Journal of the Operational Research Society (2011) 62, 1075–1084

www.palgrave-journals.com/jors/

Multiple lot-sizing decisions in a two-stage production with an interrupted geometric yield and non-rigid demand

M-C Wu*, L-C Huang, H-M Hsu and T-S Su

National Chiao Tung University, Hsin-Chu, Taiwan

In a production system with random yield, it may be more cost effective to release lots *multiple* times towards fulfilling a customer order. Such a decision, called the multiple lot-sizing problem, has been investigated in various contexts. This paper proposes an efficient algorithm for solving a *new* multiple lot-sizing problem defined in the context of a two-stage production system with non-rigid demand when its process yields are governed by interrupted geometric distributions. We formulate this problem as a dynamic program (DP) and develop lemmas to solve it. However, solving such a DP may be computationally extensive, particularly for large-scale cases with a high yield. Therefore, this study proposes an efficient algorithm for resolving computational issues. This algorithm is designed to reduce the DP network into a much simpler algorithm by combining a group of DP branches into a single one. Extensive experiments were carried out. Results indicate that the proposed reduction algorithm is quite helpful for practitioners dealing with large-scale cases characterized by high-yield.

Journal of the Operational Research Society (2011) **62,** 1075–1084. doi:10.1057/jors.2010.39
Published online 19 May 2010

Keywords: lot-sizing; interrupted geometric distribution; dynamic programming; two-stage system; production/inventory system

1. Introduction

Multiple lot-sizing problems arise when considering variations in process yield. In a production system with a random yield, it may be more cost effective to release lots *multiple* times fulfill customer orders. In this case, it is necessary to determine the optimal lot size for each lot release. This problem has been called a Multiple Lot-sizing Production to Order (MLPO) problem (Grosfeld-Nir and Gerchak, 1996).

This paper presents a *new* MLPO problem with the following features. Firstly, completion of a product requires two stages (or two manufacturing processes). The two stages are interconnected and cannot be performed at the same time. Secondly, the yield of each stage is random and governed by an interrupted geometric (IG) distribution. A manufacturing process with an IG distribution denotes that once a defective item is produced by the process, the items produced afterwards are all defective. Thirdly, the delivery due dates for customer orders are strictly imposed; in other words, delivery after the due date is not acceptable, and supply shortages will be penalised. This feature is also called *non-rigid demand* because a customer's demand may not be completely fulfilled by the due date. Fourthly, the products are customised and the salvage values of over-supplied items are negligible.

One example of a two-stage MLPO problem is a capital-intensive machine tool (eg, a steel rolling mill) equipped with two shape-forming dies. Each die is used to perform, can only perform a particular shaping process. The manufacturing of a product involves two shaping operations through the successive use of both dies. Yet, at any instance, only one die can be equipped on the machine. Therefore, the two shaping operations cannot be simultaneously carried out—fulfilling the first requirement above. Secondly, the process yields of both dies are governed by IG distributions, since once a die becomes worn out, all the parts it produces afterwards are defective. Thirdly, the product is highly customised and has a short life cycle. Therefore, customers are highly concerned with the due date and any delivery after the due date is not acceptable.

At each period, the two-stage MLPO problem involves two decisions: (1) which die or manufacturing process to select, and (2) how large to make the lot-size for the selected die. These two decisions are made at each period until the customer order is completely fulfilled or the decision point reaches the due date. In the event that the customer order is only partially fulfilled by the due date, the manufacturer will incur a penalty cost for shortage.

Numerous studies on the MLPO problem with IG distributions have been published. Most of these studies assume

Correspondence: M-C Wu, Department of Industrial Engineering and Management, National Chiao Tung University, Hsin-Chu, Taiwan.
E-mail: mcwu@mail.nctu.edu.tw

a *rigid demand* scenario in which a customer order must be completely fulfilled. A few studies examine non-rigid demand scenarios, but limit their scope to a single-stage production system (Guu and Zhang, 2003). To date, the two-stage interlocked MLPO problem described above has not been investigated in the literature.

We formulated the MLPO problem as a dynamic program (DP) and developed lemmas to solve it. For a small-scale case, solving the DP efficiently obtains an optimum solution. However, for large-scale cases with a high yield, solving this DP becomes computationally prohibitive. To address this problem, this study proposes an efficient method of solving the MLPO problem. Our idea is to reduce the complex DP network into a much simpler algorithm by combining a group of DP branches into a single one. Extensive experiments were carried out. Results indicate that the proposed method is quite helpful for practitioners dealing with large-scale cases characterized by a high yield.

The remainder of this paper is organised as follows. Section 2 reviews previous multiple-stage MLPO studies. Section 3 introduces how the MLPO problem is formulated as a DP model and develops lemmas to determine the upper bounds of lot-sizes for the DP. Section 4 describes the DP reduction method. Section 5 presents numerical experiments, followed by concluding remarks in Section 6.

2. Research background

Two comprehensive survey papers (Yano and Lee, 1995; Grosfeld-Nir and Gerchak, 2004) on MLPO problems have been published. Prior studies on MLPO problems can be categorised according to four perspectives.

The first perspective is associated with the number of *production/inspection* stages, and includes two options: *single-stage* and *multiple-stage*. Single-stage studies assume there is *only one inspection station* at the end of a production line, and that inspection results are utilised to make lot-sizing decisions for the following period. Alternatively, multiple-stage studies assume that the completion of a product must involve a sequence of manufacturing stages, and that each stage is equipped with its own inspection station. In this case, the inspection results of a particular stage are utilised to make subsequent lot-sizing decisions for upstream and downstream stages.

Examples of single-stage studies include Beja, 1977; Sepheri *et al*, 1986; Pentico, 1988; Grosfeld-Nir and Gerchak, 1990; Anily, 1995; Grosfeld-Nir and Gerchak, 1996; Zhang and Guu, 1997, 1998; Guu, 1999; Guu and Liou, 1999; Anily *et al*, 2002; Guu and Zhang, 2003. Example of multiple-stage studies include Lee and Yano, 1988; Wein, 1992; Grosfeld-Nir and Ronen, 1993; Pentico, 1994; Grosfeld-Nir, 1995, 2005; Grosfeld-Nir and Robinson, 1995; Barad and Braha, 1996, Barad, 1999; Grosfeld-Nir and Gerchak, 2002; Grosfeld-Nir *et al*, 2006; Ben-Zvi and Grosfeld-Nir, 2007. In practice, the steel engraving process is like a single-stage system, while the steel drawing and coating processes are like a multiple-stage system.

The second perspective is associated with delivery requirement, which includes two options: *rigid-demand* and *non-rigid demand*. In the *rigid-demand* scenarios, partial fulfillment of an order is not acceptable (Wein, 1992; Grosfeld-Nir and Ronen, 1993; Pentico, 1994; Grosfeld-Nir, 1995, 2005; Grosfeld-Nir and Robinson, 1995; Grosfeld-Nir and Gerchak, 2002; Grosfeld-Nir *et al*, 2006; Ben-Zvi and Grosfeld-Nir, 2007). In the *non-rigid demand* scenarios, though customers will not accept any products after their due dates, partial fulfillment of an order is acceptable (Lee and Yano, 1988; Barad and Braha, 1996; Braha, 1999).

Practical examples of rigid-demand scenarios involve the order of making components for use in an assembly line. Partial fulfillment of such an order would result in partial fulfillment of its final assembly, and might idle some other types of components in the assembly. Alternatively, non-rigid demand scenarios emphasise customers' *eager concern* in meeting due dates. For example, at the due date, customers have booked a scarce capacity to undergo a *batch operation* on the products they received. Therefore, products delivered after the due-date will not be accepted.

The third perspective is associated with the probability distributions of process yield. These probability distributions include the general discrete distribution (Grosfeld-Nir, 1995; Grosfeld-Nir and Robinson, 1995; Grosfeld-Nir and Gerchak, 2002; Grosfeld-Nir *et al*, 2006), the binomial distribution (Grosfeld-Nir and Ronen, 1993; Pentico, 1994; Grosfeld-Nir, 1995, 2005; Grosfeld-Nir and Robinson, 1995; Barad and Braha, 1996; Braha, 1999; Grosfeld-Nir and Gerchak, 2002; Grosfeld-Nir *et al*, 2006; Ben-Zvi and Grosfeld-Nir, 2007), the IG distribution (Grosfeld-Nir and Robinson, 1995; Grosfeld-Nir and Gerchak, 2002; Grosfeld-Nir *et al*, 2006; Ben-Zvi and Grosfeld-Nir, 2007), the all-or-nothing distribution (Grosfeld-Nir and Robinson, 1995; Grosfeld-Nir, 1995; Grosfeld-Nir and Gerchak, 2002; Grosfeld-Nir *et al*, 2006; Ben-Zvi and Grosfeld-Nir, 2007), and the stochastically proportional distribution (Lee and Yano, 1988; Wein, 1992; Grosfeld-Nir, 1995).

The following section describes the various applications of the aforementioned probability distributions. An IG distribution refers to a process that uses a tool or die to manufacture part one at a time. However, the tool/die gradually wears out over time. As soon as the tool/die becomes severely worn, none of the parts produced afterwards will not the specifications. A general discrete distribution also refers to a process which produces parts one by one, but the quality of each part is *independent* due to some uncontrollable factors. This leads to various discrete probabilistic distributions in yield-modelling, one of which is the Binomial distribution. The all-or-nothing distribution refers to a *batch process*, in which all parts in a lot are processed together; therefore, all the parts in a batch either pass or fail. A stochastically proportional distribution is utilised to model continuous production process.

The fourth perspective is associated with the solution approach. Most prior formulations of the MLPO problems include recursive formulas that have been widely interpreted as DP problems. Therefore, DP has been widely used to solve the MLPO problem. However, this approach can be very demanding computationally. Some researchers have proposed the use of lemmas to reduce the solution space (Beja, 1977; Anily, 1995; Zhang and Guu, 1998), while others attempt to develop near optimal heuristic rules (Sepheri *et al*, 1986; Pentico, 1988). A few authors have roughly modelled the DP problem using a simpler non-DP problem for cases with extremely large or small demand quantities (Anily *et al*, 2002).

3. DP model

This section introduces a way to model the two-stage MLPO problem as a DP. We first assume that the production lead times of both processes being equal (ie, one period). Then, we describe the MLPO problem, analyse the outcomes of a lot-sizing decision, and model its associated cost function using a recursive formula, which reveals that the MLPO problem is a DP. Finally, we discuss how this DP can be used to model scenarios in which the production lead-times of the two processes are different.

Notation

T: total number of periods in the time horizon.

t: period index for $t=T, \ldots, 2, 1, 0$, where T denotes the starting period and $t=0$ denotes the due date.

i: index of production stage M_i, $i=1, 2$.

$\alpha^{(i)}$: setup cost required for releasing a lot to stage M_i, $i=1, 2$.

$\beta^{(i)}$: variable production cost per unit at stage M_i, $i=1, 2$.

D: initial demand (number of demand units at $t=T$).

D_t: remaining demand at period t (number of demand units not fulfilled at t).

h: inventory holding cost of finished goods (\$/unit-period).

m: shortage cost per unit.

$\theta^{(i)}$: yield parameter of the IG distribution at stage M_i, $i=1, 2$.

$k_t^{(i)}$: size of the lot released to M_i at period t, $i=1, 2$ (note that $k_t^{(1)} \cdot k_t^{(2)}=0$ due to the interlock constraints imposed on the two processes.)

$W_t^{(i)}$: binary variable indicating whether or not a setup is required at stage M_i

$$W_t^{(i)} = \begin{cases} 0 & if \quad k_t^{(i)} = 0 \\ 1 & if \quad k_t^{(i)} > 0 \end{cases}.$$

$Y_{k_t^{(i)}}$: random variable denoting the number of good units produced from lot $k_t^{(i)}$, $y_{k_t^{(i)}}$ denotes the possible outcome of $Y_{k_t^{(i)}}$, $y_{k_t^{(i)}} = 0, 1, \ldots, k_t^{(i)}$, $i=1, 2$.

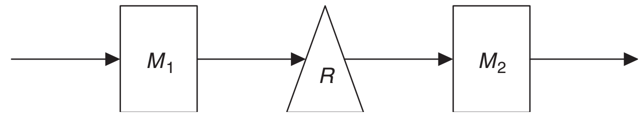

Figure 1 The two-stage production system.

$p(y_{k_t^{(i)}})$: probability of producing exactly $y_{k_t^{(i)}}$ good units from lot $k_t^{(i)}$, where

$$\sum_{y_{k_t^{(i)}}=0}^{k_t^{(i)}} p(y_{k_t^{(i)}}) = 1, i = 1, 2.$$

B_t: inventory level of work-in-process (WIP) buffer R at period t, where

$$B_t = B_{t+1} - k_{t+1}^{(2)} + Y_{k_{t+1}^{(1)}}$$

$s_t = (D_t, B_t)$: production system status at period t, also called state s_t.

$N_t(s_t) = (k_t^{(1)}, k_t^{(2)})$: decision made at state s_t; that is, $k_t^{(1)}$ units are released to M_1 and $k_t^{(2)}$ are released to M_2. Note that $k_t^{(1)} \cdot k_t^{(2)} = 0$ due to the interlock constraints imposed on the two processes.

$C_t(N_t(s_t))$: total expected cost incurred from period t to period 0, for a particular decision $N_t(s_t)$.

$C_t^*(s_t) = \underset{(k_t^{(1)},\ k_t^{(2)})}{Min} \{C_t(N_t(s_t))\}$ minimum total expected cost incurred from period t to period 0, for all possible decisions at state s_t.

$N_t^*(s_t) = (k_t^{(1)^*}, k_t^{(2)^*})$, an optimal decision made at state s_t; that is,

$$C_t(N_t^*(s_t)) = \underset{(k_t^{(1)}, k_t^{(2)})}{Min} \{C_t(N_t(s_t))\} = C_t^*(s_t)$$

3.1. MLPO problem

The production system of the MLPO problem involves two stages/processes, as Figure 1 shows: M_1 and M_2. Both of these processes are equipped with a work-in-process (WIP) buffer R, and inspection is performed at the end of each process. Units determined to be defective are scrapped. The production lead-time (including inspection) for each process is one period.

To meet a non-rigid demand quantity D, lots can be periodically released to M_1 and proceed to M_2. At the end of any period, when the inspection results are revealed, we must first select the process (either M_1 or M_2) to proceed with, and then determine the lot-size for that selected process. These two decisions can be aggregated into a single decision determining $N_t(s_t) = (k_t^{(1)}, k_t^{(2)})$. However, the condition $k_t^{(1)} \cdot k_t^{(2)} = 0$ must be maintained due to the interlock constraints imposed on the two processes.

Notice that the lot size for M_2 must be less than the currently available WIP in R. Lot-sizing decisions must be made until the due date arrives or the customer order has been completely fulfilled. Finally, the MLPO problem only considers the inventory holding cost of finished goods.

In the MLPO problem, the process yield is an IG distribution. Such a probability distribution is defined as follows:

$$P(Y_{k_t^{(i)}} = y_{k_t^{(i)}}) = \begin{cases} (1-\theta^{(i)})\cdot(\theta^{(i)})^{y_{k_t^{(i)}}} & y_{k_t^{(i)}} = 0, 1, 2, \ldots, k_t^{(i)}-1 \\ (\theta^{(i)})^{k_t^{(i)}} & y_{k_t^{(i)}} = k_t^{(i)} \end{cases}$$

This equation indicates that each process operates in two possible states: *in-control* and *out-of-control*. While the process is in-control, all products produced are good; while out-of-control, all products produced are defective. To produce exactly $y_{k_t^{(i)}}$ good units, the system must be in-control for the first $y_{k_t^{(i)}}$ units, until it reaches the out-of-control state at the $(y_{k_t^{(i)}}+1)^{th}$ unit. A mold-based shaping process is a good example of process yield characterised by the IG distribution. Once the mold is worn out, the output produced afterwards will be defective.

3.2. Outcomes of a lot-sizing decision

As stated above, the condition $k_t^{(1)}\cdot k_t^{(2)}=0$ must be maintained due to the interlock constraints imposed on the two processes. Figure 2 illustrates the possible outcomes at *t*-1 of a lotsizing decision $(k_t^{(1)}, k_t^{(2)})$ made at *t*. At *t*, the state of the production system (called *state* hereafter) is $s_t=(D_t, B_t)$, which implies that the remaining demand is D_t and the WIP level is B_t. Now either $k_t^{(1)}$ or $k_t^{(2)}$ units are released to M_1 and M_2; their respective outcomes are $y_{k_t^{(1)}}$ and $y_{k_t^{(2)}}$. With this input/output information, we can derive that $s_{t-1}=(D_t-y_{k_t^{(2)}}, B_t-k_t^{(2)}+y_{k_t^{(1)}})$.

The occurrence probabilities for $y_{k_t^{(1)}}$ and $y_{k_t^{(2)}}$ are $p(y_{k_t^{(1)}})$ and $p(y_{k_t^{(2)}})$, respectively. With $0\leqslant y_{k_t^{(1)}}\leqslant k_t^{(1)}$ and $0\leqslant y_{k_t^{(2)}}\leqslant k_t^{(2)}$, we an easily infer that the number of possible outcomes for a lot-sizing decision $(k_t^{(1)}, k_t^{(2)})$ is $(k_t^{(1)}+1)\cdot(k_t^{(2)}+1)$ where $k_t^{(1)}\cdot k_t^{(2)}=0$. That is, the occurrence of each outcome has a probability of $p(y_{k_t^{(1)}})\cdot p(y_{k_t^{(2)}})$, and $\sum_{y_{k_t^{(1)}}=0}^{k_t^{(1)}}\sum_{y_{k_t^{(2)}}=0}^{k_t^{(2)}} p(y_{k_t^{(1)}})\cdot p(y_{k_t^{(2)}})=1$.

3.3. Cost function and DP

Suppose a lot-sizing decision $N_t(s_t)=(k_t^{(1)}, k_t^{(2)})$ is made at $s_t=(D_t, B_t)$. In this case, the total expected cost incurred from period t to period 0 (the due date) can be formulated as follows.

$$C_t(N_t(s_t)) = H_1 + H_2 + \sum_{y_{k_t^{(1)}}=0}^{k_t^{(1)}}\sum_{y_{k_t^{(2)}}=0}^{k_t^{(2)}} p(y_{k_t^{(1)}}) \cdot p(y_{k_t^{(2)}})\cdot H_3 + \sum_{y_{k_t^{(2)}}=0}^{k_t^{(2)}} p(y_{k_t^{(2)}})\cdot H_4 \quad (1)$$

where

$$k_t^{(1)}\cdot k_t^{(2)}=0$$

$$H_1=\alpha^{(1)}W_t^{(1)}+\beta^{(1)}k_t^{(1)}$$

$$H_2=\alpha^{(2)}W_t^{(2)}+\beta^{(2)}k_t^{(2)}$$

$$H_3=C_{t-1}^*(s_{t-1}),\ \text{where } s_{t-1}=(D_t-y_{k_t^{(2)}}, B_t-k_t^{(2)}+y_{k_t^{(1)}})$$

$$H_4=h\cdot(t-1)\cdot y_{k_t^{(2)}}$$

In Equation (1), the terms H_1 and H_2 respectively denote the production cost incurred at s_t for M_1 and M_2. The third term $\sum_{y_{k_t^{(1)}}=0}^{k_t^{(1)}}\sum_{y_{k_t^{(2)}}=0}^{k_t^{(2)}} p(y_{k_t^{(1)}})\cdot p(y_{k_t^{(2)}})\cdot H_3$ denotes the minimum total expected cost for all possible states of s_{t-1}. The fourth term $\sum_{y_{k_t^{(2)}}=0}^{k_t^{(2)}} p(y_{k_t^{(2)}})\cdot H_4$ denotes the *expected* inventory holding cost for all possible outcomes of $k_t^{(2)}$.

Equation (1) is a recursive formula due to the inclusion of $H_3=C_{t-1}^*(s_{t-1})$. To make an optimal lot-sizing decision at state s_t, we must know the optimal decision at s_{t-1}. This recursive feature indicates that the MLPO problem is a DP.

The DP has two boundary conditions (BCs). The first BC is defined at $s_t=(0, B_t)$. When the order is completely fulfilled, no more lots are must be released and the corresponding cost is zero. That is,

$$N_t^*(s_t=(0, B_t))=(0,0) \quad (2)$$

$$C_t^*(s_t=(0, B_t))=0 \quad (3)$$

The second BC is defined at $s_0=(D_0, B_0)$. At the due date, any further production is useless, and only the shortage cost is incurred. That is,

$$C_0^*(s_0=(D_0, B_0))=mD_0 \quad (4)$$

3.4. Upper bounds for the DP

To solve the DP, we developed three lemmas that give upper bounds for $k_t^{(1)^*}$ and $k_t^{(2)^*}$. Lemma 1 states that $k_t^{(2)^*}\leqslant\min(D_t, B_t)$. Lemmas 2 and 3 as a whole state that $k_t^{(1)^*}\leqslant\min\{(\lceil\frac{\ln\beta^{(1)}-\ln m}{\ln\theta^{(1)}}\rceil-1), ((t-1)D_t-B_t)\}$. The Appendix provides the proofs of these three lemmas as well as one supporting proposition.

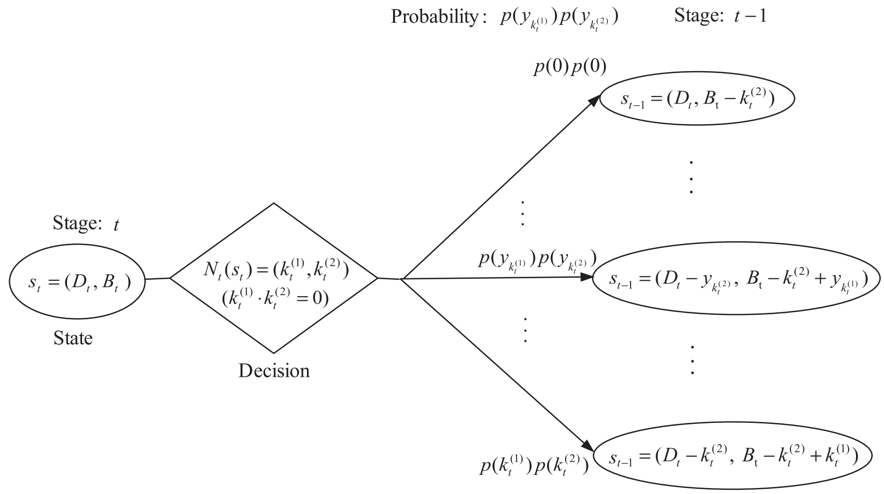

Figure 2 A decision structure in the DP model.

3.5. Embellishment of the DP model

The DP model can be embellished to model a more *complex* MLPO problem, in which the production lead-times of the two stages are *different*. Denote the production lead-time of the first stage by p_1 and that of the second stage by p_2. In such a scenario, not all periods may be used to release a lot because the machine is not always free at the end of each period. This embellished DP model assumes that a lot-sizing decision can be made *at each period* but some additional constraints must be imposed on $k_t^{(1)}$ and $k_t^{(2)}$, as Equations (5)–(8) indicate Moreover, H_4 in Equation (1) should be modified as in Equation (9).

$$\text{If } k_t^{(1)}>0, \text{ then } k_{t-\tau}^{(1)}=0 \text{ and } k_{t-\tau}^{(2)}=0 \text{ for } \tau=1,\ldots,p_1-1 \quad (5)$$

$$\text{If } k_t^{(2)}>0, \text{ then } k_{t-\tau}^{(1)}=0 \text{ and } k_{t-\tau}^{(2)}=0 \text{ for } \tau=1,\ldots,p_2-1 \quad (6)$$

$$k_t^{(1)}=0, \text{ for period } t=p_1-1,\ldots,1 \quad (7)$$

$$k_t^{(2)}=0, \text{ for period } t=p_2-1,\ldots,1 \quad (8)$$

$$H_4=h\cdot(t-p_2)\cdot y_{k_t^{(2)}} \quad (9)$$

Equations (5) and (6) ensure that whenever a manufacturing stage is triggered, no more lots can be released until the machine is free again. Equations (7) and (8) show that we are not allowed to any lot if the remaining time to the due date is less than the production lead-time. Equation (9) highlights the reduction of inventory-holding cost while the production lead-time of stage M_2 is increased.

Owing to inclusion of additional constraints on $k_t^{(1)}$ and $k_t^{(2)}$, the embellished DP network is a subset of the original DP network. Therefore, the lemmas in the Appendix, which define the upper bounds for $k_t^{(1)}$ and $k_t^{(2)}$, are also valid in the embellished DP model.

4. Reduction of DP network

This study proposes a method for reducing the DP network into a much simpler one by *consolidating* the outcomes of a lot-sizing decision. The proposed method includes three major steps. First, we cluster all the possible outcomes of a lot-sizing decision into several groups. Second, we define a representative outcome for each group. Third, we simplify the DP network by including only the representative outcomes of each group.

The first step is to *group outcomes*. For a lot-sizing decision $(k_t^{(1)}, k_t^{(2)})$, the total number of possible outcomes is $(k_t^{(1)}+1)\cdot(k_t^{(2)}+1)$. For the outcomes of $k_t^{(1)}$, there are $k_t^{(1)}+1$ options (ie, 0, 1, $\ldots$, $k_t^{(1)}$). By consolidating every n *consecutive* outcome into a group, we obtain $\lceil\frac{k_t^{(1)}+1}{n}\rceil$ groups. For example, when $n=2$ and $k_t^{(1)}=6$, we would create four groups: {0, 1}, {2, 3}, {4, 5}, {6}. Likewise, for the outcomes of $k_t^{(2)}$ we would create $\lceil\frac{k_t^{(2)}+1}{n}\rceil$ groups.

The second step is to *define each group's representative outcome*. Define $Q_j=\{y|n(j-1)\leqslant y\leqslant nj-1\}$ as the *jth* group that comprises n consecutive outcomes of either $k_t^{(1)}$ or $k_t^{(2)}$, where y denotes the possible outcome and its probability is denoted by $p(y)$. The representative outcome of group Q_j is $Round(\sum_{y=n(j-1)}^{nj-1}y\cdot p(y)/\sum_{y=n(j-1)}^{nj-1}p(y))$, and its corresponding probability is $\sum_{y=n(j-1)}^{nj-1}p(y)$. For example, consider group $j=2$ $Q_2=\{2,3\}$ with $p(2)=0.1$ and $p(3)=0.2$. Then, the representative outcome of group Q_2 is $Round(\frac{(2\cdot 0.1)+(3\cdot 0.2)}{(0.1+0.2)})=Round(2.7)=3$, and its corresponding probability is 0.1+0.2 $=0.3$.

The third step is to *simplify the DP network* by modelling each group with its representative outcomes. After this simplification method, $k_t^{(1)}$ has only $\lceil\frac{k_t^{(1)}+1}{n}\rceil$ outcomes and $k_t^{(2)}$ has

Table 1 C_p and T_p of the solutions obtained by the comprehensive algorithm

$(\theta^{(1)}, \theta^{(2)})$		(T, D)					
		(10 100)		(10 200)		(20 100)	
		C_p($)	T_p (sec)	C_p ($)	T_p (sec)	C_p ($)	T_p (sec)
0.9	0.9	8070.76	220	18 070.80	1019	6228.48	1875
0.9	0.99	5058.79	169	15 026.30	707	2151.69	821
0.9	0.999	3578.58	158	13 304.30	710	1437.32	789
0.99	0.9	5016.30	2394	14 972.70	19 489	2629.97	23 309
0.99	0.99	1018.96	2412	2561.08	22 080	1017.97	23 057
0.99	0.999	689.91	2324	1363.89	20 789	689.90	22 979
0.999	0.9	3984.39	2754	13 742.00	39 414	2343.91	57 688
0.999	0.99	847.77	2738	1846.54	43 191	847.77	57 377
0.999	0.999	541.07	2703	1057.13	41 408	541.07	57 056

Table 2 $R_C$and R_T of the solutions obtained by the reduction algorithm for $n = 5$

$(\theta^{(1)}, \theta^{(2)})$		(T, D)					
		(10 100)		(10 200)		(20 100)	
		R_C(%)	R_T(%)	R_C(%)	R_T(%)	R_C(%)	R_T(%)
0.9	0.9	0.28	3.29	0.13	2.05	0.72	2.36
0.9	0.99	0.05	5.19	0.02	4.48	0.33	4.70
0.9	0.999	0.07	4.91	0.00	4.36	0.29	4.45
0.99	0.9	0.58	3.68	0.06	3.34	1.18	3.16
0.99	0.99	0.31	3.75	0.14	3.50	0.17	3.32
0.99	0.999	0.06	3.64	0.03	3.38	0.06	3.15
0.999	0.9	0.33	3.67	0.64	3.24	4.30	2.94
0.999	0.99	0.15	3.75	0.21	3.42	0.15	3.03
0.999	0.999	0.01	3.57	0.03	3.31	0.01	3.03

Table 3 R_C and R_T of the solution obtained by the reduction algorithm for $n = 10$

$(\theta^{(1)}, \theta^{(2)})$		(T, D)					
		(10 100)		(10 200)		(20 100)	
		R_C(%)	R_T(%)	R_C(%)	R_T(%)	R_C(%)	R_T(%)
0.9	0.9	0.27	2.18	0.12	1.25	0.56	1.39
0.9	0.99	0.14	4.45	0.04	2.85	1.70	3.22
0.9	0.999	0.19	3.35	0.03	2.77	2.61	2.92
0.99	0.9	0.92	2.15	0.08	1.92	11.51	1.76
0.99	0.99	1.01	2.23	0.22	2.01	1.11	1.76
0.99	0.999	0.09	2.13	0.07	1.91	0.09	1.73
0.999	0.9	6.10	2.14	0.12	1.77	6.71	1.59
0.999	0.99	0.77	2.23	0.16	1.91	0.77	1.61
0.999	0.999	0.03	2.10	0.17	1.82	0.03	1.60

only $\lceil \frac{k_t^{(2)}+1}{n} \rceil$ outcomes. The total number of outcomes for $(k_t^{(1)}, k_t^{(2)})$ is $\lceil \frac{k_t^{(1)}+1}{n} \rceil \cdot \lceil \frac{k_t^{(2)}+1}{n} \rceil$, which greatly reduces the DP network.

Notice that while $n = 1$, the reduced DP becomes the original DP. To facilitate comparison, the method of solving the original DP is called the *comprehensive algorithm*, while that of solving the reduced DP is called the *reduction algorithm*.

5. Numerical experiments

This study proposes two methods for solving the MLPO problem—the *comprehensive* algorithm and the *reduction*

Table 4 R_C and R_T of the solution obtained by the reduction algorithm for $n = 15$

$(\theta^{(1)}, \theta^{(2)})$		(T, D)					
		(10 100)		(10 200)		(20 100)	
		R_C(%)	R_T(%)	R_C(%)	R_T(%)	R_C(%)	R_T(%)
0.9	0.9	0.59	1.49	0.26	0.75	1.06	0.99
0.9	0.99	0.29	2.59	0.07	1.91	3.83	2.23
0.9	0.999	0.54	2.48	0.03	1.89	18.44	2.10
0.99	0.9	0.87	1.65	0.22	1.44	14.67	1.23
0.99	0.99	1.28	1.66	0.58	1.48	1.27	1.27
0.99	0.999	0.85	1.61	0.26	1.41	0.85	1.25
0.999	0.9	0.87	1.61	10.65	1.28	12.36	1.05
0.999	0.99	3.10	1.68	0.92	1.36	3.10	1.12
0.999	0.999	0.05	1.61	0.24	1.32	0.05	1.12

Table 5 R_C and R_T obtained by the reduction algorithm for $n = k_t + 1$

$(\theta^{(1)}, \theta^{(2)})$		(T, D)					
		(10 100)		(10 200)		(20 100)	
		R_C(%)	R_T(%)	R_C(%)	R_T(%)	R_C(%)	R_T(%)
0.9	0.9	1.31	0.18	0.58	0.04	2.96	0.21
0.9	0.99	4.67	0.73	1.73	0.24	30.71	0.26
0.9	0.999	3.88	0.62	1.61	0.42	53.22	0.26
0.99	0.9	3.30	0.38	1.00	0.12	18.22	0.18
0.99	0.99	103.94	0.35	32.27	0.19	104.14	0.20
0.99	0.999	54.80	0.33	90.92	0.17	54.80	0.19
0.999	0.9	2.36	0.56	0.76	0.18	20.47	0.15
0.999	0.99	87.78	0.59	60.76	0.28	87.78	0.18
0.999	0.999	25.02	0.56	91.44	0.26	25.02	0.17

algorithm. Extensive experiments were carried out to justify both algorithms using personal computers equipped with a 3.4 G CPU and 504MB of RAM.

We measured the effectiveness and efficiency of the reduction algorithm using two performance indices: $R_C = \frac{C_r - C_p}{C_p}$ and $R_T = \frac{T_r}{T_p}$, where C_r is the total expected cost of the lot-sizing decision suggested by the reduction algorithm and T_r is its computation time; C_p and T_p are defined accordingly for the comprehensive algorithm.

Experimental results suggest that the *reduction algorithm* should be used to solve a large-scale problem (ie, large *TD*) with a high yield in the first stage (eg, $\theta^{(1)} \geqslant 0.99$). For other scenarios, we suggest the use of the comprehensive algorithm.

According to Table 1, when $\theta^{(1)} \geqslant 0.99$ and $(T, D) \in \{(10, 200), (20, 100)\}$, solving the MLPO problem using the comprehensive algorithm is computationally extensive taking about 5.4–16.0 hours. This is essentially due to that $k_t^{(1)^*} \leqslant \min\{(\lceil \frac{\ln \beta^{(1)} - \ln m}{\ln \theta^{(1)}} \rceil - 1), ((t-1)D_t - B_t)\}$, which indicates that higher values of $\theta^{(1)}$, D, and T increase the upper bound of $k_t^{(1)^*}$, which in turn increases computational complexity.

Table 6 Average computation time required by the comprehensive algorithm for solving some low-yield MLPO problems (unit: sec)

T	D			
	10	50	100	200
10	0.13	4.17	13.13	21.20
15	0.34	12.74	43.03	167.95
20	0.33	26.32	123.21	496.78

In contrast, such complex MLPO problems can be efficiently and satisfactorily solved using the reduction algorithm. As shown in Table 2, $n = 5$ R_T ranges from 2.94% to 3.50%, while R_c ranges only from 0.06% to 4.30%. For the scenario with $(TD) = (20{,}100)$ and $(\theta^{(1)}, \theta^{(2)}) = (0.999, 0.999)$, the computation time is greatly reduced from 16 hours to 30 min. Moreover, as Tables 3–5 indicate, increasing the value of n tends to decrease R_T at the cost of increasing R_c.

Alternatively, the comprehensive algorithm is better-suited to scenarios with a relatively low yield in the first process (ie, $\theta^{(1)} \leqslant 0.9$). Table 6 shows the average computation times for solving some MLPO problems using the comprehensive algorithm. Each (T, D) scenario in this table includes 16 instances, which are generated by setting $\theta^{(1)}, \theta^{(2)} \in \{0.6, 0.8\}$, $\beta^{(1)} = \beta^{(2)} \in \{1, 2\}$, $m \in \{100, 200\}$, and $\alpha^{(1)} = \alpha^{(2)} = 50$.

In practice, a manufacturing processes with a yield higher than 0.99 (ie, $\theta^{(1)} \geqslant 0.99$) is not unusual. In shaping processes, the yield of a die is generally higher than 0.99. Otherwise the die would be frequently changed, increasing processing/production costs. Therefore, the proposed reduction algorithm can help practitioners quickly and effectively solve the MLPO problem.

6. Conclusions

This paper examines a novel two-stage MLPO problem with an IG yield distribution and non-rigid demand. The processing times for the two stages may be different. Once a stage completes the manufacturing of a lot, we need to decide which of the two stages to proceed further and make the lot-size decision.

We first formulate this MLPO problem as a DP. Then by developing lemmas forr the upper bounds of lot-sizes, we propose an *exact* approach (ie, the comprehensive algorithm) to solve the DP. However, solving the DP using an exact approach may be computationally extensive, and particularly for large-scale, high-yield scenarios. To resolve this computational difficulty, we further develop an efficient algorithm—an *approximation* approach—that reduces the DP network into a much simpler algorithm by consolidating a group of DP branches into a single one.

Numerical experiments indicate that *the approximation approach* (*ie, the reduction algorithm*) can effectively and efficiently solve large-scale, high-yield problems. For a practical MLPO problem requiring 16 hours of computation to solve, the reduction algorithm can obtain a satisfactory solution in only 30 min. Yet, for small-scale cases, we suggest solving the DP using the *comprehensive algorithm* because it requires only a few minutes or less of computation.

A possible extension of this research is to apply the reduction method to solve other MLPO problems. For example, many other MLPO problems in the literature are modelled as a DP, and various heuristic methods have been developed for dealing with large-scale cases. Comparing the performance of these heuristic methods and the reduction method would be an interesting topic for further research.

References

Anily S (1995). Single-machine lot-sizing with uniform yields and rigid demands: robustness of the optimal solution. *IIE Trans* **27**: 625–633.

Anily S, Beja A and Mendel A (2002). Optimal lot sizes with geometric production yield and rigid demand. *Opns Res* **5**: 424–432.

Barad M and Braha D (1996). Control limits for multi-stage manufacturing processes with binomial yield (single and multiple production runs). *J Opl Res Soc* **47**: 98–112.

Beja A (1977). Optimal reject allowance with constant marginal production efficiency. *Naval Res Logist Quart* **24**: 21–33.

Ben-Zvi T and Grosfeld-Nir A (2007). Serial production systems with random yield and rigid demand: a heuristic. *Opns Res Lett* **35**: 235–244.

Braha D (1999). Manufacturing control of a serial system with binomial yields, multiple production runs, and non-rigid demand: a decomposition approach. *IIE Trans* **31**: 1–9.

Grosfeld-Nir A (1995). Single bottleneck systems with proportional expected yields and rigid demand. *Eur J Opl Res* **80**: 297–307.

Grosfeld-Nir A (2005). A two-bottleneck system with binomial yields and rigid demand. *Eur J Opl Res* **165**: 231–250.

Grosfeld-Nir A and Gerchak Y (1990). Multiple lotsizing with random common-cause yield and rigid demand. *Opns Res Lett* **9**: 383–388.

Grosfeld-Nir A and Gerchak Y (1996). Production to order with random yields: single-stage multiple lot-sizing. *IIE Trans* **28**: 669–676.

Grosfeld-Nir A and Gerchak Y (2002). Multistage production to order with rework capability. *Mngt Sci* **48**: 652–664.

Grosfeld-Nir A and Gerchak Y (2004). Multiple lotsizing in production to order with random yields: review of recent advances. *Ann Opns Res* **126**: 43–69.

Grosfeld-Nir A and Robinson LW (1995). Production to order on a two machine line with random yields and rigid demand. *Eur J Opl Res* **80**: 264–276.

Grosfeld-Nir A and Ronen B (1993). A single bottleneck system with binomial yields and rigid demand. *Mngt Sci* **39**: 650–653.

Grosfeld-Nir A, Anily S and Ben-Zvi T (2006). Lot-sizing two-echelon assembly systems with random yields and rigid demand. *Eur J Opl Res* **173**: 600–616.

Guu SM (1999). Properties of the multiple lot-sizing problem with rigid demand, general cost structures, and interrupted geometric yield. *Opns Res Lett* **25**: 59–65.

Guu SM and Liou FR (1999). An algorithm for the multiple lot sizing problem with rigid demand and interrupted geometric yield. *J Math Anal Appl* **234**: 567–579.

Guu SM and Zhang AX (2003). The finite multiple lot sizing problem with interrupted geometric yield and holding costs. *Eur J Opl Res* **145**: 635–644.

Lee HL and Yano CA (1988). Production control in multistage systems with variable yield losses. *Opns Res* **36**: 269–278.

Pentico DW (1988). An evaluation and proposed modification of the Sepheri-Silver-New heuristic for multiple lot sizing under variable yield. *IIE Trans* **20**: 360–363.

Pentico DW (1994). Multistage production systems with random yield: heuristics and optimality. *Int J Prod Res* **32**: 2455–2462.

Sepheri M, Silver EA and New C (1986). A heuristic for multiple lot sizing for an order under variable yield. *IIE Trans* **18**: 63–69.

Wein AS (1992). Random yield, rework and scrap in a multistage batch manufacturing environment. *Opns Res* **40**: 551–563.

Yano CA and Lee HL (1995). Lot sizing with random yields: a review. *Opns Res* **43**: 311–334.

Zhang AX and Guu SM (1997). Properties of the multiple lot-sizing problem with rigid demands and general yield distributions. *Comput Math Appl* **33**: 55–65.

Zhang AX and Guu SM (1998). The multiple lot sizing problem with rigid demand and interrupted geometric yield. *IIE Trans* **30**: 427–431.

Appendix

Proposition 1 *$C^*_{t-1}(s_{t-1} = (D_t, B_t - n)) \geqslant C^*_{t-1}(s_{t-1} = (D_t, B_t))$ for $B_t \geqslant n \geqslant 1$ and n is an integer.*

Proof Consider a case in which we intentionally hold n WIP units in station R until the due date. In our cost model, the n WIP units will be scrapped after the due date, and therefore add no more cost. By intentionally scrapping the n WIP units, the lot-sizing decision for $s_{t-1} = (D_t, B_t)$ in this case becomes the lot-sizing decision for $s_{t-1} = (D_t, B_t - n)$. That is, $C^*_{t-1}(s_{t-1} = (D_t, B_t - n))$ can always be obtained from the lot-sizing decision at $s_{t-1} = (D_t, B_t)$. This implies that $C^*_{t-1}(s_{t-1} = (D_t, B_t - n)) \geqslant C^*_{t-1}(s_{t-1} = (D_t, B_t))$ □

Lemma 1 *Given $t \geqslant 1$, $D_t \geqslant 1$, $s_t = (D_t, B_t)$, we have $k_t^{(2)^*} \leqslant D_t$ and $k_t^{(2)^*} \leqslant B_t$.*

Proof $k_t^{(2)^*} \leqslant B_t$ is trivial because B_t is the available WIP that can be released at stage 2.

To prove $k_t^{(2)^*} \leqslant D_t$, consider two cases.

Case 1 $k_t^{(1)^*} > 0$

We have $k_t^{(2)^*} = 0 \leqslant D_t$ because $k_t^{(1)^*} \cdot k_t^{(2)^*} = 0$

Case 2 $k_t^{(1)^*} = 0$

Let $N'_t(s_t) = (0, D_t)$ and $N''_t(s_t) = (0, D_t + n)$

Then

$$C_t(N'_t(s_t)) = \alpha^{(2)} + \beta^{(2)} \cdot (D_t) + h(t-1) \cdot E[Y_{D_t}] + \sum_{y_{D_t}=0}^{D_t} p(y_{D_t}) \cdot C^*_{t-1}(s_{t-1}{=}(D_t{-}y_{D_t}, B_t{-}D_t))$$

and

$$C_t(N''_t(s_t)) = \alpha^{(2)} + \beta^{(2)} \cdot (D_t + n) + h(t-1) E[Y_{D_t+n}] + \sum_{y_{D_t+n}=0}^{D_t+n} p(y_{D_t+n}) \cdot C^*_{t-1}(s_{t-1}{=}(D_t{-}y_{D_t+n}, B_t{-}(D_t{+}n)))$$

$$C_t(N''_t(s_t)) - C_t(N'_t(s_t)) = n \cdot \beta^{(2)} + h \cdot (t-1) \cdot (E[Y_{D_t+n}] - E[Y_{D_t}]) + A - B, \text{ where}$$

$$A = \sum_{y_{D_t+n}=0}^{D_t-1} p(y_{D_t+n}) \cdot C^*_{t-1}(s_{t-1} = (D_t - y_{D_t+n}, B_t - (D_t + n)))$$

$$B = \sum_{y_{D_t}=0}^{D_t-1} p(y_{D_t}) \cdot C^*_{t-1}(s_{t-1} = (D_t - y_{D_t}, B_t - (D_t)))$$

Based on Proposition 1, $A \geqslant B$. Therefore,

$$C_t(N''_t(s_t)) - C_t(N'_t(s_t)) \geqslant n \cdot \beta^{(2)} + h \cdot (t{-}1) \cdot (E[Y_{D_t+n}] {-} E[Y_{D_t}]) > 0$$

This implies that $k_t^{(2)^*} \leqslant D_t$. □

Lemma 2 $k_t^{(1)^*} \leqslant (t-1) \cdot D_t - B_t$.

For any t, $D_t \geqslant D_{t-1}$. Based on Lemma 1, we can infer $\sum_{\tau=1}^{t-1} k_\tau^{(2)^*} \leqslant (t-1) \cdot D_t$. That is, for stage 2, the sum of lot-sizes from period $t-1$ to 1 are at most $(t-1) \cdot D_t$. This implies that stage 2 at period $t{-}1$ at needs a WIP level of $(t-1) \cdot D_t$ at most. At period t, the WIP level at stage 2 is B_t. Therefore, the units to be produced from stage 1 at period $t{-}1$ are at most $(t-1) \cdot D_t - B_t$. Consider stage 2 as a customer, the customer's demand at period $t{-}1$ is at most $(t-1) \cdot D_t - B_t$. According to a prior study (Guu and Zhang, 2003), in a single-stage production system with an IG distribution, the optimal lot-size will not exceed the demand. This implies that the lot-size released from stage 1 at period t is at most $(t-1) \cdot D_t - B_t$. That is, the upper bound for $k_t^{(1)^*}$ is $(t-1) \cdot D_t - B_t$.

Lemma 3 *If $t \geqslant 1$, $D_t \geqslant 1$, and $s_t = (D_t, B_t)$ where $B_t \geqslant 0$ then $k_t^{(1)^*} \leqslant \lceil \frac{\ln \beta^{(1)} - \ln m}{\ln \theta^{(1)}} \rceil - 1$.*

Proof Let $N'_t(s_t) = (k, 0)$ for $k \geqslant 1$

$$C_t(N'_t(s_t)) = \alpha^{(1)} + \beta^{(1)} \cdot (k) + \sum_{y_k=0}^{k} p(y_k) \cdot C^*_{t-1}(s_{t-1} = (D_t, B_t + y_{y_k}))$$

Let $N''_t(s_t) = (k+1, 0)$ for $k \geqslant 1$

$$C_t(N''_t(s_t)) = \alpha^{(1)} + \beta^{(1)} \cdot (k+1) + \sum_{y_{k+1}=0}^{k+1} p(y_{k+1}) \cdot C^*_{t-1}(s_{t-1} = (D_t, B_t + y_{k+1}))$$

$$C_t(N''_t(s_t)) - C_t(N'_t(s_t)) = \beta^{(1)} - (\theta^{(1)})^{k+1} \cdot [C^*_{t-1}(s_{t-1} = (D_t, B_t + k)) - C^*_{t-1}(s_{t-1} = (D_t, B_t + k + 1))]$$

Consider a particular state $(D_t, B_t + k)$, which is changed by adding one more unit of WIP, becoming a new state $(D_t, B_t + k + 1)$. This newly added WIP unit could become a unit of finished goods before the due date. In the best case, this decreases the shortage cost by one unit. In other words,

$$C^*_{t-1}(s_{t-1} = (D_t, B_t + k)) - C^*_{t-1}(s_{t-1} = (D_t, B_t + k + 1)) \leqslant m$$

Accordingly, $C_t(N''_t(s_t)) - C_t(N'_t(s_t)) \geqslant \beta^{(1)} - (\theta^{(1)})^{k+1} \cdot m$

Namely, if $\beta^{(1)} \geqslant (\theta^{(1)})^{k+1} \cdot m$ we conclude $C_t(N_t''(s_t)) - C_t(N_t'(s_t)) \geqslant 0$.

This implies that if lot-size k is in the range imposed by $\beta^{(1)} \geqslant (\theta^{(1)})^{k+1} \cdot m$ then we cannot achieve a better solution by increasing one more unit in lot-size.

Therefore, $\beta^{(1)} = (\theta^{(1)})^{k+1} \cdot m$ defines the upper bound of lot-size k.

That is, $k = \frac{\ln \beta^{(1)} - \ln m}{\ln \theta^{(1)}} - 1$ is the upper bound of $k_t^{(1)^*}$.

Thus, we conclude that $k_t^{(1)^*} \leqslant \lceil \frac{\ln \beta^{(1)} - \ln m}{\ln \theta^{(1)}} \rceil - 1$. □

Received December 2008;
accepted February 2010 after two revisions