A Stochastic Multi-item Lot-sizing Problem with Bounded Number of Setups

Etienne de Saint Germain[1,2], Vincent Leclère[2] and Frédéric Meunier[2]
[1]Argon Consulting, 122 Rue Édouard Vaillant, 92300 Levallois-Perret, France
[2]CERMICS, Cité Descartes, 6-8 Avenue Blaise Pascal, 77455 Champs-sur-Marne, France

Keywords: Lot-sizing, Stochastic Optimization, Sample Average Approximation, Simulation.

Abstract: Within a partnership with a consulting company, we address a production problem modeled as a stochastic multi-item lot-sizing problem with bounded numbers of setups per period and without setup cost. While this formulation seems to be rather non-standard in the lot-sizing landscape, it is motivated by concrete missions of the company. Since the deterministic version of the problem is NP-hard and its full stochastic version clearly intractable, we turn to approximate methods and propose a repeated two-stage stochastic programming approach to solve it. Using simulations on real-world instances, we show that our method gives better results than current heuristics used in industry. Moreover, our method provides lower bounds proving the quality of the approach. Since the computational times are small and the method easy to use, our contribution constitutes a promising response to the original industrial problem.

1 INTRODUCTION

Fixing the production level for the forthcoming week is a basic decision to be taken when managing a production line. Usually, a demand has to be satisfied at due dates but the limited capacity of the line prevents last minute production. On the other hand, too early productions may lead to unnecessary high inventory costs. The challenge of this kind of problems, known as *lot-sizing problems* in the operational research community, consists in finding a trade-off between demand satisfaction and holding costs. This is a well-studied topic, with many variations (deterministic/stochastic, single/multi item, etc.). Recent surveys have been proposed: see (Gicquel et al., 2008, Quadt and Kuhn, 2008) for the deterministic version and (Mula et al., 2006, Aloulou et al., 2014, Díaz-Madroñero et al., 2014) for the stochastic version. When several references can be produced on a same line – the so-called *multi-item* lot-sizing problem –, the capacity is often all the more reduced as the number of distinct references produced over the current week is high. Indeed, changing a reference in production stops the line for a moment. This additional capacity reduction is usually modeled by setup costs contributing to the total cost.

The present work introduces a stochastic multi-item lot-sizing problem met by the authors within a partnership with a consulting company. A non-standard feature of the problem is that the capacity reduction due to reference changes is not modeled by setup costs but instead by an explicit upper bound on the total number of references that can be produced over a week. According to the consulting company, many clients aim at minimizing mainly their inventory costs while keeping the number of distinct references produced over each week below some threshold. This is essentially because, contrary to inventory costs, setup costs are hard to quantify and a maximal number of possible changes per week is easy to estimate. To the best of authors' knowledge, the problem addressed in the present work is original and such a bound on the number of distinct references produced over a week has not been considered by academics yet, with the notable exception of (Rubaszewski et al., 2011) but, contrary to our problem, their bound is an overall bound for the whole horizon and they still consider setup costs.

We consider two versions of the problem. In a first one, backorder costs are present and the objective is to minimize the overall inventory costs (sum of holding and backorder costs). In a second version, there are only holding costs and there is a service level constraint to be satisfied over the whole horizon. We propose for these two versions a method that can be easily used and maintained in practice. The efficiency

de Saint Germain E., Leclère V. and Meunier F.
A Stochastic Multi-item Lot-sizing Problem with Bounded Number of Setups.
DOI: 10.5220/0006622501060114
In *Proceedings of the 7th International Conference on Operations Research and Enterprise Systems (ICORES 2018)*, pages 106-114
ISBN: 978-989-758-285-1

of the method is proved via extensive numerical experiments on real industrial data. In particular, we compare our results with current heuristics used in industry and with lower bounds.

2 PROBLEM FORMULATION AND MODEL

The assembly line produces a set $\mathcal{R}$ of references over T weeks. The number of distinct references produced over a week cannot exceed N. There is also an upper bound on the total week production (summed over all references). We normalize all quantities so that this upper bound is equal to 1.

The production must satisfy a random demand. The demand of reference r over week t is a random parameter $\mathbf{d}_t^r$, whose realization is known at the end of week t. When production of a reference r is not used to satisfy the demand, it can be stored but incurs a unit holding cost $h^r > 0$ per week. When a demand is not satisfied by the production of the current week or by inventory, it can be satisfied later but incurs a unit backorder cost γ per week for some coefficient $\gamma > 0$. Note that there is no setup cost, as discussed in the introduction. For each reference r, there is an initial inventory $s_0^r \in \mathbb{R}_+$.

At the beginning of each week, before the demand of each reference is revealed, the production of the week has to be fixed. The objective is to minimize the total expected inventory cost (holding cost plus backorder cost) over the whole horizon of T weeks.

Regarding randomness, we assume that for any r and t, realizations of $(\mathbf{d}_t^r, \mathbf{d}_{t+1}^r, \ldots, \mathbf{d}_T^r)$ have finite expectation and can be efficiently sampled, knowing a realization of $\mathbf{d}_{t-1}^r$.

Before turning to the modeling of the problem, we make two remarks. The first one is about its position in the literaure. Our problem (in its deterministic version) is a variation of the *Capacitated Lot-Sizing Problem* (see (Karimi et al., 2003) for a review of models and algorithms). The single differences with the usual basic version of that problem is that there is no setup cost and that there is this upper bound on the total number of distinct references that can be produced over a week. The second remark is about the somehow non-conventional assumption that the demand is revealed at the end of the week. There are indeed industrial cases where this occurs, like in computer manufacturing. However, the methods developed in this paper can be easily adapted to the case where the demand is revealed at the beginning of the week.

The production problem at the beginning of week t can be modeled as the following stochastic program (S):

$$
\begin{array}{ll}
\min \ \mathbb{E}\left[\displaystyle\sum_{t'=t}^{T}\sum_{r\in\mathcal{R}}(h^r\tilde{\mathbf{s}}_{t'}^r+\gamma\mathbf{b}_{t'}^r)\right] & \\
\text{s.t.} & \\
\mathbf{s}_{t'}^r = \tilde{\mathbf{s}}_{t'}^r - \mathbf{b}_{t'}^r & t' \in [\![t,T]\!],\ r\in\mathcal{R} \\
\mathbf{s}_{t'}^r = \mathbf{s}_{t'-1}^r + \mathbf{q}_{t'}^r - \mathbf{d}_{t'}^r & t' \in [\![t,T]\!],\ r\in\mathcal{R} \\
\displaystyle\sum_{r\in\mathcal{R}} \mathbf{q}_{t'}^r \le 1 & t' \in [\![t,T]\!] \\
\mathbf{q}_{t'}^r \le \mathbf{x}_{t'}^r & t' \in [\![t,T]\!],\ r\in\mathcal{R} \\
\displaystyle\sum_{r\in\mathcal{R}} \mathbf{x}_{t'}^r \le N & t' \in [\![t,T]\!] \\
\mathbf{x}_{t'}^r \in \{0,1\} & t' \in [\![t,T]\!],\ r\in\mathcal{R} \\
\mathbf{q}_{t'}^r,\ \tilde{\mathbf{s}}_{t'}^r,\ \mathbf{b}_{t'}^r \ge 0 & t' \in [\![t,T]\!],\ r\in\mathcal{R} \\
\sigma(\mathbf{q}_{t'}^r) \subset \sigma\Big((\mathbf{d}_t^{r'},\ldots,\mathbf{d}_{t'-1}^{r'})_{r'\in\mathcal{R}}\Big) & t' \in [\![t,T]\!],\ r\in\mathcal{R},
\end{array}
$$

where the variable $\tilde{\mathbf{s}}_{t'}^r$ (resp. $\mathbf{b}_{t'}^r$) models the inventory (resp. the backorder) of reference r at the end of week t', the variable $\mathbf{q}_{t'}^r$ models the quantity of reference r produced over week t', and the variable $\mathbf{x}_{t'}^r$ takes the value 1 if the reference r is produced over week t' and 0 otherwise. All these variables are random. The last constraint of the program, written as a *measurability constraint*, means that the values of the variables $\mathbf{q}_{t'}^r$ can only depend on the values taken by the demand before time t' (the planner does not know the future). The reader would have probably expected similar constraints for the other variables, but it is easy to see that adding them changes neither the optimal value, nor the production decisions. Note that in this program the inventory $\mathbf{s}_{t-1}^r$ is not a variable and has a known deterministic value at the beginning of week t.

A feasible solution $(\mathbf{q}_{t'}^r)_{t'\ge t, r\in\mathcal{R}}$ of (S) provides a deterministic production $(\mathbf{q}_t^r)_{r\in\mathcal{R}}$ for the current week t.

3 FOCUS ON THE DETERMINISTIC VERSION

The deterministic version of (S) is obtained by removing the measurability constraint and by considering all $\mathbf{d}_t^r$'s deterministic, denoted then simply by d_t^r. The variables are then also denoted with non-bold characters. The main message of this section is that this version is already difficult.

The deterministic version is NP-hard in the strong sense for any fixed $N \ge 3$, since there is a straightforward reduction from 3-PARTITION. Given a set of $3n$ positive integers $\{a_1, \ldots, a_{3n}\}$, this latter problem consists in deciding whether this set admits a partition into n triples of same sum. It is a notorious strongly NP-complete problem (Garey and Johnson, 1979). The problem 3-PARTITION can be reduced to the case $N = 3$ where each index in $[\![1, 3n]\!]$ is a distinct reference r whose demand d_t^r is equal to 0 for all period t, except for the last period $T = n$ where it is equal to $\frac{n}{\sum_{r'=1}^{3n} a_{r'}} a_r$. (The cases $N \geq 4$ are dealt with via a similar reduction.) The complexity status of the deterministic version when $N = 1$ or $N = 2$ seems to be a challenging open question.

It is also worth noting that the optimal value of the continuous relaxation is independent of N. This can be seen by considering the mixed integer program obtained by removing the constraints $\sum_{r\in\mathcal{R}} x_{t'}^r \leq N$: any feasible solution of that program provides a feasible solution of the original program of same value, simply by redefining $x_{t'}^r = q_{t'}^r$ for all t' and all r if necessary. In an attempt of improving the quality of the continuous relaxation, one may consider the extended formulation with the binary variables y_t^p for each $p \in \binom{\mathcal{R}}{N}$ and each $t \in [\![1, T]\!]$ in place of the x_t^r's (indicating the references produced on period t) but it can be shown with a nontrivial proof that it does not improve the quality of the continuous relaxation. Similarly, the extended formulation with the binary variables z_τ^r for each $\tau \subseteq [\![1, T]\!]$ (indicating the periods of production of the reference r) does not improve the quality of the relaxation (this time, with an easy proof).

4 METHOD

As explained in Section 3, the deterministic version of (S) is difficult. Therefore, we cannot expect a quick algorithm solving exactly the problem, and this holds especially for the full stochastic version. Moreover, one of the requests of the partner was to have an easy to understand method, which can be used and maintained in practice, with short computation times. We propose a two-stage approximation consisting in replacing the measurability constraint by

$$\left\{\begin{array}{ll} \sigma(\mathbf{q}_t^r) \subset \sigma(\varnothing) & r \in \mathcal{R} \\ \sigma(\mathbf{q}_{t'}^r) \subset \sigma\big((\mathbf{d}_t^{r'}, \ldots, \mathbf{d}_T^{r'})_{r'\in\mathcal{R}}\big) & t' \geq t+1, r \in \mathcal{R}, \end{array}\right.$$

which provides a relaxation of the initial program: the production decisions for the current week t can still not depend on the future, but now the subsequent production decisions depend on the future demand. We denote this relaxation by (2SA).

This approximation is a *two-stage approximation* as we distinguish between two levels of information over the uncertainty: production decisions for the first week are the *first stage* variables, while all other decisions are *second stage* variables. Three-stages or more generally multistage approximation would give better approximations of (S) but increases exponentially the number of variables. We chose for practicability reasons to stick to the two-stage approximation.

The (2SA) relaxation is then solved by a classical *sample average approximation*, see (Kleywegt et al., 2002) for a presentation of the method. We build a set Ω of m scenarios sampled uniformly at random. Each of these scenarios is a possible realization of $(\mathbf{d}_t^r, \mathbf{d}_{t+1}^r, \ldots, \mathbf{d}_T^r)$ for each r. The parameter m is fixed prior to the resolution.

We get the following mixed integer program (2SA-m), solved by any standard MIP solver.

$$\min \frac{1}{m} \sum_{\omega\in\Omega} \sum_{t'=t}^{T} \sum_{r\in\mathcal{R}} \left(h^r \tilde{s}_{t',\omega}^r + \gamma b_{t',\omega}^r\right)$$

s.t.

$$\begin{array}{ll}
s_{t',\omega}^r = \tilde{s}_{t',\omega}^r - b_{t',\omega}^r & t' \in [\![t, T]\!], r \in \mathcal{R}, \omega \in \Omega \\
s_{t',\omega}^r = s_{t'-1,\omega}^r + q_{t',\omega}^r - d_{t',\omega}^r & t' \in [\![t, T]\!], r \in \mathcal{R}, \omega \in \Omega \\
\sum_{r\in\mathcal{R}} q_{t',\omega}^r \leq 1 & t' \in [\![t, T]\!], \omega \in \Omega \\
q_{t',\omega}^r \leq x_{t',\omega}^r & t' \in [\![t, T]\!], r \in \mathcal{R}, \omega \in \Omega \\
\sum_{r\in\mathcal{R}} x_{t',\omega}^r \leq N & t' \in [\![t, T]\!], \omega \in \Omega \\
x_{t,\omega}^r = x_t^r & r \in \mathcal{R}, \omega \in \Omega \\
q_{t,\omega}^r = q_t^r & r \in \mathcal{R}, \omega \in \Omega \\
x_t^r,\ x_{t',\omega}^r \in \{0, 1\} & t' \in [\![t, T]\!], r \in \mathcal{R}, \omega \in \Omega \\
q_t^r,\ q_{t',\omega}^r,\ \tilde{s}_{t',\omega}^r,\ b_{t',\omega}^r \geq 0 & t' \in [\![t, T]\!], r \in \mathcal{R}, \omega \in \Omega
\end{array}$$

At week t, the production is then set to be the solution $(q_t^r)_{r\in\mathcal{R}}$ found by the solver.

The validity of this method for solving (2SA) is supported by the following proposition.

Proposition 1. *The following three properties hold when m goes to infinity:*

(i) *The value of* (2SA-m) *converges almost surely to the optimal value of* (2SA).

(ii) *For every m, we consider the values $(\hat{q}_{t,m}^r, \hat{x}_{t,m}^r)_{r\in\mathcal{R}}$ of the decision variables for week t of an optimal solution of* (2SA-m). *Any limit point of these values is an optimal solution of* (2SA).

(iii) *Let $\varepsilon > \delta > 0$. Assume that the random demand $(\mathbf{d}_{t'}^r)_{t'\geq t, r\in\mathcal{R}}$ is such that*

$$\exists C, K, \quad \forall u \in \mathbb{R}, \qquad \mathbb{E}[e^{u\|\mathbf{d}\|}] \leq Ce^{u^2 K}. \qquad (1)$$

Denote by Q_m^δ (resp. Q^ε) the set of all possible values of $(\hat{q}_{t,m}^r)_{r\in\mathcal{R}}$ in a δ-optimal solution of (2SA-m) *(resp. in an ε-optimal solution of* (2SA)*). Then for every $\alpha \in (0, 1)$, we have $\mathbb{P}(Q_m^\delta \subseteq Q^\varepsilon) > 1 - \alpha$ for m large enough.*

If the random demand $\mathbf{d}$ is bounded or Gaussian then it satisfies (1). (Some companies work with possibly negative demands. Assuming a normal distribution is hence not irrelevant.)

The proof of part (iii) relies on the following technical lemma.

Lemma 2. *Consider $g(d) = \inf_{y \in Y} G(y,d)$, where Y is non-empty and where the function G is non-negative and κ-Lipschitz with respect to d. If the random variable $\mathbf{d}$ satisfies (1), then $g(\mathbf{d})$ also satisfies (1) with $C' = \max\{1, e^{G(y_0,0)}C\}$ and $K' = K\kappa + G(y_0,0)$, for $y_0 \in Y$.*

Proof. For $u \geq 0$ we take $y_0 \in Y$ yielding $g(\mathbf{d}) \leq \mathbf{G}(\mathbf{y_0},\mathbf{0}) + \kappa\|\mathbf{d}\|$. We then have $\mathbb{E}[e^{ug(\mathbf{d})}] \leq \mathbb{E}[e^{uG(y_0,0)}e^{\kappa u\|\mathbf{d}\|}] \leq C'e^{K'u^2}$. For $u \leq 0$, by non-negativity of G we have $g(\mathbf{d}) \geq 0$, hence $\mathbb{E}[e^{ug(\mathbf{d})}] \leq 1$. □

Proof of Proposition 1. Let Q_t be the (bounded) set of feasible values for the first-stage variables $q_t = (q_t^r)_{r\in\mathcal{R}}$ in (2SA). Denote by $F(q_t,d)$ the minimal cost of (2SA) that can be reached when the first stage-variables are fixed to $q_t \in Q_t$ and the realization of the demand is $d = (d_{t'}^r)_{t'\geq t, r\in\mathcal{R}}$. We introduce the map $f\colon q_t \mapsto \mathbb{E}[F(q_t,\mathbf{d})]$, which associates to a given choice of $q_t \in Q_t$ the expected minimal cost, and similarly the map $\hat{f}_m$, which associates to a $q_t \in Q_t$ the minimal cost of (2SA-m) when the first-stage variables are set to this q_t.

The map f is continuous and Q_t is compact. Thus f is bounded, and, by (Shapiro et al., 2009, Theorem 7.48), we have that $(\hat{f}_m(q_t))_{m\in\mathbb{Z}_+}$ converges to $f(q_t)$ uniformly on Q_t. Then (i) and (ii) are direct consequences of (Shapiro et al., 2009, Theorem 5.3).

By Lemma 2, there exist K and C such that for any $q_t \in Q_t$, $F(q_t,\mathbf{d})$ satisfy (1). Consequently there exists $\sigma > 0$ such that for all $q_t, q_t' \in Q_t$, the random variable $[F(q_t,\mathbf{d}) - f(q_t)] - [F(q_t',\mathbf{d}) - f(q_t')]$ is σ-subgaussian, (see e.g., (Vershynin, 2010)). Furthermore, for any demand d, the map $F(\cdot,d)$ is Lipschitz-continuous on Q_t. Then, according to (Shapiro et al., 2009, Theorem 5.18), for every $\alpha \in (0,1)$, there exists $M \in \mathbb{Z}_+$ such that for $m \geq M$, we have $\mathbb{P}(Q_m^\delta \subseteq Q^\varepsilon) > 1-\alpha$. □

5 NUMERICAL EXPERIMENTS

5.1 Instances

C++11 has been chosen for the implementations and Gurobi 6.5.1 was used to solve the model on a PC with Intel(R) Core(TM) i7-3770 CPU @ 3.40GHz and 8Go RAM.

The instances used are realistic and have been provided by a client of the partner. The client gave actually the figures of seven assembly lines but we give the results for only two of them: L2 and L6. The line L2 experiences overcapacity: the ratio expected demand over total capacity is smaller than 1 (actually equal to 0.7). The line L6 experiences undercapacity: the ratio expected demand over total capacity is larger than 1 (actually equal to 1.1). The horizon T is the typical one used in practice by this client, namely $T = 13$ weeks (a quarter). The demand is obtained via $\mathbf{d}_t^r = (\bar{d}_t^r + \mathbf{e}_t^r)^+$, where the $\bar{d}_t^r$'s are historical data and where $\mathbf{e}_t^r$ is a *generalized autoregressive process* so that $\mathbf{e}_t^r = 0.25\mathbf{e}_{t-1}^r + 0.75\varepsilon_t$, where ε_t is a Gaussian white noise process with zero mean and standard deviation equal to $v\bar{d}_t^r$ where v is the "volatility" and is chosen in $\{0.2, 0.5\}$. The initial inventory is set to $s_0^r = \frac{1}{3}(\bar{d}_1^r + \bar{d}_2^r + \bar{d}_3^r)$. The other parameters are provided in Table 1. In particular, for each value of v, we have considered three possible values of the unit backorder cost γ, which have been determined following a procedure described in Section 6. At the moment, we do not discuss further these values and take them as part of the input, as required by the problem formulation. The parameter C is the capacity of the line before normalization. (Recall that problem and the model have been formulated in Section 2 after normalization.) In the column $\tilde{h}^r$, we indicate the range of the holding costs before normalization. We obtain the h^r's by dividing these costs by C.

The number m of scenarios used to solve (2SA-m) is fixed to 20, determined by preliminary experiments showing that it is a good trade-off between accuracy and tractability. The time limit of the solver has been set to 90 seconds.

5.2 Other Heuristics

Our method is compared with three other heuristics.

The first heuristic is the deterministic version of (S), where the random demand is replaced by its expectation.

The second one, the *lot-size heuristic*, consists in determining before the first week once and for all a value ℓ_r^* for each reference $r \in \mathcal{R}$. At time t, if the inventory of reference r is below a precomputed safety level, the quantity q_t^r is chosen so that the inventory of reference r exceeds the safety level of exactly ℓ_r^*. In case of capacity issues, the production is postponed and thus backorder costs appear. In addition, if some capacity issues are easily anticipated, the production of a reference r can be activated even if the inventory

Table 1: Instance characteristics.

Instances	Instance characteristics						
	$\lvert\mathcal{R}\rvert$	$\max(\bar{d}_t^r)$	C	N	$\tilde{h}^r$	v	γ
L2_v20_13	21	4992	10562	7	35–61	0.2	13
L2_v20_81							81
L2_v20_203							203
L2_v50_48						0.5	48
L2_v50_154							154
L2_v50_341							341
L6_v20_3	22	8640	13299	8	16–23	0.2	3
L6_v20_19							19
L6_v20_55							55
L6_v50_11						0.5	11
L6_v50_42							42
L6_v50_98							98

is not below the safety level.

The third one, the *cover-size heuristic* is almost the same, but instead of precomputing a fixed quantity for each reference, a duration T_r^* is fixed before the first week. When the inventory of reference r is below the safety level, the quantity q_t^r is computed so that the inventory of reference r exceeds the safety level of the expected demand for the next T_r^* weeks.

The values ℓ_r^* and T_r^* are determined as follows. $(T_r^*)_{r\in\mathcal{R}}$ is actually chosen to be the optimal solution of the following convex program, which somehow considers the problem at a "macroscopic" level. (Similar convex programs in the same context have been considered in the literature; see (Ziegler, 1982) for example.)

$$\begin{array}{lll} \min & \displaystyle\sum_{r\in\mathcal{R}} h^r \bar{d}^r T_r & \\ \text{s.t.} & \displaystyle\sum_{r\in\mathcal{R}} \frac{1}{T_r} \leq N & \\ & T_r > 0 & r \in \mathcal{R}, \end{array}$$

where $\bar{d}^r = \mathbb{E}\left[\sum_{t=1}^{T} \mathbf{d}_t^r\right]$.

The parameter ℓ_r^* of the lot-size heuristic is then set to $\bar{d}^r T_r^*$.

The safety levels have been provided by the partner and are those used in practice.

The cover-size heuristic adapts the production to the realization of the demand, contrary to the lot-size heuristic. According to our partner, it makes the cover-size heuristic more suitable for situations with low short term volatility of demand or for overcapacitated lines, while the lot-size heuristic is expected to behave better with high short term volatility of demand or for undercapacitated lines.

Notice that the backorder costs are not taken into account at all for determining the values of the parameters ℓ_r^* and T_r^*. But playing with safety levels allows to prevent too large backorder costs. However, in real life it is usually the other way round: the company does not associate costs to backorder and aims at keeping the total amount of unsatisfied demand below some predetermined level. We come back to this point later in Section 6.

5.3 Results

The results are provided in Table 2. All quantities are in M€ and are given with a confidence interval at 95%.

The column LB provides the lower bound obtained by the optimal value at time $t = 1$ of program (2SA-m) (with $m = 1000$ and a time limit of 24 hours for the solver).

The column 2SA-m is the cost of the method proposed in Section 4. (We remind the reader that we propose $m = 20$ in this case.) The three next columns provide the results for the three heuristics described in Section 5.2. For these four columns, between 27 and 30 runs have been used for each instance.

Complementary numerical results of these experiments are provided in Tables 3 and 4, discussed in Section 6 where other criteria are considered. Table 3 can already by of interest since it deals with holding costs obtained for the L2 instances.

5.4 Comments

Our method clearly outperforms lot-size and cover-size heuristics and is better than the deterministic approximation for all but one instance. By running our method instead of a usual heuristic at the beginning of each week, the inventory costs can be reduced often by more than 50%. For the instance L2_v20_13, the inventory costs have been divided by more than 6 (which corresponds to several M€).

Table 2: Results - Inventory costs (in M€).

Instances	LB	2SA-*m*	Det.	Cover-size	Lot-size
L2_v20_13	0.53	0.89 ± 0.03	1.17 ± 0.10	6.95 ± 0.17	7.79 ± 0.14
L2_v20_81	0.94	2.29 ± 0.06	2.36 ± 0.07	8.12 ± 0.19	9.65 ± 0.14
L2_v20_203	1.00	3.05 ± 0.07	3.25 ± 0.08	9.35 ± 0.29	10.99 ± 0.19
L2_v50_48	0.97	2.73 ± 0.11	3.06 ± 0.21	8.03 ± 0.26	8.37 ± 0.21
L2_v50_154	1.36	4.54 ± 0.20	5.06 ± 0.33	10.83 ± 0.53	11.20 ± 0.38
L2_v50_341	1.51	5.91 ± 0.25	7.90 ± 0.66	15.17 ± 1.21	14.65 ± 0.77
L6_v20_3	0.54	0.61 ± 0.01	0.70 ± 0.02	1.71 ± 0.08	1.74 ± 0.08
L6_v20_19	1.41	1.81 ± 0.06	1.86 ± 0.06	3.51 ± 0.12	3.20 ± 0.08
L6_v20_55	2.67	3.57 ± 0.24	3.71 ± 0.30	7.49 ± 0.39	6.24 ± 0.34
L6_v50_11	1.33	2.00 ± 0.11	2.14 ± 0.12	3.42 ± 0.15	3.03 ± 0.13
L6_v50_42	2.99	4.45 ± 0.53	4.48 ± 0.51	7.99 ± 0.62	6.57 ± 0.60
L6_v50_98	6.13	8.29 ± 1.23	7.94 ± 1.04	16.34 ± 1.61	12.96 ± 1.45

We may note that the cover-size heuristic behaves better than the lot-size heuristic on the L2 instances, while the converse holds on the L6 instances. It is in line with our partner's opinion regarding their behaviors with respect to the capacity of the line (see Section 5.2). Regarding their behaviors with respect to the volatility, it does not seem to be possible to draw any concrete conclusion.

The 2SA-*m* algorithm requires 90 seconds to output a solution, while lot-size and cover-size heuristics take less than a second and the deterministic approximation less than 10 seconds. The 2SA-*m* algorithm is thus slower, but note that 90 seconds to be run only once at the beginning of each week remains very short. Moreover, even with improved computers, the lot-size and cover-size heuristics and the deterministic approximation will not change their output (in all our experiments, Gurobi always found the optimal solution of the deterministic approximation). This is not the case for 2SA-*m*, which means that it would benefit from improved computational capacities.

Otherwise, complementary figures not provided here show that the deterministic approximation and 2SA-*m* use almost every available setups (between 97% and 100%) whereas lot-size and cover-size heuristics cannot take full advantage of this flexibility (only 84% to 98% of setups are used).

Our method dramatically reduces the inventory costs, while remaining a quite simple approach. It provides without any doubt a positive answer to the industrial request. Moreover, as explained in the next section, our method can be used to address the problem of minimizing the holding costs, while keeping the backorder at a reasonable level.

6 CONTROLLING THE FILL RATE SERVICE LEVEL

6.1 The Problem

In practice, except when they are enshrined through contracts with the client, backorder costs can be hard to estimate. Thus, we consider here an alternative version of the problem, which often meets the objectives in industry. Everything remains the same, except that there is no backorder cost anymore, which means that we remove the corresponding part in the objective function of (S) and that the following constraint is added:

$$\mathbb{E}\left[\frac{\sum_{r\in\mathcal{R}}\sum_{t=1}^{T}\min(\mathbf{d}_t^r, \mathbf{q}_t^r + \tilde{\mathbf{s}}_{t-1}^r)}{\sum_{r\in\mathcal{R}}\mathbf{d}^r}\right] \geq \beta, \qquad (2)$$

where $\beta \in [0,1]$ is the *desired fill rate service level* defined by the company, and where $\mathbf{d}^r = \sum_{t=1}^{T}\mathbf{d}_t^r$. This constraint implies that, in expectation, the proportion of the production delivered on-time is at least β. Indeed, at week t, we deliver on-time the minimum between the demand and the sum of the real inventory at the end of week $t-1$ and of the production of week t.

First, note that this constraint does not make a difference between a one-week delay and a two-week delay. Second, this constraint is a global constraint to satisfy over the whole horizon: it does not depend on the current week. This is a matter of modeling point of view since we were not able to get a precise formulation from our partner and from its clients. A drawback of this formulation with respect to the one given in Section 2 is that this additional constraint cannot always be satisfied. In the next section, we propose a way to address this new constraint in an approximate way, while being able to provide solutions in any case.

6.2 Surrogate Backorder Costs

We address the alternative formulation by defining a surrogate backorder coefficient γ before the first week, with the idea to use the algorithm of Section 4 and to heuristically entice it to choose solutions satisfying constraint (2).

$$\gamma := \max_{r\in\mathcal{R}} \gamma^r \tag{3a}$$

$$\gamma^r := \frac{\mathbb{P}[\mathbf{d}^r \leq q^r(\beta)]}{\mathbb{P}[\mathbf{d}^r > q^r(\beta)]} \tag{3b}$$

$$q^r(\beta) := \min\left\{ q \in \mathbb{R}_+ \;\middle|\; \mathbb{E}\left[\frac{\min(\mathbf{d}^r, q)}{\mathbf{d}^r}\right] \geq \beta \right\} \tag{3c}$$

where $\mathbf{d}^r = \sum_{t=1}^T \mathbf{d}_t^r$ is the demand of reference r aggregated over time. Since $\mathbf{d}^r$ is integrable, $q^r(\beta)$ is well-defined (we set $\frac{0}{0} = \beta$ so that references with no demand would not impact the constraint). Computing an approximate value of $q^r(\beta)$ at an arbitrary precision can easily be performed by binary search.

To justify this choice, consider the original problem of Section 2 with only one reference and for a horizon of one week. Assuming no initial inventory, it takes then the form of the famous *newsvendor problem* (see e.g., (Shapiro et al., 2009, Chapter 1))

$$\min_{q\geq 0} \mathbb{E}\left[h^r(q-\mathbf{d}^r)^+ + \gamma^r(\mathbf{d}^r - q)^+\right], \tag{4}$$

where γ^r is a unit backorder cost specific to reference r. The next proposition means that with the right choice for γ^r, the alternative version with a desired fill rate service level β (which takes the form of (3b) since $h^r > 0$) is equivalent to the original one (4). Of course, it holds only for the case with one reference and a horizon of one week. Many references lead to several possible values γ^r for the surrogate backorder cost. The choice in (3a) of taking the maximum of them for our actual γ aims at facing the additional uncertainty induced by the limited number of setups per week and the length of the horizon.

Proposition 3. *Define γ^r as in* (3b). *Then $q^r(\beta)$ is the smallest optimal solution to* (4).

Proof. The aggregated production problem (4) is known to have the optimal solution $q^{r*} = F_{\mathbf{d}^r}^{-1}(\gamma^r/(\gamma^r + h^r))$, where $F_{\mathbf{d}^r}^{-1}$ is the left-inverse of the cumulative distribution function of $\mathbf{d}^r$, i.e., $F_{\mathbf{d}^r}^{-1}(\kappa) = \inf\{q \mid \mathbb{P}(\mathbf{d}^r \leq q) \geq \kappa\}$. Since we have set $\gamma^r = \frac{\mathbb{P}[\mathbf{d}^r \leq q^r(\beta)]}{\mathbb{P}[\mathbf{d}^r > q^r(\beta)]}$, we have $q^{r*} = \inf\{q \mid \mathbb{P}(\mathbf{d}^r \leq q) \geq \mathbb{P}(\mathbf{d}^r \leq q^r(\beta)\}$, which implies that $q^{r*} \leq q^r(\beta)$. Now if this inequality were strict, then it would mean that $\mathbb{P}(\mathbf{d}^r \in (q^{r*}, q^r(\beta))) = 0$, which contradicts the minimality assumption in the definition of $q^r(\beta)$ (Equation (3c)). □

Note that this formulation does not take into account the capacity constraint. It is therefore probably better suited to overcapacited production lines.

Remark. If instead of controlling the *fill rate service level*, we want to control the *cycle service level*, defined as the probability of satisfying the whole demand, then we can choose

$$\gamma = \frac{\beta}{1-\beta} \max_{r\in\mathcal{R}} h^r. \tag{5}$$

Indeed, in this case, the optimal solution q^{r*} of (4) satisfies $\mathbb{P}(q^{r*} \geq \mathbf{d}^r) = \beta$. Interestingly, Equation (5) does not depend on the distribution of the demand, which contrasts with the fill rate service level.

6.3 Numerical Results

We discuss briefly the results of the experiments described in Section 5 in the new context of this section. Tables 3 and 4 provide the numerical results: the first table shows the holding costs obtained when using the method proposed in this work; the second table shows the average (over the 30 runs) of the *fill rate service level* measured ex post, which we define by the realization of

$$\frac{\sum_{r\in\mathcal{R}} \sum_{t=1}^T \min(\mathbf{d}_t^r, \mathbf{q}_t^r + \tilde{\mathbf{s}}_t^r)}{\sum_{r\in\mathcal{R}} \mathbf{d}^r}.$$

We plot on Figure 1 the values of the holding costs obtained for each of the 30 runs for the instance L2 with $\beta = 95\%$ and $v = 0.5$. On Figure 2, we plot the average of these values for three desired fill rate service levels (85%, 95%, and 98%).

Again, the inventory costs (which are in this version also the holding costs) are dramatically better than the ones obtained by the heuristics used by the clients of our partner. These two heuristic are however able to provide almost always very good fill rate service level but it comes at the price of very high holding costs

When $\beta \in \{95\%, 98\%\}$, the deterministic approximation reaches better holding costs than 2SA-*m* but worse fill rate service levels. This is especially true when the volatility is high ($v = 0.5$). Further experiments would probably be necessary to understand better how the deterministic approximation and 2SA-*m* compare but since they both use the surrogate backorder cost, their behavior regarding the fill rate service level shows that the way we compute it is relevant.

For high desired fill rate service levels ($\geq 95\%$), the method 2SA-*m* behaves well: the fill rate service level measured in practice is close to the desired level. Since constraint (2) is an approximation of what is sought in practice, these results can be considered as

Table 3: Results - Holding costs (in M€).

Instances		Results				
name	β	γ	2SA-*m*	Det.	Cover-size	Lot-size
L2_v20	85%	13	0.44 ± 0.04	0.82 ± 0.12	6.85 ± 0.18	7.70 ± 0.15
	95%	81	1.58 ± 0.07	1.56 ± 0.11	7.49 ± 0.20	9.14 ± 0.17
	98%	203	2.34 ± 0.10	1.97 ± 0.10	7.79 ± 0.21	9.73 ± 0.19
L2_v50	85%	48	1.55 ± 0.12	1.78 ± 0.18	6.96 ± 0.30	7.58 ± 0.24
	95%	154	2.94 ± 0.19	2.55 ± 0.19	7.47 ± 0.32	8.80 ± 0.28
	98%	341	3.87 ± 0.17	2.68 ± 0.19	7.75 ± 0.34	9.30 ± 0.34

Table 4: Results - Fill rate service level (in %).

Instances		Results				
name	β	γ	2SA-*m*	Det.	Cover-size	Lot-size
L2_v20	85%	13	76.3 ± 0.7	80.3 ± 0.8	96.2 ± 0.4	94.7 ± 0.6
	95%	81	92.5 ± 0.5	92.0 ± 0.7	96.2 ± 0.4	95.1 ± 0.6
	98%	203	96.8 ± 0.4	94.9 ± 0.5	96.2 ± 0.4	95.1 ± 0.5
L2_v50	85%	48	80.5 ± 1.1	81.2 ± 1.6	92.0 ± 1.1	90.7 ± 1.0
	95%	154	90.9 ± 0.9	87.5 ± 1.2	92.1 ± 1.1	91.3 ± 0.9
	98%	341	94.6 ± 0.6	88.2 ± 1.2	92.1 ± 1.1	91.3 ± 0.9

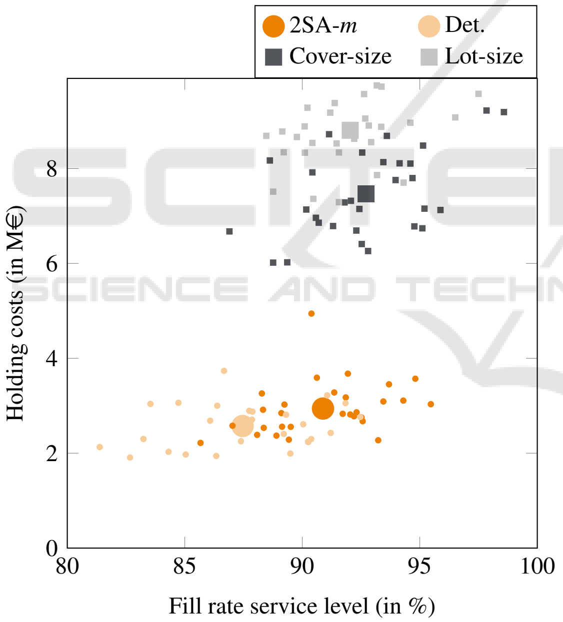

Figure 1: Results for the 30 runs of L2 with $v = 0.5$ and $\beta = 95\%$.

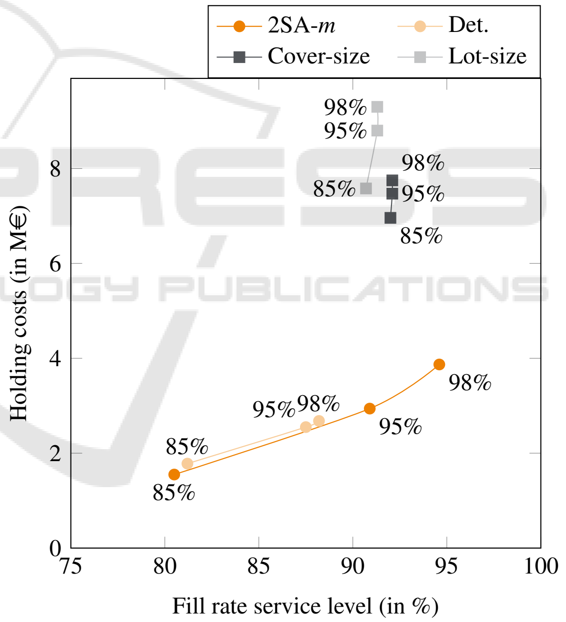

Figure 2: Average results for L2 for various values of desired service level β and $v = 0.5$.

positive and the method of Section 4 combined with our rule of thumb for determining γ forms an effective way to compute the production levels in real-world settings.

7 CONCLUSION

In this paper, we focus on a multi-item lot-sizing problem with a constraint not yet addressed by academic works, while being often met in practice: an explicit upper bound on the weekly number of setups. We model this problem as a stochastic program that includes this constraint and we propose a repeated two-stage approximation to solve it. Our method proves its efficiency on real-word instances and outperforms the heuristics currently used.

The costs in the previous model originate from holding and backorder. According to the consulting partner we work with, companies approach their production decisions by considering only holding costs, while trying to keep the backorder below some threshold. Building on the method proposed for the previous model, we propose a rule of thumb for determining surrogate backorder costs that bias the solutions towards the desired service levels. This rule relies on the famous newsvendor problem, which is the special case of our problem when there is one reference and one week. There exist generalizations of the newsvendor problem with more references or more weeks; it would be worth checking whether they could improve the way this surrogate cost is computed.

REFERENCES

Aloulou, M. A., Dolgui, A., and Kovalyov, M. Y. (2014). A bibliography of non-deterministic lot-sizing models. *International Journal of Production Research*, 52(8):2293–2310.

Díaz-Madroñero, M., Mula, J., and Peidro, D. (2014). A review of discrete-time optimization models for tactical production planning. *International Journal of Production Research*, 52(17):5171–5205.

Garey, M. and Johnson, D. (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness*. Series of books in the mathematical sciences. W. H. Freeman.

Gicquel, C., Minoux, M., and Dallery, Y. (2008). Capacitated lot sizing models: a literature review. working paper or preprint.

Karimi, B., Ghomi, S. F., and Wilson, J. (2003). The capacitated lot sizing problem: a review of models and algorithms. *Omega*, 31(5):365 – 378.

Kleywegt, A. J., Shapiro, A., and de Mello, T. H. (2002). The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, 12(2):479–502.

Mula, J., Poler, R., García-Sabater, J. P., and Lario, F.-C. (2006). Models for production planning under uncertainty: A review. *International Journal of Production Economics*, 103(1):271–285.

Quadt, D. and Kuhn, H. (2008). Capacitated lot-sizing with extensions: a review. *4OR: A Quarterly Journal of Operations Research*, 6(1):61–83.

Rubaszewski, J., Yalaoui, A., Amodeo, L., Yalaoui, N., and Mahdi, H. (2011). A capacitated lot-sizing problem with limited number of changes. In *International Conference on Industrial Engineering and Systems Management, IESM, Metz, France, May 25-27*.

Shapiro, A., Dentcheva, D., and Ruszczyński, A. (2009). *Lectures on stochastic programming: modeling and theory*. SIAM.

Vershynin, R. (2010). Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*.

Ziegler, H. (1982). Solving certain singly constrained convex optimization problems in production planning. *Operations Research Letters*, 1(6):246 – 252.