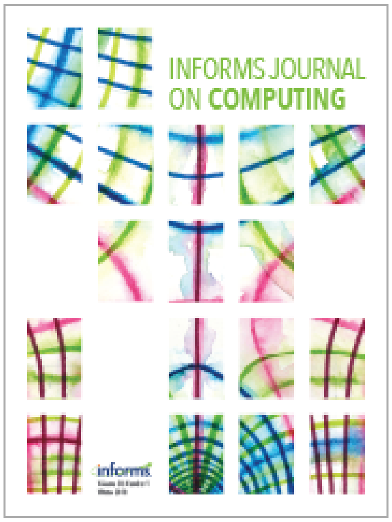

INFORMS Journal on Computing

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

An Extended Mixed-Integer Programming Formulation and Dynamic Cut Generation Approach for the Stochastic Lot-Sizing Problem

Huseyin Tunc, Onur A. Kilic, S. Armagan Tarim, Roberto Rossi

informs
http://pubsonline.informs.org/journal/ijoc/

INFORMS JOURNAL ON COMPUTING
Vol. 30, No. 3, Summer 2018, pp. 492–506
ISSN 1091-9856 (print), ISSN 1526-5528 (online)

An Extended Mixed-Integer Programming Formulation and Dynamic Cut Generation Approach for the Stochastic Lot-Sizing Problem

Huseyin Tunc,[a] **Onur A. Kilic,**[b] **S. Armagan Tarim,**[c,d] **Roberto Rossi**[e]

[a] Department of Policy and Strategy Studies, Hacettepe University, Ankara, Turkey; [b] Department of Operations, University of Groningen, 9700 AV Groningen, Netherlands; [c] Department of Management, Cankaya University, Ankara, 06800 Turkey; [d] Cork University Business School, University College Cork, T12 K8AF, Ireland; [e] Business School, University of Edinburgh, Edinburgh EH8 9JS, United Kingdom

Contact: huseyin.tunc@hacettepe.edu.tr, http://orcid.org/0000-0001-5508-3702 (HT); o.a.kilic@rug.nl, http://orcid.org/0000-0003-2136-8157 (OAK); armtar@yahoo.com, http://orcid.org/0000-0001-5601-3968 (SAT); roberto.rossi@ed.ac.uk, http://orcid.org/0000-0001-7247-1010 (RR)

Received: November 11, 2015
Revised: November 15, 2016; September 27, 2017
Accepted: September 28, 2017
Published Online: September 21, 2018

https://doi.org/10.1287/ijoc.2017.0792

Abstract. We present an extended mixed-integer programming formulation of the stochastic lot-sizing problem for the static-dynamic uncertainty strategy. The proposed formulation is significantly more time efficient as compared to existing formulations in the literature and it can handle variants of the stochastic lot-sizing problem characterized by penalty costs and service level constraints, as well as backorders and lost sales. Also, besides being capable of working with a predefined piecewise linear approximation of the cost function—as is the case in earlier formulations—it has the functionality of finding an optimal cost solution with an arbitrary level of precision by means of a novel dynamic cut generation approach.

History: Accepted by Jean-Paul Watson, Area Editor for Modeling: Methods and Analysis.
Supplemental Material: The online appendix is available at https://doi.org/10.1287/ijoc.2017.0792.

Keywords: stochastic lot sizing • static-dynamic uncertainty • extended formulation • dynamic cut generation

1. Introduction

The lot-sizing problem aims at determining a minimum cost inventory plan so as to meet demand over a finite discrete planning horizon. The lot-sizing problem and its variants are traditionally studied under the assumption of deterministic demands. However, there is a growing body of work on more realistic lot-sizing problems where demands are modeled as random variables. This has a significant impact on lot-sizing problems since the inventory levels as well as the costs incurred in later periods now become random variables following random demands. As a result, if demands are random, then the solution to a lot-sizing problem is no longer a deterministic inventory plan but an inventory policy. This policy defines, for any given period and inventory level, whether to order, and if so how much to order.

Bookbinder and Tan (1988) provide a broad classification of policies that can be employed in stochastic lot-sizing problems. Among these is the so-called static-dynamic uncertainty strategy, which we investigate in this paper. Following this strategy, one determines the number as well as the timing of all orders at the beginning of the planning horizon, and then, at each replenishment epoch decides the order quantity upon observing the inventory level. The static-dynamic uncertainty is an appealing strategy since it eases the coordination between supply chain players (Kilic and Tarim 2011) and facilitates managing joint replenishments (Silver et al. 1998) and shipment consolidations (Mutlu et al. 2010). As such, an expanding line of research has emerged on computational methods for the static-dynamic uncertainty strategy.

The static-dynamic uncertainty strategy essentially requires finding a static replenishment schedule defining in which periods an order will be placed, and a control policy that dynamically determines order quantities for each of those order periods. Özen et al. (2012) showed that once a replenishment schedule has been set, the optimal order quantity at each replenishment period follows a base-stock policy. However, finding the optimal replenishment schedule and associated base-stock levels is challenging, and there is no known efficient method to serve this purpose.

The literature yet suggests a rich variety of heuristic methods designed to handle different variants of the problem, that is, where unsatisfied demand is backordered or lost and penalized by a cost parameter or bounded by a service level measure such as α (ready rate), β_c (cycle fill rate), and β (fill rate) service levels. Nevertheless, as we discuss in the following, they can be classified into two major categories as to how they approach the underlying problem.

First, we have tailor-made algorithms with varying levels of complexity that are designed for specific variants of the problem. Silver (1978) and Bookbinder and Tan (1988) proposed stochastic adaptations of Silver and Meal's (1973) well-known heuristic. Rossi et al. (2011) developed a state space augmentation algorithm that makes use of a stochastic version of Wagner and Whitin's (1958) algorithm. Tarim et al. (2011) introduced a branch-and-bound procedure where a relaxed version of the original problem is solved at each node of the search tree. Özen et al. (2012) design two heuristics where a dynamic program is solved for each period to construct a solution to the overall problem. A number of other studies made use of constraint programming techniques. Here, the main difficulty stems from the need of effectively reducing the domains of decision variables through dedicated routines that are used in the search process. To this end, Tarim and Smith (2008) employed a preprocessing approach whereas Rossi et al. (2008), Tarim et al. (2009), and Rossi et al. (2012) made use of various filtering techniques. All of the aforementioned methods require customized computer programs, which are not always easy to obtain. Also, their effectiveness is positively correlated with their complexity.

Second, there are mathematical programming models that can be solved with off-the-shelf solvers. Because they do not require dedicated algorithms, these models are more accessible and easy to use. All models that fall in this category are mixed-integer programming (MIP) models. A major challenge in MIP models is to embed the nonlinear cost function into the formulation. Tarim and Kingsman (2004) resolved this issue by approximating the original cost function with a linear function. This is, however, a valid approach only for systems where the service level requirement is very high—as is usual for the α service level. Tarim and Kingsman (2006) instead used a piecewise linear relaxation of the objective function. This is a reliable approach for systems where unsatisfied demand is penalized by a cost parameter and those characterized by an arbitrary level of service requirement. Rossi et al. (2015) extended this framework to account for different service level constrains and the case where unsatisfied demand is lost rather than backordered. The models mentioned above provide high-quality solutions, however often at the expense of large computational times. Tunc et al. (2014) recently proposed an MIP formulation that makes use of the network flow structure of the problem. This formulation has a tighter linear relaxation and in turn has a superior computational performance. However, it is designed solely for problems characterized by α service levels.

1.1. Our Contribution

In this paper, we contribute to the literature by presenting an extended MIP formulation of the stochastic lot-sizing problem for the static-dynamic uncertainty strategy. We build on and blend heuristic methods originally introduced by Tunc et al. (2014) and Rossi et al. (2015). As a result, our formulation enjoys the computational efficiency of the Tunc et al. (2014) network flow formulation, yet it also offers the modeling flexibility of the Rossi et al. (2015) formulation and accounts for all variants of the problem previously studied in the literature.

The proposed formulation is essentially designed to approximate the original nonlinear cost function with an a priori piecewise linear function, following the standard methodology in the literature. Nonetheless, we also develop a dynamic cut generation approach to deploy the model with no prior approximation of the cost function. This approach establishes an approximation of the cost function on the fly with any desired level of precision, while not imposing a computational burden. The results of our computational study reveal that the proposed formulation (with and without the dynamic cut generation approach) is very efficient, as it can optimally solve problem instances with planning horizons over 100 periods in a few minutes.

The advantages of our formulation are summarized as follows:

1. It is computationally far superior to the state-of-the-art MIP formulations in the literature (i.e., Tarim and Kingsman 2006, Rossi et al. 2015).
2. It provides a unified modeling framework to address variants of the stochastic lot-sizing problem characterized by penalty costs and service level constraints as well as backorders and lost sales (c.f. Rossi et al. 2015).
3. Besides being capable of working with a predefined piecewise linear approximation of the cost function—as is the case in earlier formulations, it has the functionality of finding a minimum cost solution with an arbitrary level of precision by means of a novel dynamic cut generation approach.

1.2. Outline

The remainder of the paper is organized as follows. In Section 2, we present the notation and introduce some preliminary analysis. In Section 3, we provide a complete MIP formulation of the problem. In Section 4, we show how the proposed formulation can be deployed without a predefined piecewise linear approximation of the cost function. In Section 5, we discuss how to extend our analysis for different variants of the problem characterized by service level constraints and lost sales. In Section 6, we conduct an extensive computational study and report our findings.

2. Preliminaries

We begin by introducing the notation that will be used in the rest of the paper and setting the grounds for the analysis to follow. Here, we assume that demands that arrive when the system is out of stock are backordered and incur a backorder cost. Later on, we shall discuss how to extend our analysis for service level constraints and lost sales.

We consider a finite planning horizon of N discrete time periods. The demands over these periods are independent but not necessarily identically distributed random variables. We denote the random demand over a given time interval $[i, j]$ by D_{ij}. We assume that D_{ij} has a known cumulative distribution function $F_{ij}(\cdot)$ and a first-order loss function $L_{ij}(\cdot)$. There are three cost components: a fixed setup cost K that is incurred for each order, a holding cost h that is incurred for each unit of inventory carried forward from one period to the next, and a backorder cost p that is incurred for each unit of backordered demand per period. For convenience, and without loss of generality, we assume that unit procurement cost is zero and order lead time is negligible.

Following the static-dynamic uncertainty strategy, a replenishment schedule is determined at the beginning of the planning horizon. Let $T_1, \ldots, T_m$ be the replenishment periods over the planning horizon. Here, m is the number of scheduled replenishments, and T_n is the period where the nth order is scheduled. We refer to the interval $[T_n, T_{n+1})$ between any two successive replenishment periods T_n and T_{n+1} as a replenishment cycle. We assume, for the sake of notational brevity, that $T_1 = 1$ and $T_{m+1} = N + 1$. Then, the planning horizon becomes a disjoint union of m replenishment cycles.

Let us consider a replenishment cycle $[i, j)$, and assume that the postreplenishment (i.e., after an order is placed and received) inventory level at the beginning of the cycle is y. Then, the expected total cost to be incurred over this replenishment cycle can be written as

$$K + \sum_{t=i}^{j-1} (h\,\mathbb{E}(y - D_{it})^+ + p\,\mathbb{E}(y - D_{it})^-),$$

where $\mathbb{E}$ is the expectation operator, and x^+ and x^- stand for $\max\{0, x\}$ and $\max\{0, -x\}$, respectively. If z is a random variable, then $\mathbb{E}(y - z)^-$ is its loss function evaluated at y. Also, we have that $\mathbb{E}\,z = \mathbb{E}\,z^+ - \mathbb{E}\,z^-$. Thus, after rearranging the terms, we can write the cost expression as

$$K + \sum_{t=i}^{j-1} (h(y - \mathbb{E}\,D_{it}) + (h + p) L_{it}(y)). \tag{1}$$

In (1), the first term inside the summation is linear and increasing in y, whereas the second term is the loss function that is convex, nonlinear and decreasing on y (see e.g., Silver et al. 1998, Porteus 2002).

Because we aim at developing an MIP model of the stochastic lot-sizing problem, we will approximate the total cost by replacing the loss function with a piecewise linear approximation. Let us consider a loss function $L(\cdot)$ and assume that we are given a set of pairs $W = \{(a_1, b_1), (a_2, b_2), \ldots\}$ containing the intercept and slope of a finite number of linear functions. Then, a linear approximation can be written as $L(y) \approx \max_{(a,b)\in W}\{a + by\}$. Notice that the approximation is also convex since the maximum operator preserves the convexity. The quality of such an approximation strongly depends on the selection of linear functions. The problem of finding a good set of linear functions for approximating a nonlinear convex function has received considerable attention in the literature (see, e.g., Gavrilovic 1975, Frenzen et al. 2010, Rossi et al. 2014), and it is left out of the scope of this paper.

In the following sections, we make heavy use of the idea of piecewise linearization of the loss function. In Section 3, we will assume that a set of linear functions is readily available for approximating a loss function whenever necessary. In Section 4, however, we will show how our formulation can also be used in the absence of a predefined set of linear functions.

3. Model

Having established the preliminaries, we now construct an extended MIP model for the stochastic lot-sizing problem. If the static-dynamic uncertainty strategy is being used, the solution to this problem is a replenishment schedule indicating replenishment periods and a base-stock level for each replenishment period. An MIP model of the problem should be able to find the optimal values of these in concert so as to minimize the expected total cost that will be incurred over the planning horizon.

The extended formulation relies on the idea of determining the costs to be incurred in a particular period in connection with the replenishment cycle it lies within. Based on this idea, we introduce three sets of decision variables:

- x_{ij} indicator variable that takes value of 1 if $[i, j)$ is a replenishment cycle, and 0 otherwise;
- q_{ij} expected cumulative order quantity up-to and including period i if $[i, j)$ is a replenishment cycle, and 0 otherwise;
- H_{ijt} approximate loss function value at period t of replenishment cycle $[i, j)$.

Recall that our solution should specify a replenishment schedule as well as a base-stock level for each replenishment period. We use variables $x = \{x_{ij}\}$ to construct a replenishment schedule. There is no explicit set of decision variables that express base-stock levels. Instead, we use variables $q = \{q_{ij}\}$ to derive the base-stock level of the first period of a replenishment cycle $[i, j)$ as $q_{ij} - \mathbb{E}\,D_{1i-1}$. We use variables $H = \{H_{ijt}\}$ to assert the approximated loss function value of period $t \in [i, j)$—given that $[i, j)$ is a replenishment cycle. We will be bounding the values of these variables from below by a set of constraints abiding the piecewise linear approximation.

Let us once again consider the expected cost of the replenishment cycle $[i, j)$ assuming that the inventory is replenished up-to the base-stock level $q_{ij} - \mathbb{E}D_{1i-1}$ at the beginning of the cycle. Then, building on (1), we can write the expected total cost over the replenishment cycle as

$$K + \sum_{t=i}^{j-1}(h(q_{ij} - \mathbb{E}D_{1t}) + (h+p)H_{ijt}). \tag{2}$$

Because the planning horizon is a disjoint union of replenishment cycles, we can use (2) to derive an expression for the expected total cost over the planning horizon by conditioning on whether $[i,j)$ is a replenishment cycle.

Let us assume, for now, that if $x_{ij} = 0$, then both q_{ij} and H_{ijt} take the value of zero—later on we introduce a set of constraints to make sure that this is indeed the case. Then, we can write the objective function of the MIP formulation as follows:

$$\min \sum_{i=1}^{N}\sum_{j=i+1}^{N+1}\left(Kx_{ij} + \sum_{t=i}^{j-1}(h(q_{ij} - \mathbb{E}D_{1t}x_{ij}) + (h+p)H_{ijt})\right). \tag{3}$$

Here, the expression inside the outer summations gives the expected cost over the interval $[i,j)$ if $x_{ij} = 1$, and it vanishes otherwise. Therefore, given that the planning horizon is a disjoint union of replenishment cycles, (3) yields the expected total cost over the planning horizon.

We remark that the objective function given above implicitly assumes that the postreplenishment inventory level at the beginning of each replenishment cycle is equal to the base-stock level. This may not be the case if the excess inventory carried from the previous cycle exceeds the base-stock level. As in earlier studies (see, e.g., Tarim and Kingsman 2004, 2006; Rossi et al. 2015), we assume that the costs due to such occurrences are small, and ignore them in our cost computations. Nevertheless, we numerically assess the validity of this assumption in our computational study.

Next, we introduce a set of constraints that guarantee that the planning horizon is indeed a disjoint union of replenishment cycles. This can be done by means of the conventional flow conservation equations if we perceive periods as nodes and replenishment cycles as arcs, that is,

$$\sum_{i=1}^{t-1} x_{it} = \sum_{j=t+1}^{N+1} x_{tj}, \quad t \in [2,N], \tag{4}$$

$$\sum_{j=2}^{N+1} x_{1j} = 1, \tag{5}$$

$$\sum_{i=1}^{N} x_{i,N+1} = 1. \tag{6}$$

We can verbally explain flow conservation constrains as follows: (4) states that if a replenishment cycle starts at a given period, then another one ends at the same period; (5) and (6), respectively, state that the first replenishment cycle starts at the first period, and the last replenishment cycle ends right after the last period.

As we mentioned earlier, the expected cumulative order quantity q_{ij} should be positive only if $[i,j)$ is a replenishment cycle. We establish this condition with the constraint

$$q_{ij} \le Mx_{ij}, \quad i \in [1,N],\ j \in [i+1, N+1], \tag{7}$$

where M is a sufficiently large number—a possible bound could be the value of the inverse distribution function of the total planning horizon demand evaluated at the critical percentile $p/(h+p)$.

Also, we should make sure that cumulative order quantities are nondecreasing from one replenishment cycle to the next. This also implies that the expected replenishment quantity, which equals the difference between the cumulative order quantities of two consecutive replenishment cycles, is always nonnegative. We can write this condition as

is this a (R, Q) policy?

$$\sum_{i=1}^{t-1} q_{it} \le \sum_{j=t+1}^{N+1} q_{tj}, \quad t \in [2,N]. \tag{8}$$

似乎类似 (R, Q)

Finally, we turn our attention to variables $H = \{H_{ijt}\}$. Here, for each replenishment cycle $[i,j)$ and each period $t \in [i,j)$, we need to guarantee that H_{ijt} is larger than $L_{it}(q_{ij} - \mathbb{E}D_{1i-1})$, that is, the value of the loss function evaluated at the base-stock level. However, because $L_{it}(\cdot)$ is nonlinear, we use a piecewise approximation instead. Let W_{it} denote the set of intercept and slope pairs defining the piecewise linear approximation of $L_{it}(\cdot)$. Then, we have that $L_{it}(q_{ij} - \mathbb{E}D_{1i-1}) \approx \max_{(a,b)\in W_{it}}\{a + b(q_{ij} - \mathbb{E}D_{1i-1})\}$. We need to integrate this approximation into the MIP model while conditioning on whether $[i,j)$ is a replenishment cycle. This can be done as follows:

$$H_{ijt} \ge ax_{ij} + b(q_{ij} - \mathbb{E}D_{1i-1}x_{ij}) \quad i \in [1,N],$$
$$j \in [i+1, N+1],\ t \in [i, j-1],\ (a,b) \in W_{it}. \tag{9}$$

It is easy to verify that (9) is binding only if $x_{ij} = 1$ as the right-hand side vanishes otherwise (see also (7)).

This finalizes the MIP model. For convenience, we provide the entire model and specify variable domains. We refer to this model as PM:

$$\begin{aligned} &\min && (3) \\ &\text{subject to} && (4)\text{–}(9), \\ &&& H_{ijt} \ge 0 \quad i \in [1,N], j \in [i+1,N+1], t \in [i,j-1], \\ &&& q_{ij} \ge 0, x_{ij} \in \{0,1\} \quad i \in [1,N], j \in [i+1,N+1]. \end{aligned}$$

4. Dynamic Cut Generation Approach

The PM formulation introduced in Section 3 can be used to solve stochastic lot-sizing instances of realistic sizes in reasonable computational times. However, as is the case in earlier formulations in the literature, it requires a predefined piecewise linear approximation. Here, we introduce a dynamic cut generation approach that can be used to deploy PM with no prior approximation of the cost function while guaranteeing an arbitrary level of precision. This approach resembles the so-called outer approximation method (Duran and Grossmann 1986) in the sense that both approaches use gradient-based algorithms to solve mixed-integer nonlinear programming models. Nevertheless, outer approximation solves the mixed-integer nonlinear problem by iteratively solving a master mixed-integer linear problem and a continuous nonlinear subproblem. These problems generate upper and lower bounds that eventually converge and yield an optimal solution. The proposed dynamic cut generation approach, on the other hand, is a routine integrated into a conventional mixed-integer programming solution algorithm that gradually builds a more precise mixed-integer linear approximation of the original nonlinear problem by generating additional cuts as necessary.

The dynamic cut generation approach makes use of an initial MIP model—which we refer to as RM. This model is identical to PM, except (9) is replaced with the following:

初始 cut

- l

$$H_{ijt} \geq -(q_{ij} - \mathbb{E}\, D_{1t} x_{ij}),$$
$$i \in [1, N],\ j \in [i+1, N+1], t \in [i, j-1]. \quad (10)$$

The constraint above is binding only if $x_{ij} = 1$, and in this case it reads $H_{ijt} \geq -(q_{ij} - \mathbb{E}\, D_{1t})$. The following property shows that this expression is always a valid lower bound to the exact loss function value.

Property 1. *Let z be a random variable and $L(\cdot)$ be its loss function. Then,*

(i) $L(y) \geq (y - \mathbb{E}\, z)^-$ *for all* y,
(ii) $\lim_{y\to-\infty} L(y) = (y - \mathbb{E}\, z)^- = -(y - \mathbb{E}\, z)$, *and*
(iii) $\lim_{y\to\infty} L(y) = (y - \mathbb{E}\, z)^- = 0$.

Proof. (i) It follows the definition of the loss function, that is, $L(y) = \mathbb{E}(y - z)^- = \mathbb{E}(y - z)^+ - (y - \mathbb{E}\, z)$. Because $\mathbb{E}(y - z)^+$ is nonnegative, we have that $L(y) \geq -(y - \mathbb{E}\, z)$. (ii) Observe that $\lim_{y\to-\infty}(y - z)^- = -(y - z)$. This implies that $\lim_{y\to-\infty} L(y) = \mathbb{E}(y - z)^- = -(y - \mathbb{E}\, z)$. (iii) Observe that $\lim_{y\to\infty}(y - z)^- = 0$. This implies that $\lim_{y\to\infty} L(y) = \mathbb{E}(y - z)^- = 0$.

If we work out Property 1 with $z := D_{it}$ and $y := q_{ij} - \mathbb{E}\, D_{1i-1}$, from (i) we obtain $L_{it}(q_{ij} - \mathbb{E}\, D_{1i-1}) \geq (q_{ij} - \mathbb{E}\, D_{1t})^-$. Thus, if $q_{ij} - \mathbb{E}\, D_{1t} \leq 0$, then we have $L_{it}(q_{ij} - \mathbb{E}\, D_{1i-1}) \geq -(q_{ij} - \mathbb{E}\, D_{1t})$, and otherwise we have $L_{it}(q_{ij} - \mathbb{E}\, D_{1i-1}) \geq 0$. The former immediately translates into $H_{ijt} \geq -(q_{ij} - \mathbb{E}\, D_{1t})$. The latter, holds by definition since H_{ijt} is nonnegative. Therefore, (10) is a valid constraint. Furthermore, we have, from (ii) and (iii), that the lower bounds provided by (10) are exact in the limit.

The intuition behind (10) is to use the deterministic equivalent of the loss function as a limiting case. That is, if D_{it} were deterministic and equal to the expectation $\mathbb{E}\, D_{it}$, then the value of the loss function evaluated at the base-stock level $q_{ij} - \mathbb{E}\, D_{1i-1}$ would be $(q_{ij} - \mathbb{E}\, D_{1t})^-$. Property 1 shows that this value is a lower bound to the exact loss function value. Therefore, we can conclude that RM is a valid relaxation of the original problem under consideration.

We now present the details of the dynamic cut generation approach. Here, we assume that we have on hand an MIP solver that can be controlled through a callback routine when updating the incumbent—as it is in modern solvers. The procedure is initiated by invoking an MIP solver on RM. At each iteration of the procedure, the following steps are executed. First, a candidate feasible solution is obtained for RM. Then, the solution is evaluated to see whether it is also a feasible solution to the original problem. If so, the incumbent is updated. Otherwise, the solution is discarded, and a set of feasibility cuts are added to RM. These cuts can be generated analytically without solving a subproblem. Then, the next iteration starts. The procedure terminates when RM (with added cuts) is solved to optimality.

Next, we explain how feasibility cuts can be derived from a given solution. Let us denote a solution as $\bar{x} = \{\bar{x}_{ij}\}$, $\bar{q} = \{\bar{q}_{ij}\}$, and $\bar{H} = \{\bar{H}_{ijt}\}$. First, we verify that this solution is also a feasible solution to the original problem. To this end, for all $[i, j)$ and $t \in [i, j)$ we compute the difference between the real loss value and the approximate loss value returned by the solution, that is, $L_{it}(\bar{q}_{ij} - \mathbb{E}\, D_{1i-1}) - \bar{H}_{ijt}$. If the difference is less than some arbitrarily small constant ϵ, then we conclude that the solution is (sufficiently) feasible and update the incumbent. Otherwise, we generate a feasibility cut to repair infeasibility. A feasibility cut is a line separating $\bar{H}_{ijt}$ from the loss function. Here, we use the tightest possible cut—the tangent line of $L_{it}(\cdot)$ at point $\bar{q}_{ij} - \mathbb{E}\, D_{1i-1}$. This cut is valid by definition because the loss function is convex. Let us denote the intercept and the slope of this cut as a and b, respectively. Then, we have that $a = L_{it}(\bar{q}_{ij} - \mathbb{E}\, D_{1i-1}) - b(\bar{q}_{ij} - \mathbb{E}\, D_{1i-1})$ and $b = L'_{it}(\bar{q}_{ij} - \mathbb{E}\, D_{1i-1})$, where $L'_{it}(\cdot)$ stands for the derivative of $L_{it}(\cdot)$. Property 2 illustrates that $L'_{it}(\cdot)$ can easily be obtained from the cumulative distribution function $F_{it}(\cdot)$. This enables us compute the slope as $F_{it}(\bar{q}_{ij} - \mathbb{E}\, D_{1i-1}) - 1$. Once we establish the feasibility cut by means of a and b, we add the constraint $H_{ijt} \geq a x_{ij} + b(q_{ij} - \mathbb{E}\, D_{1i-1} x_{ij})$ into the model.

Property 2. *Let z be a random variable, and $F(\cdot)$ and $L(\cdot)$ be its cumulative distribution function and loss function, respectively. Then, the derivative $L'(y)$ equals $F(y) - 1$ for all y.*

Proof. Let $f(\cdot)$ be the probability distribution function of z. The loss function can be written as $L(y) = \int_y^\infty (z - y) f(z)\, \mathrm{d}z$. Then, applying the Leibniz rule, we obtain $L'(y) = \int_y^\infty -1 f(z)\, \mathrm{d}z = F(y) - 1$.

Figure 1. Dynamic Cut Generation Procedure

invoke solver on RM
repeat
  find a candidate solution $\bar{x}$, $\bar{q}$, and $\bar{H}$
  for $i \in [1, N]$ *and* $j \in [i+1, N+1]$ *such that* $\bar{x}_{ij} = 1$ **do**
    for $t \in [i, j)$ **do**
      if $L_{it}(\bar{q}_{ij} - \mathbb{E}\,D_{1i-1}) - \bar{H}_{ijt} \geq \epsilon$ **then** slope and intercept
        $b = F_{it}(\bar{q}_{ij} - \mathbb{E}\,D_{1i-1}) - 1$
        $a = L_{it}(\bar{q}_{ij} - \mathbb{E}\,D_{1i-1}) - b(\bar{q}_{ij} - \mathbb{E}\,D_{1i-1})$
        add cut $H_{ijt} \geq a x_{ij} + b(q_{ij} - \mathbb{E}\,D_{1i-1} x_{ij})$ to RM
until *solver cannot find a candidate solution.*

The proposed approach follows the idea of dynamically adding cuts into a relaxed version of the original problem for each approximated loss value that exceeds the real loss function value. As such, it can be regarded as a means to establish a piecewise linear approximation on the fly while solving the problem. The pseudocode of the overall cut generation procedure is provided in Figure 1.

5. Extensions

The extended formulation presented so far assumes that demands that are not immediately satisfied from stock are backordered and penalized by a cost factor. Here, we illustrate how this formulation can be adapted for variants of the stochastic lot-sizing problem. First, we show how different service level constraints can be embedded into the formulation. Then, we revisit the formulation to account for the case where demand that cannot immediately be satisfied from stock is lost rather than backordered.

5.1. Service Level Constraints

The service level constrained inventory models aim at minimizing system costs while ensuring a sufficient level of service quality—defined by a specific service level measure. Here, we consider three commonly employed service level measures, namely, α service level, β_c service level, and β service level; and illustrate how service level constraints associated to each of these measures can be accounted for in the proposed formulation.

5.1.1. α Service Level. The α service level constraint is a lower bound α on the nonstockout probability in any period over the planning horizon. This constraint should, in principle, be imposed on each period. However, in the context of the static-dynamic uncertainty strategy, it is sufficient to impose the constraint on the last period of each replenishment cycle since the nonstockout probability is decreasing over the periods of a replenishment cycle. Let us consider a replenishment cycle $[i, j)$. The inventory level at the end of this cycle can be expressed as $(q_{ij} - \mathbb{E}\,D_{1i-1}) - D_{ij-1}$. As such, the α service level constraint is satisfied only if $\mathbb{P}\{(q_{ij} - \mathbb{E}\,D_{1i-1}) - D_{ij-1} \geq 0\} \geq \alpha$, or alternatively, $F_{ij-1}(q_{ij} - \mathbb{E}\,D_{1i-1}) \geq \alpha$. This immediately translates to a lower bound on the base-stock level, such that, $q_{ij} - \mathbb{E}\,D_{1i-1} \geq F^{-1}_{ij-1}(\alpha)$. We can therefore express the α service level constraint as follows:

$$q_{ij} \geq (\mathbb{E}\,D_{1i-1} + F^{-1}_{ij-1}(\alpha))x_{ij}, \quad i \in [1, N],\ j \in [i+1, N+1]. \tag{11}$$

Notice that (11) is active only if $[i, j)$ is a replenishment cycle. Otherwise, both sides of the constraint become zero.

5.1.2. β_c Service Level. The β_c service level constraint is an upper bound on the expected number of backorders in each replenishment cycle. The bound is specified as a fraction $1 - \beta_c$ of expected total demand over the replenishment cycle. Because backorders accumulate over the replenishment cycle, β_c service level constraint is satisfied if the expected backorders in the last period of each replenishment cycle falls below the associated upper bound. We can easily formulate this constraint by means of the readily available variable H_{ijt} that expresses the expected backorders at period t of any replenishment cycle $[i, j)$. Because we are only interested in the last period of each replenishment cycle, we can write β_c service level constraints as follows:

$$H_{ijj-1} \leq (1 - \beta_c)\mathbb{E}\,D_{ij-1} x_{ij}, \quad i \in [1, N],\ j \in [i+1, N+1]. \tag{12}$$

5.1.3. β Service Level. The β service level constraint is an upper bound on the expected total number of backorders over the planning horizon. The bound is defined as a fraction $1 - \beta$ of expected total demand. As is evident from its definition, the β service level is an aggregate version of the β_c service level. Thus, it can be expressed by an expression that is analogous to (12). Here, the difference is that the upper bound should be imposed on the sum of backorders over all replenishment cycles, rather than each individual replenishment cycle. Therefore, the β service level constraint can be written as

$$\sum_{i=1}^{N} \sum_{j=i+1}^{N+1} H_{ijj-1} \leq (1 - \beta)\mathbb{E}\,D_{1N}. \tag{13}$$

5.2. Lost Sales

The proposed formulation can easily be augmented to account for the case where demands realized when the system is out-of-stock are lost rather than backordered. Because the basics of the formulation remain unchanged, here we present a brief sketch of this extension and highlight its differences with the formulation with backorders. From a modelling point of view, there are two consequences of lost sales. First, the objective function should be revised to account for the revenue loss because of lost sales. We undertake this by penalizing expected lost sales by a unit selling price v. Second, the inventory conservation dynamics should be altered to capture the fact that expected inventory level equals the expected on-hand stock. For this purpose, we use decision variables $s = \{s_{ij}\}$ to express base-stock levels; such that s_{ij} equals the base-stock level at period i if $[i, j)$ is a replenishment cycle, and 0 otherwise. These variables enable us to express expected on-hand inventory at period t of a replenishment cycle $[i, j)$ as $s_{ij} - \mathbb{E}\, D_{it} + H_{ijt}$. Then, we can rewrite the MIP model as follows:

$$\min \sum_{i=1}^{N} \sum_{j=i+1}^{N+1} \left(K x_{ij} + \sum_{t=i}^{j-1} h(s_{ij} - \mathbb{E}\, D_{it} x_{ij} + H_{ijt}) + v H_{ijj-1} \right) \tag{14}$$

subject to

(4)–(6),

$$s_{ij} \le M x_{ij}, \quad i \in [1, N],\ j \in [i+1, N+1], \tag{15}$$

$$\sum_{i=1}^{t-1} (s_{it} - \mathbb{E}\, D_{it-1} x_{it} + H_{itt-1}) \le \sum_{j=t+1}^{N+1} s_{tj}, \quad t \in [2, N], \tag{16}$$

$$H_{ijt} \ge a x_{ij} + b s_{ij}, \quad i \in [1, N],\ j \in [i+1, N+1],\ t \in [i, j-1],\ (a, b) \in W_{it}, \tag{17}$$

$$H_{ijt} \ge 0 \quad i \in [1, N],\ j \in [i+1, N+1],\ t \in [i, j-1],$$

$$s_{ij} \ge 0,\ x_{ij} \in \{0, 1\} \quad i \in [1, N],\ j \in [i+1, N+1].$$

The objective function given in (14) minimizes the expected total costs over the planning horizon. Here, the expression inside the outer summations stands for the expected total cost over a replenishment cycle and it is comprised of three terms. The first term is the setup cost. The second term is the sum of expected holding costs over all periods over the replenishment cycle. The third term is the expected cost of unmet demand, accounted for at the end of the replenishment cycle. The constraints are as follows: (15) states that the base-stock level should be zero if the corresponding interval is not a replenishment cycle, (16) ensures that the expected on-hand stock at the end of a replenishment cycle cannot exceed the base-stock level of the next replenishment cycle, and (17) bounds the approximate loss function value from below by means of the piecewise linear approximation.

The lost sales model presented here can also be expressed as an equivalent expected profit maximization model by subtracting the objective function from the constant $v\, \mathbb{E}\, D_{1N}$, which is the expected revenue that would have been gained if all demand were met.

6. Computational Study

The purpose of the computational study is to demonstrate the efficiency of the extended formulation with and without the dynamic cut generation approach. We are interested, in particular, to analyze the following:

1. How does the extended formulation perform as compared to the state-of-the-art formulations in the literature?

2. How does the extended formulation scale when deployed without and with the dynamic cut generation approach?

3. How reliable is the expected cost figure obtained by the extended formulation?

In the remainder of this section, we first explain the design of the computational study, and then we discuss our findings with respect to the previously raised questions.

6.1. Experiment Design

In our computational study, we conduct experiments on three different formulations, namely, the state-of-the-art benchmark formulation (i.e., Tarim and Kingsman 2006, Rossi et al. 2015), abbreviated as BM; and the extended formulation deployed without and with the dynamic cut generation approach, respectively, abbreviated as PM and RM-Cut. We consider all model variants addressed in the present study. There are four models where unmet demand is backordered, that is, the penalty cost model and α, β_c, and β service level models; and one model where unmet demand is lost, that is, the lost sales model.

We make use of two sets of instances, namely, Set-A and Set-B. These instance sets are described as follows. Set-A is designed to assess the performance of the formulations under different parameter settings. Here, we generate problem instances using a full factorial design where we take into account all factors that could possibly have an impact on the effectiveness of the proposed formulations. We use the same holding cost $h = 1$ in all experiments and consider three setup costs $K = \{225, 900, 2{,}500\}$. For the penalty cost model, service level models, and the lost sales models, we use the respective parameter values $p = \{2, 5, 10\}$, $\alpha = \{0.90, 0.95, 0.99\}$, $\beta = \{0.80, 0.90, 0.95\}$, $\beta_c = \{0.80, 0.90, 0.95\}$, and $v = \{10, 20, 40\}$. We consider three different planning horizon lengths $N = \{20, 30, 40\}$. We assume that demands over the planning horizon are normally distributed with a fixed coefficient of variation and consider three coefficient of variation values $\rho \in \{0.1, 0.2, 0.3\}$. We randomly generate mean demands using two demand patterns $\pi \in \{\text{Erratic}, \text{Lumpy}\}$. For the erratic pattern, mean demands are drawn from a uniform distribution on the interval $[0, 100]$. For the lumpy pattern, mean demands are drawn from a uniform distribution on the interval $[0, 420]$ with probability 0.2 and from a uniform distribution on the interval $[0, 20]$ with probability 0.8. For each combination of planning horizon length and demand pattern, we generate 10 random problem instances. This setting leads to 8,100 test instances. Set-B, on the other hand, is designed to assess the scalability of the formulations. Here, we consider problem instances with longer planning horizon lengths $N = \{50, 60, 70, 80, 90, 100\}$. We assume π = Erratic and $\rho = 0.3$, and generate 10 random instances for each planning horizon length. We use fixed cost parameters $h = 1$ and $K = 225$ for all instances. For the penalty cost model, service level models, and the lost sales model, we use the respective parameter values $p = 10$, $\alpha = 0.99$, $\beta_c = 0.95$, $\beta = 0.95$, and $v = 40$. Set-B has 300 problem instances in total.

We carry out all computational experiments on a single thread of a 3.40 GHz Intel Core i7-3770 CPU with 16 GB RAM. We use Gurobi v6.5 as an MIP solver. For BM and PM, we use Rossi et al. (2014) 11-piece lower bound approximation of the normal distribution loss function. We deploy RM-Cut with an ϵ value, which guarantees that the total approximation error over the planning horizon is at most one cost unit (a negligible error bound considering the cost parameters used). This setting require model-dependent ϵ values as objective functions of model variants are different. These are $\epsilon = 1/N(h + p)$ for the penalty cost model, $\epsilon = 1/Nh$ for all service level models, and $\epsilon = 1/N(h + v)$ for the lost sales model.

For the sake of completeness, all data on problem instances and computational results are made available as an online appendix available at http://www.hips.hacettepe.edu.tr/pss/joc_extended.zip.

6.2. Effectiveness of the Extended Formulation as Compared to the State-of-the-Art Formulations

In the first part of the computational study, we assess the computational performance of the extended formulation PM against the state-of-the-art benchmark formulation BM. We make use of Set-A instances to investigate the performance of these formulations under a large variety of parameter settings. For both formulations, we use default solver settings and impose a time limit of half an hour.

Tables 1–5 present the results of the computational study for each model variant. Here, S-GAP is the relative integrality gap and E-GAP is the relative optimality gap at termination, both in percentages. TIME is the solution time in seconds and NODES is the number of explored nodes. All these solution statistics are averaged over all problem instances characterized by the same pivot parameter.

We summarize the findings of this part of the computational study as follows. The results show without doubt that PM is more time efficient than BM. This finding is consistent for all model variants and parameter settings. PM solves all instances to optimality and averages a TIME of 3.41 seconds. BM, on the other hand, fails to solve one out of ten instances within the time limit, and averages a TIME of 326.34 seconds. There are settings where BM performs relatively better; for instance, when demand follows a lumpy pattern with a low level of coefficient of variation. Also, it yields a much better performance specifically for the α service level model. But even for such favorable settings, it is dominated by PM in terms of computational performance. BM becomes computationally prohibitive as the planning horizon gets longer. Its average TIME is 3.32 seconds for instances with a planning horizon of 20 periods, whereas it reaches up to 779.02 seconds for instances with 40 periods. This is not the case for PM, as it averages a TIME of only 6.92 seconds for instances with 40 periods. The dominance of PM stems from its

Table 1. Solution Statistics of BM and PM for the Penalty Cost Model

	BM				PM			
Parameters	S-GAP	E-GAP	TIME	NODES	S-GAP	E-GAP	TIME	NODES
π								
Erratic	67.40	5.06	637.89	$7.8\,10^4$	0.00	0.00	3.61	0.00
Lumpy	53.15	0.03	39.77	$7.3\,10^4$	0.01	0.00	3.17	0.02
N								
20	51.27	0.00	2.47	$2.5\,10^3$	0.00	0.00	0.28	0.00
30	61.91	0.01	136.34	$5.0\,10^4$	0.01	0.00	1.77	0.02
40	67.65	7.62	877.67	$7.5\,10^4$	0.01	0.00	8.12	0.01
ρ								
0.1	63.70	2.62	331.03	$4.8\,10^4$	0.00	0.00	3.04	0.00
0.2	59.91	2.58	341.10	$4.2\,10^4$	0.00	0.00	3.40	0.01
0.3	57.22	2.44	344.35	$3.9\,10^4$	0.01	0.00	3.73	0.03
K								
225	65.61	2.39	422.13	$5.6\,10^4$	0.02	0.00	3.64	0.03
900	62.17	4.12	363.38	$3.4\,10^4$	0.00	0.00	3.38	0.00
2,500	53.04	1.12	230.97	$3.7\,10^4$	0.00	0.00	3.16	0.00
p								
2	59.73	2.55	313.13	$4.5\,10^4$	0.00	0.00	5.08	0.00
5	60.73	2.69	346.79	$4.5\,10^4$	0.00	0.00	3.06	0.00
10	60.36	2.40	356.55	$3.8\,10^4$	0.01	0.00	2.03	0.03
Average	60.28	2.55	338.83	$4.3\,10^4$	0.01	0.00	3.39	0.01

Table 2. Solution Statistics of BM and PM for the α Service Level Model

	BM				PM			
Parameters	S-GAP	E-GAP	TIME	NODES	S-GAP	E-GAP	TIME	NODES
π								
Erratic	57.80	0.00	22.41	$3.9\,10^3$	0.00	0.00	1.27	0.00
Lumpy	41.05	0.00	2.76	$6.2\,10^2$	0.24	0.00	1.33	0.24
N								
20	41.61	0.00	0.28	$2.7\,10^2$	0.05	0.00	0.19	0.01
30	50.22	0.00	3.35	$1.3\,10^3$	0.17	0.00	0.84	0.15
40	56.45	0.00	34.12	$5.1\,10^3$	0.14	0.00	2.87	0.21
ρ								
0.1	50.63	0.00	5.03	$1.1\,10^3$	0.01	0.00	1.25	0.00
0.2	49.77	0.00	11.08	$2.1\,10^3$	0.11	0.00	1.31	0.11
0.3	47.88	0.00	21.65	$3.6\,10^3$	0.23	0.00	1.35	0.25
K								
225	53.71	0.00	3.32	$7.8\,10^2$	0.28	0.00	1.36	0.25
900	52.29	0.00	18.24	$3.2\,10^3$	0.08	0.00	1.30	0.11
2,500	42.28	0.00	16.20	$2.8\,10^3$	0.00	0.00	1.24	0.00
α								
0.90	50.08	0.00	9.28	$1.6\,10^3$	0.06	0.00	1.48	0.04
0.95	49.72	0.00	11.19	$2.1\,10^3$	0.11	0.00	1.48	0.11
0.99	48.48	0.00	17.29	$3.0\,10^3$	0.20	0.00	0.95	0.22
Average	49.43	0.00	12.59	$2.2\,10^3$	0.12	0.00	1.30	0.12

tight linear relaxation. PM averages an S-GAP of 0.05% and a NODES of 0.11. That is, most of the time an optimal solution can be found at the root node of the search tree.

We now turn our attention to the computational performance of PM for different model variants and parameter settings. The results demonstrate that the performance of PM varies among model variants. Its average TIME is largest for the β service level model—as it reaches up to 5.62 seconds. The main determinant of PM's computational performance is the length of the planning horizon. PM averages a TIME of 0.31, 1.66, and 6.92 seconds for instances with 10, 20, and 40 periods, respectively. PM yields a rather stable performance for different demand patterns, coefficient of variations, and setup costs. It is yet possible to observe, for all model variants, that PM's average TIME is decreasing on the model-dependent parameter that relates to stockouts (i.e., backorder penalty cost, service levels, and selling price). Thus, it is possible to conclude that PM's performance is positively correlated with the criticality of stockouts.

Table 3. Solution Statistics of BM and PM for the β_c Service Level Model

	BM				PM			
Parameters	S-GAP	E-GAP	TIME	NODES	S-GAP	E-GAP	TIME	NODES
π								
Erratic	78.09	0.71	295.85	$9.4\,10^3$	0.00	0.00	2.50	0.00
Lumpy	61.53	0.00	14.71	$8.4\,10^2$	0.00	0.00	2.30	0.00
N								
20	60.20	0.00	1.01	$5.2\,10^2$	0.00	0.00	0.34	0.00
30	71.52	0.00	24.89	$2.0\,10^3$	0.00	0.00	1.54	0.00
40	77.72	1.07	439.94	$1.3\,10^4$	0.00	0.00	5.33	0.01
ρ								
0.1	70.84	0.15	120.04	$4.7\,10^3$	0.00	0.00	2.31	0.00
0.2	70.10	0.37	162.29	$5.1\,10^3$	0.00	0.00	2.35	0.00
0.3	68.50	0.55	183.51	$5.4\,10^3$	0.00	0.00	2.54	0.01
K								
225	78.92	0.54	197.29	$5.6\,10^3$	0.00	0.00	2.69	0.01
900	70.64	0.46	186.89	$5.4\,10^3$	0.00	0.00	2.36	0.00
2,500	59.87	0.07	81.67	$4.3\,10^3$	0.00	0.00	2.16	0.00
β_c								
0.80	69.62	0.80	214.49	$6.0\,10^3$	0.00	0.00	2.37	0.00
0.90	70.50	0.26	176.61	$6.1\,10^3$	0.00	0.00	2.30	0.00
0.95	69.32	0.00	74.74	$3.2\,10^3$	0.00	0.00	2.54	0.01
Average	69.81	0.36	155.28	$5.1\,10^3$	0.00	0.00	2.40	0.00

Table 4. Solution Statistics of BM and PM for the β Service Level Model

	BM				PM			
Parameters	S-GAP	E-GAP	TIME	NODES	S-GAP	E-GAP	TIME	NODES
π								
Erratic	79.89	8.72	758.99	$9.0\,10^4$	0.10	0.00	6.35	0.46
Lumpy	64.83	0.55	157.04	$1.5\,10^4$	0.17	0.00	4.89	0.26
N								
20	62.37	0.00	5.70	$3.0\,10^3$	0.19	0.00	0.49	0.24
30	74.54	1.39	280.19	$7.4\,10^4$	0.12	0.00	2.88	0.32
40	80.17	12.52	1,088.16	$8.1\,10^4$	0.10	0.00	13.48	0.52
ρ								
0.1	70.38	3.53	377.02	$4.8\,10^4$	0.22	0.00	5.73	0.47
0.2	72.35	4.59	454.98	$4.9\,10^4$	0.12	0.00	5.59	0.34
0.3	74.35	5.79	542.05	$6.0\,10^4$	0.07	0.00	5.53	0.27
K								
225	83.20	7.71	669.40	$9.8\,10^4$	0.08	0.00	7.30	0.27
900	72.83	4.24	402.97	$4.0\,10^4$	0.11	0.00	5.45	0.29
2,500	61.06	1.96	301.68	$2.0\,10^4$	0.22	0.00	4.11	0.52
β								
0.80	69.15	6.91	471.83	$6.5\,10^4$	0.32	0.00	7.01	0.83
0.90	72.82	4.77	482.03	$5.5\,10^4$	0.07	0.00	5.24	0.19
0.95	75.12	2.23	420.19	$3.7\,10^4$	0.02	0.00	4.61	0.07
Average	72.36	4.64	458.01	$5.3\,10^4$	0.14	0.00	5.62	0.36

Table 5. Solution Statistics of BM and PM for the Lost Sales Model

	BM				PM			
Parameters	S-GAP	E-GAP	TIME	NODES	S-GAP	E-GAP	TIME	NODES
π								
Erratic	2,158.30	3.42	908.87	$1.2\,10^5$	0.00	0.00	2.11	0.00
Lumpy	4,419.41	0.70	425.11	$4.8\,10^4$	0.02	0.00	2.09	0.07
N								
20	2,579.01	0.00	7.16	$6.4\,10^3$	0.01	0.00	0.28	0.01
30	3,019.54	0.30	538.59	$1.2\,10^5$	0.01	0.00	1.25	0.04
40	4,268.01	5.88	1,455.21	$1.3\,10^5$	0.01	0.00	4.77	0.06
ρ								
0.1	3,311.64	1.19	548.99	$7.4\,10^4$	0.00	0.00	1.90	0.00
0.2	3,159.59	2.04	695.94	$8.5\,10^4$	0.01	0.00	2.10	0.04
0.3	3,395.33	2.95	756.04	$9.3\,10^4$	0.02	0.00	2.30	0.07
K								
225	1,799.42	1.41	909.62	$1.4\,10^5$	0.03	0.00	1.69	0.10
900	2,212.45	2.20	702.95	$7.4\,10^4$	0.00	0.00	2.09	0.01
2,500	5,854.70	2.58	388.39	$3.5\,10^4$	0.00	0.00	2.52	0.00
v								
10	5,714.83	3.56	593.19	$7.0\,10^4$	0.00	0.00	3.49	0.01
20	2,321.89	1.70	688.94	$8.1\,10^4$	0.01	0.00	1.48	0.02
40	1,829.84	0.92	718.83	$1.0\,10^5$	0.02	0.00	1.33	0.08
Average	3,288.85	2.06	666.99	$8.4\,10^4$	0.01	0.00	2.10	0.04

6.3. Scalability of the Extended Formulation Without and With the Dynamic Cut Generation Approach

In the second part of the computational study, we conduct experiments to investigate the scalability of PM and RM-Cut, that is, extended formulation deployed without and with the dynamic cut generation approach. Here, we make use of Set-B instances that are characterized by longer planning horizons, that is, from 50 to 100 periods. We adopt default solver settings in our experiments, with the exception of using the barrier method to solve linear relaxations as it yields significantly better results as compared to solver's default method. We implement the cut generation approach by Gurobi's lazy constraint callback routines. We set the feasibility threshold to 10^{-9} to make sure that the cut generation procedure can indeed provide the desired precision. We deploy both formulations with no computational time limit, and solve all instances to optimality.

Tables 6–10 present the average computation times of PM and RM-Cut in seconds for different model variants and varying planning horizon lengths. There are two main findings that can be derived from our numerical results. First, we confirm that PM and RM-Cut are capable of solving realistic-sized instances to optimality in very reasonable computational times. Second, we observe that RM-Cut dominates PM in terms of computational performance. These findings are consistent among all model variants. PM solves all problem instances in less than ten minutes and RM-Cut in less than two minutes. The largest gap between PM and RM-Cut is observed for the α service level model. Here, for instances with a planning horizon of 100 periods, PM and RM-Cut average a computational time of 323.24 and 6.40 seconds, respectively. This is an important result as it highlights that one can bypass the offline piecewise linearization of the cost function and establish an even more precise approximation on the fly without sacrificing computational efficiency. The importance of this result becomes more evident if we recall that finding a good piecewise linear approximation is a challenging optimization problem itself.

Table 6. Computational Times of PM and RM-Cut for the Penalty Cost Model

	50	60	70	80	90	100
PM	10.46	30.53	47.83	97.91	264.86	401.98
RM-Cut	2.05	3.74	7.48	11.15	18.97	29.22

Table 7. Computational Times of PM and RM-Cut for the α Service Level Model

	50	60	70	80	90	100
PM	11.06	25.99	53.52	78.94	183.23	323.24
RM-Cut	0.43	0.78	1.72	2.82	4.05	6.40

Table 8. Computational Times of PM and RM-Cut for the β_c Service Level Model

	50	60	70	80	90	100
PM	12.52	29.61	70.92	107.19	239.54	414.18
RM-Cut	2.75	5.14	9.61	16.12	28.87	44.65

Table 9. Computational Times of PM and RM-Cut for the β Service Level Model

	50	60	70	80	90	100
PM	15.38	34.73	76.23	148.38	256.92	480.59
RM-Cut	5.40	9.87	18.07	28.35	47.57	77.46

Table 10. Computational Times of PM and RM-Cut for the Lost Sales Model

	50	60	70	80	90	100
PM	7.25	16.40	32.94	47.15	93.05	127.04
RM-Cut	2.95	5.03	8.06	12.62	19.61	30.51

6.4. Reliability of the Extended Formulation

In the last part of the computational study, we analyze the reliability of the expected cost figures obtained by PM and RM-Cut. As discussed earlier, these formulations are not exact (as are the earlier formulations in the literature) because of the approximation error resulting from piecewise linearization of the original nonlinear objective function and the systematic error due to the assumption that the postreplenishment inventory level at the beginning of each replenishment cycle always equals the base-stock level. As such, the optimal costs of these formulations are always lower than the expected costs that would realize when the associated policies are deployed. Here, we assess the extent of this mismatch under different parameter settings on Set-A instances. We adopt the following procedure. We solve all instances with PM and RM-Cut and record optimal policies and costs. Then, first we compute the ex post approximation error, that is, the difference between the cost obtained when policy parameters are plugged into the original nonlinear objective function and the cost obtained by the formulation. Second, we compute the simulation error, that is, the difference between the expected cost obtained when the policy is simulated and the cost obtained by the formulation. We conduct 10^5 simulation runs for all instances.

Tables 11–15 present the error rates for each model variant. Here, A-ERR and S-ERR are the relative approximation and simulation errors in percentages, and A-ERR* and S-ERR* are their nominal values, respectively. These values are averaged over all problem instances characterized by the same pivot parameter. It should be clear that it makes better sense to concentrate on relative errors rather than nominal ones when comparing problem instances with different planning horizon lengths. Nevertheless, we provide both error rates for the sake of completeness.

We first discuss the approximation error. Because PM's piecewise linear approximation is model independent, its approximation error significantly varies over model variants. Its approximation quality is better for service level models since their objective functions are less sensitive to the expected loss value. PM's approximation quality is lowest for the lost sales model, but even for this model it averages an A-ERR of 0.49%. Thus, it is fair to say that the static 11-piece

Table 11. Error Rates of PM and RM-Cut for the Penalty Cost Model

	PM				RM-Cut			
Parameters	A-ERR*	A-ERR	S-ERR*	S-ERR	A-ERR*	A-ERR	S-ERR*	S-ERR
π								
Erratic	18.86	0.21	19.16	0.21	0.24	0.00	0.55	0.01
Lumpy	16.07	0.24	84.23	1.90	0.27	0.01	68.55	1.66
N								
20	10.52	0.21	21.57	0.70	0.24	0.01	11.56	0.50
30	18.30	0.23	61.57	1.32	0.27	0.00	43.69	1.09
40	23.59	0.23	71.96	1.15	0.27	0.00	48.40	0.91
ρ								
0.1	8.47	0.14	9.85	0.18	0.22	0.00	1.85	0.05
0.2	17.61	0.23	44.20	0.95	0.26	0.00	26.94	0.72
0.3	26.33	0.31	101.04	2.05	0.28	0.00	74.87	1.73
K								
225	16.20	0.36	99.38	2.62	0.31	0.01	83.40	2.26
900	17.89	0.20	36.76	0.43	0.25	0.00	19.30	0.23
2,500	18.32	0.12	18.96	0.12	0.21	0.00	0.95	0.01
p								
2	8.14	0.12	27.96	0.72	0.28	0.01	19.02	0.56
5	16.05	0.21	51.16	1.06	0.26	0.00	35.52	0.87
10	28.22	0.34	75.97	1.39	0.23	0.00	49.10	1.06
Average	17.47	0.23	51.70	1.06	0.26	0.00	34.55	0.83

Table 12. Error Rates of PM and RM-Cut for the α Service Level Model

	PM				RM-Cut			
Parameters	A-ERR*	A-ERR	S-ERR*	S-ERR	A-ERR*	A-ERR	S-ERR*	S-ERR
π								
Erratic	1.06	0.01	3.36	0.04	0.03	0.00	2.48	0.03
Lumpy	1.37	0.02	116.40	2.13	0.04	0.00	115.76	2.12
N								
20	0.65	0.01	28.09	0.84	0.04	0.00	27.54	0.83
30	1.29	0.01	76.17	1.33	0.04	0.00	76.06	1.33
40	1.71	0.01	75.37	1.08	0.03	0.00	73.77	1.07
ρ								
0.1	0.51	0.01	7.55	0.20	0.04	0.00	7.18	0.20
0.2	1.20	0.01	50.50	1.05	0.04	0.00	49.30	1.04
0.3	1.94	0.02	121.59	2.00	0.03	0.00	120.89	1.99
K								
225	1.05	0.02	115.50	2.56	0.04	0.00	114.53	2.54
900	1.28	0.01	51.72	0.60	0.04	0.00	51.52	0.60
2,500	1.32	0.01	12.42	0.09	0.03	0.00	11.31	0.08
α								
2	8.14	0.12	27.96	0.72	0.28	0.01	19.02	0.56
5	16.05	0.21	51.16	1.06	0.26	0.00	35.52	0.87
10	28.22	0.34	75.97	1.39	0.23	0.00	49.10	1.06
Average	1.22	0.01	59.88	1.08	0.04	0.00	59.12	1.08

linear approximation yields a reasonable approximation error on the overall. RM-Cut, on the other hand, is deployed with a dynamic piecewise approximation scheme designed to guarantee a nominal approximation error of at most one cost unit. We observe that its error rate is indeed below this upper bound. RM-Cut averages an A-ERR* less than 0.3 cost units for all model variants. This leads to an A-ERR far less than 0.01%. As such, we can conclude that RM-Cut completely eliminates the approximation error.

Next, we assess the systematic error. It is important to remark that the simulation error accommodates both the approximation error and the systematic error. As such, we can infer that the difference between S-ERR and A-ERR reflects the relative systematic error. A first observation we can derive from our results is that the systematic errors of PM and RM-Cut are almost the same. For instance, let us consider the penalty cost model. Here, PM and RM-Cut, respectively, average an S-ERR's of 1.06% and 0.83%

Table 13. Error Rates of PM and RM-Cut for the β_c Service Level Model

	PM				RM-Cut			
Parameters	A-ERR*	A-ERR	S-ERR*	S-ERR	A-ERR*	A-ERR	S-ERR*	S-ERR
π								
Erratic	1.75	0.02	1.95	0.03	0.10	0.00	0.15	0.00
Lumpy	2.39	0.05	16.78	0.50	0.11	0.00	14.94	0.46
N								
20	1.17	0.03	3.84	0.18	0.10	0.00	2.75	0.15
30	2.21	0.04	10.84	0.32	0.11	0.00	9.14	0.29
40	2.84	0.04	13.42	0.29	0.10	0.00	10.74	0.25
ρ								
0.1	0.90	0.02	0.92	0.02	0.07	0.00	0.08	0.00
0.2	2.10	0.04	4.36	0.12	0.11	0.00	2.39	0.08
0.3	3.23	0.05	22.81	0.64	0.13	0.00	20.16	0.60
K								
225	1.86	0.07	21.30	0.70	0.13	0.00	20.11	0.66
900	2.20	0.03	4.34	0.06	0.10	0.00	2.32	0.03
2,500	2.16	0.02	2.46	0.02	0.09	0.00	0.20	0.00
β_c								
0.8	2.27	0.05	2.56	0.07	0.09	0.00	0.36	0.02
0.9	2.13	0.04	6.90	0.20	0.10	0.00	4.71	0.17
0.95	1.82	0.03	18.64	0.51	0.11	0.00	17.57	0.50
Average	2.07	0.04	9.37	0.26	0.10	0.00	7.54	0.23

Table 14. Error Rates of PM and RM-Cut for the β Service Level Model

	PM				RM-Cut			
Parameters	A-ERR*	A-ERR	S-ERR*	S-ERR	A-ERR*	A-ERR	S-ERR*	S-ERR
π								
Erratic	2.71	0.04	2.75	0.04	0.22	0.00	0.30	0.01
Lumpy	2.86	0.07	27.60	0.98	0.25	0.01	25.89	0.95
N								
20	1.57	0.05	6.78	0.38	0.20	0.01	5.39	0.34
30	2.91	0.06	19.21	0.67	0.25	0.01	17.28	0.64
40	3.88	0.06	19.54	0.49	0.26	0.00	16.62	0.45
ρ								
0.1	1.21	0.03	1.18	0.03	0.19	0.01	0.17	0.01
0.2	2.82	0.06	6.31	0.18	0.25	0.01	4.27	0.15
0.3	4.33	0.08	38.04	1.32	0.26	0.01	34.84	1.28
K								
225	2.82	0.10	38.48	1.44	0.30	0.01	37.25	1.40
900	2.86	0.04	4.33	0.07	0.23	0.00	1.95	0.03
2,500	2.69	0.02	2.72	0.02	0.18	0.00	0.09	0.00
β								
0.8	2.89	0.07	5.62	0.25	0.23	0.01	3.07	0.19
0.9	2.82	0.05	14.57	0.52	0.25	0.01	12.48	0.49
0.95	2.65	0.04	25.34	0.76	0.23	0.00	23.74	0.75
Average	2.79	0.05	15.18	0.51	0.24	0.01	13.09	0.48

and an A-ERR of 0.23% and less than 0.01%. Thus, for both PM and RM-Cut the systematic error is around 0.83%. This result is consistent for all model variants and parameter settings. Based on this observation, we restrict the remainder of our discussion to RM-Cut's systematic error—which can directly be reflected by its S-ERR. The results demonstrate that RM-Cut's average S-ERR is 0.66%, which is relatively small, but not negligible. The systematic error varies among model variants. The average S-ERR is largest for the α service level model—as it reaches up to 1.08%. It also follows a very clear pattern with respect to parameter settings. More specifically, S-ERR is significantly lower (approaches zero) for the erratic pattern as compared to the lumpy pattern. Also, it tends to increase with the coefficient of the variation of the demand. This is not unexpected as the assumption on postreplenishment inventory levels is valid when demand is

Table 15. Error Rates of PM and RM-Cut for the Lost Sales Model

	PM				RM-Cut			
Parameters	A-ERR*	A-ERR	S-ERR*	S-ERR	A-ERR*	A-ERR	S-ERR*	S-ERR
π								
Erratic	49.30	0.53	49.86	0.54	0.15	0.00	0.74	0.01
Lumpy	28.21	0.45	85.47	1.80	0.09	0.00	58.85	1.37
N								
20	24.73	0.47	37.81	0.99	0.12	0.00	14.03	0.54
30	39.23	0.50	74.94	1.32	0.13	0.00	36.90	0.83
40	52.32	0.50	90.24	1.20	0.12	0.00	38.45	0.70
ρ								
0.1	19.34	0.30	21.75	0.36	0.12	0.00	2.69	0.07
0.2	38.60	0.51	63.66	1.17	0.12	0.00	25.34	0.66
0.3	58.33	0.67	117.59	1.98	0.13	0.00	61.34	1.35
K								
225	48.46	0.93	124.18	2.82	0.19	0.00	77.27	1.92
900	38.07	0.37	48.85	0.51	0.11	0.00	12.07	0.16
2,500	29.73	0.17	29.96	0.18	0.07	0.00	0.04	0.00
v								
10	16.74	0.23	34.90	0.73	0.12	0.00	18.20	0.50
20	33.24	0.43	60.19	1.09	0.12	0.00	28.15	0.69
40	66.30	0.81	107.90	1.69	0.13	0.00	43.03	0.89
Average	38.76	0.49	67.66	1.17	0.12	0.00	29.79	0.69

deterministic. Finally, S-ERR is decreasing on the setup cost and increasing on the model-dependent parameter that relates to stockouts. These observations are consistent among all model variants.

7. Conclusions

In this paper, we presented an extended formulation of the stochastic lot-sizing problem for the static-dynamic uncertainty strategy. The formulation has a number of advantages; it is computationally efficient, it is very flexible as it can accommodate variants of the problem characterized by penalty costs and service level constraints, as well as backorders and lost sales, and it can find a minimum cost solution with any given level of precision by means of a novel dynamic cut generation approach. We concluded our study with a large-scale computational experiment where we demonstrated that the extended formulation is far more time efficient as compared to the state-of-the-art formulations and it is able to solve realistic-sized instances in reasonable computational times when deployed without and with the dynamic cut generation approach. We also showed that the dynamic cut generation approach improves the computational performance of the extended formulation, and we thereby concluded that one can bypass a priori piecewise linearization of the cost function and establish a better approximation on the fly without sacrificing computational efficiency. Finally, because the extended formulation is not exact, we critically assessed and verified its reliability.

There are several avenues for further research. An important research direction is to develop exact methods for the static-dynamic uncertainty strategy. The extended formulation presented in the current study is not exact because of the approximation error and the systematic error, and although our dynamic cut generation approach can effectively eliminate the approximation error, the systematic error still prevails. As we illustrated, this error is small but not negligible. Also, it is relatively large under some parameter settings. For instance, practitioners should put serious consideration into the systematic error when the underlying lot sizing problem involves slow moving products with highly variable demands, in particular when the cost of placing an order is very low. Another interesting research direction is to analyze the problem with limited distributional information on random demands. The assumption of complete distributional information of demand does not hold in general and it also becomes very restrictive especially for longer planning horizon lengths. This provides a motivation for investigating distributionally robust approaches for the stochastic lot-sizing problem.

References

Bookbinder JH, Tan JY (1988) Strategies for the probabilistic lot-sizing problem with service-level constraints. *Management Sci.* 34(9):1096–1108.

Duran MA, Grossmann IE (1986) An outer-approximation algorithm for a class of mixed-integer nonlinear programs. *Math. Programming* 36(3):307–339.

Frenzen CL, Sasao T, Butler JT (2010) On the number of segments needed in a piecewise linear approximation. *J. Comput. Appl. Math.* 234(2):437–446.

Gavrilovic MM (1975) Optimal approximation of convex curves by functions which are piecewise linear. *J. Math. Anal. Appl.* 52(2):260–282.

Kilic OA, Tarim SA (2011) An investigation of setup instability in non-stationary stochastic inventory systems. *Internat. J. Production Econom.* 133(1):286–292.

Mutlu F, Cetinkaya S, Bookbinder JH (2010) An analytical model for computing the optimal time-and-quantity-based policy for consolidated shipments. *IIE Trans.* 42(5):367–377.

Özen U, Doğru MK, Tarim SA (2012) Static-dynamic uncertainty strategy for a single-item stochastic inventory control problem. *Omega* 40(3):348–357.

Porteus EL (2002) *Foundations of Stochastic Inventory Theory* (Stanford University Press, Palo Alto, CA).

Rossi R, Kilic OA, Tarim SA (2015) Piecewise linear approximations for the static–dynamic uncertainty strategy in stochastic lot-sizing. *Omega* 50:126–140.

Rossi R, Tarim SA, Hnich B, Prestwich S (2008) A global chance-constraint for stochastic inventory systems under service level. *Constraints* 13(4):490–517.

Rossi R, Tarim SA, Hnich B, Prestwich S (2011) A state space augmentation algorithm for the replenishment cycle inventory policy. *Internat. J. Production Econom.* 133(1):377–384.

Rossi R, Tarim SA, Hnich B, Prestwich S (2012) Constraint programming for stochastic inventory systems under shortage cost. *Ann. Oper. Res.* 195(1):49–71.

Rossi R, Tarim SA, Prestwich S, Hnich B (2014) Piecewise linear lower and upper bounds for the standard normal first order loss function. *Appl. Math. Comput.* 231:489–502.

Silver E (1978) Inventory control under a probabilistic time-varying, demand pattern. *IIE Trans.* 10(4):371–379.

Silver EA, Meal HC (1973) A heuristic for selecting lot size quantities for the case of a deterministic time-varying demand rate and discrete opportunities for replenishment. *Production Inventory Management* 14(2):64–74.

Silver EA, Pyke DF, Peterson R (1998) *Inventory Management and Production Planning and Scheduling* (Wiley, New York).

Tarim SA, Kingsman BG (2004) The stochastic dynamic production/inventory lot-sizing problem with service-level constraints. *Internat. J. Production Econom.* 88(1):105–119.

Tarim SA, Kingsman BG (2006) Modelling and computing (R^n, S^n) policies for inventory systems with non-stationary stochastic demand. *Eur. J. Oper. Res.* 174(1):581–599.

Tarim SA, Smith BM (2008) Constraint programming for computing non-stationary (R, S) inventory policies. *Eur. J. Oper. Res.* 189(3):1004–1021.

Tarim SA, Dogru MK, Ozen U, Rossi R (2011) An efficient computational method for a stochastic dynamic lot-sizing problem under service-level constraints. *Eur. J. Oper. Res.* 215(3):563–571.

Tarim SA, Hnich B, Rossi R, Prestwich S (2009) Cost-based filtering techniques for stochastic inventory control under service level constraints. *Constraints* 14(2):137–176.

Tunc H, Kilic OA, Tarim SA, Eksioglu B (2014) A reformulation for the stochastic lot sizing problem with service-level constraints. *Oper. Res. Lett.* 42(2):161–165.

Wagner HM, Whitin TM (1958) Dynamic version of the economic lot size model. *Management Sci.* 5(1):89–96.