p^s \Big[b\, y^s - (1-\alpha) q \sum_{i=1}^{I} t_i \left(x_i - z_i^s \right) \Big] \tag{20}$$

$$\text{s.t.} \qquad q\sum_{i=1}^{I} x_i \leq g, \tag{21}$$

$$l_0 + \sum_{i=1}^{I} qz_i^s + y^s - d^s \geq 0\ , \qquad s \in \mathscr{S}, \tag{22}$$

$$l_0 + \sum_{i=1}^{I} qz_i^s + y^s - d^s \leq l_{\max}\ , \qquad s \in \mathscr{S}, \tag{23}$$

$$z_i^s \leq x_i\ , \qquad i \in \mathscr{I},\ s \in \mathscr{S}, \tag{24}$$

$$qz_i^s \leq a_i^s\ , \qquad i \in \mathscr{I},\ s \in \mathscr{S}, \tag{25}$$

$$x_i \in \mathbb{N}\ , \qquad i \in \mathscr{I}, \tag{26}$$

$$y^s \geq 0\ , \qquad s \in \mathscr{S}, \tag{27}$$

$$z_i^s \in \mathbb{N}\ , \qquad i \in \mathscr{I},\ s \in \mathscr{S}. \tag{28}$$

The first sum in the objective function (20) is the booking costs of the vehicles, while the second sum represents the expected cost associated to the recourse actions, consisting of buying extra clinker (y^s) and canceling unwanted vehicles. Constraint (21) guarantees that the number of booked vehicles from the suppliers to the customer is not greater than g/q. Constraints (22) and (23) ensure that the second-stage storage level is between zero and $l_{\max}$. Constraints (24) guarantee that the number of vehicles serving supplier i is at most equal to the number of vehicles booked in advance, and constraints (25) control that the quantity of clinker delivered from supplier i does not exceed its production capacity a_i^s. Finally, (26)–(28) define the decision variables of the problem (both for the first and second stages).

The goal is to find, for each supplier, the number of vehicles to book at the beginning of the first period.

A.2 Power generation scheduling

This real-case problem is based on an economic scheduling model formulated in Williams (2013) and Garver (1962) as a deterministic mixed integer program. Power generation scheduling involves the selection of generating units to be put into operation and the allocation of the power demand among the units over a set of time periods. In the problem considered, there are two types of generating units available (i.e., four units of type 1 and four units of type 2). Each type is defined according to specific technical characteristics and operational costs. Therefore, a generating unit will run at a level that is between a minimum and a maximum threshold, these threshold values being type specific. When a unit is used, there is a base hourly cost that is charged for running it at the minimum level. In the case where a unit runs above the minimum threshold, an extra hourly cost is applied for each additional megawatt. There is also a starting up cost that is charged each time a new generating unit is used. Once again, all specific cost values vary according to the unit types.

At any considered time period, there must be a sufficient number of operating generators to meet a possible increase in the overall demand of up to 15%. In the event of an increase, the running levels of the used units are simply adjusted to meet the new demand requirements. In the present problem, two time periods are considered. While the demands of the first time period are assumed known, the demands in the second time period are stochastic. Therefore, the problem is formulated as a two-stage stochastic model. In the first stage, a set of generating units are chosen and their operating levels are fixed for the two time periods defined in the problem (an estimate is used here for the demands in the second period). In the second stage, the actual values of the demands in the second period are observed and the number units and their operating levels are adjusted accordingly. Production decisions are thus made after the demands have been revealed. Instead of writing the model in terms of scenarios, we consider a node formulation defined on the structure of the scenario tree (see Table 20 in Annex). Therefore, nodes $n = 1, 2$ represent the first stage of the model, while nodes $n = 3, \ldots, 22$ define the 20 considered scenarios that can be observed in the second stage. For each node n in the scenario tree, value $pa(n)$ defines its predecessor.

We now define the model that is considered. To do so, let us first define the general notation that is used:

$$\mathscr{I} = \{i : i = 1, \ldots, I\} : \text{types of generating units;}$$
$$\mathscr{N} = \{n : n = 1, \ldots, N\} : \text{ordered set of nodes of the scenario tree;}$$

m_i : minimum output level for generator of type $i \in \mathscr{I}$;
M_i : maximum output level for generator of type $i \in \mathscr{I}$;
D^n : demand in node $n \in \mathscr{N}$;
p^n : probability of node $n \in \mathscr{N}$;
C_i : cost per hour per megawatt (mw) of unit $i \in \mathscr{I}$ for operating above minimum level;
E_i : cost per hour per megawatt (mw) of unit $i \in \mathscr{I}$ for operating at minimum level;
F_i : start-up cost of unit $i \in \mathscr{I}$;
$u_{i,max}$: upper bound on the total number of generators of type $i \in \mathscr{I}$;
u_i^0 : starting value of open units of type $i \in \mathscr{I}$;

The decision variables are:

u_i^n : number of generating units of type $i \in \mathscr{I}$ working in node $n \in \mathscr{N}$;
s_i^n : number of generators of type $i \in \mathscr{I}$ started up in node $n \in \mathscr{N} \setminus \{1\}$;
x_i^n : total output rate from generators of type $i \in \mathscr{I}$ in node $n \in \mathscr{N}$;

The formulation of the generator scheduling problem as an integer program including start-up costs is now defined as follows:

$$\min \sum_{n \in \mathscr{N}} p^n \left[\sum_{i \in \mathscr{I}} C_i \left(x_i^n - m_i u_i^n\right) + \sum_{i \in \mathscr{I}} E_i u_i^n + \sum_{i \in \mathscr{I}} F_i s_i^n \right] \tag{29}$$

$$\text{s.t.} \quad \sum_{i \in \mathscr{I}} x_i^n \geq D^n, \qquad n \in \mathscr{N}, \tag{30}$$
$$x_i^n \geq m_i u_i^n \ , \qquad i \in \mathscr{I},\ n \in \mathscr{N}, \tag{31}$$
$$x_i^n \leq M_i u_i^n \ , \qquad i \in \mathscr{I},\ n \in \mathscr{N}, \tag{32}$$
$$\sum_{i \in \mathscr{I}} M_i u_i^n \geq \frac{115}{110} D^n \ , \qquad n \in \mathscr{N}, \tag{33}$$
$$s_i^n \geq u_i^n - u_i^{pa(n)} \ , \ i \in \mathscr{I} \ , \ n \in \mathscr{N} \setminus \{1\}, \tag{34}$$
$$u_i^1 = u_i^0 \ , \qquad i \in \mathscr{I}, \tag{35}$$
$$u_i^n \leq u_{i,max} \ , \qquad i \in \mathscr{I},\ n \in \mathscr{N}, \tag{36}$$
$$x_i^n \geq 0 \ , \qquad i \in \mathscr{I},\ n \in \mathscr{N}, \tag{37}$$
$$s_i^n \in \mathbb{N} \ , \qquad i \in \mathscr{I},\ n \in \mathscr{N} \setminus \{1\}, \tag{38}$$
$$u_i^n \in \mathbb{N} \ , \qquad i \in \mathscr{I},\ n \in \mathscr{N}. \tag{39}$$

The objective function (29) consists in the minimization of the total costs, which include the starting up costs of units and their operational costs (both at the minimum level and above it) for each period. Constraints (30) guarantee that the demand in each period is met, whereas (31) and (32) make sure that the output lies within the limits of the operating generators at all times. Constraints (33) guarantee that, for each period, the additional load requirement of 15% is met without the need to resort to additional generators, while constraints (34) ensure that the number of generators that are started up in node n be equal to the increase in the number of operating units with respect to the node $pa(n)$. Finally, constraints (35)-(36) define the starting values and upper bounds for the number of started up units and (37)-(39) impose the necessary non-negativity and integrality requirements on the decision variables of the problem.

A.3 Supply transportation problem

This problem is inspired by a real case of *gypsum* replenishment in Italy, provided by the primary Italian cement producer, see Maggioni et al. (2017) for more details. The logistic system is organized as follows: 24 suppliers, each of them having several plants located all around Italy, are used to satisfy the demand for gypsum of 15 cement factories belonging to the same company. The demands for gypsum at the 15 cement factories are considered stochastic. As in the first problem considered, shipments are performed by capacitated vehicles, which have to be booked in advance, before the demand is revealed. When the demands become known, there is the option to discount vehicles that were booked but not actually used. However, if the quantity shipped from the suppliers using the booked vehicles is not enough to satisfy the observed demands, vehicle services to transport the extra demand of gypsum directly to the factories can be purchased from an external company at a premium price. The problem is to determine for each of the supplier plants, the number of vehicles to book to replenish in gypsum the factories in order to minimize the total cost. The total cost is defined as the sum of the booking costs for the vehicles used to perform the distribution operations between the plants and the factories (including the discount for vehicles booked but not used), and the costs of the extra vehicles added to satisfy the observed demand. It should be noted that, in all cases, the cost of a vehicle is obtained by multiplying its capacity by a unit cost that either reflects the transportation cost between a plant and a factory (for the vehicles booked in advance), or, the premium rate charged by the external company for a direct transportation service to a factory. Regarding the discount for the vehicles booked but not used, it is expressed as a fixed percentage of the cost associated to the number of unused vehicles.

The notation adopted is the following:

$$\begin{aligned}
\mathscr{K} = \{k : k = 1, \dots, K\} &: \text{set of suppliers;}\\
\mathscr{O}_k = \{i : i = 1, \dots, O_k\} &: \text{set of plant locations of supplier } k \in \mathscr{K};\\
\mathscr{D} = \{j : j = 1, \dots, D\} &: \text{set of cement factories (destinations);}\\
\mathscr{S} = \{s : s = 1, \dots, S\} &: \text{set of scenarios;}
\end{aligned}$$

$$\begin{aligned}
t_{ij} &: \text{unit transportation cost from plant } i \in \mathscr{O}_k, k \in \mathscr{K} \text{ to factory } j \in \mathscr{D};\\
b_j &: \text{the premium rate charged by the external company for a vehicle assigned to factory } j \in \mathscr{D};\\
q &: \text{the capacity of a vehicle;}\\
g_j &: \text{maximum capacity which can be booked for the factory } j \in \mathscr{D};\\
v_k &: \text{maximum requirement capacity of supplier } k \in \mathscr{K};\\
a_k &: \text{minimum requirement capacity of supplier } k \in \mathscr{K};\\
l_{\max} &: \text{storage capacity at the factories;}\\
\alpha &: \text{discount;}\\
p^s &: \text{probability of scenario } s \in \mathscr{S};\\
d_j^s &: \text{demand of factory } j \text{ in scenario } s \in \mathscr{S}.
\end{aligned}$$

The decision variables are

$x_{ij} \in \mathbb{N}$: number of vehicles booked between plant $i \in \mathscr{O}_k,\ k \in \mathscr{K}$ and factory $j \in \mathscr{D}$;
$z^s_{ij} \in \mathbb{N}$: number of vehicles actually used between plant $i \in \mathscr{O}_k,\ k \in \mathscr{K}$ and factory $j \in \mathscr{D}$, for scenario $s \in \mathscr{S}$;
$y^s_j \in \mathbb{N}$: number of extra vehicles used from the external company for factory $j \in \mathscr{D}$, for scenario $s \in \mathscr{S}$.

The two-stage integer stochastic programming model with recourse can now be defined as follows:

$$\min q\sum_{k=1}^{K}\sum_{i=1}^{O_k}\sum_{j=1}^{D} t_{ij}x_{ij} + \sum_{s=1}^{S} p^s \left[\sum_{j=1}^{D} qb_j\, y^s_j - \alpha q\sum_{k=1}^{K}\sum_{i=1}^{O_k}\sum_{j=1}^{D} t_{ij}\left(x_{ij} - z^s_{ij}\right)\right] \tag{40}$$

$$s.t. \qquad q\sum_{k=1}^{K}\sum_{i=1}^{O_k} x_{ij} \le g_j\ , \quad j \in \mathscr{D}, \tag{41}$$

$$0 \le l^0_j + q\Big(\sum_{k=1}^{K}\sum_{i=1}^{O_k} z^s_{ij} + y^s_j\Big) - d^s_j \le l_{\max}\ , \quad j \in \mathscr{D}, \quad s \in \mathscr{S}, \tag{42}$$

$$z^s_{ij} \le x_{ij}\ , i \in \mathscr{O}_k, k \in \mathscr{K}, j \in \mathscr{D}, s \in \mathscr{S}, \tag{43}$$

$$a_k \le q\sum_{i=1}^{O_k}\sum_{j=1}^{D} z^s_{ij} \le v_k\ , \quad k \in \mathscr{K}, \quad s \in \mathscr{S}, \tag{44}$$

$$x_{ij} \in \mathbb{N}\ , i \in \mathscr{O}_k, k \in \mathscr{K}, j \in \mathscr{D}, \tag{45}$$

$$y^s_j \in \mathbb{N}\ , j \in \mathscr{D}, s \in \mathscr{S}, \tag{46}$$

$$z^s_{ij} \in \mathbb{N}\ , i \in \mathscr{O}_k, k \in \mathscr{K}, j \in \mathscr{D}, s \in \mathscr{S}. \tag{47}$$

The first sum in the objective function (40) denotes the booking costs of the vehicles between the plants and the factories, while the second sum represents the expected recourse costs, which include the cost of the extra vehicles provided by the external company and the discount for the unused booked vehicles. Constraints (41) guarantee that, for each factory $j \in \mathscr{D}$, the number of booked vehicles from the suppliers to the factory does not exceed g_j/q. Constraints (42) ensure that the storage levels of factories $j \in \mathscr{D}$ are between zero and $l_{\max}$. Constraints (43) guarantee that the number of vehicles used by the suppliers are at most equal to the number vehicles booked in advance. Constraints (44) ensure that, for all suppliers $k \in \mathscr{K}$, the number of vehicles used allow the volume of product transported to be between the minimum (i.e., a_k) and maximum (i.e., v_k) established requirements. Finally, (45)–(47) define the decision variables of the problem.

B Numerical Data

B.1 A single-sink transportation problem

Problem data are presented in Tables 16–17. Table 16 presents the production and transportation costs for each supplier, together with its distance from the customer in Catania, while Table 17 reports the monthly production capacity of each supplier in the considered period (zero entries represent production site closures due to equipment failure or maintenance).

We used in our computational experiment, the vehicle capacity $q = 30$ tonnes (t), the storage capacity $l_{max} = 35$ kilotonnes (kt) and the *daily* unloading capacity of 1800 t, giving

Table 16 Production costs c_i and transportation costs t_i from Catania.

Supplier	c_i (€/t)	t_i (€/t)
Porto Empedocle (AG)	18.79	11.40
Castrovillari (CS)	9.55	33.00
Isola d. Femmine (PA)	11.00	14.10
Vibo Valentia (VV)	11.54	18.50

Table 17 Monthly production capacity a_i of suppliers $i \in \mathcal{I}$, January 2003 to May 2007, in kilotonnes (kt).

i	Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
	'03	9.1	4.0	11.1	14.6	21.7	14.2	17.4	8.4	24.9	17.4	12.3	13.0
	'04	0.0	4.1	9.0	10.5	9.3	12.2	11.6	13.6	9.4	11.0	9.7	0.0
AG	'05	0.0	9.1	8.3	21.1	15.0	15.1	12.1	13.2	11.3	13.0	7.1	1.2
	'06	1.7	9.5	4.5	14.0	12.5	15.2	11.3	15.9	6.2	11.9	7.2	9.0
	'07	13.0	13.0	19.0	4.0	10.0							
	'03	10.9	14.0	13.9	19.1	14.1	13.0	4.5	0.0	4.0	13.7	9.1	4.5
	'04	8.3	6.3	3.0	0.0	16.2	14.2	12.3	14.4	19.8	19.3	20.0	15.2
CS	'05	15.1	10.8	21.9	19.7	15.3	10.8	6.3	0.0	9.1	23.2	11.7	0.9
	'06	18.7	0.0	8.9	16.0	17.6	13.9	4.8	5.0	14.1	24.3	14.5	8.1
	'07	17.0	8.0	0.0	0.0	10.0							
	'03	15.5	18.1	23.3	12.4	0.5	5.7	12.5	13.5	12.3	10.2	8.3	12.0
	'04	27.1	10.0	12.8	13.8	13.7	14.0	10.6	1.4	10.3	12.6	11.5	16.9
PA	'05	16.0	3.8	10.6	16.6	23.0	27.7	16.7	13.4	16.8	11.1	19.0	22.4
	'06	27.5	21.5	18.6	20.4	0.0	14.0	14.3	11.2	18.4	16.9	9.4	11.1
	'07	11.0	9.0	7.0	6.0	10.0							
	'03	4.9	1.2	12.7	2.7	19.3	11.9	5.4	3.0	14.6	3.4	15.2	2.5
	'04	4.0	9.4	18.3	10.5	13.9	8.6	6.2	4.3	7.2	12.4	9.5	0.0
VV	'05	3.5	21.1	20.8	13.0	23.5	19.1	8.2	8.6	4.6	9.2	16.2	16.0
	'06	8.5	22.3	21.7	15.1	7.4	10.3	0.0	2.5	4.3	5.2	18.3	6.3
	'07	0.0	0.0	0.0	0.0	10.0							

us the *monthly* unloading capacity $g = 21 \times 1800\,\text{t} = 37.8\,\text{kt}$, or 1260 full vehicles. The cost of clinker from an external source was set to $b =$ €45/t and the cancellation fee to $\alpha = 0.5$. For the initial inventory level l_0 at the customer, we have taken the value at the beginning of January 2007, that is $l_0 = 2000\,\text{t}$.

We run the model for 10 different instances with a demand randomly generated in the interval $\left[d^{min}, d^{max}\right]$, where $d^{min} = 20000$ and $d^{max} = 30000$ are respectively the minimum and maximum demand observed in the historical data.

The deviations (in %) from the optimal solutions of the stochastic model are reported in Table 18. Table 19 displays for instance 9 the optimal solution (optimal number of booked vehicles for each supplier and total optimal cost) for the deterministic and the stochastic models, as well as for the various problem types (similar observations and arguments apply to the other instances).

The deterministic model, EV, always books the exact numbers of vehicles needed ($\bar{x}_i = \bar{z}_i^k$, $i \in \mathscr{I}$, $k \in \mathscr{K}$); it sorts the suppliers according to the transportation costs and books a full production capacity from the cheapest one (AG), followed by the next-cheapest (PA). The deterministic model thus books much fewer vehicles than the stochastic one, RP, resulting in a solution costing two-thirds of the stochastic counterpart. The EEV is much higher (it is 481 484.25 € instead of the predicted cost 287 874 €). So,

$$VSS = 481\,484.25 - 410\,573 = 70910.36, \tag{48}$$

which shows that we can save about 17% of the cost by using the stochastic model, compared to the deterministic one.

Table 18 Results for the SSTP

Instance	VSS	% from RP $LRCVF(p,2)$ $p=1$	$p=2$
1	12.28	5.41	0
2	16.18	6.95	0
3	12.93	6.03	0
4	13.40	6.05	0
5	18.11	4.37	0
6	14.27	7.32	0.17
7	17.46	9.13	0.03
8	12.97	4.95	0
9	17.27	8.11	0
10	13.91	3.95	0
Mean	14.88	6.23	0.02

Table 19 Optimal solutions for different problem types of SSTP instance 9

Instance	Problem Type	AG	CS	PA	VV	Objective value (€)
9	EV	206	0	514	0	287 874
	RP	377	0	533	200	410 573
	EEV	206	0	514	0	481 484.25
	$RCVF(1,2) = ESSV$	390	0	633	0	443 881.93
	$RCVF(2,2)$	377	0	533	200	410 573

Why is the deterministic solution bad? Is it due to a shortsighted guess on the randomness (leading to too *few* booked vehicles for the four suppliers) or, can it be explained by the fact that the *wrong suppliers* were chosen? We compute the $LUSS$ following the skeleton solution from the deterministic model and, thus, not allowing vehicles to be booked from both CS and VV. The *Expected Skeleton Solution Value* $ESSV$ is then €443 88.93, still higher than RP with a consequent *Loss Using the Skeleton Solution* of

$$LUSS = 443\,881.93 - 410\,573 = 33\,308.04, \tag{49}$$

which measures the loss when vehicles are booked exclusively from suppliers AG and PA as suggested by the deterministic model. We can thus conclude that the deterministic solution is bad because it books the wrong number of vehicles from the wrong suppliers. It should be noted that this approach simply requires solving a MIP of smaller dimension when compared to the original problem.

Could we infer from the deterministic solution the variables to fix? Since the skeleton solution from the deterministic model sets to zero only the number of vehicles from CS and VV (formulation (13)), we compute their reduced costs in the continuous relaxation, $r_{CS} = 495$ and $r_{VV} = 277.5$. Set $\mathscr{R}_1 = \{r_{VV}\}$ and $\mathscr{R}_2 = \{r_{CS}\}$. Then, fixing at zero only the variables at zero in the stochastic model in the expected-value solution with the highest reduced cost, we obtain $\mathscr{R}_2 = \{r_{CS}\}$, so that, the *Reduced Costs-based Variable Fixing* $RCVF(2,N) = 410\,573$ with a consequent *Loss of Reduced Costs-based Variable Fixing*:

$$LRCVF(2,N) = RCVF(2,N) - RP = 0. \tag{50}$$

Not allowing vehicles to be booked both from CS and VV (i.e., fixing at zero the variables at zero in the expected-value solution, with associated reduced costs $\mathscr{R}_1 = \{r_{VV}\}$, and $\mathscr{R}_2 = \{r_{CS}\}$), we again compute the *Expected Skeleton Solution Value* $RCVF(1,N) = ESSV = 443\,881.93$ and

$$LRCVF(1,N) = LUSS = 33\,308.04. \tag{51}$$

It should be noticed that, since the number of variables at zero in the deterministic solution is 2, the maximum number of classes is $N = 2$.

We can conclude from the measures computed above hat the deterministic solution does not perform well in a stochastic environment. This is explained by the insufficient number of vehicles that are booked in the first stage (720 instead of 1110) from the suppliers AG and PA. We reach the stochastic solution, however, by following the proposed approach and focusing on the skeleton variable with highest reduced costs (i.e., not booking vehicles from CS). However, the optimality of the solution is not detected with the elements considered above. So in Section 5 a proposed procedure for identifying the potential good variables to be inherited by the stochastic model is presented.

B.2 Power generation scheduling

Table 20 reports energy demands at the nodes $n \in \mathscr{N}$ of the scenario tree, while the characteristics of the two types of generators are shown in Table 21. Value $\bar{D}$ is the mean demand considered in the deterministic model. We assume that the number of running units as we enter the modelling period is u_i^0, $i \in \mathscr{I}$. These units have a capacity of 800 mw, which is well above the expected need of $\bar{D} = 300$ mw during the first time period. Consequently, no generators are started up in period one ($s_i^1 = 0$, $i \in \mathscr{I}$) independently of the considered start up cost. The aim of the model is to select and allocate the power demands among an optimal number of operating units of types 1 and 2.

Table 20 Predecessor $pa(n)$, energy demand D^n and probability p^n at node $n \in \mathscr{N}$ of the two-period scenario tree.

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
$pa(n)$	-	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
D^n	300	300	605	630	580	650	600	520	100	180	130	100	120	102	50	41	100	102	125	69	600	596
p^n	1	1	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{20}$
$\bar{D}$	300	300	300																			

Table 21 Costs and production characteristics for generators of type $i \in \mathscr{I}$.

	C_i (€)	E_i (€)	F_i (€)	m_i (mw)	M_i (mw)	u_i^0	$u_{i,max}$
$i = 1$	100	2500	14000	20	80	4	4
$i = 2$	150	5000	16000	30	120	4	4

B.3 Supply transportation problem

Deterministic and stochastic parameter values are reported below. Table 22 lists the set of suppliers $\mathscr{K}$ and the sets of their plants $\mathscr{O}_k$, $k \in \mathscr{K}$. The list of destinations (cement factories) is shown in Table 23 with the premium rates charged by the external company and the unloading capacities (expressed in tons of gypsum). Table 24 provides the minimum and maximum requirements for suppliers $k \in \mathscr{K}$ (again expressed in tons of gypsum). It is assumed that an initial inventory level of $l_j^0 = 0$ is available for all the destinations $j \in \mathscr{D}$. The capacity for all vehicles is fixed to $q = 31$ tons. The discount α is set to the value 0.7. The values of the transportation costs t_{ij} over all origins and destinations are in the following range: $[t_{ij}^{min}, t_{ij}^{max}] = [10.80, 73.52]$. Finally, the demand scenarios were obtained using historical data. Scenarios were built using the weekly demand values for the months of March, April, May and June of 2011, 2012 and 2013. Thus, a set of 48 weekly demand scenarios were obtained and assumed to be equiprobable.

Table 22 Set of suppliers $\mathscr{K}$ with their sets of plants $O_k,\ k \in \mathscr{K}$.

Supplier $k \in \mathscr{K}$	Plant $i \in \mathscr{O}_k$
1	$1, \ldots, 6$
2	7
3	8
4	4
5	9
6	$6, 10$
7	1
8	$1, 2$
9	11
10	12
11	13
12	14
13	15
14	12
15	8
16	16
17	17
18	9
19	$5, 15$
20	5
21	18
22	19
23	7
24	12

Table 23 List of destinations (cement factories) with emergency costs b_j and unloading capacities g_j, $j \in \mathscr{D}$.

Destination $j \in \mathscr{D}$	emergency cost b_j	Unloading capacity g_j
1	72.61	422.95
2	70.58	2054.55
3	68.01	1330.67
4	64.94	453.64
5	73.52	613.41
6	58.57	695.24
7	69.83	443.14
8	66.32	815.36
9	62.63	933.33
10	68.22	319.79
11	48.92	443.11
12	50.04	760.11
13	73.07	381.20
14	59.93	498.33
15	55.63	232411.75

Table 24 Minimum a_k and maximum v_k requirement capacity of supplier $k \in \mathscr{K}$.

Supplier $k \in \mathscr{K}$	a_k	v_k
1	1057.69	-
4	0	96.15
6	0	576.92
10	0	194.23
11	0	480.76
15	0	192.30
16	0	384.61