Analysis of postponement strategy by EPQ-based models with planned backorders

Jian Li [a,c], Shouyang Wang [a] and T.C. Edwin Cheng [b,1]

[a] *Institute of Systems Science, Academy of Mathematics and Systems Sciences, Chinese Academy of Sciences,Beijing, 100080, China*

[b] *Department of Logistics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong*

[c] *Department of Mathematics, Qufu Normal University, Qufu, Shandong, 273165, China*

Abstract

This paper develops economic production quantity (EPQ)-based models with planned backorders to evaluate the impact of the postponement strategy on a manufacturer in a supply chain. We derive the optimal total average costs per unit time for producing and keeping n end-products in a postponement system and a non-postponement system, respectively. By comparing the optimal total average costs of the two systems, we evaluate the impact of postponement on the manufacturer under four circumstances. Our results show that postponement strategy can give a lower total average cost under certain circumstances. We also find that the key factors in postponement decisions are the variance of the machine utilization rates and the variance of the backorder costs.

Key words: Postponement; Economic production quantity (EPQ); Backorder; Supply chain management (SCM)

1 Introduction

In order to meet the needs of increasingly demanding customers for more diverse products, many companies have reconfigured their supply chain structures to accommodate mass customization. The new structures often involve either delaying the delivery of the products until after orders arrive or delaying differentiation of the products to later production stages of the supply chain. This is known as the postponement strategy [11]. Postpone-

[1] Corresponding author. Tel.: (852)-2766-5216; Fax:(852)-2364-5245.
E-mail address: LGTcheng@polyu.edu.hk(Edwin Cheng).

ment, also known as late customization or delayed product differentiation, was first discussed by Alderson [1]. van Hoek [26] pointed out that postponement is one of the central features of mass customization. Li et al. [19] selected postponement as one of five major SCM practices that have discernible impact on competitive advantage and organizational performance. Practical examples of postponement can be found in the high-tech industry, food industry and other industries that require high differentiation. For example, Hewlett-Packard produces generic printers at its factories and distribute them to local distribution centers, where power plugs with appropriate voltage and user manuals in the right language are packed. They have saved a lot of money every year by adopting the postponement strategy ([10, 17]). Similarly, many food manufacturers have moved their labeling or branding processes closer to customers.

The empirical analysis of Li et al. [19] indicated that postponement may not be a an evident SCM practice compared to the other four practices. This can be true. Postponement has both advantages and disadvantages. The advantages include following the JIT principles, reducing end-product inventory [5], making forecasting easier [7, 9] and pooling risk [12]. The high cost of designing and manufacturing generic components is the main drawback of postponement [17, 20]. The implementation of postponement is dependent on a firm's market characteristics and the type of the products it produces and therefore may not be applicable in all situations. Thus, evaluation of postponement structures is an important issue and there have been many qualitative and quantitative models for analyzing the postponement strategy under different scenarios. A comprehensive review of this stream of research can be found in van Hoek [26] and Wan et al. [27]. Recent quantitative models include, but are not limited to, those by Lee [16], Lee and Tang [18], Garg and Tang [12], Garg and Lee [11], Ernst and Kamrad [9], Aviv and Federgruen [3], Ma [20], and Su [25]. They evaluated the cost and benefits of applying postponement in a large variety of stochastic settings. In their models, the product demands were assumed to be stochastic and independent across time. Under these assumptions, the benefits of postponement, in terms of inventory related performance measures, are confined to two factors: statistical economies of scale and risk pooling via a common buffer [2]. The common buffer can reduce the magnitude of systemwide safety stocks.

If the demand is deterministic, e.g., because there is a long-term supply contract between the manufacturer and the customers, the benefits due to economies of scope and risk pooling do not exit. Recent deterministic models include, among others, those by Wan et al. [28, 29]. They developed multi-product models with deterministic demand to analyze postponement. There are numerous studies addressing deterministic multi-product inventory models. The classical inventory models are discussed in well-known books and others, e.g., [30]. Deterministic multi-product models can be classified into two types: EOQ-based models and EPQ-based models. The multi-product EOQ model originated from Shu [23], and Nocturns [21]. They gave the optimal order frequency of two and multi-product models, respectively. Silver [24] developed a simple cycle policy based on the EOQ model with n groups of products. Ben-Daya and Hariga [4] reviewed some efficient heuristics for solving the joint replenishment problem. EPQ models for multiple products can be classified into two types: single machine and multiple machines. In this paper we only focus on the single machine case. Goyal [13] presented a procedure for obtaining the optimal production frequencies for a number of items that are manufactured jointly. He [14] also used a search procedure to determine the EPQs of two items. Eilon [8] classified the production of several products by a single machine as a multi-product batch scheduling problem. Goyal and Satir [15] presented a survey of the solution procedures. Recent research on the multi-product inventory problem has focused on models with a large variety of constraints and considerations of the coordination of a supply chain. Most of the above papers focused on deriving the optimal ordering strategy. None of them compare the total average cost between a postponement system and a non-postponement system, except Wan et al. [28, 29]. Wan et al. showed that postponed customization of the end-products will result in a lower total average cost and a lower EOQ in EOQ-based models [28]. In EPQ-based models, they showed that postponed customization always results in a lower average cost when the demands are met after production is finished [29]. They assumed that there are no backorders in their EPQ models. However, backorder plays an important role both in theoretical analysis and actual practice. It is natural to consider EPQ-based models with backorders to analyze postponement. There are some papers addressing the EPQ model with planned backorders when the end-product demand is met continuously by the current production batch [6, 22]. In these studies, the EPQ and the optimal total average cost per unit time for producing and keeping one end-product were given. To the best of our knowledge, there has been no study that addresses the EPQ model with backorders when the demand is met after production is finished.

Motivated by the above observations, we develop models in this paper to fill this gap in the literature. We give the cost function and the optimal strategy of an EPQ-based model with planned backorders when the demand is met after production is finished. We derive the optimal total average costs per unit time of a postponement system and a non-postponement system under four different circumstances, respectively. By comparing the optimal total average costs of the two systems, we evaluate the impact of postponement on the manufacturer. Our results show that the postponement strategy can yield a lower total average cost under certain circumstances. We also find that the key factors in postponement decisions are the variance of the machine utilization rates and the variance of the backorder costs. The results presented in this paper provide insights for managers to find a tradeoff between postponement and non-postponement.

The rest of this paper is organized as follows. In the next section the EPQ model with planned backorders and some assumptions are explained in detail. In Sections 3 and 4, four cases for analyzing postponement based on the total average cost per unit time are discussed. Numerical examples are presented in Section 5. Conclusions are given in Section 6.

2 Notation and Assumptions

Consider a supply chain with a manufacturer and n customers. The manufacturer produces n different products in response to the demands of the n customers. These products are manufactured from the same type of raw material and the end products have only slight differences. These products are independent without any supply-demand links between them. The customers demand rates and the manufacturers production rates are deterministic and constant. All the demands are ultimately filled, although perhaps after some delay. That is, demands not filled immediately are backordered. The manufacturer can produce the n products independently on n machines under different production schedules such that there are n EPQ decisions. It is viewed as a non-postponement system. However, if the customization process can be delayed, the manufacturer can first produce a generic product. Then the production of the generic product can be carried out under the same production schedule such that there is only one EPQ decision. It is viewed as a form postponement system. Our objective is to apply the EPQ-based model with backorders to examine whether postponement outperforms non-postponement. There are two scenarios to describe the model. In the first scenario, we assume that the end-product demands are met continuously by the current production batch. In the second scenario, we assume that the end-product demands are only met after production is finished. In addition, we consider two cases of backorder costs. In the general case, we assume different backorder costs for different end-products. In the special case, we assume the backorder costs are the same for all the end-products. In sum, there are four cases to be discussed, and we investigate the following four hypotheses.

- H1. Postponement leads to a lower optimal total average cost per unit time for the manufacturer when the demand is met continuously and the planned backorder costs are not all equal.
- H2. Postponement leads to a lower optimal total average cost per unit time for the manufacturer when the demand is met continuously and the planned backo-

rder costs are the same for all the end-products.

H3. Postponement leads to a lower optimal total average cost per unit time for the manufacturer when the demand is met after production is finished and the planned backorder costs are not all equal.

H4. Postponement leads to a lower optimal total average cost per unit time for the manufacturer when the demand is met after production is finished and the planned backorder costs are the same for all the end-products.

Definitions of the notation of this paper are introduced below.

- i = end-product ($i = 1, 2, \cdots, n$),
- λ_i = demand rate of end-product $i, \lambda_i > 0$,
- μ_i = production rate of end-product $i, \mu_i > 0$,
- c = common unit variable production cost, $c > 0$,
- k = common fixed setup cost, $k > 0$,
- h = common unit holding cost per unit time, $h > 0$,
- p = common extra unit customization cost, $p \geq 0$,
- b_i = unit backorder cost per unit time for end-product $i, b_i \geq 0$,
- v_i = planned backorder quantity for end-product $i, v_i \leq 0$,
- L_i = total cycle time for end-product $i, L_i > 0$,
- L_i' = backorder lead-time for end-product $i, L_i' > 0$,
- q_i = production quantity for end-product $i, q_i > 0$,
- $C(q_i, v_i)$ = total average cost per unit time for producing and keeping end-product i with production quantity q_i and planned backorder quantity v_i, $C(q_i, v_i) > 0$,
- TC = total average cost per unit time for producing and keeping end-products $1, 2, \cdots, n$ in the non-postponement system with production quantities $q_1, q_2, \cdots, q_n$ respectively,
- TCP = total average cost per unit time for producing and keeping end-products $1, 2, \cdots, n$ in the postponement system with production quantity $q_1 + q_2 + \cdots + q_n$, $TCP > 0$,
- IP = inventory position over time.

In addition, the following assumptions are made:

A1. A production cycle means the time between the production of two consecutive batches. The end-product demand rate λ_i and the production rate μ_i are deterministic and constant. To avoid unrealistic and trivial cases, we assume $\mu_i > \lambda_i, i = 1, 2, \cdots, n$. When production starts, inventory accumulates until it is enough for the cycle. Then, production stops and inventory starts to decline. When inventory drops below zero, the product is backordered (Fig. 1-4).

A2. Demands not filled immediately are backordered and all the demands are ultimately filled. The manufacturer always uses any inventory on hand to fill the demands. Backorders accumulate only when the manufacturer runs out of stock entirely, which means that all the products will be backordered synchronously in the postponement system (Fig. 1-4).

A3. The inventory holding cost for raw materials is ignored. In the non-postponement system, we only consider the holding cost for the end products. In the postponement system, we only consider the holding cost for the generic products. Because the generic product and all the end-products have only slight differences, we assume that the holding cost for the generic product and all the end-products are the same.

A4. The manufacturer incurs a common setup cost for setting up a production run and an item-specific setup cost for each product. Because all the end-products have only slight differences, the item-specific setup cost is usually much less than the common setup cost, we can assume that all the item-specific setup costs are the same. For simplicity of analysis, we further assume that all the item-specific setup cost are zero and the fixed setup cost is the only common setup cost. The manufacturer incurs common fixed set-up cost, k, in each production cycle when production starts in the postponement system or in the non-postponement system.

A5. Because all the end-products have only slight differences, we assume that c and p are the same for all the end-products, respectively. Moreover, an extra customization process cost is incurred only if the customization process is delayed. In practice, the time for customization is very short. For example, an apparel manufacturer can postpone its color dyeing process at the very end of the production. The dyeing process can be finished quickly after the order are received. So the lead time of customization can be assumed to be negligible for simplicity of analysis.

In Section 3 and Section 4, we discuss the four hypotheses in detail by examining whether or not the postponement system is more cost-effective than the non-postponement system.

3 Demand is met continuously

We first assume that the customer demands are met continuously by the current production batch. For example, if the customers are downstream stations of the manufacturer in a production line of a factory, the customer demands may be met continuously. Alternatively, if the product transportation time from the manufacturer to the customers is short and the transportation frequency is high, we can also assume that the demand is met continuously.

3.1 Different backorder costs

First, we consider the general case in which the backorder costs are not all equal. A graph of the inventory position of end-product i over time in the non-postponement system is illustrated in Fig. 1. The horizontal axis t denotes time. The vertical axis IP denotes the inventory position of end-product i over time. Each cycle consists of an active period when production occurs and an idle period following production. In the active period, inventory increases with a slope of $\mu_i - \lambda_i$. In the idle period, inventory decreases with a slope of $-\lambda_i$. When production starts, the manufacturer incurs a fixed setup cost. When the inventory position is positive, there is inventory on hand and the manufacturer incurs holding cost. When the inventory is negative, the product is backordered and the manufacturer incurs backorder cost. The production quantity q_i and the planned backorder quantity v_i for end-product i in each cycle are our decision variables, which also determine the production cycle time L_i and the backorder lead-time L_i'. The objective of the EPQ model is to find the optimal q_i^* and v_i^* to minimize the average cost per unit time for producing and keeping end-product i.

The average cost per unit time for producing and keeping end-product i is obtained, as follows

$$C(q_i, v_i)$$

$$= c\lambda_i + \frac{k\lambda_i}{q_i} + \frac{h\left(\rho_i q_i + v_i\right)^2}{2\rho_i q_i} + \frac{b_i v_i^2}{2\rho_i q_i}, \tag{1}$$

where $\rho_i = 1 - \dfrac{\lambda_i}{\mu_i}$.

The first two terms are the variable production cost and the fixed production cost, the third term is the average inventory holding cost, while the last term is the average backorder cost.

Minimizing Eq. (1), we obtain the EPQ and the optimal backorder quantity as follows

$$q_i^* = \sqrt{\frac{2k\lambda_i(b_i + h)}{h\rho_i b_i}}, \ v_i^* = \frac{-\rho_i h}{b_i + h} q_i^*.$$

The optimal average cost per unit time for producing and keeping end-product i is

$$C\left(q_i^*, v_i^*\right) = c\lambda_i + \sqrt{\frac{2kh\lambda_i b_i \rho_i}{b_i + h}}. \tag{2}$$

In the non-postponement system, the production of the end-products is processed independently by n different machines with production rate μ_i, on which the end-products are customized. The optimal total average cost for producing and keeping the n end-products is the sum of all the costs of products i and is given by

$$TC^* = \sum_{i=1}^{n} C\left(v_i^*, q_i^*\right)$$

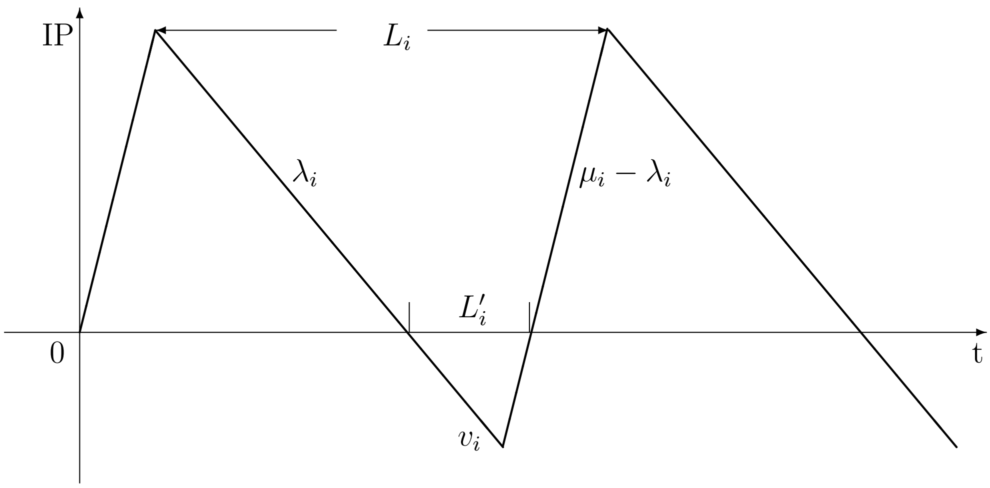

Fig. 1. Inventory position of end-product i over time in the non-postponement system when the demand is met continuously

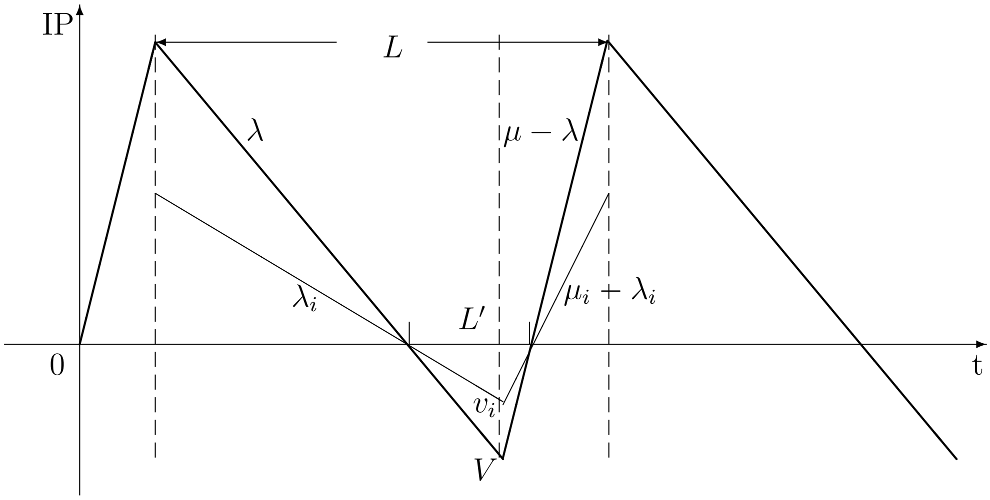

Fig. 2. Inventory position of the generic product over time in the postponement system when the demand is met continuously

$$= c\sum_{i=1}^{n}\lambda_i + \sum_{i=1}^{n}\sqrt{\frac{2kh\lambda_i b_i \rho_i}{b_i + h}}. \quad (3)$$

In the postponement system, the customization process is delayed. According to Assumptions A3, A4 and A5, the production of the generic product can be viewed as being processed by a single machine whose production rate is $\mu = \mu_1 + \mu_2 + \cdots + \mu_n$. The unit variable production cost is c, the fixed set-up cost is k, the unit holding cost per unit time is h, and the extra unit customization cost is p. A graph of the inventory position of the generic product over time in the postponement system is illustrated in Fig. 2. In this figure the thick lines denote the inventory position of the total generic product and the thin lines denote the inventory position of the generic product used to produce product i.

According to Assumptions A2, the production cycle of the total generic product and that of the generic product used to produce product i are the same and so is the backorder leadtime. There is only one EPQ problem in this system. The decision variables are the production quantity Q and the planned backorder quantity V of the generic product in one production cycle, which also determine the cycle time L and the backorder lead time L'. The objective of the EPQ model is to find the optimal Q^* and V^* to minimize the total average cost per unit time in the postponement system. For simplicity, we do not consider the extra unit customization cost p, which will be discussed later.

In the postponement system the total average cost per unit time (excluding the customization cost) for producing and keeping the genetic product is

$$
\begin{aligned}
&TCP \\
&= c\lambda + \frac{k\lambda}{Q} + \frac{h\left(\rho Q + V\right)^2}{2\rho Q} + \frac{\hat{B}V^2}{2\rho Q} \\
&= C\left(Q, V\right), \qquad (4)
\end{aligned}
$$

where $\hat{B} = \sum_{i=1}^{n} \frac{b_i \lambda_i}{\lambda}$.

The detailed derivation of Eq. (4) is given in the Appendix.

Minimizing Eq. (4), we obtain the EPQ and optimal backorder quantity, respectively, as follows

$$
Q^* = \sqrt{\frac{2k\lambda(\hat{B} + h)}{h\rho\hat{B}}},\; V^* = \frac{-\rho h}{\hat{B} + h} Q^*.
$$

The optimal total average cost per unit time for producing and keeping the generic product is

$$
TCP^* = c\lambda + \sqrt{\frac{2kh\lambda\hat{B}\rho}{\hat{B} + h}}. \qquad (5)
$$

The difference in the optimal total average cost per unit time between the two systems is defined as Z^*, given by (5) − (3), as follows:

$$
\begin{aligned}
Z^* = \frac{2kh}{A} \Bigg[&\sum_{i=1}^{n} \lambda_i \left(\frac{\rho \hat{B}}{\hat{B} + h} - \frac{\rho_i b_i}{b_i + h} \right) \\
&-2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \sqrt{\frac{\lambda_i \lambda_j b_i b_j \rho_i \rho_j}{\left(b_i + h\right)\left(b_j + h\right)}} \Bigg], \qquad (6)
\end{aligned}
$$

where

$$
A = \sqrt{\frac{2kh\lambda\hat{B}\rho}{\hat{B} + h}} + \sum_{i=1}^{n} \sqrt{\frac{2kh\lambda_i b_i \rho_i}{b_i + h}}.
$$

It should be noted that the term $\sum_{i=1}^{n} \lambda_i \left(\frac{\rho \hat{B}}{\hat{B}+h} - \frac{\rho_i b_i}{b_i+h} \right)$ and (6) can be positive. For example, when $i = 2$, $h = 2$, $k = 50$, $c = 20$, $b_1 = 50$, $\lambda_1 = 990$, $\mu_1 = 1000$, $b_2 = 500$, $\lambda_2 = 1$, and $\mu_2 = 1000$, the difference in the optimal total average cost (6) $=$ $206.5 > 0$ and $(6)/(3) = 1.03\% > 0$. Therefore, (6) can be positive, zero or negative (Table 1). It implies that the postponement system does not always give a lower optimal total average cost per unit time. Thus, H1 is not supported.

3.2 Same backorder costs

Now we consider a special case in which the backorder cost b_i is the same for all the end-products. Letting $b_1 = b_2 = \cdots = b_n = b$ in Eqs. (3), (5) and (6), we obtain the following results.

The optimal total average cost of the non-postponement system is

$$TC^* = c\sum_{i=1}^{n}\lambda_i + \sum_{i=1}^{n}\sqrt{\frac{2kh\lambda_i b\rho_i}{b+h}}. \quad (7)$$

The optimal total average cost (excluding the customization cost) in the postponement system is

$$TCP^* = c\lambda + \sqrt{\frac{2khb\lambda\rho}{b+h}}. \quad (8)$$

The difference in the optimal total average cost per unit time of the two systems, Z^*, is given by (8) – (7), as follows:

$$Z^* = \frac{\sqrt{\frac{2khb}{b+h}}}{B}\left(\sum_{i=1}^{n}\frac{\lambda_i^2}{\mu_i} - \frac{\lambda^2}{\mu} - 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sqrt{\lambda_i\lambda_j\rho_i\rho_j}\right), \quad (9)$$

where $B = \sqrt{\rho\lambda} + \sum_{i=1}^{n}\sqrt{\rho_i\lambda_i}$.

By the Cauchy-Schwarz inequality, we have $\sum_{i=1}^{n}\frac{\lambda_i^2}{\mu_i} - \frac{\lambda^2}{\mu} \geq 0$ and $\sum_{i=1}^{n}\frac{\lambda_i^2}{\mu_i} - \frac{\lambda^2}{\mu} = 0$ if and only if $\frac{\lambda_1}{\mu_1} = \frac{\lambda_2}{\mu_2} = \cdots \frac{\lambda_n}{\mu_n}$. It should be noted that Eq. (9) can be positive. For example, when $i = 2$, $h = 2$, $k = 50$, $c = 20$, $b = 100$, $\lambda_1 = 50$, $\mu_1 = 1000$, $\lambda_2 = 950$, and $\mu_2 = 1000$, the difference in the optimal total average costs (9) = 120.1 > 0, and (9)/(7) = 0.59% > 0. Thus, (9) can be positive, zero or negative (Table 2). H2 is not supported.

In the first scenario, H1 and H2 are not supported. The postponement system does not always give a lower optimal total average cost per unit time when the customer demands are continuously met by the current production batch. But we can observe that if $b_1 = b_2 = \cdots = b_n = b$ and $\frac{\lambda_1}{\mu_1} = \frac{\lambda_2}{\mu_2} = \cdots = \frac{\lambda_n}{\mu_n}$, then $Z^* < 0$. It implies that the variance of the backorder costs $b_1, b_2, \cdots, b_n$ and the variance of the machine utilization rates $\frac{\lambda_1}{\mu_1}, \frac{\lambda_2}{\mu_2}, \cdots, \frac{\lambda_n}{\mu_n}$ are key factors in a postponement decision. When there are a large number of end-products, we can group them into different product families based on machine utilization rates and backorder costs. Those products whose machine utilization rates and backorder costs are equal or close can share a single lot size and gain a lower total average cost.

4 Demands are met after production is finished

Now we assume that the product demands are met only after a whole production batch is finished. This scenario is more appropriate for describing the inventory level of the end-products that need to be moved to another warehouse in batches or to be further processed in batches, or for which instant consumption is not possible.

4.1 Different backorder costs

First, we consider the general case in which the backorder costs are not all equal. A graph of the inventory position of end-product i over time in the non-postponement system is illustrated in Fig. 3. In one production cycle, there are two lines. One increases with a slope of μ_i, which denotes the inventory position over time when product i is being produced. The other line decreases with a slope of $-\lambda_i$, which denotes the inventory position over time when product i is being consumed. The production quantity q_i and the planned backorder quantity v_i for end-product i in each cycle are our decision variables, which also determine the production cycle time L_i and the backorder lead-time L_i'. The objective is to find the optimal q_i^* and v_i^* to minimize the average cost per unit time for producing and keeping end-product i.

The total average cost per unit time for producing and keeping end-product i is as follows

$$C(q_i, v_i) = c\lambda_i + \frac{k\lambda_i}{q_i} + \frac{b_i v_i^2}{2q_i} + \frac{h}{2}\left(\frac{q_i\lambda_i}{\mu_i} + \frac{(q_i+v_i)^2}{q_i}\right). \quad (10)$$

Minimizing Eq. (10), we obtain the EPQ and the optimal backorder quantity, respectively, as follows:

$$q_i^* = \sqrt{\frac{2k\lambda_i}{h\left(\frac{b_i}{b_i+h} + \frac{\lambda_i}{\mu_i}\right)}}, \; v_i^* = \frac{-b_i}{b_i+h} q_i^*.$$

The optimal total average cost per unit time for producing and keeping n end-products in the non-postponement system is given by

$$\begin{aligned} TC^* &= \sum_{i=1}^{n} C(q_i^*, v_i^*) \\ &= c\lambda + \sum_{i=1}^{n} \sqrt{2kh\lambda_i\left(\frac{b_i}{b_i+h} + \frac{\lambda_i}{\mu_i}\right)}. \end{aligned} \quad (11)$$

Similarly, in the postponement system we assume that the core production is carried out by a single machine whose production rate is $\mu = \mu_1 + \mu_2 + \cdots + \mu_n$, the unit variable production cost is c, the fixed setup cost is k, the unit holding cost per unit time is h, and the extra unit customization cost is p. A graph of the inventory position of the generic product over time in the postponement system is illustrated in Fig. 4. The production quantity Q and the planned backorder quantity V for generic product in each cycle are our decision variables, which also determine the production cycle time L and the backorder lead-time L'. The objective is to find the optimal Q^* and V^* to minimize the average cost per unit time for producing and keeping generic product in the postponement system. For simplicity, we

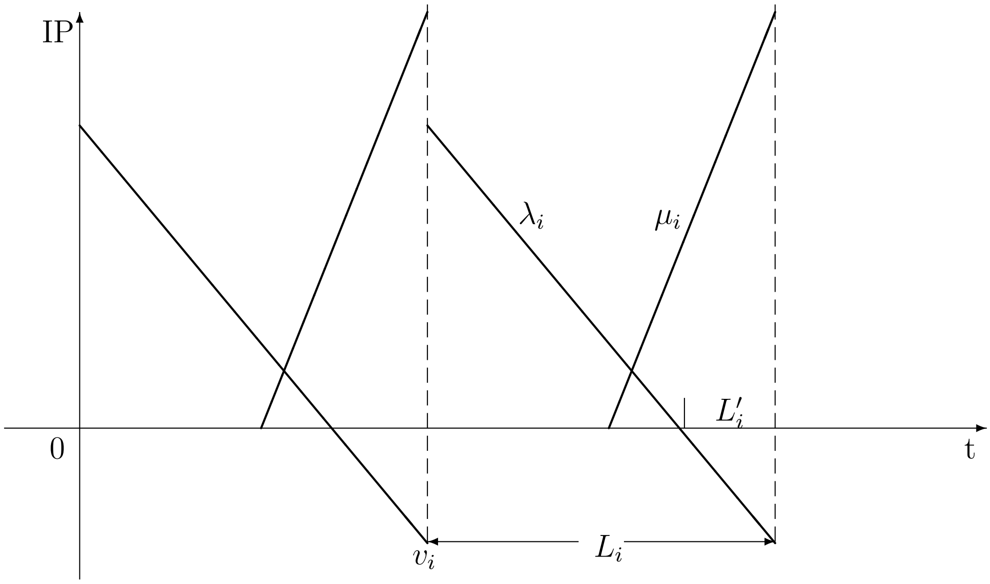

Fig. 3. Inventory position of end-product i over time in the non-postponement system when the demand is met after production is finished

again do not consider the extra unit customization cost p, which will be discussed later.

Similar to the derivation of Eq. (4) given in the Appendix, we can obtain the total average cost (excluding the customization cost) for producing and keeping the generic product in the postponement system when the demand is met after production is finished as follows

$$
\begin{aligned}
&TCP \\
&= c\lambda + \frac{k\lambda}{Q} + \frac{hQ\lambda}{2\mu} + \frac{h\left(Q+V\right)^2}{2Q} \\
&\quad + \frac{\hat{B}V^2}{2Q} \\
&= C\left(Q,V\right), \qquad (12)
\end{aligned}
$$

where $\hat{B} = \sum_{i=1}^{n} \frac{b_i\lambda_i}{\lambda}$.

Minimizing (12), we obtain the EPQ and the optimal backorder quantity, respectively, as follows

$$
Q^* = \sqrt{\frac{2k\lambda}{h\left(\frac{\hat{B}}{\hat{B}+h} + \frac{\lambda}{\mu}\right)}}, V^* = \frac{-\hat{B}}{\hat{B}+h}Q^*.
$$

The optimal total average cost per unit time in the postponement system is

$$
\begin{aligned}
&TCP^* \\
&= c\lambda + \sqrt{2kh\lambda\left(\frac{\hat{B}}{\hat{B}+h} + \frac{\lambda}{\mu}\right)}. \qquad (13)
\end{aligned}
$$

The difference in the optimal total average cost per unit time between the two systems is defined as Z^*, given by (13) $-$ (11), as follows

$$
\begin{aligned}
Z^* = \frac{2kh}{D}\Bigg[&\sum_{i=1}^{n} \frac{\lambda_i h\left(\hat{B}-b_i\right)}{\left(\hat{B}+h\right)\left(b_i+h\right)} \\
&+ \left(\frac{\lambda^2}{\mu} - \sum_{i=1}^{n}\frac{\lambda_i^2}{\mu_i}\right) \\
&- 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sqrt{\lambda_i\lambda_j\left(\frac{\lambda_i}{\mu_i} + \frac{b_i}{b_i+h}\right)}
\end{aligned}
$$

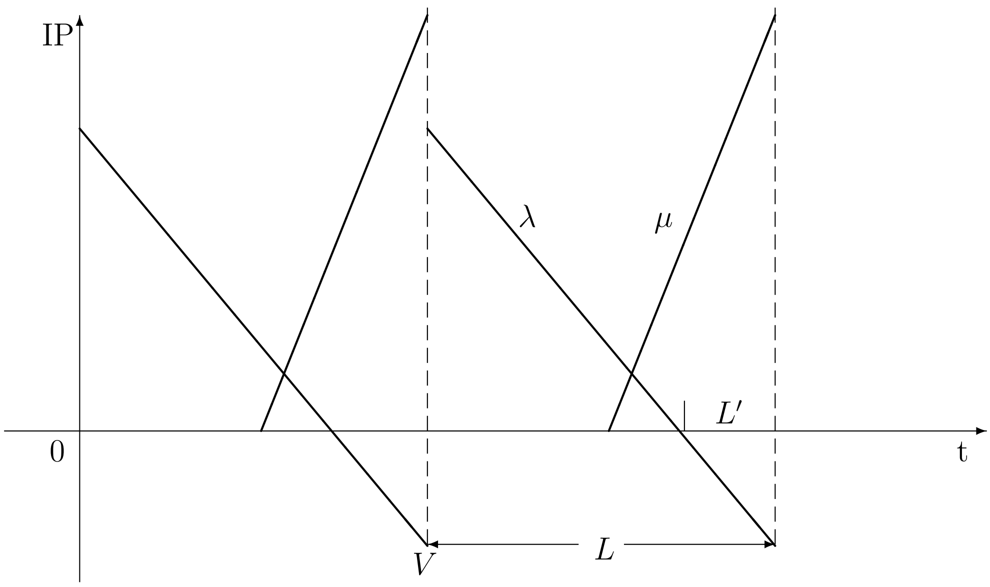

Fig. 4. Inventory position of the generic product over time in the postponement system when the demand is met after production is finished

$$\times\sqrt{\left(\frac{\lambda_j}{\mu_j}+\frac{b_j}{b_j+h}\right)}\Bigg], \qquad (14)$$

where

$$D=\sqrt{2kh\lambda(\frac{\hat{B}}{\hat{B}+h}+\frac{\lambda}{\mu})}$$
$$+\sum_{i=1}^{n}\sqrt{2kh\lambda_i(\frac{b_i}{b_i+h}+\frac{\lambda_i}{\mu_i})}.$$

It should be noted that the term $\sum_{i=1}^{n}\frac{\lambda_i h\left(\hat{B}-b_i\right)}{\left(\hat{B}+h\right)(b_i+h)}$ and Eq. (14) can be positive. For example, when $i=2$, $h=1$, $k=1$, $c=2$, $b_1=1$, $\lambda_1=1000$, $\mu_1=1200$, $b_2=1000$, $\lambda_2=10$, $\mu_2=12$, $(14)=1.7496>0$, and $(14)/(11)=0.0084\%>0$. Thus, (14) can be positive, zero or negative (Table 3). Thus, H3 is not supported.

4.2 Same backorder costs

If $b_1=b_2=\cdots=b_n=b$, according to Eq. (14), the difference in the optimal total average cost per unit time of the two systems is given by

$$Z^*=-\frac{\sqrt{2kh}}{E}\left[\sum_{i=1}^{n}\frac{\lambda_i^2}{\mu_i}-\frac{\lambda^2}{\mu}\right.$$
$$+2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sqrt{\lambda_i\lambda_j}$$
$$\left.\times\sqrt{\left(\frac{\lambda_i}{\lambda_i}+\frac{b}{b+h}\right)\left(\frac{\lambda_j}{\lambda_j}+\frac{b}{b+h}\right)}\right]$$
$$<0,$$

where

$$E=\sqrt{\lambda\left(\frac{b}{b+h}+\frac{\lambda}{\mu}\right)}$$
$$+\sum_{i=1}^{n}\sqrt{\lambda_i\left(\frac{b}{b+h}+\frac{\lambda_i}{\mu_i}\right)}.$$

We have shown that the postponement system always gives a lower optimal total average cost per unit time than the non-postponement system when $b_1=b_2=\cdots=b_n=b$ and the demand is met after production is finished. Thus, H4 is supported.

In the second scenario H3 is not supported, but H4 is supported. It implies that the variance of the backorder costs $b_1, b_2, \cdots, b_n$ is a key factor in a postponement decision when the demand is met after production is finished. If there are a large number of end-products, we can group them into different product families based on their backorder costs. Those products whose backorder costs are equal or close can share a single lot size and gain a lower total average cost.

5 Numerical Examples

We give numerical examples to illustrate how postponement and the key factors impact on the optimal total average cost of the two scenarios we have presented in this paper. We assume that the manufacturer produces five end-products. They can be produced in non-postponement system or in postponement system. The difference in the optimal total average cost per unit time between the two systems is denoted as Z^*. $Z^* < 0$ means that the postponement system outperforms the non-postponement system. $\frac{Z^*}{TC^*}$ denotes the relative difference between the two systems. For the five products, the unit common variable production cost c is 20, the common fixed setup cost k is 50, and the common unit holding cost h per unit time is 2 (all in appropriate units).

For the first scenario in which the demand is met continuously, we first assume that $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = \lambda_0 = 250$, $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_0 = 500$ and that b_i is variable to examine the impact of the backorder costs. The values of various parameters and the results are shown in Table 1, from which the following observations can be made.

- The postponement system yields savings in the total average cost.
- The absolute values of Z^* and $\frac{Z^*}{TC^*}$ become smaller when the variance of the backorder costs becomes larger. This means that the smaller the variance of the backorder costs, the more cost-effective the postponement system is when the demand is met continuously.

For the first scenario, we then assume that $b_1 = b_2 = b_3 = b_4 = b_5 = b = 100$, $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_0 = 500$ and that λ_i is variable to examine the impact of the machine utilization rates. The values of various parameters and the results are shown in Table 2, from which the following observations can be made.

- The postponement system yields savings in the total average cost.
- The absolute values of Z^* and $\frac{Z^*}{TC^*}$ become smaller when the variance of the machine utilization rates becomes larger. This means that the smaller the variance of the machine utilization rates, the more cost-effective the postponement system is when the demand is met continuously.

For the second scenario in which the demand is met after the produc-

Table 1
Impact of backorder costs on the difference in the optimal total average costs between the two systems when the demand is met continuously

c	k	h	λ_0	μ_0	b_1	b_2	b_3	b_4	b_5	Z^*	$\frac{Z^*}{TC^*}$
20	50	2	250	500	100	100	100	100	100	-432.7	-1.68%
20	50	2	250	500	80	90	100	110	120	-421.5	-1.68%
20	50	2	250	500	60	80	100	120	140	-432.0	-1.68%
20	50	2	250	500	40	70	100	130	160	-430.7	-1.67%
20	50	2	250	500	20	60	100	140	180	-427.0	-1.66%

Table 2
Impact of machine utilization rates on the difference in the optimal total average costs between the two systems when demand is met continuously

c	k	h	b	λ_1	λ_2	λ_3	λ_4	λ_5	μ_0	Z^*	$\frac{Z^*}{TC^*}$
20	50	2	100	250	250	250	250	250	500	-432.7	-1.68%
20	50	2	100	190	220	250	280	310	500	-421.3	-1.63%
20	50	2	100	130	190	250	310	370	500	-385.1	-1.50%
20	50	2	100	70	160	250	340	430	500	-315.9	-1.23%
20	50	2	100	10	130	250	370	490	500	-168.8	-0.66%

tion is finished, we first assume that $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = \lambda_0 = 250$, $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_0 = 500$ and that b_i is variable to examine the impact of the backorder costs. The values of various parameters and the results are shown in Table 3, from which the following observations can be made.

- The postponement system yields savings in the total average cost.
- The absolute values of Z^* and $\frac{Z^*}{TC^*}$ become smaller when the variance of the backorder costs becomes larger. This means that the smaller the variance of the backorder costs, the more cost-effective the postponement system is when the demand is met after production is finished.

For the second scenario, we then assume that $b_1 = b_2 = b_3 = b_4 = b_5 = b = 100$, $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_0 = 500$, and that λ_i is variable to examine the impact of the machine utilization rates. The values of various parameters and the results are shown in Table 4, from which the following observation can be made.

- The postponement system yields savings in the total average cost.
- The absolute values of Z^* and $\frac{Z^*}{TC^*}$ become smaller when the variance of the machine utilization rate becomes larger. This means that the smaller the variance of the machine utilization rates, the more cost-effective the postponement system is when the demand is met after production is finished.

Table 3
Impact of backorder costs on the difference in the optimal average costs between the two systems when the demand is met after production is finished

c	k	h	λ_0	μ_0	b_1	b_2	b_3	b_4	b_5	Z^*	$\frac{Z^*}{TC^*}$
20	50	2	250	500	100	100	100	100	100	-752.0	-2.85%
20	50	2	250	500	80	90	100	110	120	-751.7	-2.85%
20	50	2	250	500	60	80	100	120	140	-751.1	-2.85%
20	50	2	250	500	40	70	100	130	160	-749.7	-2.84%
20	50	2	250	500	20	60	100	140	180	-745.4	-2.83%

Table 4
Impact of machine utilization rates on the difference in the optimal total average costs between the two systems when the demand is met after production is finished

c	k	h	b	λ_1	λ_2	λ_3	λ_4	λ_5	μ_0	Z^*	$\frac{Z^*}{TC^*}$
20	50	2	100	250	250	250	250	250	500	-752.0	-2.85%
20	50	2	100	190	220	250	280	310	500	-749.8	-2.84%
20	50	2	100	130	190	250	310	370	500	-742.5	-2.82%
20	50	2	100	70	160	250	340	430	500	-726.8	-2.76%
20	50	2	100	10	130	250	370	490	500	-683.1	-2.60%

From the four tables and the numerical examples in Section 3 and Section 4, we can derive the following results:

1. Under most of the circumstances, the postponement system yields savings in the total average cost. But when the variance of the backorder costs is very large, or the variance of the machine utilization rates is very large in the case that the demand is met continuously, it is possible that the postponement system does not outperform the non-postponement system. So the manufacturer must be careful to find a proper tradeoff between postponement and non-postponement in such cases.
2. The smaller the variance of the backorder costs and the variance of the machine utilization rates, the more cost-effective the postponement system is in the two scenarios.
3. The cost saving in the second scenario is more than that in the first scenario. This means that it is more appropriate to apply postponement when the demand is met after production is finished.

6 Concluding Remarks

This paper examines the impact of postponement on EPQ-based models with planned backorders. Four cases are considered. In the first and the second cases, the demand is met instantly by the current production batch. The postponement system may not always outperform the non-postponement system. H1 and H2 are not supported. But when all the backorder costs are equal and all the machine utilization rates are equal, the postponement system outperforms the non-postponement system. The key factors in a postponement decision are the variance of the backorder costs and the variance of the machine utilization rates. In the third and the fourth cases, the demand is met after production is finished. H3 is not supported but H4 is supported. The key factor in a postponement decision is only the variance of the backorder costs. If the backorder costs are equal, the postponement system outperforms the non-postponement system. Our analysis and numerical examples imply that in most of the cases, the postponement system outperforms the non-postponement system. The smaller the variance of the backorder costs and the variance of the machine utilization rates, the more cost-effective the postponement system is. The end-products can be classified into different product families based on their machine utilization rates and backorder costs. The production of those products whose machine utilization rates and backorder costs are equal or close can be handled by postponement. A lower total average cost can be obtained.

Now we consider the extra customization cost in a postponement system. It is obvious that the average customization cost per unit time is $\left(\sum_{i=1}^{n} \lambda_i\right) p$. The difference in the optimal total average costs per unit time between the two systems is $Z^* + \left(\sum_{i=1}^{n} \lambda_i\right) p$. Postponement is more cost-effective if $Z^* + \left(\sum_{i=1}^{n} \lambda_i\right) p < 0$.

Although a number of simplifying assumptions are made in our model, our analysis should still be valid for more general systems. One potential future research direction is to study the impact of postponement on the entire supply chain with deterministic customer demands.

Acknowledgement

We would like to thank the editor and three anonymous referees for their valuable and constructive comments, which have led to a significant improvement of the manuscript. This research was supported in part by The Hong Kong Polytechnic University under a grant from the *Area of Strategic Development in China Business Services* and the National Science Foundation of China.

Appendix.

Derivation of Eq. (4).

Because the manufacturer always uses any inventory on hand to fill demands and because backorders accumulate only when the manufacturer runs out of stock entirely, the backorder lead time of end-product i is the same ($L_1' = L_2' = \cdots = L_n' = L'$) and the core production rate for product i becomes μ_i'. From Fig. 2, we observe that they yield the following equations

$$\mu = \mu_1 + \mu_2 + \cdots + \mu_n$$

$$= \mu_1' + \mu_2' + \cdots + \mu_n',$$
$$L' = \frac{-V}{\lambda} + \frac{-V}{\mu - \lambda} = \frac{-V}{\lambda \rho},$$
$$L' = L_i' = \frac{-v_i}{\lambda_i} + \frac{-v_i}{\mu_i' - \lambda_i} = \frac{-v_i}{\lambda_i \rho_i'},$$

where $\rho_i' = 1 - \frac{\lambda_i}{\mu_i'}$.

Since $V = v_1 + v_2 + \cdots + v_n$, we have

$$L'\lambda\rho = -v_1 - v_2 - \cdots - v_n$$
$$= L'(\lambda_1\rho_1' + \lambda_2\rho_2' + \cdots + \lambda_n\rho_n'),$$
$$\lambda\rho = \lambda_1\rho_1' + \lambda_2\rho_2' + \cdots + \lambda_n\rho_n',$$
$$\frac{\lambda^2}{\mu} = \sum_{i=1}^{n} \frac{\lambda_i^2}{\mu_i'}.$$

Let $a_i = \frac{\lambda_i}{\sqrt{\mu_i'}}, b_i = \sqrt{\mu_i'}$. By the Cauchy-Schwarz inequality

$$\left(\sum_{i=1}^{n} a_i b_i\right)^2 \leq \left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} b_i^2\right),$$

we have

$$\left(\sum_{i=1}^{n} \lambda_i\right)^2 \leq \left(\sum_{i=1}^{n} \frac{\lambda_i^2}{\mu_i'}\right)\left(\sum_{i=1}^{n} \mu_i'\right).$$

Therefore,

$$\sum_{i=1}^{n} \frac{\lambda_i^2}{\mu_i'} - \frac{\lambda^2}{\mu} \geq 0,$$
$$\sum_{i=1}^{n} \frac{\lambda_i^2}{\mu_i'} - \frac{\lambda^2}{\mu} = 0$$
$$\Longleftrightarrow \frac{\lambda_1}{\mu_1'} = \frac{\lambda_2}{\mu_2'} = \cdots = \frac{\lambda_n}{\mu_n'}.$$

So we have

$$\rho = \rho_1' = \rho_2' = \cdots = \rho_n'.$$

In the postponement system, the total average cost per unit time (excluding the customization cost) for producing and keeping the genetic product is

$$TCP$$
$$= c\lambda + \frac{k\lambda}{Q} + \frac{h(\rho Q + V)^2}{2\rho Q} + \frac{\hat{B}V^2}{2\rho Q}$$
$$= C(Q, V),$$

where $\hat{B} = \sum_{i=1}^{n} \frac{b_i \lambda_i}{\lambda}$. $\square$

References

[1] Alderson W. Marketing efficiency and the principle of postponement. Cost and Profit Outlook 1950;3:15-18.

[2] Aviv Y, Federgruen A. The benefit of design for postponement. In: Tayur S, Ganesham R, Magazine M. editors. Quantitative Models for Supply Chain Management. Boston/Dordrecht/London: Kluwer Academic Publishers, 1998.

[3] Aviv Y, Federgruen A. Capacitated multi-item inventory systems with random and seasonally fluctuating demands: implications for postponement strategies. Management Science 2001;47(4):512-531.

[4] Ben-Daya M, Hariga M. Comparative study of heuristics for the joint replenishment problem. OMEGA 1995;23(3):341-344.

[5] Brown AO, Lee HL, Petrakian R. Xilinx improves its semiconductor supply chain using prod-

uct and process postponement. Interfaces 2000;30(4):65-80.

[6] Cardenas-Barron LE. The economic production quantity (EPQ) with shortage derived algebraically. International Journal of Production Economics 2001;70(3):289-292.

[7] Christopher M. The agile supply chain: competing in volatile markets. Industrial Marketing Management 2000;29:(1):37-44.

[8] Eilon S. Multi-product batch production on a single machine–A problem revisited. OMEGA 1985;13(5):453-468.

[9] Ernst R, Kamrad B. Evaluation of supply chain structures through modularization and postponement. European Journal of Operational Research 2000;124:495-510.

[10] Feitzinger E, Lee HL. Mass customization at Hewlett-Packard: the power of postponement, Harvard Business Review 1997;January-February:116-121.

[11] Garg A, Lee HL. Managing product variety: an operations perspective. In: Tayur S, Ganesham R, Magazine M. editors. Quantitative Models for Supply Chain Management. Boston/Dordrecht/London: Kluwer Academic Publishers, 1998.

[12] Garg A, Tang CS. On postponement structures for product families with multiple points of differentiation. IIE Transactions 1997;29(8):641-650.

[13] Goyal SK. Determination of optimum packaging frequency of items joint replenished. Management Science 1974;21(4):436C443.

[14] Goyal SK. Determination of economic production quantities for a two-product single machine system. Internal Jouranl of Production Research 1984;22(1):121-126.

[15] Goyal SK, Satir AT. Joint replenishment inventory control: deterministic and stochastic models. European Journal of Operational Research 1989;38(1)2-13.

[16] Lee HL. Effective inventory and service management through product and process redesign. Operations Research 1996;44(1):151-159.

[17] Lee HL. Postponement for mass customization: satisfying customer demands for tailor-made products. In: Gattorna J, editor. Strategy Supply Chain Alignment. England: Gower, 1998.

[18] Lee HL, Tang C. Modeling the costs and benefits of delayed product differentiation. Management Science 1997;43(1):40-53.

[19] Li S, Ragu-Nathan B, Ragu-Nathan TS, Subba Rao S. The impact of supply chain management practices on competitive advantage and organizational performance. OMEGA 2006;34(2):107-124.

[20] Ma S, Wang W, Liu L. Commonality and postponement in multistage assembly systems. European Journal of Operational Research 2002;142(3):523-538.

[21] Nocture DJ. Economic ordering frequency for several items jointly replenished. Management Science 1973;19(9):1093-1096.

[22] Ronald R, Yang GK, Chu P.

Technical note: The EOQ and EPQ models with shortages derived without derivatives. International Journal of Production Economics 2004;92(2):197-200.
[23] Shu FT. Economic ordering frequency for two items jointly replenished. Management Science 1971;17(6):B406-B410.
[24] Silver EA. A simple method of determining order quantities in joint replishment under deterministic demand. Management Science 1976;22(12):1381-1361.
[25] Su JCP, Chang YL, Ferguson M. Evaluation of postponement structures to accommodate mass customization. Journal of Operations Management 2005;23(3&4):305-318.
[26] van Hoek RI. The rediscovery of postponement: a literature review and directions for research. Journal of Operations Management 2001;19:161-184.
[27] Wan JCL, Ng CT, Cheng TCE. Postponement: a review and future research directions. 2003a; Working paper, The Hong Kong Polytechnic University, Hong Kong.
[28] Wan JCL, Ng CT, Cheng TCE. An analysis of pull postponement by EOQ, 2003b; Working paper, The Hong Kong Polytechnic University, Hong Kong.
[29] Wan JCL, Ng CT, Cheng TCE. An analysis of pull postponement by an EPQ-based model, 2004; Working paper, The Hong Kong Polytechnic University, Hong Kong.
[30] Zipkin PH. Foundations of Inventory Management. Boston: McGraw-Hill, 2000.