Author’s Accepted Manuscript

www.elsevier.com/locate/omega

Linear solution schemes for mean-semivariance project portfolio selection problems: An application in the oil and gas industry

Jorge A. Sefair, Carlos Y. Méndez, Onur Babat, Andrés L. Medaglia, Luis F. Zuluaga

PII: S0305-0483(16)30277-8
DOI: http://dx.doi.org/10.1016/j.omega.2016.05.007
Reference: OME1678

To appear in: *Omega*

Received date: 18 August 2014
Revised date: 2 May 2016
Accepted date: 31 May 2016

Cite this article as: Jorge A. Sefair, Carlos Y. Méndez, Onur Babat, Andrés L. Medaglia and Luis F. Zuluaga, Linear solution schemes for mean-semivarianc project portfolio selection problems: An application in the oil and gas industry *Omega,* http://dx.doi.org/10.1016/j.omega.2016.05.007

Linear solution schemes for mean-semivariance project portfolio selection problems: An application in the oil and gas industry

Jorge A. Sefair[a], Carlos Y. Méndez[b], Onur Babat[c], Andrés L. Medaglia[b,*], Luis F. Zuluaga[c]

[a]School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, AZ, USA
[b]Centro para la Optimización y Probabilidad Aplicada (COPA), Departamento de Ingeniería Industrial, Universidad de los Andes, Bogotá (Colombia)
[c]Industrial and Systems Engineering Department, Lehigh University, Bethlehem, PA, USA

Abstract

We study the Mean-SemiVariance Project (MSVP) portfolio selection problem, where the objective is to obtain the optimal risk-reward portfolio of *non-divisible* projects when the risk is measured by the semivariance of the portfolio's Net-Present Value (NPV) and the reward is measured by the portfolio's expected NPV. Similar to the well-known mean-variance portfolio selection problem, when integer variables are present (e.g., due to transaction costs, cardinality constraints, or asset illiquidity), the MSVP problem can be solved using Mixed-Integer Quadratic Programming (MIQP) techniques. However, conventional MIQP solvers may be unable to solve large-scale MSVP problem instances in a reasonable amount of time. In this paper, we propose two *linear* solution schemes to solve the MSVP problem; that is, the proposed schemes avoid the use of MIQP solvers and only require the use of Mixed-Integer Linear Programming (MILP) techniques. In particular, we show that the solution of a class of real-world MSVP problems, in which project returns are positively correlated, can be accurately approximated by solving a single MILP problem. In general, we show that the MSVP problem can be effectively solved by a sequence of MILP problems, which allow us to solve large-scale MSVP problem instances faster than using MIQP solvers. We illustrate our solution schemes by solving a real MSVP problem arising in a Latin American oil and gas company. Also, we solve instances of the MSVP problem that are constructed using data from the PSPLIB library of project scheduling problems.

Keywords: semivariance; project selection; project portfolio optimization; Benders decomposition; mean-semivariance; risk; petroleum industry.

1. Introduction

The selection of the best investment projects within a set of alternatives is crucial to any firm facing competition. Moreover, the ability to build portfolios that efficiently allocate scarce resources contributes to the achievement of corporate goals in the long run. Typically, a portfolio's expected

*Corresponding author
Email addresses: jorge.sefair@asu.edu (Jorge A. Sefair), cy-mendez@uniandes.edu.co (Carlos Y. Méndez), onur.babat@lehigh.edu (Onur Babat), amedagli@uniandes.edu.co (Andrés L. Medaglia), luis.zuluaga@lehigh.edu (Luis F. Zuluaga)

profit is considered the single most important corporate goal to be maximized; however, it is not the only one: the fitness of a firm's portfolio should also involve a measure of the portfolio's volatility or risk. For instance, a portfolio with very attractive expected profits might expose the company to a large loss with high probability, whereas a low-risk portfolio might secure the company lower but more certain profits. For these reasons, the problem of selecting projects to create an optimal risk-reward portfolio has been actively considered in the literature [cf., 10, 36, 38].

A keystone economic sector where the problem of selecting an appropriate portfolio of project investments arises is the upstream oil and gas industry. In this sector, the project investment's returns are subject to high uncertainty, mainly driven by factors like geology, equipment costs, oil selling price, well production levels, and oil quality, among others. In a typical project, the profit's probability distribution is usually asymmetrical (skewed), exhibiting a high probability of low profits and a low probability of high profits [37]. Moreover, given the significant amount of investment required to carry out a project, managers and investors in this industry have a strong bias against underperforming portfolios [26, 30, 34, 35], leaning towards downside-risk measures to quantify the risk of investment [33].

Although different downside-risk measures are available in the literature [cf., 6, 16, 25, 31], in this paper we focus on the semivariance risk measure. Through this measure, projects with a high probability of having returns lower than a critical value (e.g., the expected value or any other value specified by the decision maker) are considered risky. In other words, the semivariance does not consider values beyond the critical value (i.e., gains) as risk; thus, it is a more appropriate measure when investors are worried about portfolio underperformance [25].

The semivariance is a widely used measure of risk in the oil and gas industry. For example, Orman and Duggan [29] propose an optimization routine in which the portfolio's semistandard deviation (square root of the semivariance) is minimized, subject to budget constraints and a target value for the expected Net Present Value (NPV). By varying this target, the authors construct an efficient frontier. Then, they find the optimal investment level for each project based on a predetermined set of projects. In a more recent work, Sira [33] uses scatter search to heuristically approximate an efficient portfolio frontier in the petroleum industry. This approach is used to determine how much investment must be allocated in a fixed set of projects. After comparing portfolios that minimize both variance and semivariance of the project portfolio's return, the author argues that the latter is preferable as a measure of risk in petroleum projects. Similar to Sira [33], we consider the problem of finding a portfolio of projects; that is, *non-divisible* assets with minimum semivariance, but where the projects to be included in the portfolio, rather than fixed, can be selected from a set of available investment projects.

To address this problem we first consider the more common portfolio allocation problem where the portfolio assets are divisible. In his seminal work on risk-reward portfolio selection, Markowitz [24] proposed the use of the portfolio returns' variance as a measure of risk, and developed an optimization problem, together with a solution method, to obtain the portfolio selection that has minimum risk among those with a required expected return. This problem is now commonly referred as the Mean-Variance (MV) portfolio selection problem. Similar to the classical MV problem, Markowitz et al. [25] proposed a quadratic programming formulation for the Mean-SemiVariance (MSV) portfolio selection problem, which is obtained using a *sampling approach* to estimate the problem parameters; that is, an estimation of the asset return distributions is obtained from a finite number of samples. These samples are typically obtained from historical data, simulations, or a combination of both. Thus, these portfolio selection problems have the characteristic that no specific distributional assumption about the asset return distributions is required to be able to formulate or solve the corresponding selection problem.

The Mean-SemiVariance Project (MSVP) portfolio selection problem, a MSV problem with non-divisible assets, can be formulated as a Mixed-Integer Quadratic Programming (MIQP) problem for which specialized MIQP solvers can be used. However, unlike the MV problem formulation whose size only depends on the number of assets, the size of the MSVP problem formulation grows with both the number of non-divisible assets, and the number of samples used to estimate the problem's parameters, thus leaving open some concerns regarding scalability and solvability of the MSVP problem via MIQP solvers. Although existing solution methods for Quadratic Programming (QP) are quite competitive, the introduction of integer variables significantly increases the complexity of solving a MIQP problem and limits the size of the problems that can be solved [23]. Similar challenges have been addressed for MV problems with integer variables (due to, e.g., transaction costs, cardinality constraints, lot size) by proposing solution approaches that avoid using MIQP solvers [cf., 21, 2, 15]

To tackle the inherent difficulty in solving the MSVP problem, we propose two *linear* solution schemes that avoid the use of QP methods and only require the use of Mixed-Integer Linear Programmming (MILP) techniques. These approaches are useful alternatives to the MIQP when either because of problem size, solution time requirements, software requirements, or expertise, it is not suitable to directly use a MIQP solver. The first scheme is obtained from a natural approximation of the portfolio's semivariance that can be reformulated as a MILP problem. This MILP approximation is (formally) shown to work as an accurate proxy of the MSVP problem when the projects' NPVs are positively correlated, which is the case in our oil and gas industry problem. Furthermore, we develop a second linear solution scheme that requires the solution of a series of MILP problems for general instances of the MSVP problem. This scheme works even in the case of NPVs having arbitrary correlations (i.e., not all are positively correlated).

The proposed schemes have both practical and computational advantages. They might be more suitable for practitioners that are well acquainted with MILP techniques [5], but not with more advanced MIQP techniques. Also, the software required to solve the corresponding MIQP may require an additional investment over regular software required to solve MILP problems. More importantly, both solution schemes have the ability to solve instances of the MSVP problem that might not be possible to solve efficiently using MIQP solvers. Our linear solution schemes also contribute to the rich literature on using linear methods for portfolio allocation problems [cf., 23, for a recent review].

The remainder of the article is organized as follows. In Section 2, we formally introduce the MSVP problem. In Section 3.1, we present a linear approximation of the MSVP problem that requires the solution of a single MILP problem. Also, we quantitatively characterize the MILP approximation's effectiveness. In Section 3.2, we present a linear solution scheme capable of solving general MSVP instances by iteratively solving a series of MILP problems. Our computational results are presented in Section 4, where we illustrate the effectiveness of the linear solution schemes by solving a MSVP problem arising in a Latin American oil and gas company. In Section 5 we solve general instances of the MSVP problem that are constructed using data from the PSPLIB library of project scheduling problems [18]. In Section 6, we conclude the paper with some final remarks.

2. Mean-semivariance project portfolio selection problem

In this section we formally introduce the MSVP problem. Consider n risky *non-divisble* investment projects. Let $r = (r_1, \dots, r_n)^\intercal \in \mathbb{R}^n$ denote the uncertain NPV of the n risky projects, which is calculated over a time horizon of T periods. Let $x = (x_1, \ldots, x_n)^{\mathrm{T}} \in \{0,1\}^n$ denote a portfolio on these projects; that is, the binary variables x_i take the value of 1 if the company invest in project i and 0 otherwise, for $i = 1, \ldots, n$. Thus, the portfolio's NPV is given by

$$r^{\mathrm{T}}x = x^{\mathrm{T}}r = \sum_{i=1}^{n} x_i r_i.$$

A (single-period) MSVP problem aims at finding the portfolio of projects $x \in \{0,1\}^n$ at time $t = 0$ that minimizes the semivariance of the project portfolio's NPV, subject to a given minimum expected NPV. Formally, the MSVP problem can be written as the following optimization problem:

$$\begin{array}{ll} \min & \mathbb{E}(\min\{0, x^{\mathrm{T}}r - \mathbb{E}(x^{\mathrm{T}}r)\}^2) \\ \text{s.t.} & \mathbb{E}(x^{\mathrm{T}}r) \geq \mu_0 \\ & x \in \mathcal{X} \cap \{0,1\}^n, \end{array} \tag{1}$$

where $\mathbb{E}(\cdot)$ denotes expectation; $\mu_0 \in \mathrm{I\!R}$ is the given minimum expected portfolio NPV; and $\mathcal{X} \subseteq \mathrm{I\!R}^n$ is a given set defined by linear constraints, which might be used to enforce some relevant business conditions such as a budget constraint (*i.e.*, $\sum_{i=1}^{n} c_i x_i \leq B$, where c_i is the investment required for i-th project and B is the total available budget). For the MSVP problem in the oil and gas industry considered here, a detailed description of the set $\mathcal{X}$ is provided in Section 4. Here, we choose 0 as the *critical value* [e.g., 25] for the semivariance calculation. That is, portfolios with a negative return will be considered risky. However, our results extend in straightforward fashion to other choices of the critical value, such as a market benchmark [cf., 25].

It is clear from (1) that the MSVP problem is analogous to a classical risk-reward portfolio allocation problem with illiquid assets in which the risk is measured by the portfolio returns' semivariance, and the reward is the expected portfolio's return.

In order to solve (1), we use a *sampling approach* [cf., 4, 19, 21, 25, 31], in which an estimation of the distribution of the random variables of interest is obtained from a finite number of samples $r^1, \ldots, r^m \in \mathrm{I\!R}^n$. These samples are typically obtained from historical data, simulations, or a combination of both. Using this sampling approach, the MSVP problem in (1) can be written as:

$$\begin{array}{ll} \min & \displaystyle\frac{1}{m}\sum_{j=1}^{m} \min\{0, x^{\mathrm{T}}r^j - x^{\mathrm{T}}\mu\}^2 \\ \text{s.t.} & x^{\mathrm{T}}\mu \geq \mu_0 \\ & x \in \mathcal{X} \cap \{0,1\}^n, \end{array} \tag{2}$$

where the vector $\mu = (\mu_1, \ldots, \mu_n)^{\mathrm{T}} \in \mathrm{I\!R}^n$ of mean project return estimates is obtained by letting

$$\mu_i = \frac{1}{m}\sum_{j=1}^{m} r_i^j, \tag{3}$$

for $i = 1, \ldots, n$.

For ease of exposition, we will use (3) to obtain $\mu \in \mathrm{I\!R}^n$ in our numerical experiments; however, our results are independent of this choice, and a variety of other estimation methods can be used. Also, note that to obtain an asymptotically unbiased and strongly consistent estimator of the semivariance [17], we should use the factor $\frac{m}{(m-1)^2}$ instead of $\frac{1}{m}$ in the objective function of (2).

However, for the sake of clarity, we will use the latter, as changing this factor does not affect the composition of the optimal project selection.

After introducing the auxiliary variable y_j, which captures the value $\min\{0, x^{\mathrm{T}}(r^j - \mu)\}$ for each $j = 1, \ldots, m$, the MSVP problem in (2) can be written as an optimization problem with a convex quadratic objective, linear constraints, and binary variables. This result is formalized in Proposition 1.

Proposition 1 (Markowitz et al. [25]). *The mean-semivariance project portfolio selection problem* (2) *is equivalent to:*

$$
\begin{array}{lll}
\min & \displaystyle\frac{1}{m}\sum_{j=1}^{m} y_j^2 & \\
\text{s.t.} & y_j \leq x^{\mathrm{T}}(r^j - \mu) & j = 1, \ldots, m \\
& y_j \leq 0 & j = 1, \ldots, m \\
& x^{\mathrm{T}}\mu \geq \mu_0 & \\
& x \in \mathcal{X} \cap \{0,1\}^n. &
\end{array} \tag{4}
$$

Furthermore, the objective function in (4) *is convex.*

Proposition 1 shows that the MSVP problem is a MIQP problem, that is, an optimization problem with a convex quadratic objective and linear constraints, with the additional constraint of some of its variables being integer (more specifically, in (4) variables are required to be binary). Thus, the MSVP formulation in (4) can be solved using branch-and-bound [cf., 8, 27] in conjunction with QP techniques [cf., 3]. In particular, `CPLEX`, `Gurobi`, and `Xpress-MP` are among the commercial optimization solvers that offer special solution algorithms for MIQP problems based on such techniques.

3. Linear solution schemes

In this section we show that the MSVP problem in (4) can be efficiently solved without using QP solvers; that is, it can be solved using branch-and-bound in conjunction with linear programming techniques. We refer to these solution methodologies as *linear* solution schemes. Besides substantially enlarging the number of optimization solvers that can be used to solve the MSVP problem, these linear solution schemes allow us to solve large-scale instances of the MSVP problem much faster than with a MIQP approach.

Note that the MIQP problem in (4) can easily become a large-scale problem when either the number of projects, n, or the number of samples used to estimate the distribution of the projects' NPVs, m, is large. Clearly, this behavior results from n and m being the dimension of the x- and y-variables in (4), respectively. In this regard, we emphasize the difference in the project portfolio selection problem when the variance is used as a measure of risk. As opposed to semivariance, using the variance implies the solution of a single MIQP problem whose size depends on the number of candidate projects, but not on the number of samples used to estimate the mean and the variance-covariance matrix of project's NPVs. Further, in order to solve a single MIQP problem, it is necessary (loosely speaking) to solve a large number of (potentially large) QP problems (relaxed MIQP problems), obtained by branching on the corresponding binary variables.

For the reasons discussed above, in Section 3.1 we first introduce a MILP formulation that accurately approximates the solution of the MSVP problem when the project's NPVs are positively correlated and the total number of projects is moderate. Next, in Section 3.2 we show that a general class of the MSVP problem, and in particular instances of the problem with a large number of projects and samples, can be solved efficiently by solving a sequence of MILP problems using a *Benders decomposition* approach in which the *Benders cuts* (cf. Nemhauser and Wolsey [27, Sections II.3.7 and II.5.4], and Freund [14]) are computed in closed-form.

3.1. MILP approximation for MSVP portfolio selection problem

In this section we present an approximation for the MSVP problem in (4), which is obtained by solving a *single* MILP problem with as many binary variables as the corresponding MIQP. We begin by stating the following optimization problem related to (2):

$$\begin{array}{ll} \min & \dfrac{1}{m}\displaystyle\sum_{i=1}^{n}\sum_{k=1}^{n}\widetilde{\sigma}_{ik}x_i x_k \\ \text{s.t.} & x^{\mathrm{T}}\mu \geq \mu_0 \\ & x \in \mathcal{X} \cap \{0,1\}^n, \end{array} \tag{5}$$

where

$$\widetilde{\sigma}_{ik} = \sum_{j=1}^{m} \min\{0, r_i^j - \mu_i\}\min\{0, r_k^j - \mu_k\}. \tag{6}$$

for $i = 1,\ldots,n$, $k = 1,\ldots,n$. First, we will show that (5) is a pessimistic approximation to (2); that is, (5) overestimates the semivariance of the project's portfolio in (2), making the project selection process more conservative. Then, we will show that the more positively correlated the projects in the portfolio are, the better (5) works as an approximation to (2). Even though this condition seems overly restrictive, there is strong evidence that positive correlations are ubiquitous in the oil and gas industry, in part, because most projects are influenced by the same economic and market conditions (e.g. interest rates, oil prices, and gas prices). Further evidence of this will be given in Section 4. Finally, we will show that (5) can be rewritten as a MILP problem by introducing appropriate extra continuous variables.

To see that (5) provides a pessimistic approximation to (2), let $u \in \mathbb{R}^n$ be given, and define $I^- = \{i : u_i < 0, i = 1,\ldots,n\}$, and $I^+ = \{i : u_i \geq 0, i = 1,\ldots,n\}$. Clearly,

$$0 \geq \min\left\{0, \sum_{i=1}^{n} u_i\right\} = \begin{cases} \displaystyle\sum_{i\in I^-} u_i + \sum_{i\in I^+} u_i, & \text{if } \left|\displaystyle\sum_{i\in I^-} u_i\right| \geq \left|\displaystyle\sum_{i\in I^+} u_i\right|; \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

Also

$$\sum_{i=1}^{n} \min\{0, u_i\} = 0 + \sum_{i\in I^-} u_i. \tag{8}$$

Using (7) and (8) in the two cases $\left|\sum_{i\in I^-} u_i\right| \geq \left|\sum_{i\in I^+} u_i\right|$ and $\left|\sum_{i\in I^-} u_i\right| \leq \left|\sum_{i\in I^+} u_i\right|$, it follows that

$$0 \geq \min\left\{0, \sum_{i=1}^{n} u_i\right\}, \text{ and } \min\left\{0, \sum_{i=1}^{n} u_i\right\} \geq \sum_{i=1}^{n} \min\{0, u_i\}, \tag{9}$$

and therefore:

$$\left(\min\left\{0,\sum_{i=1}^{n}u_i\right\}\right)^2 \le \left(\sum_{i=1}^{n}\min\{0,u_i\}\right)^2. \tag{10}$$

With (10), and letting $\widetilde{r}^j := r^j - \mu$, $j = 1,\dots,m$, we have that the objective function of (2) can be bounded from above as follows:

$$\begin{aligned} \frac{1}{m}\sum_{j=1}^{m}\min\left\{0,\sum_{i=1}^{n}x_i\widetilde{r}_i^j\right\}^2 \quad & \le \frac{1}{m}\sum_{j=1}^{m}\left(\sum_{i=1}^{n}\min\{0,x_i\widetilde{r}_i^j\}\right)^2 = \\ & \frac{1}{m}\sum_{j=1}^{m}\sum_{i=1}^{n}\sum_{k=1}^{n}\min\{0,x_i\widetilde{r}_i^j\}\min\{0,x_k\widetilde{r}_k^j\}) = \\ & \frac{1}{m}\sum_{i=1}^{n}\sum_{k=1}^{n}x_ix_k\left(\sum_{j=1}^{m}\min\{0,\widetilde{r}_i^j\}\min\{0,\widetilde{r}_k^j\}\right) = \\ & \frac{1}{m}\sum_{i=1}^{n}\sum_{k=1}^{n}\widetilde{\sigma}_{ik}x_ix_k. \end{aligned} \tag{11}$$

The first inequality follows from (10), and the second to last equality follows from $x_i \ge 0$, $i = 1,\dots,n$. Hence, (5) is a pessimistic approximation to (2) because its objective overestimates the expected squared downside deviations; that is, the semivariance.

Notice that (5) will be equivalent to (2) whenever the second inequality in (9) holds with equality when replacing $u = x^{\mathrm{T}}\widetilde{r}^j$, for all $j = 1,\dots,m$. The second inequality in (9) holds with equality when $I^- = \{1,\dots,n\}$ or $I^+ = \{1,\dots,n\}$. That is, problems (5) and (2) will be equivalent if ${I^j}^+ := \{i \in \{1,\dots,n\} : x_i\widetilde{r}_i^j \ge 0,\} = \{1,\dots,n\}$ or ${I^j}^- := \{i \in \{1,\dots,n\} : x_i\widetilde{r}_i^j < 0\} = \{1,\dots,n\}$, for all $j = 1,\dots,m$. Clearly, for all the samples to satisfy that the deviations in a sample be either all above the mean or all below the mean, the NPVs of each project must be highly correlated. As will be discussed in Section 4, for MSVP problems arising in the oil and gas industry, it is reasonable to expect (real-world) scenarios with high correlations where this approximation works remarkably well.

The objective function of (5) can be linearized by introducing appropriate extra continuous variables. Let $I_\sigma^+ := \{(i,k) : \widetilde{\sigma}_{ik} > 0, i = 1,\dots,n, k = 1,\dots,n\}$, and $I_\sigma^- := \{(i,k) : \widetilde{\sigma}_{ik} \le 0, i = 1,\dots,n, k = 1,\dots,n\}$. Then problem (5) is equivalent to the following MILP problem:

$$\begin{aligned} \min \quad & \frac{1}{m}\sum_{(i,k)\in I_\sigma^+}\widetilde{\sigma}_{ik}y_{ik} \\ \text{s.t.} \quad & x^{\mathrm{T}}\mu \ge \mu_0 \\ & y_{ik} \ge x_i + x_k - 1 \quad \text{for all } (i,k)\in I_\sigma^+ \\ & y_{ik} \ge 0 \quad \text{for all } (i,k)\in I_\sigma^+ \\ & x \in \mathcal{X}\cap\{0,1\}^n. \end{aligned} \tag{12}$$

The equivalence between (5) and (12) follows from the next observations. First, from (6) it follows that $I_\sigma^- = \{(i,k) : \widetilde{\sigma}_{ik} = 0, i = 1,\dots,n, k = 1,\dots,n\}$. Second, if $(i,k) \in I_\sigma^+$, then $y_{ik} \ge x_i + x_k - 1$ and $y_{ik} \ge 0$ imply that $y_{ik} \ge x_ix_k$, but since $\widetilde{\sigma}_{ik} > 0$, then at any optimal solution of (12), y_{ik} would be at its lower bound $y_{ik} = x_ix_k$.

3.2. Benders-based linear solution scheme for MSVP portfolio selection problems

In this section, we present a linear solution scheme for the MSVP problem that is based on a suitable use of the *Benders decomposition* technique (cf. Nemhauser and Wolsey [27, Sections II.3.7 and II.5.4], and Freund [14]). To make the presentation more succinct, we re-state (4) as follows:

$$\begin{aligned} \min \quad & \frac{1}{m} y^{\mathrm{T}}(I)y \\ \text{s.t.} \quad & y \le \widetilde{R}x \\ & y \le 0 \\ & x \in \mathcal{X}' \cap \{0,1\}^n, \end{aligned} \tag{$\mathcal{S}$}$$

where $y := [y_j]_{j=1,\dots,m}$, I is the $m \times m$ identity matrix, $\widetilde{R}$ is a $m \times n$ matrix, whose row j is given by $[\widetilde{R}]_j := (r^j - \mu)^{\mathrm{T}}$, $j = 1, \dots, m$, and $\mathcal{X}' := \mathcal{X} \cup \{x \in \mathbb{R}^n : x^{\mathrm{T}}\mu \ge \mu_0\}$.

The idea of a Benders decomposition approach is to divide the problem variables into two groups: the *complicating* and the *non-complicating* variables. One begins by fixing the complicating variables in the original problem to a particular value. The resulting problem –so-called *Benders subproblem*– should be solvable to optimality, and in particular, the *dual* (see, e.g. Fang and Puthenpura [12, Chapter 9.1.2]) of the Benders subproblem should be solvable to optimality. The dual solution of the Benders subproblem is then used to construct a *Benders master problem* on the complicating variables of the original problem. Solving iteratively both the Benders subproblem and master problem leads to a solution of the original problem that might be obtained faster than by solving the (full) original problem. For the MSVP problem, next we show that with an appropriate choice of the complicating variables, the Benders subproblem can be solved in closed-form.

To address problem ($\mathcal{S}$) via a Benders decomposition approach, we choose the x variables as the complicating variables in ($\mathcal{S}$). After fixing the x variables to a value $\hat{x} \in \mathcal{X}' \cap \{0,1\}$, and (for convenience) making the change of variable $y \to -y$, we obtain the problem:

$$\begin{aligned} \min \quad & \frac{1}{m} y^{\mathrm{T}}(I)y \\ \text{s.t.} \quad & y \ge -\widetilde{R}\hat{x} \quad (u) \\ & y \ge 0 \qquad\quad (u_0), \end{aligned} \tag{13}$$

where $u \in \mathbb{R}^m$ are the dual variables associated to the return constraints and $u_0 \in \mathbb{R}^m$ are the dual variables associated to the non-negativity constraints in (13). The (convex) quadratic program in (13) corresponds to the Benders subproblem, whose *Wolfe dual* is given by (see, e.g., Nocedal and Wright [28, Chapter 12]):

$$\begin{aligned} \max \quad & -u^{\mathrm{T}}\widetilde{R}\hat{x} - \frac{1}{m} y^{\mathrm{T}}(I)y \\ \text{s.t.} \quad & -\frac{2}{m} y + u + u_0 = 0 \\ & u, u_0 \ge 0. \end{aligned} \tag{14}$$

Problem (14) is equivalent to:

$$\begin{aligned} \max \quad & -u^{\mathrm{T}}\widetilde{R}\hat{x} - \frac{m}{4}(u+u_0)^{\mathrm{T}}(u+u_0) \\ \text{s.t.} \quad & u, u_0 \ge 0. \end{aligned} \tag{15}$$

In any optimal solution of (15) we have $u_0 = \vec{0}$, so (15) is equivalent to:

$$\begin{aligned} \max \quad & \sum_{j=1}^{m} \left(-(\hat{x}^{\mathrm{T}}\widetilde{r}^j)u_j - \frac{m}{4}u_j^2 \right) \\ \text{s.t.} \quad & u_j \geq 0, j = 1, \ldots, m. \end{aligned} \tag{16}$$

Notice that problem (16) decomposes into m independent problems:

$$\begin{aligned} \max \quad & -(\hat{x}^{\mathrm{T}}\widetilde{r}^j)u_j - \frac{m}{4}u_j^2 \\ \text{s.t.} \quad & u_j \geq 0, \end{aligned} \tag{17}$$

for $j = 1, \ldots, m$; which can be solved by inspection: If $(\hat{x}^{\mathrm{T}}\widetilde{r}^j) \geq 0$, then the optimal solution of (17) is $u_j^* = 0$. If $(\hat{x}^{\mathrm{T}}\widetilde{r}^j) < 0$, then we get a concave quadratic objective in (17):

$$|(\hat{x}^{\mathrm{T}}\widetilde{r}^j)|u_j - \frac{m}{4}u_j^2$$

that has a maximum at $u_j^* = \frac{2}{m}|(\hat{x}^{\mathrm{T}}\widetilde{r}^j)|$. So the optimal solution $u^*(\hat{x}) \in \mathbb{R}^m$ of the Benders dual subproblem (14) can be obtained in closed-form as follows:

$$u_j^*(\hat{x}) = \begin{cases} 0 & \text{if } (\hat{x}^{\mathrm{T}}\widetilde{r}^j) \geq 0, \\ \frac{2}{m}|(\hat{x}^{\mathrm{T}}\widetilde{r}^j)| & \text{if } (\hat{x}^{\mathrm{T}}\widetilde{r}^j) < 0, \end{cases} \tag{18}$$

for $j = 1, \ldots, m$. With the Benders dual subproblem solution, the Benders master problem is constructed as follows. Given a finite index set $\mathcal{K}$, and a set of feasible portfolios $\hat{\mathcal{X}}'_{\mathcal{K}} = \{\hat{x}_k \in \mathcal{X}' \cap \{0,1\}^n : k \in \mathcal{K}\}$, consider the Benders master problem

$$\begin{aligned} \min \quad & q \\ \text{s.t.} \quad & q \geq \sum_{j=1}^{m} -(x^{\mathrm{T}}\widetilde{r}^j)u_j^*(\hat{x}_k) - \frac{m}{4}u_j^*(\hat{x}_k)^2; \quad \forall \hat{x}_k \in \hat{\mathcal{X}}'_{\mathcal{K}} \\ & x \in \mathcal{X}' \cap \{0,1\}^n. \end{aligned} \tag{19}$$

Note that the right-hand side of the first set of constraints in (19) is closely related to the objective function of the Benders dual subproblem (15).

With a closed-form expression for the solution of the Benders dual subproblem, and with the construction of the Benders master subproblem given in (19), we can now state in Algorithm 1, a Benders-based solution algorithm for the MSVP problem.

After execution, Algorithm 1 returns an ϵ-optimal portfolio solution x_ϵ^*. That is, if we let $x^* := \operatorname{argmin}_x\{\mathcal{S}\}$ be the optimal mean-semivariance project portfolio, then

$$\frac{\mathrm{SV}(x_\epsilon^*) - \mathrm{SV}(x^*)}{\mathrm{SV}(x^*)} < \epsilon, \tag{20}$$

where SV represents the semivariance of any portfolio of projects $x \in \{0,1\}^n$, given by

$$\mathrm{SV}(x) := \frac{1}{m}\sum_{j=1}^{m} \min\{0, x^{\mathrm{T}}r^j - x^{\mathrm{T}}\mu\}^2.$$

Algorithm 1 Benders linear solution scheme for the MSVP problem

1: **procedure** MSVP_BENDERS($\epsilon > 0$)
2: $\mathcal{K} \leftarrow \emptyset$, $k = 1$, GAP$= \infty$
3: **while** GAP$> \epsilon$ **do**
4: **compute** $\hat{x}_k$, z_k, the optimal solution and objective of (19)
5: **compute** $u^*(\hat{x}_k)$ using (18)
6: $\mathcal{K} \leftarrow \mathcal{K} \cup k$, $k \leftarrow= k + 1$
7: UPPBOUND $\leftarrow \sum_{j=1}^{m} -(\hat{x}_k^{\mathrm{T}} \widetilde{r}^j) u_j^*(\hat{x}_k) - \frac{m}{4} u_j^*(\hat{x}_k)^2$, LOWBOUND$_k \leftarrow z_k$
8: GAP$\leftarrow$ $|$UPPBOUND $-$ LOWBOUND$_k|/|$LOWBOUND$_k|$
9: **end while**
10: **return** $x_\epsilon^* = \hat{x}_k$
11: **end procedure**

The correctness of Algorithm 1 follows from the theory behind the Benders decomposition technique [cf., 13].

Note that the Benders-based linear solution scheme for the MSVP problem outlined in this section requires, at its core, the iterative solution of MILPs in `Step 4` of the algorithm. This is because the non-linearity of the original problem's objective is handled in closed-form in `Step 5` of the algorithm. It is worth to mention that a *regularized* version [cf., 32] of the Benders-based algorithm outlined here for the MSVP problem can be implemented without changing the complexity of the Master problem in (19). Namely, following [32], the objective function in (19) can be changed to $c(q, x) := q + \frac{1}{\sigma}\|x - \hat{x}_k\|^2$ with $\sigma > 0$. Moreover, taking advantage of the fact that both $x, \hat{x}_k \in \{0, 1\}^n$ it follows that $c(q, x)$ is equivalent to the following linear function $c(q, x) = q + \frac{1}{\sigma}(x - 2x\hat{x}_k + \hat{x}_k)$. This means that the MSPV problem can be solved via a Benders decomposition approach where the regularized Benders Master problem remains a MILP and the Benders cuts are found in closed-form. Although experiments were carried out with this regularized version of the Benders algorithm, the performance difference with the classical Benders Algorithm 1 are not significant, and in Section 5, we report results using the non-regularized Benders Algorithm.

4. Case study: project selection in an upstream oil and gas company

In this section we consider an instance of the MSVP problem arising in the oil and gas industry. After giving a detailed description of the problem in Section 4.1, in Section 4.2 we report the computational results of the linear solution scheme presented in Section 3.1.

4.1. Data and detailed model

The case study is based on our experience with an upstream oil and gas company operating in Latin America, which is one of the top 40 largest companies in the world. We consider a division of the company with 27 non-divisible candidate projects for investment along a 30-year planning horizon with an available budget of US$ 100 million per year and expected production for the first year of at least 40,000 barrels. This division is one of the six divisions in charge of prioritizing exploration projects in different geographical areas of Latin America. In 2011-2016, the average annual budget for exploration in this division was around 12% of the company's exploration budget.

Besides the known capital investment requirements and the production and operational costs, the projects are subject to precedence relations. For example, the execution of some projects require the execution of other complementary projects. The NPV calculation for each project involves deterministic elements like the capital investment requirements and the production and operational costs. It also involves more volatile and stochastic components, like the project's production level –modeled by triangular distributions for pessimistic, moderate, and optimistic scenarios– and the international trade petroleum price (WTI), forecasted by a mean-reversion model [cf., 9]. It should be emphasized that, according to Sira [33], the uncertain production levels and the oil prices account for 80% of the NPV's volatility in a typical petroleum project (for literature on forecasting petroleum prices, see [1]). We use Monte Carlo simulation to model the uncertainties, considering a variance reduction technique known as common random numbers [cf., 20] to ensure that the same realizations for the key underlying random variable, namely the WTI price, were used to calculate the NPV for all projects. These values are used to construct the vector μ used in the expected return constraint in (2); that is, μ_i corresponds to the average NPV of project i, for $i = 1, \ldots, n$ where $n = 27$. In Appendix A, Table A.4 displays the average NPV of the projects when estimated with different sample sizes. In addition, Figure 1 shows the skewed nature of the NPV for a typical oil and gas project (i.e., low profits are more likely to occur than high profits). In this case, 1,000 NPV realizations are produced using Monte Carlo simulation. Due to confidentiality agreements, the average NPV for each project has been modified by adding a constant. However, although this shift affects the probability of loss and the mean return of the projects, it does not affect the deviations from the mean used to measure the risk of the project's portfolio.

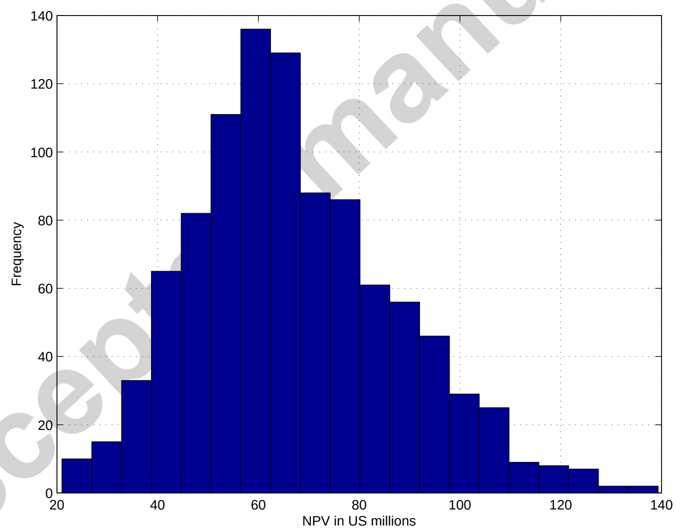

Figure 1: Histogram of 1,000 random realizations of the NPV (in US millions) of a typical oil and gas project obtained using Monte Carlo simulation.

The NPVs of the considered oil and gas projects are highly correlated, given that they belong to the same industry and are affected by the same market conditions. Figure 2 shows a histogram of the upper triangular portion of the correlation matrix (excluding the diagonal) where it is worth noting that more than 75% of the pairwise correlations are higher than 0.80, all correlations are positive, and only 8% of the correlations are less than 0.1. Although the calculated correlations appear to be overly high, evidence of positive and strong correlation between the projects in the same industry is ubiquitous in the literature. For instance, in [7] it is stated that correlations between security returns in the same industry tend to be positive because they are influenced by the same economic and market conditions. Thus, changes in economic factors such as interest rates, labor, and raw material cost affect simultaneously the performance of all companies and their projects in the same sector.

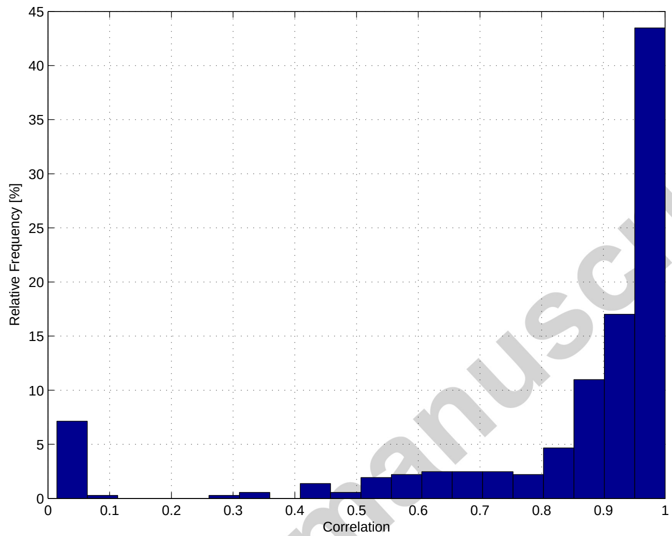

Figure 2: Histogram of the 27 project NPV correlations (excluding the diagonal) used in Table 1.

The linear constraints defining the set $\mathcal{X}$ in (1) for the oil and gas MSVP portfolio selection problem include a required minimum production level per period of the planning horizon; limiting budget constraints per time period; limits on the total production and operational cost per time period; and precedence relations between projects. Specifically, in this case we have

$$\mathcal{X} = \left\{ x \in \{0,1\}^n : \begin{array}{ll} \sum_{i=1}^{n} q_{it} x_i \geq w_t, & t \in T; \\ \sum_{i=1}^{n} k_{it} x_i \leq b_t, & t \in T; \\ \sum_{i=1}^{n} c_{it} x_i \leq h_t, & t \in T; \\ x_i \leq x_j, & (i,j) \in A \end{array} \right\}. \tag{21}$$

In (21), set T represents the time periods within the planning horizon. Parameters q_{it}, c_{it}, and k_{it} are the expected barrel production, the production and operational costs, and the capital investment requirements of project i in time period t, respectively. Parameters w_t, h_t, and b_t are the minimum production level, the maximum allowable production and operational costs, and the available budget for investment in period t, respectively. Note that, although variable x_i is not indexed in t, the time is implicitly considered in the expected barrel production, the production and operational costs, and the capital investment requirements for each project per period of the planning horizon (i.e., parameters q_{it}, c_{it}, and k_{it}, respectively). That is, if project i is selected (i.e., variable x_i equals 1), its expected oil production and costs are accounted in the left-hand side of the constraints, in order to satisfy the minimum production level and the costs limits for each period of the planning horizon. Further, set A defines the precedence relations between projects; that is, if selecting project i implies the selection of project j, then $(i, j) \in A$. The complete list of precedence relations between the projects used in the case study is given in equation (A.1) in Appendix A.

Our algorithms are implemented in MATLAB and executed on a 64-bit workstation with AMD Opteron 2.0 GHz CPU and 32 GB of RAM. We use `CPLEX 12.5` to solve both the MILP approximation and the MIQP formulation to optimality.

4.2. Numerical results

In this section, computational experiments are conducted to show the accuracy and efficiency of the MILP approximation proposed in Section 3.1 to solve the oil and gas industry MSVP problem. Figure 3 displays the *semivariance efficient frontier* (i.e., plots the optimal project portfolio's semivariance for different values of μ_0) obtained after solving the MILP problem defined in (12) and the MIQP formulation in (4) with the side constraints $\mathcal{X}$ defined in (21). The number of projects and number of samples in the problems solved are $n = 27$ and $m = 1000$, respectively. Results in Figure 3 show that, thanks to the strong positive correlations of the projects in this case study, the MILP approximation effectively finds the set of non-dominated portfolios in the frontier. For practical purposes, this result implies that the MILP approximation in (12) can help decision makers to create a semivariance efficient frontier showing the tradeoff between risk and profitability, without the use of nonlinear programming techniques. The total time required to compute the efficient frontier in Figure 3 using the MIQP approach is 84.04 s, whereas the total time required to compute it using the MILP approach is 20.79 s.

To further illustrate the performance of the MILP in (12) when the NVPs are highly correlated, we generate additional instances of the MSVP problem based on the original oil and gas data. Namely, we generate instances of $n = 27$ projects with sample sizes $m = 100$, 500, 1000, 3000, 5000, 7000, 9000, and 10000. To reach the desired value of m, additional samples are randomly drawn from the original data. Regardless of the sample size, we use $\mu_0 = 698$, as in the original oil and gas case study. We compare the proposed MILP approximation with the default `CPLEX 12.5` MIQP solver.

Table 1 shows the results obtained by the MILP and the MIQP models. The first column shows the sample size used to estimate the projects' NPVs return distributions. The resulting portfolio's semivariance (i.e., objective function values) are shown in the second and third columns, whereas the execution times are reported in the fourth and fifth columns. The last column shows the computation time speedup, which is calculated as the ratio between the MIQP to the MILP execution times. The last row reports the geometric mean of the speedups for all the instances. In this case, we use the geometric mean because it avoids being overly optimistic with good ratios obtained on few instances [22].

The results summarized in Table 1 show that the MILP approximation finds the optimal portfolio's semivariance (i.e., the MIQP solution). As before, good accuracy performance of the MILP approximation is due to the positive correlated nature of the project's NPVs. Although the MIQP

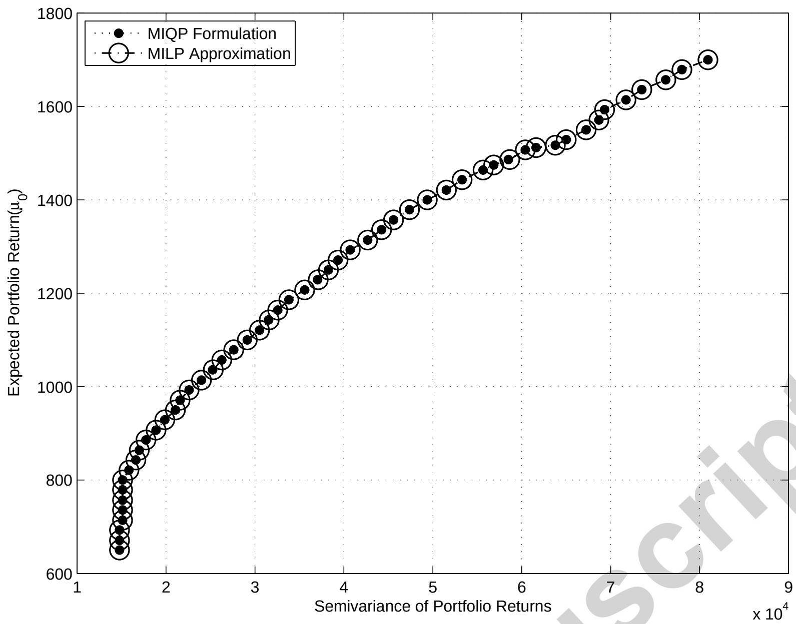

Figure 3: Semivariance efficient frontier of MSVP portfolios for fifty (50) different benchmark NPV return values, computed using the (exact) MIQP formulation (4) and the MILP approximation (12)

approach does slightly better than the MILP approach in the instance with the smallest number of samples; overall, the geometric mean shows that the MILP approach is roughly eight times faster than the MIQP approach. This result is expected, given that the size of the MILP does not increase as the number of samples grows.

No. Samples	**Portfolio's Semivariance MIQP**	**Portfolio's Semivariance MILP Apx.**	**Time (s) MIQP**	**Time (s) MILP Apx.**	**Speedup**
100	14887	14887	0.47	0.75	0.62
500	15746	15746	1.21	0.48	2.54
1000	14769	14769	3.44	0.58	5.93
3000	14767	14767	4.79	0.60	7.94
5000	14768	14768	8.91	0.52	17.09
7000	14764	14764	15.23	0.55	27.48
9000	14769	14769	21.50	0.59	36.21
10000	14769	14769	34.66	1.53	22.66
Geo. Mean					8.55

Table 1: Computational results for instances of the MSVP problem based on the oil and gas case study with 27 projects, a minimum NPV benchmark return of 698, and number of samples ranging from 100 to 10000.

We further explore the quality of the MILP approximation scheme in (12) for the case in which the projects's NPVs are less correlated. To do so, we use the same instance of Figure 3 and multiply the sample NPVs of thirteen (13) projects (randomly selected) by -1. The results are shown in Figure 4, in which the quality of the MILP approximation decreases compared to the MIQP. However, even in this case the MILP approximation scheme provides a fair approximation of the MSVP efficient frontier. Note that due to this change on the instance data, the maximum expected NPV that can be obtained from the projects is now lower than in the original instance shown in Figure 3. In this case, the total time required to compute the efficient frontier in Figure 4 using the MIQP approach is 69.22 s, whereas the total time required to compute it using the MILP approach is 4.49 s.

Although it provides an accurate approximation to the semivariance when NPVs are positively correlated, the MILP scheme in (12) is limited by the fact that the number of continuous variables grows quadratically with the number of projects in the problem (i.e., as n^2).

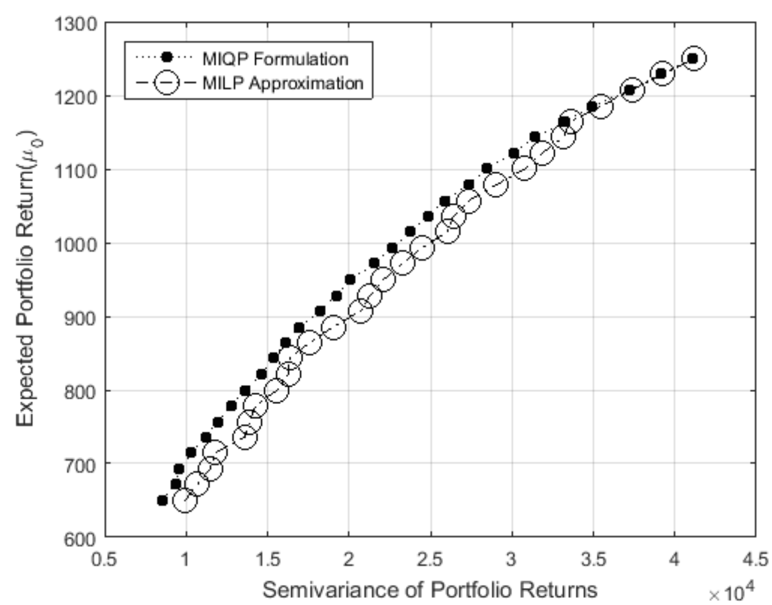

Figure 4: Semivariance efficient frontier of MSVP portfolios computed using the (exact) MIQP formulation and the MILP approximation (12) when the NPVs of the projects are less highly positively correlated.

5. General MSVP instances

In this section we study the accuracy and efficiency of the Benders-based solution approach proposed in Section 3.2 to solve general instances of the MSVP problem. We test the limits of our approaches by considering instances with a large number of projects (n) and samples (m), and with multiple correlation levels among project's NPVs. This analysis is motivated by the fact that some oil and gas companies may have a large number of candidate projects.

To see this, note that the case study considered in Section 4 arises from a project selection problem in one of the six divisions of an oil and gas company operating in Latin America. For the particular year of this analysis, the division's exploration budget was around US$100 M, which was 20% of the company's total exploration budget. To put the project selection problem of this division in context, in 2014 the top-ranked capital expenditures in exploration of some larger oil and gas companies ranged between US$1400 M - US$2500 M [11]. Thus, from a budget perspective and considering the worldwide scale of operations of larger companies, MSVP problem instances with possibly hundreds of candidate projects may arise in practice. Additionally, the number of drilling permits approved by environmental authorities could be an estimate of the number of candidate projects in a company's portfolio. In 2014, the Oil and Gas Conservation Commission of the state of Wyoming alone 3786 different drilling permits, with some companies requesting permits for the exploration of more than 300 and up 923 different wells [40].

Also, the consideration of different correlation profiles among NPVs is motivated by the fact that current petroleum prices may encourage oil and gas companies to bring new types of projects into their portfolios (e.g., enhanced oil recovery, alternative refining processes, biofuels), which can be less (or even negatively) correlated with the traditional exploration and production projects. Additionally, the datasets used in this section include realistic features arising in the oil and gas industry such as resource and precedence constraints, as well as skewed NPV's distributions. Section 5.1 describes the dataset generation procedure used to test our algorithms and Section 5.2 presents the computational results of the Benders-based solution approach described in Section 3.2.

5.1. Data

Given the absence of datasets for the MSVP problem in the literature, we generate our test instances based on the well known PSPLIB library [18, url: `http://www.om-db.wi.tum.de/psplib/library.html`]. The PSPLIB library contains problem sets for single- and multi-mode resource-constrained project scheduling problems. In particular, we use the PSPLIB single-mode datasets listed in Table 2.

No. Projects	Filename	Location: `www.wiwi.tu-clausthal.de/fileadmin/...`
100	psp1.sch	`...Produktion/Benchmark/RCPSP/testset_ubo100.zip`
200	psp1.sch	`...Produktion/Benchmark/RCPSP/testset_ubo200.zip`
500	PSP1.sch	`...Produktion/Benchmark/RCPSP/testset_ubo500.zip`
1000	PSP1.sch	`...Produktion/Benchmark/RCPSP/testset_ubo1000.zip`

Table 2: PSPLIB instances used to construct different instances of the MSVP problem.

Although PSPLIB does not contain instances for the MSVP problem, we use both the precedence and resource constraints provided in its instances. To construct an instance for the MSVP problem, we split the set of projects in a PSPLIB instance into 10 subsets. These subsets represent the time periods in which a project demand resources in the MSVP problem formulation (cf., (21)). For example, if a project belongs to the second subset, then this project demands resources in the second time period in the MSVP problem. This procedure defines the left-hand side coefficients of the resource constraints in (21). To vary the complexity of the MSVP instances, we set the right-hand side of the resource constraints to be equal to a fraction of the sum of the left-hand side coefficients. This fraction ranges from the smallest value that results on a feasible instance of the problem to 1.00 (i.e., the resource constraint is redundant).

To generate instances of different sample size, additional samples for the NPVs are generated by adding noise and re-sampling the oil and gas project's NPVs described in Section 4. Following the same procedure as in Section 4.2, we also generate different correlation levels among the NPVs. These NPVs correlations range from -1 to 1 as shown in Figure 5. Precedence constraints are included without modifications.

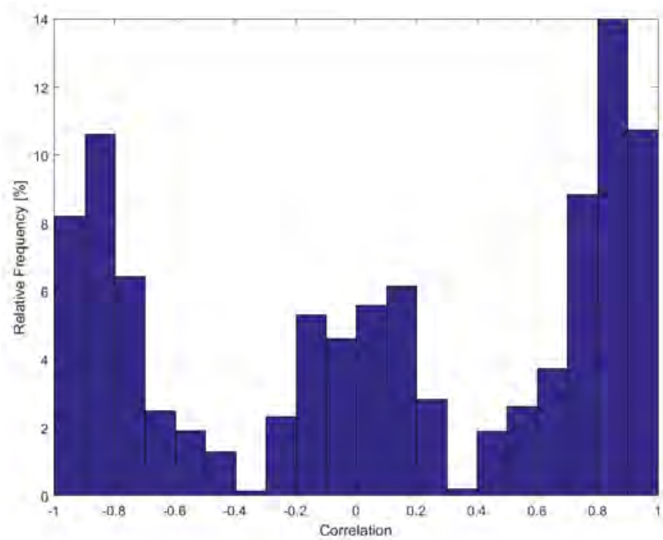

Figure 5: Histogram of the NPV correlations (excluding the diagonal) in a general instance of the MSVP problem.

To run our computational experiments, we use MATLAB on a 64-bit workstation with AMD Opteron 2.0 GHz CPU and 32 GB of RAM. We use `CPLEX 12.5` to solve both the MIQP formulation and the MILP iterations in the Benders-based algorithm. In Algorithm (1), we use $\epsilon = 0.5\%$, and, for fairness of the comparison, we also set the `CPLEX` relative optimality gap to stop the MIQP to $\epsilon = 0.5\%$. We impose a time limit of 3600 seconds for each experiment.

5.2. Numerical results

In this section we compare the performance of the Benders-based solution approach described in Section 3.2 with the MIQP formulation of the MSVP problem. Table 3 shows the results of our experiments for different instances that are generated using the procedure described in Section 5.1. In total, we generate 452 instances that include different number of projects and samples, as reported in columns 1–2. For each instance, we use a minimum expected portfolio NPV, μ_0, within the range shown in columns 4–5. Also, the range of the factor that modifies the right-hand-side of the resource constraints is reported in columns 6–7.

To illustrate the variability existing in our test instances, columns 8 and 9 in Table 3 summarize the minimum and maximum number of projects selected in the optimal solution of the MSVP problem. In this case we see that variations in the input parameters, besides number of projects and sample size, lead to instances of the MSVP problem with very different solutions. The computational

No.	No.	No.	μ_0		Resource		Projects		Time (s)			
Projects	Samples	Instances	min	max	min	max	min	max	MIQP	Benders	iter.	Speedup
100	1000	32	10	500	0.64	1.00	16	82	1.32	0.15	3.34	8.07
100	5000	20	10	2000	0.70	1.00	11	74	13.08	0.16	2.00	80.07
100	10000	17	10	1500	0.70	1.00	7	79	43.35	0.27	2.00	158.80
200	1000	99	50	10000	0.27	1.00	4	135	3.15	0.13	2.00	23.97
200	5000	29	50	15000	0.70	1.00	4	200	26.42	0.29	2.00	93.32
200	10000	28	50	15000	0.70	1.00	4	200	104.75	0.47	2.00	224.56
500	1000	63	100	10000	0.04	1.00	2	475	12.46	0.93	2.75	15.24
500	5000	19	100	10000	0.70	1.00	21	364	154.31	1.04	2.00	153.73
500	10000	20	100	10000	0.70	1.00	2	457	341.50	1.63	2.32	210.25
1000	1000	86	100	15000	0.04	1.00	3	852	13.89	1.37	2.64	10.87
1000	5000	19	100	15000	0.70	1.00	22	785	263.92	3.28	2.95	87.78
1000	10000	20	100	15000	0.70	1.00	10	746	571.08	8.80	6.20	76.08
Geo. Mean												59.17

Table 3: Comparison between MIQP and Benders-based linear solution scheme for general instances of the MSVP problem generated from PSPLIB instances of the resource constrained project scheduling problems. The column Resource indicates the range of the factor used to constraint the resources available in the instance. The column Projects indicate the range of number of projects selected in the optimal solution of the instances. In all instances, differences between the semivariance values of MIQP and Benders algorithms are within a 0.5% margin of error, and the Benders algorithm is faster than the MIQP approach.

time of the MIQP and the Benders-based solution scheme are reported in columns 10 and 11, respectively. All the tested instances are solved within the time limit, implying that the Benders solution approach obtains the optimal semivariance within a 0.5% margin of error. As shown in column 10 (MIQP) and column 11 (Benders), the average solution time of the Benders solution approach is much lower than the MIQP approach. This difference increases as the number of samples in the problem increase. This is not surprising, given that the size of the master problem used in the Benders solution approach does not change with the number of samples. Instead, the number of samples only affects the computation of the Benders cuts, which is done through a closed-form calculation. The average number of iterations required by the Benders solution approach and the average solution time speedup are reported in columns 12 and 13, respectively. The reported speedups show that the Benders approach is on average 59 times faster than the MIQP, demonstrating the efficiency of using this approach for general large-scale instances of the MSVP problem.

6. Concluding Remarks

In this paper, we studied the MSVP problem. After presenting a convex quadratic formulation of the problem, we proposed two alternative linear solution schemes that effectively solve this problem. These schemes have both practical and computational advantages over a direct MIQP approach to solve the MSVP. The first scheme is based on a MILP approximation that overestimates the project's portfolio NPV semivariance (prone for risk-averse decision makers) by solving a single

MILP. Aside from providing a formal proof of this overestimation, the computational tests show that the MILP approximation is very accurate when dealing with projects with positively correlated NPVs. Moreover, for instances of the MSVP problem with a moderate number of projects in which it is desired to use a large number of samples to accurately estimate the project's portfolio NPV semivariance, the MILP approximation solution approach is shown to consistently outperform the default `CPLEX 12.5` MIQP solver that can be used to directly solve the MSVP problem.

In a more general setting, we proposed a Benders-based linear solution scheme that allows the decision maker to solve the MSVP problem for any positive or negative level of correlation among the NPVs. This approach has proven to be effective, consistently outperforming the default `CPLEX 12.5` MIQP solver for general large-scale instances of the MSVP problem.

The proposed methods have a very broad potential of being applied to other problems. In particular, note that some of the key characteristics of oil and gas project selection problems such as: non-divisible assets, skewed NPV distributions, resource and precedence constraints, preference for downside-risk measures, etc. are common to project selection problems in other industries. Also, both linear solution schemes can be easily extended to solve MSVP problems with additional combinatorial constraints which provide real features on the projects (cf. transaction costs [39], transaction lots [21], cardinality constraints [2]). Moreover, recent approaches have focused on efficiently solving the mean-variance portfolio allocation problems with integrality constraints [2, 21, to name a few]. Thus, extending the Benders solution scheme to address this type of problems will be a promising topic for future research work.

Acknowledgements

The authors would like to thank Oscar Bravo for his guidance and helpful comments throughout the case study development. We also thank the two anonymous reviewers for his comments that greatly improved this paper.

7. References

[1] Al-Harthy, M. H. (2007). Stochastic oil price models: comparison and impact. *The Engineering Economist*, 52(3):269–284.

[2] Bertsimas, D. and Shioda, R. (2009). Algorithm for cardinality-constrained quadratic optimization. *Computational Optimization and Applications*, 43(1):1–22.

[3] Bienstock, D. (1996). Computational study of a family of mixed-integer quadratic programming problems. *Mathematical Programming*, 74(2):121–140.

[4] Birge, J. R. and Louveaux, F. (2011). *Introduction to Stochastic Programming*. Springer Series in Operations Research and Financial Engineering. Springer, 2nd edition.

[5] Bixby, R. E. (2002). Solving real-world linear programs: A decade and more of progress. *Operations Research*, 50(1):3–15.

[6] Boasson, V., Boasson, E., and Zhou, Z. (2011). Portfolio optimization in a mean-semivariance framework. *Investment management and financial innovations*, 8(3):58–68.

[7] Bodie, Z., Kane, A., and Markus, A. J. (2005). *Investments*. McGraw-Hill.

[8] Borchers, B. and Mitchell, J. E. (1994). An improved branch and bound algorithm for mixed integer nonlinear programs. *Computers and Operations Research*, 21(4):359–367.

[9] Dixit, A. K. and Pindyck, R. S. (1994). *Investment Under Uncertainties*. Princeton University Press.

[10] Doerner, K. F., Gutjahr, W. J., Hartl, R. F., Strauss, C., and Stummer, C. (2006). Pareto ant colony optimization with ilp preprocessing in multiobjective project portfolio selection. *European Journal of Operational Research*, 171(3):830–841.

[11] Ernst & Young (2015). US oil and gas reserves study. Technical report, Ernst & Young. Available at `http://www.ey.com/Publication/vwLUAssets/EY-us-oil-and-gas-reserves-study-2015/$FILE/EY-us-oil-and-gas-reserves-study-2015.pdf`.

[12] Fang, S. and Puthenpura, S. (1993). *Linear Optimization and Extensions: Theory and Algorithms*. Prentice Hall.

[13] Flippo, O. E. and Rinnooy-Kan, A. H. G. (1993). Decomposition in general mathematical programming. *Mathematical Programming*, 60:361–382.

[14] Freund, R. M. (2004). Benders decomposition methods for structured optimization, including stochastic optimization. Technical report, Massachusetts Institute of Technology. Available at `http://ocw.mit.edu/courses/sloan-school-of-management/15-094j-systems-optimization-models-and-computation-sma-5223-spring-2004/lecture-notes/benders_art.pdf`.

[15] Huang, X., Zhao, T., and Kudratova, S. (2016). Uncertain mean-variance and mean-semivariance models for optimal project selection and scheduling. *Knowledge-Based Systems*, 93:1 – 11.

[16] Jarrow, R. and Zhao, F. (2006). Downside loss aversion and portfolio management. *Management Science*, 52(4):558–566.

[17] Josephy, N. H. and Aczel, A. D. (1993). A statistically optimal estimator of semivariance. *European Journal of Operations Research*, 67(2):267–271.

[18] Kolisch, R. and Sprecher, A. (1997). PSPLIB-a project scheduling problem library: OR software–ORSEP operations research software exchange program. *European Journal of Operational Research*, 96(1):205–216.

[19] Konno, H. and Yamazaki, H. (1991). Mean-absolute deviation portfolio optimization model and its applications to Tokyo Stock Market. *Management Science*, 37(5):519–531.

[20] Law, A. M. and Kelton, W. D. (2000). *Simulation modeling and analysis*. McGraw-Hill.

[21] Lejeune, M. (2013). Portfolio optimization with combinatorial and downside return constraints. In Zuluaga, L. F. and Terlaky, T., editors, *Modeling and Optimization: Theory and Applications*, volume 62 of *Springer Proceedings in Mathematics & Statistics*, pages 31–50. Springer New York.

[22] Lozano, L. and Medaglia, A. L. (2013). On an exact method for the constrained shortest path problem. *Computers & Operations Research*, 40(1):378 – 384.

[23] Mansini, R., Ogryczak, W., and Speranza, M. G. (2013). Twenty years of linear programming based portfolio optimization. *European Journal of Operational Research*, 234(2):518–535.

[24] Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1):77–91.

[25] Markowitz, H., Todd, P., Xu, G., and Yamane, Y. (1993). Computation of mean-semivariance efficient sets by the critical line algorithm. *Annals of Operations Research*, 45(1):307–317.

[26] Merritt, D., de San Miguel, A., et al. (2000). Portfolio optimization using efficient frontier theory. In *SPE Asia Pacific Conference on Integrated Modelling for Asset Management*. Society of Petroleum Engineers.

[27] Nemhauser, G. L. and Wolsey, L. A. (1988). *Integer and combinatorial optimization*. Wiley.

[28] Nocedal, J. and Wright, S. (2006). *Numerical optimization*. Springer Science & Business Media.

[29] Orman, M. M. and Duggan, T. E. (1999). Applying modern portfolio theory to upstream investment decision making. *Journal of petroleum technology*, 51(03):50–53.

[30] Quintino, A., Lourenço, J. C., and Catalão-Lopes, M. (2013). An integrated risk management model for an oil and gas company. *Business Administration*, page 144.

[31] Rockafellar, R. T. and Uryasev, S. (2000). Optimization of Conditional Value-at-Risk. *Journal of Risk*, 2:21–41.

[32] Ruszczyński, A. (1997). Decomposition methods in stochastic programming. *Mathematical programming*, 79(1-3):333–353.

[33] Sira, E. (2006). Semivariance as real project portfolio optimisation criteria–an oil and gas industry application. *International journal of global energy issues*, 26(1):43–61.

[34] Suslick, S. and Schiozer, D. (2004). Risk analysis applied to petroleum exploration and production: an overview. *Journal of Petroleum Science and Engineering*, 44(1):1–9.

[35] Tyler, P., McVean, J., et al. (2001). Significance of project risking methods on portfolio optimization models. *Society of Petroleum Engineers Inc.*

[36] Vetschera, R. and De Almeida, A. T. (2012). A promethee-based approach to portfolio selection problems. *Computers & Operations Research*, 39(5):1010–1020.

[37] Walls, M. R. (2004). Combining decision analysis and portfolio management to improve project selection in the exploration and production firm. *Journal of Petroleum Science and Engineering*, 44(1):55–65.

[38] Wang, J. and Hwang, W. L. (2007). A fuzzy set approach for R&D portfolio selection using a real options valuation model. *Omega*, 35(3):247–257.

[39] Woodside-Oriakhi, M., Lucas, C., and Beasley, J. (2013). Portfolio rebalancing with an investment horizon and transaction costs. *Omega*, 41(2):406 – 420.

[40] Wyoming Oil and Gas Conservation Commission (2015). State of Wyoming approved permits to drill. Technical report, Wyoming Oil and Gas Conservation Commission. Available at `http://wogcc.state.wy.us/DumpApds.cfm`.

Appendix A. Precedence constraints and NPV data for the oil and gas case study

Below, we provide the expected project NPV values used in the MSVP problem arising in the oil and gas industry and discussed in Section 4.1.

Projects	Avg. NPV		
	$m = 1000, \dots, 10000$	$m = 500$	$m = 100$
1	305.897	300.329	293.160
2	460.993	454.685	446.099
3	0.107	0.103	0.098
4	67.265	65.698	64.094
5	97.901	95.899	93.731
6	0.036	0.034	0.028
7	125.989	123.594	120.963
8	106.834	104.655	102.824
9	6.128	5.979	5.766
10	8.877	8.708	8.156
11	30.411	29.568	28.955
12	159.684	156.401	153.327
13	6.681	6.562	6.471
14	0.057	0.049	0.040
15	101.636	100.084	98.133
16	37.272	36.460	35.702
17	33.162	32.495	31.192
18	18.138	17.841	17.388
19	66.076	64.928	63.245
20	10.962	10.734	10.428
21	19.332	18.883	18.413
22	17.926	17.435	16.924
23	18.780	18.284	17.795
24	20.957	20.491	19.857
25	330.199	330.508	314.499
26	1.693	1.662	1.628
27	5.574	5.551	5.258

Table A.4: Average project NPV of the 27 projects used in Section 4 for different sample sizes

Next, we provide the precedence constraints used in the MSVP problem arising in the oil and gas industry and discussed in Section 4.1.

$$\text{Precedence Constraints} = \left\{ x \in \{0,1\}^n : \begin{array}{r} x_5 \leq x_{25}; \\ x_8 \leq x_{24}; \\ x_5 \leq x_{24}; \\ x_8 \leq x_{21}; \\ x_4 \leq x_{20}; \\ x_2 \leq x_{19}; \\ x_2 \leq x_{17}; \\ x_1 \leq x_{16}; \end{array} \right\} \tag{A.1}$$