Managing flexibility in Stochastic Multi-level Lot Sizing Problem with Service Level Constraints

ARTICLE INFO

Keywords:
Stochastic lot sizing
Multi-level lot sizing
Demand uncertainty
Service level
Sample average approximation

ABSTRACT

We investigate the stochastic multi-level lot sizing problem with a service level and in a general setting in which it is possible to have independent demand for the components as well. In this work, we present a systematic approach to evaluate the value of adding flexibility in such context. To this end, the problem with uncertain demand is modeled as a two-stage stochastic program considering different demand scenarios. We first consider at all levels a static strategy in which both the setup decisions and the production quantities are determined in the first stage before the demand is realized. We also model a more adaptive strategy to be more responsive to the realized demand when production quantity decisions of some items can be treated as recourse decisions. We investigate the value of applying such an adaptive strategy and adding more flexibility in the system under different settings. Three different bill of material (BOM) structures (serial, assembly, and general) are considered. We numerically show that adding flexibility to the system results in a cost savings depending on where we add the flexibility in the BOM. While controlling the variation in the plans is very important in the multi-level system, this research show that even having a small degree of flexibility may result in a reasonable amount of cost savings.

1. Introduction

Being cost efficient is a crucial imperative in a competitive business environment. For manufacturing companies, having an efficient production plan in the context of a material requirements planning (MRP) system is important to minimize different costs of production and inventory control. In MRP, time-phased production and inventory plans are crucial decisions to make a balance between customers' demand satisfaction and cost management. While insufficient inventory will lead to shortages, unnecessary stocks will increase the inventory holding cost.

The standard lot sizing problem aims to determine the optimal timing and production quantities in order to satisfy known demand over a finite and discrete time horizon [18]. One of the extensions to the standard lot sizing problem is to consider the multi-level product structure which is common in MRP systems. While only independent demand exists for each of the products in the single level lot sizing problem, there is also dependent demand due to the bill of material (BOM) structure in a multi-level lot sizing problem. There are different product structures in the literature including serial, assembly, and general structures [18].

Within the optimization models for production planning, where all the levels of the BOM are optimized simultaneously, the decision variable related to the backlog only exists for the items that have independent demand. This is due to the fact that, in order to produce the items at the lower levels, their components need to be available at the required time, and it is hence not possible to have backlog for the dependent demand [14]. Indeed, if we would allow backlog at the component level, then a solution can exist in which there is some backlog at the component level, while there is no backlog at the end item level. In this research, we address a more general setting in which, in addition to the end items, each of the components in the BOM may also have an independent demand. Therefore, it is possible to have backlog for them due to this independent portion. This problem with demand at multiple levels in the BOM structure has practical relevance in industries with production and aftermarket services, which require spare parts [36]. A good example is the aerospace industry where, in addition to the demand for end items, the components also have independent demand, which has to be taken into account in the planning process.

Even though demand is typically stochastic in nature, the calculations used in the MRP systems are based on the deterministic demand assumption while safety stocks of items with independent demand are separately determined to hedge against demand uncertainty. The use of safety stocks in such context can potentially lead to sub-optimal solutions since the calculations are performed in isolation under different assumptions than the deterministic model [29]. Unlike this approach which can potentially result in sub-optimal decisions, in this research, we will use stochastic optimization models to deal with demand uncertainty in a single framework. In these models, the lot sizing and safety stock level decisions are jointly determined as the demand's probability distributions are considered in the model [13, 31].

Different forms of service levels are widely used in the calculations of safety stock to deal with demand uncertainty in stochastic lot-sizing problems. However, most of the research has been focused on the single level problem. In the multi-level problem, as it is not possible to have backlog for the dependent demand in the BOM, the service levels are defined only for the independent demand of the end products and the components. The service level which we use in this research is closely related to the δ service level proposed by Helber et al., which considers both the sizes of backorders and the customer waiting time [13]. Here, instead of limiting average backlog, we limit the maximum proportion, taken over all demand realizations, of total backlog to the total possible backlog over the whole planning horizon.

In the single level lot sizing problem, there are three main strategies to deal with multi-period lot sizing problems with stochastic demand and these have a different approach for the setup and production decisions, namely the static, dynamic, and static-dynamic strategy [8]. In the static strategy, the setup and production decisions will be defined at the beginning of the planning horizon and they remain unchanged with the demand realization. In the dynamic strategy, both setups and production decisions may be modified after the demand realization. The static-dynamic strategy is the combination of these two strategies in which the setups are fixed at the beginning of the planning horizon and the production decisions are made after the demand realization. These strategies can also be applied in multi-level lot sizing problems. In the system in which we have independent demand for the components as well, we can apply different strategies at different levels in the BOM to increase the responsiveness in the system, while keeping the nervousness under control. By allowing some production level decision to be made in the second stage, we gain more flexibility and hence lower costs. We provide an illustrative example to demonstrate the benefits of such flexibility, later in this section.

In this research, we model the stochastic multi-level lot sizing problem as a two-stage stochastic programming model which is solved using the sample average approximation (SAA) formulation [15, 23]. The contributions of this research can be stated as follows. First, we investigate the stochastic multi-level lot sizing problem with service level constraints. We specifically consider the case where independent demand exists not only for the end items but at the component level as well. Second, we model different variants of the problem which allow a different level of flexibility (i.e., static strategy versus static-dynamic strategy) at different levels in the BOM structure. Third, we apply the SAA method to empirically evaluate the solution quality with different number of scenarios. Fourth, we propose a systematic way to calculate the cost savings. Based on that, we perform extensive computational experiments to empirically validate the value of flexibility and derive managerial insights for this problem under different settings.

1.1. An illustrative example

The aim of this short section is to offer some intuition on how we may benefit from the production recourse in a multi-level lot sizing problem. We will use a small example which only considers one period. Assume we have two items in the BOM, one end item and one component. The average external demand for each of the items is 100 units. The stochastic demand for each item is represented by 3 independent demand scenarios with equal probability and the demand values of 50, 100, 150. Therefore, we have 9 scenarios in total based on different combinations of external demand for the two items as illustrated in Table 1. Based on these demand scenarios, we consider and compare three different cases. One without any flexibility and two with flexibility in which we have recourse for the end item production. In the first case, there is no flexibility, and the production decisions for both the component and end product are taken before the demand realization. For illustrative purposes, we assume that we will produce equal to the average demand. This results in a production of 100 for the end item and 200 for the component. The 200 units for the component are based on the average external demand of the component itself and the internal demand coming from the end item. Since the production of both the end item and the component is fixed, there is no flexibility. More specifically, the 200 available units of the component will always be allocated in the same way: 100 units to satisfy the internal demand generated by the fixed production of 100 units of the end item, and the other 100 units to satisfy the external demand of the component. These allocation decisions are indicated in the table for each scenario. Because there is no flexibility, a situation might arise such as in scenario 3, where we have enough components (i.e., 200) to satisfy the external demand of the component (i.e., 50) and end item (i.e., 150), but because of the fixed production at of 100 units the end item level we end up with an inventory of 50 units at the component level while having a backlog of 50 at the end item level. The average backlog and inventory levels are calculated over the 9 different scenarios.

Table 1
An illustrative example

Demand scenarios			Solution 1 with no flexibility						Solution 2 with flexibility at the end item						Solution 3 with flexibility at the end item					
			Comp	End	Comp	End	Comp	End	Comp	End	Comp	End	Comp	End	Comp	End	Comp	End	Comp	End
Production			200	100					200	Recourse					200	Recourse				
#	Comp	End	Alloc	Alloc	Inv	Inv	Back	Back	Alloc	Alloc	Inv	Inv	Back	Back	Alloc	Alloc	Inv	Inv	Back	Back
1	50	50	100	100	50	50	0	0	150	50	100	0	0	0	100	100	50	50	0	0
2	50	100	100	100	50	0	0	0	100	100	50	0	0	0	100	100	50	0	0	0
3	50	150	100	100	50	0	0	50	50	150	0	0	0	0	50	150	0	0	0	0
4	100	50	100	100	0	50	0	0	150	50	50	0	0	0	150	50	50	0	0	0
5	100	100	100	100	0	0	0	0	100	100	0	0	0	0	100	100	0	0	0	0
6	100	150	100	100	0	0	0	50	50	150	0	0	50	0	50	150	0	0	50	0
7	150	50	100	100	0	50	50	0	150	50	0	0	0	0	150	50	0	0	0	0
8	150	100	100	100	0	0	50	0	100	100	0	0	50	0	100	100	0	0	50	0
9	150	150	100	100	0	0	50	50	50	150	0	0	100	0	100	100	0	0	50	50
Avg	100	100	100	100	16.7	16.7	16.7	16.7	100	100	22.2	0	22.2	0	100	100	16.7	5.6	16.7	5.6

In the second case, we consider some level of flexibility in which we have a production recourse for the end item. As in the first case, we will produce 200 units for the component, but how much to produce for the second item is a recourse decision and will be defined based on the observed demand. The flexibility with respect to the production quantity of the end item results in flexibility in the allocation of the 200 units of the component, to satisfy the external demand of the component or to produce end item. This flexible allocation will define what portion of the produced component should be used for its own external demand and how much should be used for the end item production. In the second case, the production quantity for the end item is determined so that it satisfies as much as possible the external demand for this end item, while avoiding any inventory for the end item. In Table 1, the allocation of the 200 available units of the component to the component and to the end product are given. These decisions are now different in each scenario because the production decision for the end item is now a recourse decision. We observe here that for scenario 3, the flexibility in the production quantity for the end product now allows a flexible allocation of the 200 components: 50 to satisfy the external demand for the component and 150 to be allocated to the end product. The result is that demand for both the end product and the component is exactly satisfied without creating any backlog or inventory. In this case, the recourse decisions taken lead to an average of 22.2 units backlog and the same amount of inventory for the component, while the average backlog and inventory for the end item is equal to zero.

Case 3 is similar to case 2, but with slightly different recourse decisions taken, resulting in an average inventory and backlog level of 16.7 units for the component and 5.6 units for the end item. This result dominates the result of the first case. As we can see, in general the flexibility can reduce the average inventory and backlog in the system, but we may have several solutions to use this flexibility, which can be defined optimally based on the structure and different costs in the system. This small illustration makes clear that the flexibility with respect to the production quantity of the end product results in a flexible decision on the allocation of the fixed production quantity of the component.

2. Literature review

We organized the literature review into two sections. The first part discusses the lot sizing problem with service level constraints, and the second one discusses the multi-level lot sizing problem.

2.1. Stochastic lot sizing problem with service level constraints

The stochastic lot sizing problem with service level constraints has been studied extensively. Several service level measures have been proposed which can be classified as event-oriented service levels, quantity-oriented service levels, and time and quantity-oriented service levels [22]. The α service level is an event-oriented service level which imposes a limit on the probability of stock out. The β service level or the fill rate is the proportion of the demand directly filled from stock and it is calculated based on the expected backorders to the expected demand. This service level is a quantity oriented service level. The γ service level limits the proportion of expected backlog to expected demand. The δ service level is based on the proportion of total expected backlog to the maximum expected backlog. Both γ and δ service levels are time and quantity oriented service levels.

Many papers studied the lot sizing problem with service level constraints using different strategies and different types of service levels [13, 26, 28, 29, 30, 34]. These service levels are commonly considered in systems with one item, or multiple end items, but they are not considered in multi-level systems where we have BOM structures. The service level which we consider in this research is closely related to the δ service level proposed by Helber et al. [13].

Table 2
Multi-level lot sizing research

Authors	Year	Capacity	Backlogging	Independent demand at the component level	Stochasticity	Service level
Templemeier and Derstrof [27]	1996	+	-	-	-	-
Hung and Chien [14]	2000	+	+	+	-	-
Stadtler [24]	2003	+	-	-	-	-
Sahling et al. [21]	2009	+	-	-	-	-
Akartunalı and Miller [4]	2009	+	+	-	-	-
Almeder [5]	2010	+	-	-	-	-
Wu et al. [37]	2011	+	+	-	-	-
Seeanner et al. [20]	2013	+	-	-	-	-
Xiao et al. [38]	2014	-	-	-	-	-
Toledo et al. [32]	2015	+	+	-	-	-
You et al. [39]	2019	-	-	-	-	-
Thevenin et al. [31]	2020	-	-	-	+	-
Quezada et al. [19]	2020	-	-	+	+	-
Gruson et al. [12]	2021	-	-	-	+	-
Our work		+	+	+	+	+

Here, instead of limiting average backlog we limit the maximum proportion, taken over all demand realisations, of total backlog, to the total possible backlog over the whole planning horizon, for each product with external demand. This service level which we denote by δ' is more strict compared to the standard δ service level. While the δ service level is defined based on the averages over all scenarios, δ' is imposed for each of the scenarios separately. This per-scenario service level is also adopted in other similar problems. For example, Alvarez et al. [7] investigate the inventory routing problem with stochastic supply and demand with a service level in which they limit the proportion of total lost sale to the total demand for each scenario.

2.2. Multi-level lot sizing problem

Several formulations have been considered for the multi-level lot sizing problem. Some formulations use the concept of echelon stock (Afentakis et al. [2, 3], Pochet and Wolsey [18] and Akartunali and Miller [4]). This reformulation allows the use of strong cuts based on the single-level lot sizing problem, which results in better bounds. Wu et al. [37] investigated the capacitated multi-level lot sizing problem with backlogging, and proposed different mathematical models. These models are the common multi-level lot sizing model, the model based on echelon variables, the model based on the facility location formulation, and the model based on the shortest path formulation. As mentioned before, it is not possible to allow backlog decisions for the dependent demand in the BOM structure, and these backlogs are defined only for the independent demands in the system. Hung and Chien [14] proposed a mathematical model for a multi-class multi-level capacitated lot sizing problem, where two classes of orders, i.e., confirmed and predicted, are considered. In this model, there are three different constraints for the inventory balance constraints; one for end products, one for components with external demand, and one for components without external demand. In addition to these constraints, there is also a set of constraints to ensure that there is no backlog for the dependent demand of the components. To solve this problem, the authors used simulated annealing, tabu search, and genetic algorithm heuristics. Table 2 shows some of the related works on multi-level lot sizing problem, and their similarities and differences with our work.

3. Mathematical formulation

In this section, we propose the mathematical models for the deterministic and stochastic multi-level lot sizing problems with a service level constraint. As mentioned, in this model, there is a possibility of having external (independent) demand for the components as well.

3.1. Deterministic model

The deterministic model is an extension of the model proposed by Hung and Chien [14] in which, for each type of product in the system, there is a different set of inventory balance constraints. Table 3 provides the list of sets,

Table 3
Notation for the multi-level deterministic lot sizing problem

Sets	Definition
$\mathcal{K}$	Set of products, indexed by $1, ..., K$
S_k	Set of immediate successors of product k
$\mathcal{T}$	Set of planning periods, indexed by $1, ..., T$
EI	Set of end items
CW	Set of components without external demand
CE	Set of components with external demand
$\mathcal{MC}$	Set of machines
$\mathcal{K}_m$	Set of products that are produced on machine m
Parameters	**Definition**
cap_{mt}	Production capacity for machine m in period t
d_{kt}	External demand of item k in period t
bc_{kt}	Backlog cost for product k in period t
hc_{kt}	Inventory holding cost for product k in period t
$\bar{M}_{kt}, M'_{kt}$	A sufficiently large number
pt_{kt}	Unit production time for product k in period t
r_{ki}	Number of units of item k required to produce one unit of the immediate successor item i
oc_{mt}	Overtime cost for machine m in period t
sc_{kt}	Setup cost for product k in period t
st_{kt}	Setup time for product k in period t
δ, δ'	Target service level in deterministic and stochastic models, respectively
Decision variables	**Definition**
y_{kt}	Binary variable which is equal to 1 if there is a setup for product k in period t, 0 otherwise
x_{kt}	Amount of production for product k in period t
o_{mt}	Overtime for machine m in period t
B_{kt}	Amount of backlog for product k at the end of period t
I_{kt}	Amount of physical inventory for product k at the end of period t

parameters and decision variables in the model. In this model, the structure of the BOM is considered by the successors of each item. We have multiple machines and the capacity is also defined for each machine separately. Furthermore, we consider the possibility of overtime.

$$\text{Min} \sum_{t\in\mathcal{T}} \sum_{k\in\mathcal{K}} (sc_{kt}y_{kt} + hc_{kt}I_{kt}) + \sum_{t\in\mathcal{T}} \sum_{m\in\mathcal{MC}} oc_{mt}o_{mt} + \sum_{k\in\mathcal{K}} bc_{kT}B_{kT} \tag{1a}$$

Subject to:

$$I_{k,t-1} + x_{kt} + B_{kt} = I_{kt} + B_{k,t-1} + d_{kt} \qquad \forall t \in \mathcal{T}, \forall k \in EI \tag{1b}$$

$$I_{k,t-1} + x_{kt} = I_{kt} + \sum_{i\in S_k} r_{ki}x_{it} \qquad \forall t \in \mathcal{T}, \forall k \in CW \tag{1c}$$

$$I_{k,t-1} + x_{kt} + B_{kt} = I_{kt} + B_{k,t-1} + d_{kt} + \sum_{i\in S_k} r_{ki}x_{it} \qquad \forall t \in \mathcal{T}, \forall k \in CE \tag{1d}$$

$$B_{kt} - B_{k,t-1} \leq d_{kt} \qquad \forall t \in \mathcal{T}, \forall k \in CE \tag{1e}$$

$$x_{kt} \leq \bar{M}_{kt}y_{kt} \qquad \forall t \in \mathcal{T}, \forall k \in \mathcal{K} \tag{1f}$$

$$\sum_{k\in\mathcal{K}_m} (st_{kt}y_{kt} + pt_{kt}x_{kt}) \leq cap_{mt} + o_{mt} \qquad \forall m \in \mathcal{MC}, \forall t \in \mathcal{T} \tag{1g}$$

$$\frac{\sum_{t\in T} B_{kt}}{\sum_{t\in T}(T - t + 1)d_{kt}} \leq 1 - \delta \qquad \forall k \in EI \cup CE \tag{1h}$$

$$y_{kt} \in \{0, 1\} \qquad \forall t \in \mathcal{T}, \forall k \in \mathcal{K} \tag{1i}$$

$$B \in \mathbb{R}_+^{KT}, I \in \mathbb{R}_+^{KT}, x \in \mathbb{R}_+^{KT}, o \in \mathbb{R}_+^{MCT} \tag{1j}$$

The objective function (1a) minimizes the setup cost, holding cost, overtime cost and the cost of unsatisfied demand at the end of the planning period. Constraints (1b-1d) are the inventory balance constraints. Constraints (1b) are the inventory balance constraints for each of the end items in each planning period, and constraints (1c) and (1d) are for the components without and with external demand, respectively. Constraints (1e) ensure that the amount of backorder in each planning period cannot be more than the period external demand [14]. These constraints are only imposed for the components with external demand and they ensure that any backlog is only directly related to the independent demand of the component and there is no backlog for dependent demands. Constraints (6b) are the production setup constraints. $\bar{M}_{kt}$ which is the maximum possible production is calculated using equations (2) and (3) in a similar fashion as in Toledo et al. [32]. These calculations need to be done recursively, starting from the end items. Constraints (1h) are the δ service level constraints in the deterministic setting [11].

$$D_{k(t..T)} = \sum_{u=t}^{T} d_{ku} + \sum_{i \in S_k} r_{ki} D_{i(t..T)} \qquad \forall t \in \mathcal{T}, \forall k \in \mathcal{K} \tag{2}$$

$$\bar{M}_{kt} = D_{k(1..T)} \qquad \forall t \in \mathcal{T}, \forall k \in \mathcal{K} \tag{3}$$

3.2. Static stochastic model with service level

In this section, we present the model for the stochastic capacitated multi-level lot sizing problem. In this first variant, as in [8] we assume that the strategy is static which implies that the setup and production quantity decisions are determined at the beginning of the planning horizon and cannot be changed. Figure 1 illustrates the sequence of decisions in this problem. In the same spirit as [13], we consider the case where the demand realizations for a given number of subsequent periods are revealed (confirmed) all at once. This is in line with the sales and operations planning (SOP) practice in which the data are reviewed and established for a fixed planning period [6]. In addition, for the case where demands can be sequentially revealed in a multi-stage fashion, the framework presented in this work can be applied in a rolling horizon framework as in [22].

The setup and production, and overtime decisions are the first stage variables which are defined before the demand realization. It is worthwhile to mention that in general the overtime is a part of the recourse decisions in our problem definition. However, in the static case, there is no benefit in defining overtime as a second stage variable since all the production amounts are determined in the first stage and hence the overtime does not depend on the demand realization. Thus, for the static case, we include this variable in the first-stage model for simplicity. After the demand realization (for the entire planning horizon), the resulting inventory and backlog levels are determined for each scenario in the second stage [13]. In this problem, the model guarantees that for each product with external demand, the proportion of backlog divided by the maximum possible backlog considering any realization is less than $(1 - \delta')$. As mentioned earlier, this service level is more strict than the standard δ service level which is defined based on the expected value of the backlog [13]. Without loss of generality, a backlog unit cost, if available, can also be directly incorporated in the objective function. Nevertheless, such a consideration is often omitted in the literature [13, 22] when service level constraints are imposed. In addition, since our framework deals with stochastic demand, we also account for the end-of-horizon inventory and backlog costs as in [1]. We model this problem as a two-stage stochastic mixed integer program. To account for the stochasticity, a random vector $\boldsymbol{d} = (\tilde{d}_{11}, ..., \tilde{d}_{KT})$ is considered, where $\tilde{d}_{kt}$ represents the random demand for product k, in period t. This model is represented in (4a)-(4e).

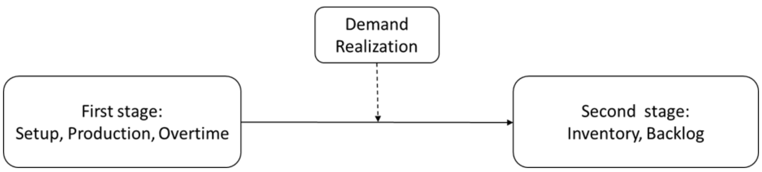


Figure 1: Sequence of events for the case with no flexibility

$$v^* := \text{Min } F(y, o) + \mathbb{E}_d[Q(x, \boldsymbol{d})] \tag{4a}$$

Subject to:

$$x_{kt} \leq M'_{kt} y_{kt} \qquad \forall t \in \mathcal{T}, \forall k \in \mathcal{K} \tag{4b}$$

$$\sum_{k \in \mathcal{K}_m} (st_{kt} y_{kt} + pt_{kt} x_{kt}) \leq cap_{mt} + o_{mt} \qquad \forall m \in \mathcal{MC}, \forall t \in \mathcal{T} \tag{4c}$$

$$y \in \{0,1\}^{KT} \tag{4d}$$

$$x \in \mathbb{R}_+^{KT}, o \in \mathbb{R}_+^{MC,T} \tag{4e}$$

In this model, the first-stage cost function is defined as (5a) which minimizes the setup and overtime costs, and the second stage cost function is represented in (5b-5h).

$$F(y,o) = \sum_{t \in \mathcal{T}} \left(\sum_{k \in \mathcal{K}} sc_{kt} y_{kt} + \sum_{m \in \mathcal{MC}} oc_{mt} o_{mt} \right) \tag{5a}$$

$$Q(x, \boldsymbol{d}) = min \sum_{t \in \mathcal{T}} \sum_{k \in \mathcal{K}} hc_{kt} I_{kt} + \sum_{k \in \mathcal{K}} bc_{kT} B_{kT} \tag{5b}$$

Subject to:

$$I_{k,t-1} + x_{kt} + B_{kt} = I_{kt} + B_{k,t-1} + \tilde{d}_{kt} \qquad \forall t \in \mathcal{T}, \forall k \in EI \tag{5c}$$

$$I_{k,t-1} + x_{kt} = I_{kt} + \sum_{i \in S_k} r_{ki} x_{it} \qquad \forall t \in \mathcal{T}, \forall k \in CW \tag{5d}$$

$$I_{k,t-1} + x_{kt} + B_{kt} = I_{kt} + B_{k,t-1} + \tilde{d}_{kt} + \sum_{i \in S_k} r_{ki} x_{it} \qquad \forall t \in \mathcal{T}, \forall k \in CE \tag{5e}$$

$$B_{kt} - B_{k,t-1} \leq \tilde{d}_{kt} \qquad \forall t \in \mathcal{T}, \forall k \in CE \tag{5f}$$

$$\frac{\sum_{t \in T} B_{kt}}{\sum_{t \in T} (T - t + 1) \tilde{d}_{kt}} \leq 1 - \delta' \qquad \forall k \in EI \cup CE \tag{5g}$$

$$I \in \mathbb{R}_+^{KT}, B \in \mathbb{R}_+^{KT} \tag{5h}$$

The objective function (5b) minimizes the holding cost, and the cost of unsatisfied demand at the end of the planning period. Constraints (5c-5e) are the inventory balance constraints. Constraints (5f) limit the maximum amount of backorder in each planning period. Constraints (5g) are the service level constraints.

3.3. Stochastic model with service level and production recourse

In the previous section, we presented the model based on the static strategy which does not allow any flexibility in the production decisions. In this section, we consider the case where we have production recourse and flexibility can be allowed for some products by assuming that they can follow a static-dynamic strategy, while other products keep following the static strategy.

In this model, the production amounts of the products with no flexibility are part of the first-stage decisions, and the production for the rest of them are part of the second stage decisions which are determined after the demand realization (Figure 2). Table (4) presents the additional set and variables required for this model.

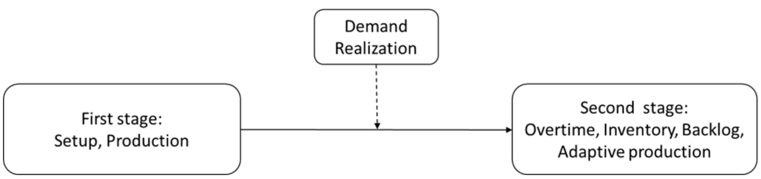


Figure 2: Sequence of events for the case with flexibility

Table 4
Additional notation for the stochastic model with production flexibility

Set	Definition
$Flex$	Set of products with flexible production
Random variables	**Definition**
X_{kt}	Amount of production for product $k \in Flex$ in period t
Decision variables	**Definition**
x_{kt}	Amount of production for product $k \in \mathcal{K} \backslash Flex$ in period t

The stochastic multi-level lot sizing problem with flexibility is presented in (6a)-(6c).

$$v^* := \text{Min } F'(y) + \mathbb{E}_d[Q'(y, x, \boldsymbol{d})] \tag{6a}$$

Subject to:

$$x_{kt} \leq M'_{kt} y_{kt} \qquad \forall t \in \mathcal{T}, \forall k \in \mathcal{K}, k \setminus Flex \tag{6b}$$
$$y \in \{0,1\}^{KT}, x \in \mathbb{R}_+^{KT} \tag{6c}$$

In model 6, the first-stage cost function is defined as (7a) and the second stage recourse model is represented in (7b-7m).

$$F'(y) = \sum_{t \in \mathcal{T}} \sum_{k \in \mathcal{K}} sc_{kt} y_{kt} \tag{7a}$$
$$Q'(y, x, \boldsymbol{d}) = min \sum_{t \in \mathcal{T}} (\sum_{k \in \mathcal{K}} hc_{kt} I_{kt} + \sum_{m \in \mathcal{MC}} oc_{mt} O_{mt}) + \sum_{k \in \mathcal{K}} bc_{kT} B_{kT} \tag{7b}$$

Subject to:

$$X_{kt} \leq M_{kt} y_{kt} \qquad \forall t \in \mathcal{T}, \forall k \in \mathcal{K}, k \in Flex \tag{7c}$$
$$\sum_{k \in \mathcal{K}_m} st_{kt} y_{kt} + \sum_{k \in \mathcal{K}_m \backslash Flex} pt_{kt} x_{kt} + \sum_{k \in \mathcal{K}_m \cap Flex} pt_{kt} X_{kt} \leq cap_{mt} + O_{mt} \qquad \forall m \in \mathcal{MC}, \forall t \in \mathcal{T} \tag{7d}$$
$$I_{k,t-1} + x_{kt} + B_{kt} = I_{kt} + B_{k,t-1} + \tilde{d}_{kt} \qquad \forall t \in \mathcal{T}, \forall k \in EI, k \setminus Flex \tag{7e}$$
$$I_{k,t-1} + X_{kt} + B_{kt} = I_{kt} + B_{k,t-1} + \tilde{d}_{kt} \qquad \forall t \in \mathcal{T}, \forall k \in EI, k \in Flex \tag{7f}$$
$$I_{k,t-1} + x_{kt} = I_{kt} + \sum_{i \in S_k \backslash Flex} r_{ki} x_{it} + \sum_{i \in S_k \cap Flex} r_{ki} X_{it} \qquad \forall t \in \mathcal{T}, \forall k \in CW, k \setminus Flex \tag{7g}$$
$$I_{k,t-1} + X_{kt} = I_{kt} + \sum_{i \in S_k \backslash Flex} r_{ki} x_{it} + \sum_{i \in S_k \cap Flex} r_{ki} X_{it} \qquad \forall t \in \mathcal{T}, \forall k \in CW, k \in Flex \tag{7h}$$
$$I_{k,t-1} + x_{kt} + B_{kt} = I_{kt} + B_{k,t-1} + \tilde{d}_{kt} + \sum_{i \in S_k \backslash Flex} r_{ki} x_{it} + \sum_{i \in S_k \cap Flex} r_{ki} X_{it} \qquad \forall t \in \mathcal{T}, \forall k \in CE, k \setminus Flex \tag{7i}$$
$$I_{k,t-1} + X_{kt} + B_{kt} = I_{kt} + B_{k,t-1} + \tilde{d}_{kt} + \sum_{i \in S_k \backslash Flex} r_{ki} x_{it} + \sum_{i \in S_k \cap Flex} r_{ki} X_{it} \qquad \forall t \in \mathcal{T}, \forall k \in CE, k \in Flex \tag{7j}$$
$$B_{kt} - B_{k,t-1} \leq \tilde{d}_{kt} \qquad \forall t \in \mathcal{T}, \forall k \in CE \tag{7k}$$
$$\frac{\sum_{t \in T} B_{kt}}{\sum_{t \in T} (T - t + 1) d_{kt}} \leq 1 - \delta' \qquad \forall k \in EI \cup CE \tag{7l}$$
$$I \in \mathbb{R}_+^{KT}, B \in \mathbb{R}_+^{KT}, X \in \mathbb{R}_+^{KT}, O \in \mathbb{R}_+^{MCT} \tag{7m}$$

Notable differences between model 5 and model 7 are in the presence of the recourse variable X_{it} which shows the adaptive production, and of the recourse variable O_{it} which shows the overtime and cannot be moved to the first stage decisions anymore. Each of the inventory balance constraints in model 5 should be separately considered for the items with and without adaptive production in model 7.

4. Sample average approximation

We cannot directly solve the two-stage programming models (4) and (6) due to the presence of the random demand vector $\boldsymbol{d}$ and the expectation terms in their objective functions (4a) and (6a). To this end, we apply the sample average approximation (SAA) method to tackle this problem. SAA is a Monte Carlo simulation-based method to solve the stochastic optimization problems in which the random distribution is replaced by a finite number of scenarios and the true expected value of the objective function is approximated by the average cost over the scenarios [23]. To generate a scenario sample for the random vector $\boldsymbol{d} = \{d_s\}_{s \in \mathcal{S}}$, Monte Carlo sampling is used and an equal probability is assigned to each scenario.

The quality of the approximation in the stochastic approach mainly depends on the number of scenarios used [16]. The challenge here is to define a proper number of scenarios which can provide a near-optimal approximation of the original problem. This number is defined based on the statistical lower bound and upper bound of the optimal solution. The SAA procedure, the definition of these bounds, and the gap between them are explained next.

In the SAA procedure, we have a set of scenario $\mathcal{S} = \{1, 2, ..., S\}$. First we choose the initial sample sizes S, and the number of SAA replications M. Then, for each replication $m = 1$ to M, an instance with S scenarios is generated and the corresponding SAA model is solved based on the chosen set of scenarios. Let the $\hat{v}_m^S$ and $\hat{\sigma}_m^S$, be the optimal objective function value and the solution for replication m, respectively. Equation (8) defines the expected value of the lower bound of v^*, the optimal objective value for the original two-stage stochastic problem, based on M replications of size S, denoted by $L_{M,S}^{mean}$.

$$L_{M,S}^{mean} = E(\hat{v}_M) = \frac{1}{M} \sum_{m=1}^{M} \hat{v}_m^S \tag{8}$$

To estimate the upper bound, we generate a large enough sample set $\mathcal{S}^{eval} = \{1, 2, ..., S^{eval}\}$, where $S^{eval} \gg S$. Having a solution $\hat{\sigma}$ as the first stage decision, $U_{M,S}$ (9) is an estimation of the upper bound, in which $\hat{g}(\mathcal{S}^{eval}, \hat{\sigma})$ is the objective function of the SAA formulation when the first stage solution, $\hat{\sigma}$, is fixed and when scenario set $\mathcal{S}^{eval}$ is used.

$$U_{M,S} = \hat{g}(\mathcal{S}^{eval}, \hat{\sigma}) \tag{9}$$

It should be noted that in this procedure, we have M different feasible solutions to calculate the upper bound. Among those we will choose the one with the smallest estimated objective value which is calculated based on a scenario set $\mathcal{S}^{eval}$, denoted by σ^* [35] (see eq. 10).

$$\sigma^* = arg\ min\{\hat{g}(\mathcal{S}^{eval}, \hat{\sigma}) : \hat{\sigma} \in \{\hat{\sigma}_1^S, ..., \hat{\sigma}_M^S\}\} \tag{10}$$

The empirical gap between the $U_{M,S}$ and $L_{M,S}^{mean}$ is used to define the proper number of scenarios to run the experiments to investigate the flexibility. In the following section we will present the SAA formulation of the two-stage models (4) and (6), as models (11) and (14), respectively. Applications of this method can also be found in Contreras et al. [10], Taş et al. [25], and Mousavi et al. [16].

4.1. Reformulating the SAA problem

In this section, we present the extensive forms of SAA formulation [16], for both the static and adaptive models. Additional parameters and decision variables are presented in Table 5 and the stochastic models are presented afterward.

4.1.1. The static SAA model

Following the static strategy, the setup and production, and overtime variables are the first stage variables and they do not have any index for the different scenarios in this model. The resulting inventory and backlog for each scenario are determined in the second stage.

Table 5
Additional parameters and variables used in the SAA formulations

Parameters	Definition
d_{kts}	Demand for product k in period t in scenario s
Decision variables	**Definition**
X_{kts}	The production amount for product $k, k \in Flex$, in period t for scenario s
O_{mts}	The backorder for machine m in period t for scenario s
B_{kts}	Backlog for product k in period t for scenario s
I_{kts}	Amount of inventory for product k at the end of period t in scenario s

$$\text{Min} \sum_{t\in\mathcal{T}}\sum_{k\in\mathcal{K}} sc_{kt}y_{kt} + \sum_{t\in\mathcal{T}}\sum_{m\in\mathcal{MC}} oc_{lt}o_{mt} + \sum_{t\in\mathcal{T}}\sum_{k\in\mathcal{K}} hc_{kt}\frac{\sum_{s\in\mathcal{S}} I_{kts}}{S} + \sum_{k\in\mathcal{K}} bc_{kT}\frac{\sum_{s\in\mathcal{S}} B_{kTs}}{S} \tag{11a}$$

Subject to constraints (4b) - (4d), and:

$$I_{k,t-1,s} + x_{kt} + B_{kts} = I_{kts} + B_{k,t-1,s} + d_{kts} \qquad \forall t\in\mathcal{T}, \forall k\in EI, \forall s\in\mathcal{S} \tag{11b}$$

$$I_{k,t-1,s} + x_{kt} = I_{kts} + \sum_{i\in S_k} r_{ki}x_{it} \qquad \forall t\in\mathcal{T}, \forall k\in CW, \forall s\in\mathcal{S} \tag{11c}$$

$$I_{k,t-1,s} + x_{kt} + B_{kts} = I_{kts} + B_{k,t-1,s} + d_{kts} + \sum_{i\in S_k} r_{ki}x_{it} \qquad \forall t\in\mathcal{T}, \forall k\in CE, \forall s\in\mathcal{S} \tag{11d}$$

$$B_{kts} - B_{k,t-1,s} \le d_{kts} \qquad \forall t\in\mathcal{T}, \forall k\in CE, \forall s\in\mathcal{S} \tag{11e}$$

$$\frac{\sum_{t\in T} B_{kts}}{\sum_{t\in T}(T-t+1)d_{kts}} \le 1-\delta' \qquad \forall k\in EI, CE, \forall s\in\mathcal{S} \tag{11f}$$

$$B\in\mathbb{R}_+^{KTS}, I\in\mathbb{R}_+^{KTS} \tag{11g}$$

The objective function (11a) is to minimize the setup cost, overtime cost, and the expected value of the inventory holding costs and the backlog in the last period. Constraints (11b-11d) are the inventory balance constraints which are defined for all products in all periods, and for all the scenarios. Constraints (11e) define the limit for the maximum amount of backorder for each component with external demand in each period for each scenario. Constraints (11f) show the service level for each product and each scenario, in which the proportion of total backlog to the maximum possible backlog for each scenario should not exceed the threshold percentage set by service level δ'. The parameter M'_{kt} in the production setup constraint is calculated recursively using equations (12) and (13).

$$D_{k(t..T)s} = \sum_{u=t}^{T} d_{kus} + \sum_{i\in S_k} r_{ki}D_{i(t..T)s} \qquad \forall t\in\mathcal{T}, \forall k\in\mathcal{K}, \forall s\in\mathcal{S} \tag{12}$$

$$M'_{kt} = \max_{s\in\mathcal{S}} D_{k(1..T)s} \qquad \forall t\in\mathcal{T}, \forall k\in\mathcal{K} \tag{13}$$

4.1.2. The SAA model with flexibility

Similar to the model without flexibility, in the SAA formulation for case with flexibility, the model (6) is reformulated as model (14).

$$\text{Min} \sum_{t\in\mathcal{T}}\sum_{k\in\mathcal{K}} (sc_{kt}y_{kt} + h_{kt}\frac{\sum_{s\in\mathcal{S}} I_{kts}}{S}) + \sum_{t\in\mathcal{T}}\sum_{m\in\mathcal{MC}} oc_{mt}\frac{\sum_{s\in\mathcal{S}} O_{mts}}{S} + \sum_{k\in\mathcal{K}} bc_k\frac{\sum_{s\in\mathcal{S}} B_{kTs}}{S} \tag{14a}$$

Subject to constraints (6b), (6c), (11e), (11f), and:

$$\sum_{k\in\mathcal{K}_m} st_{kt}y_{kt} + \sum_{k\in\mathcal{K}_m\setminus Flex} pt_{kt}x_{kt} + \sum_{k\in\mathcal{K}_m\cap Flex} pt_{kt}X_{kts} \le cap_{mt} + O_{mts} \qquad \forall s\in S, \forall m\in\mathcal{MC}, \forall t\in\mathcal{T} \tag{14b}$$

$$I_{k,t-1,s} + x_{kt} + B_{kts} = I_{kts} + B_{k,t-1,s} + d_{kts} \qquad \forall s\in S, \forall t\in\mathcal{T}, \forall k\in EI, k\setminus Flex \tag{14c}$$

$$I_{k,t-1,s} + X_{kts} + B_{kts} = I_{kts} + B_{k,t-1,s} + d_{kts} \quad \forall s \in S, \forall t \in \mathcal{T}, \forall k \in EI, k \in Flex \quad (14d)$$

$$I_{k,t-1,s} + x_{kt} = I_{kts} + \sum_{i \in S_k \setminus Flex} r_{ki} x_{it} + \sum_{i \in S_k \cap Flex} r_{ki} X_{its} \quad \forall s \in S, \forall t \in \mathcal{T}, \forall k \in CW, k \setminus Flex \quad (14e)$$

$$I_{k,t-1,s} + X_{kts} = I_{kts} + \sum_{i \in S_k \setminus Flex} r_{ki} x_{it} + \sum_{i \in S_k \cap Flex} r_{ki} X_{its} \quad \forall s \in S, \forall t \in \mathcal{T}, \forall k \in CW, k \in Flex \quad (14f)$$

$$I_{k,t-1,s} + x_{kt} + B_{kts} = I_{kts} + B_{k,t-1,s} + d_{kts} + \sum_{i \in S_k \setminus Flex} r_{ki} x_{it} + \sum_{i \in S_k \cap Flex} r_{ki} X_{its} \quad \forall s \in S, \forall t \in \mathcal{T}, \forall k \in CE, k \setminus Flex \quad (14g)$$

$$I_{k,t-1,s} + X_{kts} + B_{kts} = I_{kts} + B_{k,t-1,s} + d_{kts} + \sum_{i \in S_k \setminus Flex} r_{ki} x_{it} + \sum_{i \in S_k \cap Flex} r_{ki} X_{its} \quad \forall s \in S, \forall t \in \mathcal{T}, \forall k \in CE, k \in Flex \quad (14h)$$

$$B \in \mathbb{R}_+^{KTS}, I \in \mathbb{R}_+^{KTS}, x \in \mathbb{R}_+^{KT}, X \in \mathbb{R}_+^{KTS}, O \in \mathbb{R}_+^{MCT} \quad (14i)$$

The objective function (14a) minimizes the setup cost, plus the expected value of inventory holding cost, overtime cost, and unsatisfied demand at the end of planning period. Constraints (14b) are the capacity constraints for each production level, each scenario in each period. In this constraint the overtime is different for each demand scenario. Constraints (14c - 14h) are the inventory balance constraints. The difference between these constraints and inventory balance constraints in the model without any flexibility is that we have separate constraints for the items that have flexibility and the items that do not have it.

4.1.3. SAA implementation

In the SAA procedure, we first solve the SAA model to define the first stage decisions which we refer as the planning phase. In this phase, we solve the models using a set of sampled demand scenarios. In the second phase or evaluation phase, after fixing the first stage solution, we solve the model using a new and larger set of scenarios. Since each of the evaluation scenarios differ from the scenarios considered initially in the planning phase, this may cause infeasibility due to the service level constraint. Note that the overtime decision, even in the case of the static-dynamic strategy, does not rule out this infeasibility if the flexibility level is insufficient. More specifically, we may have some cases where we are not allowed to produce more of an item since the item's component production is fixed in the planning phase or the first stage and it is not possible to increase it. In case of full flexibility, i.e., all the production quantities are determined after the demand realization, the model will always be feasible with the larger set of scenarios.

The SAA method requires that the model must be feasible in order to calculate the SAA bounds. Thus, to alleviate this issue, we make use of a set of auxiliary variables which are associated with a penalty cost. To this end, we need to change the service level constraint from a hard one to a soft constraint in the the evaluation model. There is an additional penalty variable, ϵ_{ks}, for the violation in the service level constraint and a penalty cost, P, for this violation in the objective function. To have a consistent model in both the planning and the evaluation, we use constraints (16) instead of the service level constraints (11f) and (15) instead of the objective function (14a) in the extensive form model (14) which will be used both in planning and evaluation.

$$Min \sum_{t \in \mathcal{T}} \sum_{k \in \mathcal{K}} (sc_{kt} y_{kt} + h_{kt} \frac{\sum_{s \in S} I_{kts}}{S}) + \sum_{t \in \mathcal{T}} \sum_{m \in \mathcal{MC}} oc_{mt} \frac{\sum_{s \in S} O_{mts}}{S} + \sum_{k \in \mathcal{K}} (bc_k \frac{\sum_{s \in S} B_{kTs}}{S} + \frac{\sum_{s \in S} P\epsilon_{ks}}{S}) \quad (15)$$

$$\frac{\sum_{t \in T} B_{kts}}{\sum_{t \in T} (T - t + 1) d_{kts}} \leq 1 - \delta' + \epsilon_{ks} \quad \forall k \in \mathcal{K}, \forall s \in \mathcal{S} \quad (16)$$

In our experiments, the value of P is set high enough so that the value of ϵ in the planning phase becomes zero, which guarantees that the plan is feasible with respect to all the demand scenarios used in the planning phase, and the service level constraint (11f) is not violated in this phase.

For the evaluation of the decisions, we generate a large number of scenarios, and we solve the extensive form of the problem.

The average of the objective functions over all of these scenarios will be the result of the evaluation which is also the the upper bound of the optimal value which was explained in section 5 in the SAA procedure. In the numerical experiments, the percentage of violated scenarios and how much the service level constraint is violated are also calculated in the evaluation phase.

5. Numerical experiments

5.1. Instance generation

This research combines a multi-level lot sizing problem with external demand for the components, and the stochastic lot sizing with service level constraints. To the best of our knowledge, instances for this problem are not available in the literature. To this end, we make use of two data sets in the literature and adapt them to the problem considered in this paper. More specifically, we modified the method used by Helber et al. [13] for the stochastic lot sizing problem with service level, and adapted the instance generation method presented by Tempelmeier et al. [27] for the multi-level lot sizing and different BOM structures, as follows:

- We consider three different structures, serial, assembly and general (Figure 3). For the last two structures, we follow the Tempelmeier and Derstroff data [27].

- The holding cost of an item is equal to the sum of the holding cost of its components, multiplied by $(1+HValue)$. A High $HValue$ means higher value adding operations at each level. We will also consider a case in which all the values for $HValue$ are equal to 0 which means that the holding cost of an item is equal to the sum of the holding costs of its components, without any added value [27].

- Following Tempelmeier and Derstroff [27], five different Time Between Orders (TBO) profiles are considered (Table 6). For example, for the serial BOM, in the first profile, the TBO for all the items is equal to 1, and in the fourth profile, the TBO for the first 2 items is equal to 1, for the second item it is equal to 2 and for the last two items it is equal to 4.

- The capacity is defined for each machine. In our instances, exactly one machine is assigned to each level of BOM, and all the unit processing times are equal to 1. The capacity of each machine is equal to the sum of the average demand of the items assigned to a machine, divided by a parameter $Util$. The average demand of an item is equal to the sum of dependent and independent demand. Note that $Util$ is a parameter for data generation and it does not show the actual utilized capacity. We also considered the case without any capacity.

- The setup cost is determined based on the TBO, average demand and the holding cost, using equation (17) [13].

$$sc_{kt} = \frac{E[\overline{D}_{kt}] \times TBO^2 \times hc_{kt}}{2} \tag{17}$$

- We will have different levels of flexibility, which is defined based on the items for which production quantities are not fixed at the beginning of the planning horizon and they may be modified when the demand is observed. 0 means no flexibility, and i means flexibility for all the items until the $(i+1)^{th}$ level of BOM structure.

- Average demand profile for different structures. There are three different patterns for the external demand average. The first one is constant average independent demand for all the items at different levels. The second one is in increasing order and the third one is in decreasing orders from the end item level to the following levels (lowest to the highest level). For a given item, the average demand remains the sames over the horizon. The patterns and their detailed demand generations are summarized as follows:

    1. Constant external demand in which the average external demand ($dl = 100$) is multiplied by 1 at subsequent levels. For example if there is an external demand for any of the items at different levels it is equal to 100.
    2. The increasing order of demand in which the average external demand ($dl = 100$) is multiplied by the (level of the item +1). For example the external demand for the items at level 0 is equal to 100, and for the items at level 1 is equal to 200, is there is any.
    3. The decreasing order of demand in which the average external demand ($dl = 100$) is multiplied by (Max level - level of the item +1). For example, the external demand for the item at level 0 of the serial structure is equal to 500, and for the assembly and general structures it is equal to 300.

- The various combinations of average demand profiles and external demand profiles are provided in Table 7. In the assembly and general structure, the external demand is added level by level to some of the components. This is in line with reality in which some of the components may have external demand, and some may not. Based on

Table 6
Different TBO profiles

	Serial product structure			Assembly product structure			General product structure		
TBO profile	TBO = 1	TBO = 2	TBO = 4	TBO = 1	TBO = 2	TBO = 4	TBO = 1	TBO = 2	TBO = 4
1	1 ... 5	-	-	1 ... 10	-	-	1 ... 10	-	-
2	-	1 ... 5	-	-	1 ... 10	-	-	1 ... 10	-
3	-	-	1 ... 5	-	-	1 ... 10	-	-	1 ... 10
4	1,2	3	4,5	1	2 ... 4	5 ... 10	1 ... 4	5 ... 7	8 ... 10
5	4,5	3	1,2	5 ... 10	2 ... 4	1	8 ... 10	5 ... 7	1 ... 4

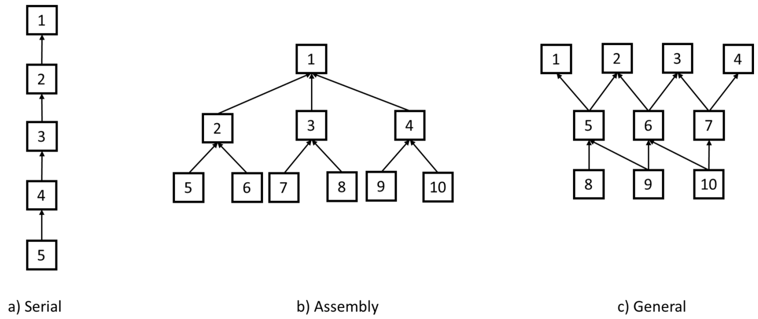


Figure 3: Different BOM structure (adapted from [27])

these classes and numbers, the demand for each item in each period is randomly generated based on the normal distribution, the average demand profiles, and a 30% coefficient of variance. For each of these settings, 5 different random replications are generated.

- We consider 5 planning periods and 1 machine at each level. Without loss of generality, we assume that the lead time is equal to 0. The processing time (pt_{kt}) and the setup times (st_{kt}) are equal to 1 and 0, respectively. The BOM coefficient (r_{ki}) is equal to 1. Table 8 illustrates the value of the parameters, in the base case (set A) and for the sensitivity analysis.

In the following sections, we provide the numerical results for different product structures and investigate the value of flexibility in the multi-level lot sizing problem. For the experiments, we used the CPLEX 12.8.1.0 and Python libraries. We performed these experiments on a 2.4 GHz Intel Gold processor with only one thread on the Compute Canada computing grid.

5.2. SAA analysis

In this section, we perform the SAA analysis on the basic set A which has been defined in Table 8, with different numbers of scenarios. We determine a reasonable number for the rest of the experiments based on the SAA Gap, solution time and memory limits. Table 9 illustrates these experiments with $M = 10$, and $S^{eval} = 10000$ for different numbers of scenarios S in each replication. To calculate the gap and its standard deviation for each instance, and to determine a proper number of scenarios in the SAA, we used the version with full flexibility. The time per replication in seconds (labeled as Time) is also reported for the version with full flexibility, which has the highest execution time compared to all versions with lower levels of flexibility. For the case with S= 1000 and full level of flexibility, for about 54% of the instances with an assembly structure and for all of the instances with a general structure, an optimal solution could NOT be found within the time limit of 7200 seconds or due to memory limitations. Considering the execution time, we will use 500 scenarios for the rest of the experiments to solve the model, and 10000 scenarios for evaluation. Among the three structures, the general structure has the highest execution time and the serial structure has the lowest one.

As discussed in Section 4.1.3, it is possible to have violated service levels for some scenarios in the evaluation phase of the SAA method, except for the case with full flexibility. We analyse the extent of these infeasibilities. In Table 10, the “Infeasibility percentage" shows the average percentage of scenarios for which a violation occurs out of

Table 7
Demand Profiles for different structures

		Average demand category 1	2	3
		Serial structure		
External demand	1	(100,0,0,0,0)	(100,0,0,0,0)	(500,0,0,0,0)
	2	(100,100,0,0,0)	(100,200,0,0,0)	(500,400,0,0,0)
	3	(100,100,100,0,0)	(100,200,300,0,0)	(500,400,300,0,0)
	4	(100,100,100,100,0)	(100,200,300,400,0)	(500,400,300,200,0)
	5	(100,100,100,100,100)	(100,200,300,400,500)	(500,400,300,200,100)
		Assembly structure		
External demand	1	(100,0,0,0,0,0,0,0,0,0)	(100,0,0,0,0,0,0,0,0,0)	(300,0,0,0,0,0,0,0,0,0)
	2	(100,100,0,0,0,0,0,0,0,0)	(100,200,0,0,0,0,0,0,0,0)	(300,200,0,0,0,0,0,0,0,0)
	3	(100,100,100,0,0,0,0,0,0,0)	(100,200,200,0,0,0,0,0,0,0)	(300,200,200,0,0,0,0,0,0,0)
	4	(100,100,0,0,100,0,0,0,0,0)	(100,200,0,0,300,0,0,0,0,0)	(300,200,0,0,100,0,0,0,0,0)
	5	(100,100,100,0,100,0,100,0,0,0)	(100,200,200,0,300,0,300,0,0,0)	(300,200,200,0,100,0,100,0,0,0)
		General structure		
External demand	1	(100,100,100,100,0,0,0,0,0,0)	(100,100,100,100,0,0,0,0,0,0)	(300,300,300,300,0,0,0,0,0,0)
	2	(100,100,100,100,100,0,0,0,0,0)	(100,100,100,100,200,0,0,0,0,0)	(300,300,300,300,200,0,0,0,0,0)
	3	(100,100,100,100,100,100,0,0,0,0)	(100,100,100,100,200,200,0,0,0,0)	(300,300,300,300,200,200,0,0,0,0)
	4	(100,100,100,100,100,0,0,100,0,0)	(100,100,100,100,200,0,0,300,0,0)	(300,300,300,300,200,0,0,100,0,0)
	5	(100,100,100,100,100,100,0,100,0,100)	(100,100,100,100,200,200,0,300,0,300)	(300,300,300,300,200,200,0,100,0,100)

Table 8
Parameter values for the base case and the sensitivity analysis

Parameter	Base case	Sensitivity analysis
Util	0.5	0.1 , 0.5 , 0.9
HValue	1	0, 1, 10
TBO profile	2	1, 2, 3, 4, 5
Service Level	95%	80%, 90%, 95% , 99%

Table 9
SAA analysis

	Serial			Assembly			General		
# Scenario	Avg Gap (%)	Std Gap (%)	Time	Avg Gap (%)	Std Gap (%)	Time	Avg Gap (%)	Std Gap (%)	Time
100	0.17	0.010	8.2	0.32	0.008	54.5	0.21	0.005	129.6
250	0.14	0.006	39.9	0.25	0.005	390.8	0.17	0.003	800.5
500	0.14	0.004	219.8	0.14	0.005	1314.7	0.16	0.003	3802.0
1000	0.14	0.006	524.7	0.13	0.004	5544.7			

the 10000 scenarios in the evaluation phase. The average value of service level violation is reported as ϵ. These two measures are calculated based on all levels of flexibility for each instance, except the full flexibility. We can see that these two measures are also acceptable for 500 scenarios.

5.3. The value of stochastic solution

In this section, we calculate the value of stochastic solution for the problem which is the cost difference between the cost of the optimal solution of the stochastic model and the deterministic model. To this end, the deterministic model (1) is solved in which the demand is equal to the expected demand. The solution of this model is then fixed as

Table 10
SAA analysis, infeasibility percentage

# Scenario	Serial		Assembly		General	
	Infeasibility percentage	ϵ (%)	Infeasibility percentage	ϵ (%)	Infeasibility percentage	ϵ (%)
100	3.35	0.07	3.79	0.08	7.14	0.13
250	1.82	0.04	1.76	0.03	3.49	0.06
500	1.11	0.02	1.10	0.02	2.19	0.04
1000	0.37	0.01	0.59	0.01	1.14	0.02

Table 11
Service level violation, $\epsilon(\%)$

LoF	Serial	Assembly	General
0	9.7	8.2	13.9
1	7.6	6.5	9
2	5.9	5.4	6.4
3	4.8	n/a	n/a
4	4.2	n/a	n/a

the first stage solution, and the cost of this model is calculated by optimally solving the second stage problem using the 10000 demand scenarios.

However, the current problem requires an additional consideration because of the service level. Although we can also consider the service level in the deterministic case, it is calculated based on the expected demand, and not several scenarios. This will result in a significant difference in the level of production between between the deterministic and stochastic case. This difference is more pronounced, when there is no flexibility in the model and all production decisions are defined in the first stage based on the expected value. In the deterministic case, based on the level of flexibility, the total evaluated cost (excluding the service level violation penalty) may be lower than the cost of the stochastic solution due to this lower level of production, but it also results in a high service level violation. Therefore, we do two separate analyses to show the value of the stochastic solution. First, for the case with full flexibility, we calculate the traditional VSS, calculated as the relative cost difference between the value of the deterministic solution and the stochastic solution. This VSS is equal to 10.6% for the serial structure, 11.9% for the assembly structure and 1.8% for the general structure. The second analysis focuses on the cases with a lower level of flexibility. For these cases, we focus on the service level violations in order to show the superiority of the stochastic model. Table 11 shows the service level violation of the mean value solution at different levels of flexibility for different structures. We conclude hence that the mean value deterministic approach cannot provide solutions that satisfy the service level.

5.4. Serial structure

In this section, we investigate the effect of adding flexibility in the serial structure. This effect is measured using the ratio of the total cost decrease ($\Delta Cost$) when we have some level of flexibility (LoF) compared to the case when there is no flexibility and every decision is fixed at the beginning of the planning horizon. To solve the models we use 500 scenarios for the first stage, and the evaluation phase is performed using 10000 randomly generated scenarios.

Figure 4 shows the effect of adding flexibility for the base case, set A, as defined in Table 8. When referring to the BOM, the lowest level (starting from 0) refers to the level of the end item and the highest level refers to the incoming components. The horizontal axis is classified in two categories, the upper one for the level of flexibility which is from 0 to 5, and the lower one for the external demand profile from 1 to 5. In demand profile 1, only the end item has the external demand. In demand profile 2, we also have the external demand for the component at level 1 of the BOM. This is explained in detail in the previous section and Table 7. Figure 4 illustrates that increasing the flexibility will result in cost reduction in all the demand profile. Even if there is external demand only for the end items (External demand profile 1), it is still beneficial (with decreasing marginal benefits) to increase the flexibility at the component levels. If external demand exists also at the component level, the benefits (in terms of relative cost reduction) will increase

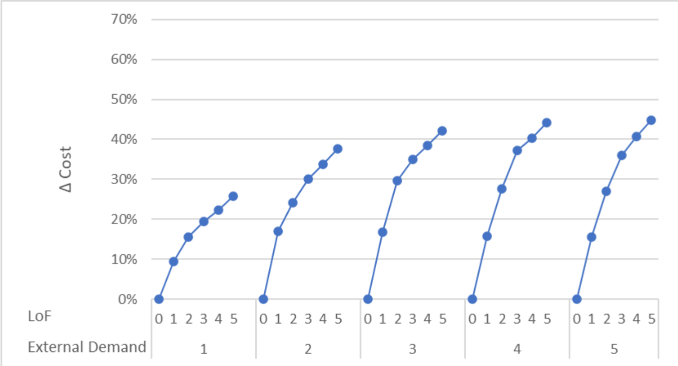


Figure 4: Effect of adding flexibility for serial structure

compared to the case with only external demand for the end item. So both an increased level of flexibility and the presence of external demand at the component level lead to larger relative cost reductions.

In the next set of experiments, we perform sensitivity analyses, in which all the parameters of the base case A, except the parameter of interest, remain fixed. Figure 5 illustrates the sensitivity analysis on the effect of adding flexibility, considering changes in the TBO profile, service level, $HValue$, and $Util$ parameters.
The smaller the TBO, the higher is the cost decrease by adding the flexibility. Low TBO means that you have many setup periods, and hence many opportunities to adjust the production (if flexible). As can be seen $\Delta Cost$ is the highest in TBO profile 1 compared to the other profiles. Comparing the TBO profiles 4 and 5, we can see that a high TBO at the lower level in the BOM (i.e., end items and lower level components) as in TBO profile 5 leads to lower benefits compared to a high TBO at the higher levels. This is logical as the production of the products at the lower level, determines the internal demand for the higher level components in the BOM and directly influences their production. TBO profile 4 has a higher rate of cost decrease as the TBO for the items at the lower level is less than the TBO of the items at the higher level.

$HValue$ is related to the amount of added value to the components at different levels of the BOM. When it is equal to 0, it means there is no difference in the holding cost of different items at different levels. Adding some flexibility only provides a limited cost decrease. Indeed, because the inventory holding cost is the same for holding an end item or for holding all its components, the total holding cost cannot be improved by a better redistribution of the inventory at different levels in the BOM. In case of full flexibility, we observe a sudden jump in the cost reduction as the production of all items is now reactive to the demand realization and the total amount of inventory in the system is reduced. This will be further explained when we provide a more detailed analysis of this at the end of this section. On the other hand, when $HValue$ is high, for example equal to 10, the inventory holding cost of the end item compared to other components is very high, and adding flexibility at higher levels where the inventory cost is much lower, will not result in a very high cost reduction. However, adding flexibility only for the end item reduces the costs by almost 30% since this flexibility with respect to the production of the end item already allows some redistribution of the inventory between levels 0 and 1. If demand for the end items is low, we do not have to produce excess end items (with a high holding cost) and we can keep inventory at the component level (where holding costs are cheaper).

At the higher service level, the value of flexibility is bigger compared to the lower ones. Having a higher service level result in more production and inventory to mitigate the uncertainty. In this case, having more reactive inventory system will cause a higher cost reduction. It is also interesting that adding only one level of flexibility will result in higher cost reduction at higher service level compared to lower ones.

The last diagram presents the effect of capacity on $\Delta Cost$. When there is no capacity limitation or when the capacity is loose, we have a slightly higher cost reduction. This is logical as the capacity limitation and overtime cost will limit the production even if it is flexible. To show the scale of the total costs in different setting, Table 12 illustrates the sensitivity analysis for the total cost. As can be seen, in the average column, by increasing the $HValue$, the total cost increases as this parameter has a direct effect on the holding cost. Regarding the TBO profiles, profile 1 which has a TBO of 1 for all the items, has the lowest cost, while TBO profile 3 which has the TBO equal to 4 for all the items has the highest cost as this value increases the setup cost. Comparing the total cost of profile 4 and 5 shows that, when the

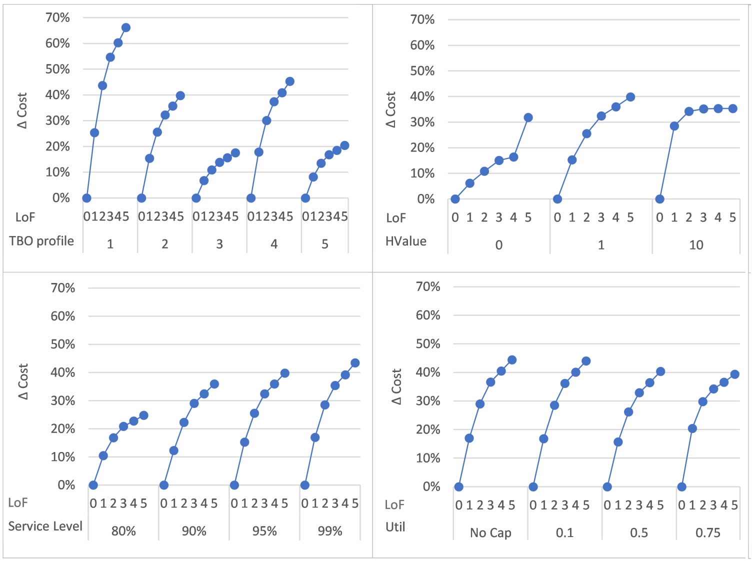


Figure 5: Sensitivity analysis for serial structure

item at the lower level of BOM has a higher TBO, the total cost is higher compared to the case when the higher TBO is at the higher level. Increasing the service level results in a total cost increase as the total production increases in the system. Increase in $Util$ means tighter capacity which imposes more cost to the production system. All the mentioned trends are valid not only for the average cost over all levels of flexibility, but also at each level of flexibility individually.

As can be seen in Figure 5, the trend of $\Delta Cost$ for different options of TBO profiles, Service levels, and $Util$ parameter are relatively similar. However, for the parameter $HValue$, the trends are also different for different options. Figure 6 presents a more detailed analysis for this parameter. In addition to the level of flexibility and $HValue$, the horizontal axis is also categorized based on the external demand profile which is shown in the second level from 1 to 5. When $HValue$ is equal to 0, and the external demand profile is equal to 1, i.e., we have only external demand for the end item at level 0, adding flexibility does not decrease the cost unless we have the full flexibility at all levels. The reason behind it is that there is no advantage in keeping inventory at a higher level of BOM, as the unit holding cost is not different. However, when we have full flexibility, the total amount of inventory decreases in the system and the total cost will decrease. When $HValue$ is equal to zero and when we have independent demand for the components as well (demand profile 2 to 5), depending on the level of these components, we have a cost decrease. The reason is that when we have proper level of flexibility, the amount of production is reactive to the demand realization, and it result in more efficient production and a cost decrease. At the levels where there is no independent demand, we see that the cost remains unchanged, until having the full flexibility. Having full flexibility, the total production is reactive to the demand realization and the the total amount of inventory is reduced in the system. When $HValue$ is equal to 10, we have a very high added value BOM. As can be seen, the marginal cost decrease is most significant when adding the first level of flexibility, compared to adding the further levels of flexibility. This is due to the fact that the cost of inventory for the item which is stored at the lowest level of the BOM is much lower compared to the same amount of inventory stored for an item at a higher level. When the $HValue$ is equal to 1, which is between the two extreme cases of $HValue = 0$ and $HValue = 10$, we can see a smooth cost reduction by adding multiple levels of flexibility.

5.5. Assembly structure

In this section, we investigate the value of flexibility for the assembly structure. Similar to the serial structure, we will compare the possible cost decrease ($\Delta Cost$) when we consider a more adaptive strategy. For this structure, flexibility will be added to the BOM level by level. Having 3 levels in the BOM for the assembly structure, we define 4 options for flexibility, from no flexibility to full flexibility. Assuming 0 as no flexibility for an item, and 1 for flexibility, Table 13 illustrates these levels of flexibility considered for this structure. The first option (0) has no flexibility, the second option (1) has flexibility for the end item only. The last option (3) has flexibility for all the items.

Table 12
Sensitivity analyses of the total cost for the serial structure

	Level of flexibility						
Parameter	0	1	2	3	4	5	Average
External demand							
1	89633	79940	73356	69428	66876	63617	73808
2	81441	67467	61108	56013	53048	49969	61508
3	91223	74407	62840	58016	55006	51873	65561
4	94821	79508	68198	59693	56895	53371	68748
5	94753	78750	67842	59697	55732	52213	68165
TBO profile							
1	48842	35491	26368	21035	18400	15657	27632
2	89267	74842	65458	59444	56495	53021	66421
3	227301	211502	201694	195318	191472	187463	202458
4	73859	60001	50601	45025	42305	39010	51800
5	192183	176345	165968	159916	157022	153779	167536
HValue							
0	12835	12045	11382	10767	10550	8484	11010
1	90427	75734	66191	59937	56842	53537	67111
10	35885935	26206487	24374210	24079449	24037381	24020933	26434066
Service Level							
80%	61672	54832	50850	48413	47256	46067	51515
90%	76737	66782	58787	53618	51149	48542	59269
95%	90277	75801	66172	59907	56812	53495	67078
99%	109196	89797	77158	69896	66069	61492	78935
Util							
No cap	81026	66436	56428	50199	47140	44005	57539
0.1	80797	66558	56626	50373	47299	44168	57637
0.5	90943	76176	66217	60010	56880	53469	67282
0.75	122711	95964	84240	78668	75814	72549	88324

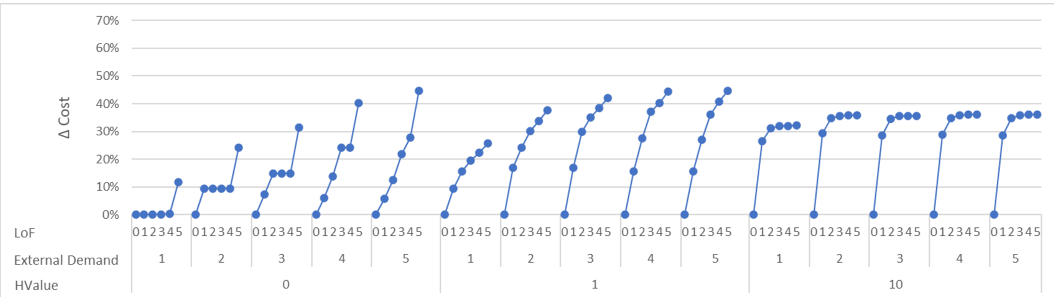


Figure 6: Effect of adding flexibility for serial structure

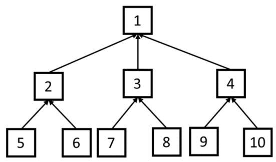
Figure 7 illustrates the cost decrease percentage ($\Delta Cost$) when increasing the flexibility level by level for different external demand profile (see Table 7). For all cases, adding flexibility only for the end item results in the highest rate of cost decrease, compared to the case of adding flexibility to the component levels. Comparing Figure 7 and Figure 4, we can see similar trends in the assembly structure compared to what has been observed for the serial structure.

Figure 8 illustrates the sensitivity analysis of $\Delta Cost$ for the assembly structure based on the TBO, $HValue$, service level and $Util$ parameters considering full flexibility options at different levels of BOM.

When the TBO is equal to 1 for all the items (TBO profile 1), we have the highest percentage of cost decrease, and when TBO is at its highest value for all the items (TBO profile 3), we have the lowest rate for the cost decrease by adding flexibility. Between these two extreme cases, having lower values of TBO for the items at the lower levels of BOM (TBO profile 4), results in higher rate of cost decrease compared to the case when we have higher value of TBO

Table 13
Levels of flexibility for assembly structure

# Flexibility	# product 1	2	3	4	5	6	7	8	9	10
0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	0
2	1	1	1	1	0	0	0	0	0	0
3	1	1	1	1	1	1	1	1	1	1
Level	0	1	1	1	2	2	2	2	2	2



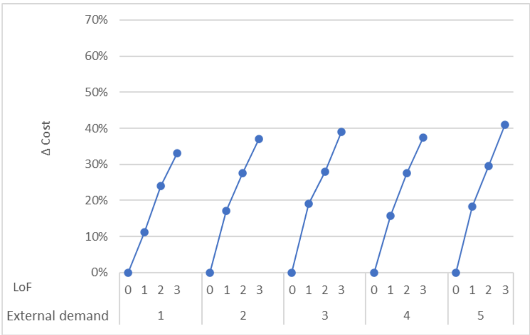


Figure 7: Analysis of adding flexibility per level for the base case in assembly structure

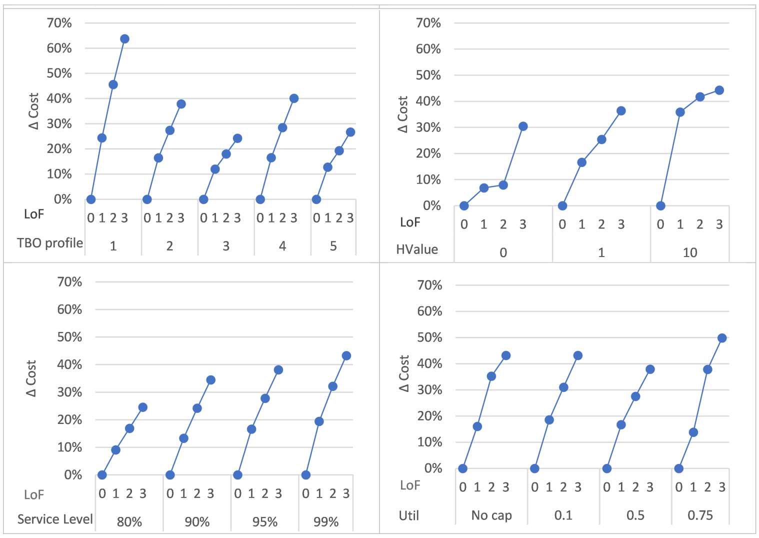


Figure 8: Sensitivity analysis for assembly structure

at these levels (TBO profile 5). Considering the $HValue$ parameter, we can see that different patterns for the added value in the product structure, result in different pattern of cost decrease, when we add flexibility. When the $HValue$ is equal to 0, there is no advantage to keep the components at the higher level of BOM to save the holding cost. In this case, when we add the flexibility at the highest level of BOM, we see a significant increase in the rate of decrease, as the total amount of inventory in the system will decrease. We hence observe a similar effect as in the serial case.

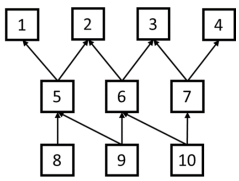
At different service levels, we have the same pattern for $\Delta Cost$ but there is a higher cost decrease at the higher service level when we add the flexibility. While at 80% of service level we have about 25% of cost decrease at the

Table 14
Sensitivity analyses of the total cost for the assembly structure

	LoF				
Parameter	0	3	5	8	Average
TBO profile					
1	41734	31217	22266	14661	27469
2	79300	65849	56977	48566	62673
3	201413	177611	165053	152651	174182
4	73000	60797	51697	43066	57140
5	148742	130453	120892	110171	127564
Hvalue					
0	28858	26828	26494	19796	25494
1	78454	64894	57787	49030	62541
10	1539484	1002998	913605	869498	1081396
Service Level					
80%	55965	51072	46423	42162	48905
90%	68036	58752	51168	44083	55509
95%	80408	66805	57606	49191	63502
99%	97996	77919	65430	54406	73937
Util					
No cap	64093	53203	40662	35488	48361
0.1	72615	58412	49203	40304	55133
0.5	80817	66925	57873	49378	63748
0.75	131791	111196	79101	62467	96138

Table 15
Levels of flexibility for general structure

# Flexibility	# product									
	1	2	3	4	5	6	7	8	9	10
0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	0	0	0	0	0	0
2	1	1	1	1	1	1	1	0	0	0
3	1	1	1	1	1	1	1	1	1	1
Level	0	0	0	0	1	1	1	2	2	2



full flexibility, the same value is about 45%, when the service level is equal to 99%. Considering different values for the $Util$ parameter, we see a similar trend for all cases. When we have a very tight capacity ($Util = 0.75$), we have a slightly higher rate of cost decrease.

Table 14 illustrates the sensitivity analysis for the total costs. The patterns we can see in assembly structure are similar to ones of the serial structure. In summary, higher $HValue$, higher TBO, and higher service levels result in higher cost. In addition, higher $Util$ which means tighter capacity imposes extra cost to the system.

5.6. General structure

In this section, we study the general structure and adding different levels of flexibility to different items in this structure. Table 15 illustrates different levels of flexibility for this structure. The levels of flexibility start from no flexibility to full flexibility. In this section, we only discuss the full flexibility per level.

Figure 9 illustrates the cost decrease percentage ($\Delta Cost$) with respect to the level of flexibility for different external demand profiles. Having more items with external demand generally results in a slightly higher cost decrease. For all cases, adding flexibility only for the end items (flexibility level 1) results in a rate of cost decrease of about 25%. In this structure, we have higher cost decrease compared to the assembly structure, and in general lower variability in different trends per external demand profiles, where there are more items with external demand, in the system.

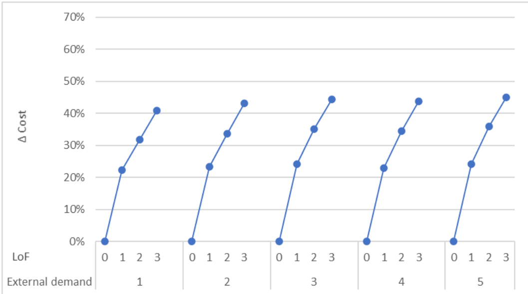


Figure 9: Analysis of adding flexibility per level for the base case in general structure

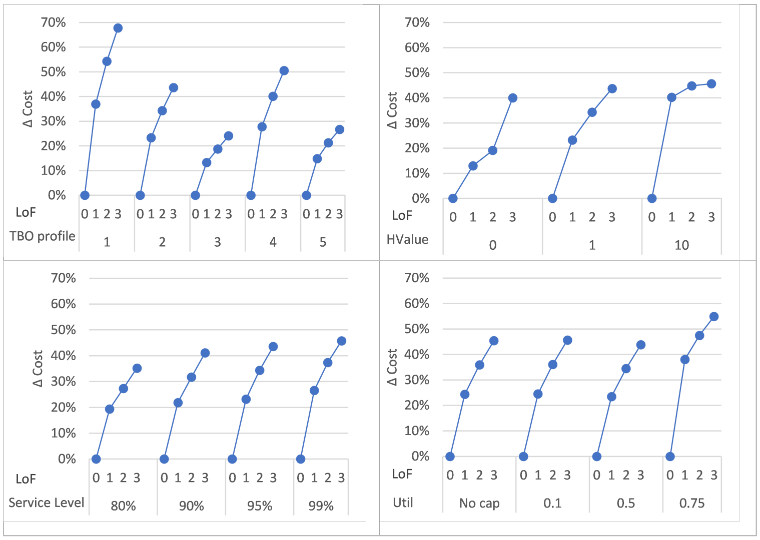


Figure 10: Sensitivity analysis for general structure

Figure 10 shows the sensitivity analysis for the general structure based on different parameters. The sensitivity analysis for the TBO shows similar patterns of cost decrease when adding flexibility compared to the serial and assembly structures. A lower TBO results in a higher cost decrease when adding flexibility. At higher service levels, it is more beneficial to add the flexibility. We can see that the diagrams for different service levels have similar patterns, but $\Delta Cost$ increases as we increase the service levels. When the capacity is very tight, where the model should use a significant amount of overtime, adding flexibility results in a higher cost reduction.

5.7. Insights

In this section we will present some insight based on our findings in the numerical experiments. First and in general, adding flexibility reduces the cost in the system. This cost reduction depends on the level where the flexibility happens and the external demand as well. In all structures, adding flexibility at the lower level, i.e., for the end items leads to more cost saving.

Second, different parameters in the problem affect the cost savings as well. The ratio between the setup cost and the inventory holding cost plays an important role in the cost savings obtained by adding flexibility. When the time between orders is low, for example one, based on the trade-off between the ordering costs and the holding costs, it is less costly that the production covers a smaller number of periods and there are hence more frequent setups. The value

Table 16
Sensitivity analyses of the total cost for the general structure

	LoF				
Parameter	0	3	5	8	Average
TBO profile					
1	60103	37252	26928	18910	35798
2	103831	79093	67675	58054	77163
3	227132	196767	184271	172151	195080
4	82730	58683	48221	39423	57264
5	188368	160128	148268	138152	158729
Hvalue					
0	32983	28602	26563	19518	26917
1	104371	79535	68062	58196	77541
10	2494477	1497842	1386434	1364212	1685741
Service Level					
80%	64460	51680	46521	41379	51010
90%	84653	65693	57226	49164	64184
95%	104102	79415	67851	58239	77402
99%	124723	90922	77581	67258	90121
Util					
No cap	96557	72533	61467	52223	70695
10%	98110	73387	62160	52826	71621
50%	104147	79254	67829	58105	77334
75%	147411	88238	74653	64337	93660

of adding production recourse is higher in this case, as you can adjust the production levels in each period. As the ratio between setup cost and holding cost may be different for different items, we should note that the end items, and the items at the lower levels of BOM have a higher impact in this matter.

Third, having flexibility results in holding cost reduction and there are two reasons behind that. First, production flexibility generally reduces the amount of inventory in the system, as the production are more responsive to the demand and there is less need for safety stock. Second, having flexibility will increase the option of where we can keep our inventory. More specifically, considering the added value in the BOM, keeping stock at higher level of BOM, where we have lower holding cost, and use them when needed, will reduce the total holding cost in the system.

6. Conclusion

In this research, we study the stochastic multi-level lot sizing problem with service level and investigate the benefits of production flexibility in the BOM based on static and adaptive strategies. The problems are modeled as two-stage stochastic models which are approximated using a finite number of scenarios and solved by the SAA method. Extensive numerical experiments and simulations are conducted for different BOM structures and under different parameter settings. The results show that increasing the production flexibility leads to a significant cost reduction, even in the case where there is no external demand for any of the components, in all BOM structures. Sensitivity analyses have been performed to demonstrate the effect of changing different parameters on the cost reduction by adding flexibility and allowing more adaptive decisions. In general, the value of adding flexibility is more significant at higher service levels compared to lower ones and with lower TBO compared to higher ones. The value added in the BOM structure also has effect on the pattern of cost decrease in different structures.

In the adaptive version of problem, we consider the static-dynamic strategy for some or all items and we modeled it as a two-stage stochastic model. Considering the dynamic strategy for these items is an interesting future research direction which needs a multi-stage stochastic model. This will make the problem more challenging to solve but it may result in more responsive plans.

References

[1] Absi, Nabil and Kedad-Sidhoum, Safia and Dauzère-Pérès, Stéphane, Uncapacitated lot-sizing problem with production time windows, early productions, backlogs and lost sales, International Journal of Production Research, 49, 9, 2551–2566, (2011)

[2] Afentakis, Panayotis and Gavish, Bezalel and Karmarkar, Uday, Computationally efficient optimal solutions to the lot-sizing problem in multistage assembly systems, Management Science, 30, 2, 222–239, (1984)

[3] Afentakis, Panayotis and Gavish, Bezalel, Optimal lot-sizing algorithms for complex product structures, Operations Research, 34, 2, 237–249, (1986)

[4] Akartunalı, Kerem and Miller, Andrew J, A heuristic approach for big bucket multi-level production planning problems, European Journal of Operational Research, 193, 2, 396–411, (2009)

[5] Almeder, Christian, A hybrid optimization approach for multi-level capacitated lot-sizing problems, European Journal of Operational Research, 200, 2, 599–606, (2010)

[6] Almeida, JFF and Conceição, SV and Pinto, LR and Oliveira, BRP and Rodrigues, LF, Optimal sales and operations planning for integrated steel industries, Annals of Operations Research, 315, 2, 773–790, (2022)

[7] Alvarez, Aldair and Cordeau, Jean-François and Jans, Raf and Munari, Pedro and Morabito, Reinaldo, Inventory routing under stochastic supply and demand, Omega, 102–304, (2020)

[8] Bookbinder, James H and Tan, Jin-Yan, Strategies for the probabilistic lot-sizing problem with service-level constraints, Management Science, 34, 9, 1096–1108, (1988)

[9] Buschkühl, Lisbeth, Multi-level capacitated lotsizing with setup carryover, (2008), (Springer)

[10] Contreras, Ivan and Cordeau, Jean-François and Laporte, Gilbert, Stochastic uncapacitated hub location, European Journal of Operational Research, 212, 3, 518–528, (2011)

[11] Gruson, Matthieu and Cordeau, Jean-François and Jans, Raf, The impact of service level constraints in deterministic lot sizing with backlogging, Omega, 79, 91–103, (2018)

[12] Gruson, Matthieu and Cordeau, Jean-François and Jans, Raf, Benders decomposition for a stochastic three-level lot sizing and replenishment problem with a distribution structure, European Journal of Operational Research, 291, 1, 206–217, (2021)

[13] Helber, Stefan and Sahling, Florian and Schimmelpfeng, Katja, Dynamic capacitated lot sizing with random demand and dynamic safety stocks, OR Spectrum, 35, 1, 75–105, (2013)

[14] Hung, Yi-Feng and Chien, Kuo-Liang, A multi-class multi-level capacitated lot sizing model, Journal of the Operational Research Society, 51, 11, 1309–1318, (2000)

[15] Kleywegt, Anton J and Shapiro, Alexander and Homem-de-Mello, Tito, The sample average approximation method for stochastic discrete optimization, SIAM Journal on Optimization, 12, 2, 479–502, (2002)

[16] Mousavi, Kianoush and Bodur, Merve and Roorda, Matthew , Stochastic Last-mile Delivery with Crowd-shipping and Mobile Depots (2021), Available online at: http://www.optimization-online.org/DB−HTML/2020/02/7619.html

[17] Pochet, Yves and Wolsey, Laurence A, Solving multi-item lot-sizing problems using strong cutting planes, Management Science, 37, 1, 53–67, (1991)

[18] Pochet, Yves and Wolsey, Laurence A, Production planning by mixed integer programming, (2006), Springer Science & Business Media

[19] Quezada, Franco and Gicquel, Céline and Kedad-Sidhoum, Safia and Vu, Dong Quan, A multi-stage stochastic integer programming approach for a multi-echelon lot-sizing problem with returns and lost sales, Computers & Operations Research, 116, 104865, (2020)

[20] Seeanner, Florian and Almada-Lobo, Bernardo and Meyr, Herbert, Combining the principles of variable neighborhood decomposition search and the fix&optimize heuristic to solve multi-level lot-sizing and scheduling problems, Computers & Operations Research, 40, 1, 303–317, (2013)

[21] Sahling, Florian and Buschkühl, Lisbeth and Tempelmeier, Horst and Helber, Stefan, Solving a multi-level capacitated lot sizing problem with multi-period setup carry-over via a fix-and-optimize heuristic, Computers & Operations Research, 36, 9, 2546–2553, (2009)

[22] Sereshti, Narges and Adulyasak, Yossiri and Jans, Raf, The value of aggregated service levels in stochastic lot sizing problems, Omega, 102335, (2020)

[23] Shapiro, Alexander and Dentcheva, Darinka and Ruszczyński, Andrzej, Lectures on stochastic programming: modeling and theory, (SIAM) (2014)

[24] Stadtler, Hartmut, Multilevel lot sizing with setup times and multiple constrained resources: Internally rolling schedules with lot-sizing windows, Operations Research, 51, 3, 487–502, (2003)

[25] Taş, Duygu and Gendreau, Michel and Jabali, Ola and Jans, Raf, A capacitated lot sizing problem with stochastic setup times and overtime, European Journal of Operational Research, 273, 1, 146–159, (2019)

[26] Tavaghof-Gigloo, Dariush and Minner, Stefan, Planning approaches for stochastic capacitated lot-sizing with service level constraints, International Journal of Production Research, 59, 17, 5087–5107, (2021)

[27] Tempelmeier, Horst and Derstroff, Matthias, A Lagrangean-based heuristic for dynamic multilevel multiitem constrained lotsizing with setup times, Management Science, 42, 5, 738–757, (1996)

[28] Tempelmeier, Horst, A column generation heuristic for dynamic capacitated lot sizing with random demand under a fill rate constraint, Omega, 39, 6, 627–633, (2011)

[29] Tempelmeier, Horst and Herpers, Sascha, Dynamic uncapacitated lot sizing with random demand under a fillrate constraint, European Journal of Operational Research, 212, 3, 497–507, (2011)

[30] Tempelmeier, Horst and Hilger, Timo, Linear programming models for a stochastic dynamic capacitated lot sizing problem, Computers & Operations Research, 59, 119–125, (2015)

[31] Thevenin, Simon and Adulyasak, Yossiri and Cordeau, Jean-François, Material Requirements Planning Under Demand Uncertainty Using Stochastic Optimization, Production and Operations Management, 30, 2, 475–493 (2021)

[32] Toledo, Claudio Fabiano Motta and da Silva Arantes, Márcio and Hossomi, Marcelo Yukio Bressan and França, Paulo Morelato and Akartunalı, Kerem, A relax-and-fix with fix-and-optimize heuristic applied to multi-level lot-sizing problems, Journal of Heuristics, 21, 5, 687–717, (2015)

[33] Tunc, Huseyin and Kilic, Onur A and Tarim, S Armagan and Eksioglu, Burak A simple approach for assessing the cost of system nervousness, International Journal of Production Economics, 141, 2, 619–625, (2013)

[34] Tunc, Huseyin and Kilic, Onur A and Tarim, S Armagan and Eksioglu, Burak, A reformulation for the stochastic lot sizing problem with service-level constraints, Operations Research Letters, 42, 2, 161–165, (2014)

[35] Verweij, Bram and Ahmed, Shabbir and Kleywegt, Anton J and Nemhauser, George and Shapiro, Alexander, The sample average approximation method applied to stochastic routing problems: a computational study, Computational optimization and applications, 24, 2, 289–333, (2003)

[36] Wagner, Stephan M and Lindemann, Eckhard, A case study-based analysis of spare parts management in the engineering industry, Production Planning & Control, 19, 4, 397–407, (2008)

[37] Wu, Tao and Shi, Leyuan and Geunes, Joseph and Akartunalı, Kerem, An optimization framework for solving capacitated multi-level lot-sizing problems with backlogging, European Journal of Operational Research, 214, 2, 428–441, (2011)

[38] Xiao, Yiyong and Zhang, Renqian and Zhao, Qiuhong and Kaku, Ikou and Xu, Yuchun, A variable neighborhood search with an effective local search for uncapacitated multilevel lot-sizing problems, European Journal of Operational Research, 235, 1, 102–114, (2014)

[39] You, Meng and Xiao, Yiyong and Zhang, Siyue and Zhou, Shenghan and Yang, Pei and Pan, Xing, Modeling the capacitated multi-level lot-sizing problem under time-varying environments and a fix-and-optimize solution approach, Entropy, 21, 4, 377, (2019)