Analysis of form postponement based on optimal positioning of the differentiation point and stocking decisions

Hartanto Wong, Joakim Wikner, M M Naim

Analysis of form postponement based on optimal positioning of the differentiation point and stocking decisions

Journal:	*International Journal of Production Research*
Manuscript ID:	TPRS-2007-IJPR-0230
Manuscript Type:	Original Manuscript
Date Submitted by the Author:	28-Mar-2007
Complete List of Authors:	Wong, Hartanto; Cardiff University, Innovative Manufacturing Research Centre Wikner, Joakim; Jonkoping University, School of Engineering Naim, M M; Cardiff University, Cardiff Business School
Keywords:	POSTPONEMENT, MASS CUSTOMIZATION, QUEUEING MODELS, SUPPLY CHAIN MANAGEMENT
Keywords (user):	

Analysis of form postponement based on optimal positioning of the differentiation point and stocking decisions

Hartanto Wong[a], Joakim Wikner[c], Mohamed Naim[a,b]

[a]*Cardiff University, Innovative Manufacturing Research Centre, Cardiff CF10 3EU, Wales - UK*

[b]*Cardiff University, Logistics Systems Dynamics Groups, Cardiff CF10 3EU, Wales - UK*

[c]*Jönköping University,School of Engineering, Jönköping, S-551 11, Sweden*

Abstract: In this paper we analyse the use of form postponement based on the positioning of differentiation point and stocking policy. Six classes of manufacturing configurations are identified based on the choice of whether or not form postponement is employed and the decision regarding the stocking policy for the final product configurations as well as for the generic component. Analytical evaluation methods based on queuing models are used to assess operational measures for each class of configuration and solution algorithms are developed to determine the optimal positioning of differentiation point and the optimal stocking levels. This allows us to compare the relative merits of all manufacturing configurations based on their respective best performances. The results of numerical experiment show how different operational parameters may influence the choice of optimal configuration, the preference of early or late postponement, and the relative cost savings obtained from employing form postponement.

Keywords: Supply chain management; Inventory control; Postponement; Stochastic model

1. Introduction

It can be argued that the increasing pressure to become more and more customer-centric has forced manufacturing firms to continuously revise their supply chain structures so that they are able to provide an ever more valuable service to customers while at the same time cut delivery times and operating costs. This has led to an increasingly fast growing attention paid to the new manufacturing paradigm called *mass customization* replacing the conventional *mass production* which is no longer suitable for today's competitive environment. Mass customization allows customers to get tailor-made products reflecting their personal preference of styles, features, and colours with reasonable prices. For more than two decades mass customization has been perceived as the future of manufacturing and for some manufacturers it probably always will be (Agrawal, 2001).

An important concept used to accommodate mass customization that has been increasingly drawing attention from researchers and practitioners in recent years is *postponement* which has also been termed as *delayed product differentiation* or *late customization*. Postponement represents a way to implement

mass customization without incurring large operating costs associated with managing proliferating product variety. This is done by properly designing the product structure and the manufacturing and supply chain process so that one can delay the point in which the final customization of the product is to be configured (Swaminathan and Lee, 2003).

There is a large body of literature on postponement. We refer the readers to van Hoek (2001), Swaminathan and Lee (2003), and Yang and Burns (2003) for a comprehensive review of research on postponement. The concept of postponement was actually introduced in the literature by Alderson (1950) as a means of reducing marketing costs. He believed that risks related to marketing operations could be reduced by postponing changes in form and identity to the latest possible point in the marketing flow or postponing change in inventory location to the latest possible point in time. Over time, a number of authors have introduced different conceptual categorisations of postponement strategies extending the understanding of where and when postponement is appropriate. In the paper by Zinn and Bowersox (1988), five different types of postponement strategies are identified. Four different strategies of form postponement (labelling, packaging, assembly and manufacturing) which, when combined with time postponement, constitute the five postponement strategies. Bowersox and Closs (1996) made a clear differentiation between logistics postponement and form or manufacturing postponement. Logistics postponement can be seen as a combination of time and place postponement (where place postponement refers to the storage of goods at central locations in the channel until customer orders are received). Pagh and Cooper (1998) provided a classification of postponement applications in the mid- to down-stream stages of the supply chain. Their classification is in fact a reworked version of the classification suggested by Zinn and Bowersox (1988). They identified four generic strategies by combining manufacturing and logistics postponement and speculation. These include: the full speculation strategy, the logistics postponement strategy, the manufacturing postponement strategy, and the full postponement strategy. Despite their differences, all the conceptual classifications discussed above actually employ a common concept. That is, all agree in referring postponement to the delaying of certain operations related to either manufacturing or logistics until customer orders are received. In the case of form postponement for example, such a concept suggests that the final differentiation process would be performed in a make-to-order fashion. Ideally, this concept would maximise the profits of form postponement as it omits the inventory of the final products. However, it is obvious that in reality it may not always be possible to employ such a postponement strategy especially in the highly responsive environments where the tolerance time that the customer is willing to wait is quite short. In such environments it may be necessary to produce the final products in a make-to-stock fashion. A classic example of form postponement application in which finished-goods inventory for each distinct product are held at the product's

respective point of customization is Hewlett-Packard's (HP) postponement for their DeskJet printers (Lee et al., 1993; Feitzinger and Lee, 1997). The company opted to customise the printers at its local distribution centres rather than at its factories. For example, instead of customising the DeskJet at its factory in Singapore before shipping them to Europe, HP has its European distribution centre near Stuttgart, Germany to perform this job. The distribution centre not only customises the product but also purchases the materials that differentiate it (power supplies, packaging, and manuals). Even though the localisation of the printers is postponed, the distribution centres still produced the localised printers in a make-to-stock fashion.

From that perspective, we argue that the concept referring postponement to the delaying of activities until customer orders are received does not always represent the best course of action. This motivated us to undertake this study looking at a more complete set of manufacturing configurations related to the implementation of postponement strategy. In particular, our primary focus is on the evaluation of form postponement in which we explicitly consider two inter-related decisions that need to be made. The first decision is dealing with the positioning of *differentiation point* (DP), which is the point where the final configuration of the product is to be processed. The second decision is dealing with the stocking policy for each final product as well as for the generic component. We may also relate the second decision with the so called Order Penetration Point (OPP) (Sharman, 1984). This is the stage in the manufacturing value chain where a particular product is linked to a specific customer order. Different manufacturing environments such as make-to-stock (MTS), make-to-order (MTO) and assemble-to-order (ATO) all relate to the different positions of the OPP.

While there are a significant number of papers in the existing literature, we are not aware of any previous work that systematically investigates postponement structures that involve and inter-relate these two types of decision. Clearly, such a study would be valuable to extend the understanding of how postponement should be implemented. It is our objective to make a contribution to this important line of inquiry. More specifically, the main goal of this research is to investigate different possible manufacturing configurations that are characterised based on the positioning of the differentiation point and stocking policy. We consider a stochastic capacitated production system as it is our interest to explicitly model the congestion effect on the system performance. We developed algorithms for determining the optimal DP and stocking policy allowing us to compare configurations based on their respective best performance.

We identify six different manufacturing configurations as presented in Fig.1. Depending on whether or not form postponement is employed, we distinguish two different systems, namely the single-stage system and the two-stage system. In the single-stage system where form postponement is not employed,

end products are processed and customised through a single-stage production. The DP position for this system can be considered being located in the beginning of the production process. This single-stage system is represented in the first two configurations (Figs 1a and 1b). These two configurations differ from each other with respect to the stocking policy employed. The first configuration depicted in Fig. 1a (MTS-1) produces the products in a MTS fashion. That is, products are produced ahead of demand and kept in stock. The OPP for this configuration is positioned at the right. In contrast, the second configuration in Fig. 1b (MTO-1) produces the products in a MTO fashion with the OPP being positioned at the left. From the modelling perspective, the MTO configuration can be seen as a special case of the MTS configuration with zero stock levels for all products.

The second system employing form postponement consists of two stages. Stage 1 produces the generic components and Stage 2 differentiates the final products. The DP in this system is located in the beginning of the second stage. The next four configurations represent the four variants of the two-stage system and they are different with respect to the stocking policy employed for the generic component and for the final products. As stated earlier, the system described by most of the existing conceptual studies in the postponement literature could be referred to the configuration shown in Fig. 1d in which the system stocks the generic component but the differentiation process is delayed until a customer order has been received. As the DP position, the OPP for this configuration is also positioned in the middle. This configuration can be seen as an assemble-to-order (ATO) system. Notice, however, that the ATO system here simply represents a system with an internal OPP and does not necessarily represent assembly operations. Alternatively, the MTS-2 configuration (Fig. 1c) with the OPP being positioned at the right should also be taken into account when looking for the best configuration. This configuration would be attractive in the situations where the system is required to be highly responsive so that it is no longer possible to process Stage 2 after receiving the order.

The next variant of the two-stage system is the MTS-3 system (Fig. 1e). Note that this configuration has the same OPP position as the MTS-2 configuration, which means that both are forecast driven. While in the MTS-2 configuration both the generic component and end products are made to stock, the MTS-3 configuration avoids keeping stock of a subassembly by producing the generic components in exactly the quantities required by the forecast of the end products. We show later in Section 5 that this configuration is particularly attractive when the product's value increases significantly at the beginning stages of production. The last variant is the MTO-2 configuration (Fig. 1f) in which no inventory is held for both the generic component and the finished products. This configuration is order driven and its OPP is positioned at the left.

Fig. 1 is about here

Several previous studies present analytical models measuring the costs and benefits of employing form postponement including e.g. Lee and Tang (1997), Garg and Tang (1997), Swaminathan and Tayur (1998), and Aviv and Federgruen (2001a and 2001b). In general, these models are different from ours in two respects. Firstly, these models do not explicitly contrast different postponement structures with respect to the decisions on the DP positioning and stocking policy for both the generic component and the final products as we do in this paper. Secondly, these models ignore the effect of congestion at the production facility while we explicitly model the queuing effect as a result of considering a capacitated production facility.

Analytical models related to the decision of manufacturing products in a MTS or MTO fashion are presented in e.g. Federgruen and Katalan (1995), Arreola-Risa and DeCroix (1998), and Rajagopalan (2002). Different from our work, they all studied the choice of MTS or MTO in a rather simple manufacturing system without considering any postponement structure. Conceptual models concerning different factors affecting the positioning of OPP were studied by e.g. Rudberg and Wikner (2004) and all the references therein.

There are two papers addressing problems more closely related to ours. Gupta and Benjaafar (2004) consider the capacitated production system and model the system employing form postponement as a two-stage system where a common product platform is produced in a MTS fashion in the first stage which is differentiated into different products in the second stage in a MTO fashion. Our work is different in that we allow a richer set of manufacturing configurations to be compared and systematically investigate postponement structures that inter-relate the DP and OPP positioning decisions. Su et al. (2005) compare two specific configurations. In the first configuration products are produced after orders arrive (MTO mode). The second configuration represents the system employing form postponement. Different from Gupta and Benjaafar, they examine the system where the second stage produces differentiated products in an MTS fashion instead of an MTO fashion. Our work differs from theirs in two ways. Firstly, as also compared to Gupta and Benjaafar, our model allows a richer set of configurations. Secondly, they do not deal with the optimization problem as we do in this paper. Their numerical results are therefore not based on the best policy within each configuration.

The rest of the paper is organised as follows. In Section 2, we describe the system's operation under study. We introduce the notation and some assumptions used in the model. In Section 3 we present the models used to assess all the relevant performance measures for all the system configurations. Section 4 presents the optimization problem formulations and the corresponding solution algorithms. In Section 5 we present and discuss our numerical findings. Finally, we summarise the results in Section 6 and conclude with directions for further research.

2. Problem description and notation

Consider a manufacturing firm that supplies a product family consisting of N different product configurations. The products are indexed by $i = 1, 2, \ldots, N$. End customer demand of product i arrives in single units according to a Poisson process with rate λ_i. We denote λ_0 as the aggregate demand rate where $\lambda_0 = \sum_{i=1}^{N} \lambda_i$.

For the single-stage system not employing form postponement, we assume that the total production lead times for all products are i.i.d. random variables and exponentially distributed with rate $1/\mu$. This helps to keep the analysis simple and represents the practical worst case for benchmarking production system performance [24]. For the two-stage system employing form postponement, the processing rates for Stage 1 and Stage 2 are defined as μ_1 and μ_2 respectively. We assume that $1/\mu_1+1/\mu_2=1/\mu$ and this applies to all products. For both systems there is a limited production capacity and the manufacturer processes items one-by-one using a single resource. To represent the position of DP, we define p ($0 \le p < 1$) as the fraction of the mean total processing time consumed by the generic component. Thus, we may write $1/\mu_1 = p \cdot 1/\mu$ and $1/\mu_2 = (1-p) \cdot 1/\mu$. Small p values represent early form postponement while large values represent late form postponement. Note that the single-stage system can actually be seen as a special case of the two-stage system with p being set to zero. But in our analysis we treat the two systems differently since it is our aim to assess the benefits of introducing form postponement by contrasting the merits of the two systems and moreover, the models used to analyse the two systems are also different. Further, because all products belong to the same product family, changeover times between products are assumed to be negligible.

We assume that a base-stock policy is used for the inventory control. Under this assumption, while in the single-stage system each demand triggers a manufacturing order of the requested product, in the two-stage system each demand triggers a manufacturing order of the requested product at Stage 2 and at the

same time a manufacturing order of the generic component at Stage1 (see Buzacott and Shathikumar, 1993 for a formal definition of a base-stock policy). We assume that raw materials are always available and can be immediately released for the manufacturing process. For all types of systems we assume that all shortages are backordered. All demands including backordered demands are served in a first-come first-served (FCFS) basis.

Let h_i denote the holding cost per unit per unit time for product i and $h_0(p)$ denote the holding cost per unit per unit time for the generic component as a function of the DP position. As the product is processed along a value chain, it is reasonable to assume that $h_0(p)$ is increasing in p. Without loss of generality, we assume that all end products have the same holding cost and that the holding cost is the same for both the single-stage system and the two-stage system.

To enable form postponement in the two-stage system, there may be a premium cost associated with the investment required for redesigning the product and/or the manufacturing processes. Lee and Tang (1997) observed three basic approaches that companies have used for the form postponement including standardisation, modular design, and process restructuring. The reader is also referred to Lee et al. (1993) for a detailed discussion on various cost drivers associated with form postponement. In our model we denote r as the amortised premium cost per period.

Further, for each product i, there is a maximum level $W_i^{\max}$ given for the expected order waiting time. In this paper we consider a service model rather than a cost model. In a service model, the objective is to minimize the total system cost subject to a set of service level constraints. In our case, the service level constraints are represented by the maximum expected waiting time constraints. Alternatively, one may also consider a cost model in which the service constraints are replaced with the penalty (backorder) cost. As quantifying the backorder cost is often difficult in real practice, we choose to use the service model. Van Houtum and Zijm (2000) present a systematic overview of possible relations between the cost and the service model for general inventory systems. We assume that all products have identical target waiting times so that it is reasonable to serve all demands in a FCFS basis.

Fig. 2 below summarises the notation used to model the single-stage and two-stage systems.

Fig. 2. is about here

The optimization problem we deal with is to determine the best system configuration that minimises the expected total cost subject to the maximum expected waiting time constraints for all the products. For the system with no form postponement, we need to decide on the stocking policy $\underline{S}^1 = (S_1, S_2, ..., S_N)$ minimising the expected total cost, which is the sum of the inventory holding costs of all the customised products. For the system with form postponement we need to decide on a policy $(p, \underline{S}^2)$ where p represents the DP position and $\underline{S}^2 = (S_0, S_1, S_2, ..., S_N)$ represents the stocking policy including the stock levels of the generic component and all the final product configurations. The expected total cost for this system is comprised of the expected total inventory holding cost and the premium cost associated with form postponement.

3. Evaluation models

In this section we present the models used to assess all the necessary performance measures for all the system configurations. The performance measures required for optimization purposes include: $\bar{I}_0$, the expected on-hand inventory level for the generic component (for the system with form postponement); $\bar{I}_i$, the expected on-hand inventory level for product i (i =1,…, N); and $\overline{W}_i$, the expected order waiting time for product i (i =1,…, N);

3.1 The single-stage system

The single-stage system can be considered as a multi-class, single server base-stock system in which the production system can be modelled as an $M/M/1$ queuing system. An MTO system can be seen as a special case of an MTS system with $S_i = 0$ for all i. Determining the performance measures in the MTO configuration is quite straightforward. Since the stock levels are zero, the expected on-hand inventory levels are zero for all products, i.e. $\bar{I}_i = 0$ for all i. By treating the system as a basic $M/M/1$ queuing system with arrival rate λ_0 and service rate μ, it can be shown that the expected waiting times for all the products are identical and can be expressed as

$$\overline{W}_i(0) = \frac{1}{\mu - \lambda_0} \ . \tag{1}$$

For the MTS configuration, the model is more complicated. Let us define $O_i(t)$ to be the outstanding orders for product *i* at time *t*. If S_i is the number of units stocked for product *i*, then $S_i - O_i(t)$ is the net inventory on hand at time *t*, where a negative value denotes a shortage level. We define $O_0(t)$ as the aggregate outstanding orders at time *t*, i.e. $O_0(t) = \sum_{i=1}^{N} O_i(t)$. As we assume that the outstanding orders are filled on a first-come first-serve basis, the apportioning of the aggregate outstanding orders $O_0(t)$ to product *i* is simply proportional to the individual demand rate λ_i. To determine the distribution of $O_i(t)$, we use the fact that, given the current aggregate outstanding orders is $O_0(t)$, the conditional distribution of $O_i(t)$ is a binomial distribution. Specifically, we obtain

$$\begin{aligned} P[O_i(t) = j] &= \sum_{k=j}^{\infty} P[O_0(t) = k] P[O_i(t) = j \mid O_0(t) = k] \\ &= \sum_{k=j}^{\infty} P[O_0(t) = k] \binom{k}{j} \left[\frac{\lambda_i}{\lambda_0}\right]^j \left[\frac{\lambda_0 - \lambda_i}{\lambda_0}\right]^{k-j} . \end{aligned} \tag{2}$$

Using a birth-death process, the steady state probability that there are *k* aggregate outstanding orders, π_k^0, can easily be determined from $\pi_k^0 = \rho^k (1-\rho)$, where $\rho = \lambda_0 / \mu$. Following from (2), the steady state probability that there are *j* outstanding orders of product *i*, π_j^i, is

$$\pi_j^i = \sum_{k=j}^{\infty} \pi_k^0 \binom{k}{j} \left[\frac{\lambda_i}{\lambda_0}\right]^j \left[\frac{\lambda_0 - \lambda_i}{\lambda_0}\right]^{k-j} . \tag{3}$$

Given the base stock level is S_i, the expected on hand inventory level for product *i* is

$$\overline{I}_i(S_i) = \sum_{j=0}^{S_i} (S_i - j) \pi_j^i \ , \tag{4}$$

and the expected backorder level for product *i* is

$$\overline{B}_i(S_i) = -\sum_{j=S_i+1}^{\infty} (S_i - j) \pi_j^i \ . \tag{5}$$

Using Little's formula, the expected order waiting time for product *i* is

$$\overline{W}_i(S_i) = \frac{\overline{B}_i(S_i)}{\lambda_i} = \frac{-\sum_{j=S_i+1}^{\infty}(S_i - j)\pi_j^i}{\lambda_i}. \quad (6)$$

3.2 The two-stage system

Recall that Stage 1 builds the generic component with rate μ_1, where $\mu_1 = \mu / p$ and Stage 2 differentiates the products with rate μ_2 where $\mu_2 = \mu/(1-p)$. We define $\rho_1 = \lambda_0/\mu_1$ and $\rho_2 = \lambda_0/\mu_2$. Compared to the single-stage system, the evaluation for the two-stage system considering form postponement is more difficult. We use an approximation scheme developed by Lee and Zipkin (1992) that treats each stage as an independent *M*/*M*/1 queuing system. Lee and Zipkin show that their approximate method is accurate to be used for determining optimal base-stock levels. This approximate method gives exact results only in the case where S_0 =0. In this case, the system behaves like two *M*/*M*/1 queues in tandem whose steady-state probabilities can also be obtained using Jackson's theorem. When $S_0 = \infty$ the two stages are completely decoupled and behave like two independent *M*/*M*/1 queuing systems. When $0 < S_0 < \infty$, the two stages are not completely decoupled i.e. there is a positive dependence between the two stages. That is, the differentiation process at the second stage can only take place when the on-hand inventory level for the generic component is positive. A delay would be incurred when the on-hand inventory level is zero.

The approximate method mentioned above works as follows. By treating stage 1 as a single-class single server base-stock system, it is easy to see that the expected waiting time in Stage 1 given p and S_0 are known, is

$$\overline{W}_0^1(p, S_0) = \frac{\rho_1^{S_0}}{\mu_1 - \lambda_0}. \quad (7)$$

The expected on-hand inventory level for the generic component is given by

$$\overline{I}_0(p, S_0) = S_0 - \left(\frac{\rho_1(1-\rho_1^{S_0})}{1-\rho_1}\right). \quad (8)$$

Stage 2 can be treated as a multi-class single server base stock system like in the single-stage system. The expected waiting time in Stage 2, $\overline{W}_i^2(p, S_i)$ which is dependent on p and S_i can be obtained in a similar way using (1)-(6) with μ being replaced by μ_2. Thus, the expected waiting time for product *i* is

$$\overline{W}_i(p, S_0, S_i) = \overline{W}_0^1(p, S_0) + \overline{W}_i^2(p, S_i). \quad (9)$$

The expected on-hand inventory level for the individual product i, $\bar{I}_i(p, S_0, S_i)$ can also be obtained in a similar way using (1)-(6).

4. Optimization

In this section we first present the formulation of the optimization problems for the single-stage system and the two-stage system and describe the solution algorithm to determine the optimal policy. Then we summarise the procedure to determine the best manufacturing configuration

4.1 The single-stage system

Let $Z(\underline{S}^1)$ denote the expected total cost per period that corresponds to the stocking policy $\underline{S}^1 = (S_1, S_2, ..., S_N)$ for the single-stage system. The optimization problem is formulated as follows:

Problem (**P1**):

Minimise $Z(\underline{S}^1) = \sum_{i=1}^{N} h_i \bar{I}_i(S_i)$

Subject to $\overline{W}_i(S_i) \le W_i^{\max}$ $\quad i = 1, 2, \ldots, N$

S_i: non-negative integer.

From (4) and (6), it is obvious that the choice of base-stock level for product i does not affect the performance measures for all the other products. Thus, Problem (**P1**) can be decomposed into N single-product sub-problems and the i-th sub-problem is to minimise $h_i \bar{I}_i(S_i)$ subject to $\overline{W}_i(S_i) \le W_i^{\max}$. From (1)-(6), it is easy to see that $\overline{W}_i(S_i)$ decreases but $\bar{I}_i(S_i)$ increases as S_i increases. This implies that the optimal base-stock level for product i, S_i^*, is the minimum S_i that meets the constraint $\overline{W}_i(S_i) \le W_i^{\max}$. This also means that if the MTO-1 configuration is able to meet the average waiting time constraints, it immediately becomes the best configuration since this system has a zero total cost.

4.2 The two-stage system

For the two-stage system let $Z(p, \underline{S}^2)$ denote the expected total cost per period that corresponds to the choice of p and the stocking policy $\underline{S}^2 = (S_0, S_1, S_2, ..., S_N)$. Recall that in this system we need to include the premium cost of form postponement in addition to the inventory holding cost. We formulate the optimization problem as follows:

Problem (**P2**):

Minimise $Z(p, \underline{S}^2) = h_0(p)\bar{I}_0(p, S_0) + \sum_{i=1}^{N} h_i \bar{I}_i(p, S_0, S_i) + r$

Subject to $\overline{W}_i(p, S_0, S_i) \le W_i^{\max}$ $\quad i = 1, 2, \ldots, N$

$0<p<1$; S_0 and S_i non-negative integers.

Since r is a constant, it can be removed when solving Problem (**P2**). However, it must be included when making the total cost comparison between the two-stage system and the single-stage system. Problem (**P2**) can be solved by applying a three-step method. The first step has the objective of determining the optimal stock level S_i given p and S_0 are fixed. In the second step we need to determine the optimal S_0 (and S_i) given p is fixed. Finally, the third step optimises p. We developed an optimization algorithm based on the following observations.

Observation 1*: Given p and S_0 are fixed, problem* (**P2**) *can be decomposed into N independent single-product sub-problems.*

Proof: From (7) to (9) it is obvious that, when p and S_0 have been chosen, the expected on-hand inventory level $\bar{I}_i$ and the average waiting time $\overline{W}_i$ for product i are influenced only by S_i and independent of S_j ($j \neq i$). In addition, the choice of S_i does not affect the expected on-hand inventory for the generic component as $\bar{I}_0$ depends only upon S_0 and ρ_1.

Observation 2*: For a fixed parameter p, (i)* $\bar{I}_0$ *is increasing and* $\overline{W}_0^1$ *is decreasing in* S_0*, and (ii)* $S_i^*(p, S_0)$ *is non-increasing in* S_0.

(i) From (7) it is straightforward to see that, as $\rho_1 < 0$, $\overline{W}_0^1$ will decrease if S_0 increases. The maximum value for $\overline{W}_0^1$ is obtained when $S_0 = 0$ and $\overline{W}_0^1 \to 0$ when $S_0 \to \infty$. Following from (8), $\bar{I}_0(p, S_0 + 1) - \bar{I}_0(p, S_0) = 1 + \frac{\rho_1}{1-\rho_1}\left(\rho_1^{S_0+1} - \rho_1^{S_0}\right)$, which can also be written as $\frac{1-\rho_1+\rho_1^{S_0+2}-\rho_1^{S_0+1}}{1-\rho_1}$. By rewriting the nominator as $(1-\rho_1) - \rho_1^{S_0+1}(1-\rho_1) = (1-\rho_1)(1-\rho_1^{S_0+1})$, we may write $\bar{I}_0(p, S_0 + 1) - \bar{I}_0(p, S_0) = 1 - \rho_1^{S_0+1} > 0$. This shows that increasing S_0 results in the increase of $\bar{I}_0$.

(ii) Since $\overline{W}_0^1(S_0 + 1) < \overline{W}_0^1(S_0)$ and $\overline{W}_i(p, S_0, S_i) = \overline{W}_0^1(p, S_0) + \overline{W}_i^2(p, S_0, S_i)$, based on Observation 1, it follows that $S_i^*(S_0 + 1) \le S_i^*(S_0)$.

Based on Observation 1, if p and S_0 are known, the optimal base-stock policy for each product i can be determined using a method similar to the one used for solving the single-stage system. That is, for each product i we need to find $S_i^*(p, S_0)$, the minimum S_i that satisfies $\overline{W}_i(p, S_0, S_i) \le W_i^{\max}$.

Based on Observation 2, we use the following technique to determine the optimal base stock levels S_0^* and S_i^* for each product i, given p is fixed. We start with $S_0 = 0$ and then increase S_0 incrementally by one. For each S_0, we determine the optimum S_i^* and keep track of the best solution obtained so far $Z^*(p)$. This procedure is continued until one of the two following stopping criteria is met: (i) $\overline{W}_0^1(S_0) \approx 0$ (e.g. $\le 10^{-6}$) or (ii) $h_0(p)\bar{I}_0(S_0) \ge Z^*(p)$. While the first criterion ensures that increasing S_0 no longer results in a significant decrease of $\overline{W}_0^1$, the second criterion ensures that no further cost savings can be achieved by increasing S_0.

There is no closed form solution available to determine the optimal value p^*. We propose to use a simple search technique as a heuristic to estimate the optimal p^*. That is, we search over p values in the range $0<p<1$ using a pre-specified incremental factor. For example, if the incremental factor used is .1 then we start with p=.1 and end with p=.9. Obviously, more accurate results would be obtained by decreasing the incremental factor. For each p value, the optimal S_0 and S_i values are obtained using the previously described algorithm. A formal description of the algorithm is provided in the Appendix.

4.3 Selecting the best configuration

Having solved the optimization problems for the single-stage and two-stage systems for a given set of parameters, we now are able to determine the best configuration. For this purpose we need to compare the optimal solutions obtained for the single-stage and for the two-stage system and choose the one with the least cost. The choice of the best configuration: MTS-1 or MTO-1 for the single-stage system or MTS-2, ATO, MTS-3 and MTO-2 for the two-stage system is then made by looking at the optimal stocking policy.

5. Numerical experiment

In this section we present and discuss our numerical findings. Our main inquiry will focus on how different system parameters may influence the decisions regarding the DP positioning and stocking policy. Our experiment involved extensive data sets with a total of 7200 problem instances being tested. Table 1 presents all the parameter values used in the experiment. We fixed the aggregate demand rate ($\lambda_0 = 40$) in this experiment. Ten values were used for the number of product configurations to see the effect of product proliferation. 12 values of μ were used that represent different utilisation rates of the production capacity. To study how the system's behaviour affected by the service level requirements, 20 levels were used for the maximum level of the expected order waiting time. In this experiment we assume that all products have the same target average waiting time. While we fixed the unit inventory holding cost at $h_i = 100$ for all the products, we used three different functions for the unit inventory holding cost of the generic component. Each function represents a different progression rate of h_0 along the product's value chain. The first function is a linear one: $h_0(p) = h_i p$. The second function is convex with $h_0(p) = h_i p^3$ representing the situation where the product's value is added with an increasing rate. In contrast to the second function, the third function is concave with $h_0(p) = h_i(1 - e^{-5p})$ representing the situation where the product's value increases with a diminishing rate. The three functions are depicted in Fig. 3. The parameter values described above do not represent any specific industrial case but are selected such that we are able to obtain some general insights from the experiment. Notice, for example, that how the system performance is influenced by the demand and production rates is actually dependent upon the ratio rather than exact values of the parameters. In this experiment we include situations ranging from a highly capacitated system ($\lambda/\mu = 40/50$) to a very loose system ($\lambda/\mu = 40/160$). The values for the maximum average order waiting time relative to the other parameters are also selected such that the circumstances in which each particular configuration is optimal are potentially observable.

With regards to the premium cost of form postponement, in this experiment we used the base case in which $r = 0$. As we are particularly interested in evaluating the relative merits of the single-stage system as opposed to the two-stage system for which r is an important factor, we introduce the use of the *threshold premium* r^*, under which the optimal total cost of the single-stage system is equal to the optimal total cost of the two-stage system. Employing form postponement would only be attractive if the actual amortised premium cost does not exceed the threshold value, i.e. $r < r^*$. Suppose we obtain an optimal total cost $Z(\underline{S}_1^*)$ for the single-stage system and $Z(p^*, \underline{S}_2^*)$ for the two-stage system under $r = 0$. Recall that changes in r will not alter the optimal policy $(p^*, \underline{S}_2^*)$. We may write $r^* = Z(\underline{S}_1^*) - Z(p^*, \underline{S}_2^*)$. As problem instances may differ in terms of their cost magnitudes, we measured a relative rather than an absolute value. That is,

$$r^*(\%) = \frac{Z(\underline{S}_1^*) - Z(p^*, \underline{S}_2^*)}{Z(\underline{S}_1^*)} \times 100\% . \quad (10)$$

The above measure also represents the relative cost savings obtained by introducing form postponement under the assumption of zero premium cost.

Fig. 3 is about here

Table 1 is about here

For each of the problem instances, we applied the algorithms described in Section 4 to determine the optimal configuration. The observations of interest include the interdependencies of some input parameters: the number of products N, the production rate μ, the target average waiting time $W_i^{\max}$, and the unit holding cost function of the generic component $h_0(p)$ with some output parameters: the average total cost, the optimal differentiation point p^*, the average threshold premium $r^*(\%)$, and the distribution of the best configuration. The results are summarised as follows.

The effect on the average optimal total cost

Figs. 4 and 5 illustrate how the average optimal total cost is affected by different input parameters. Fig. 4 depicts the average total cost as a function of the number of products and the production rate while Fig. 5 depicts the average total cost as a function of the number of products and the target expected order waiting time. Note that only the problem instances with the linear holding cost function were used in both figures. The three dimensional figures for the other two holding cost functions show similar behaviour and are left out for the sake of brevity. In Fig. 4, the average total cost value for each combination of N and μ is obtained by averaging 20 cost values across all $W_i^{\max}$ values. Similarly, the average total cost value for each combination of N and $W_i^{\max}$ in Fig. 5 is the average of 12 cost values across all μ values. As expected the average total cost decreases as the target average waiting time and the production rate increase. This is reasonable as the required stock level would be lower in both situations. The two figures also show that the average total cost tends to increase with the number of products particularly when the target waiting times or production rates are low. When $W_i^{\max}$ and μ are low, the results show that the MTS-2 configuration is dominant. Although increasing N results in a lower demand rate for each product configuration, the total stock across all products is most likely larger due to the requirement that S_i must be an integer. This explains why the average total cost is increasing. The effect of N is not shown when the target average waiting time and production rate are high. This is due to the fact that when both $W_i^{\max}$ and μ are sufficiently large, holding any stock may not be required, i.e. the MTO-2 or the MTO-1 configuration becomes optimal. Since $S_i = 0$ for both configurations and the aggregate demand rate is constant, the total cost remains unchanged by increasing N. Fig. 6 depicts the overall average of total cost as a function of N for each of the three $h_0(p)$ functions. The figure shows that the overall average total cost is increasing in N and this can be seen as a resultant of the increasing behaviour observed for the situations with low production rates and/or target waiting times and the non-increasing behaviour for the situations with high production rates and/or target waiting times.

Fig. 4 is about here

Fig. 5 is about here

Fig. 6 is about here

The effect on the optimal differentiation point

The effects of different input parameters on the optimal differentiation point are illustrated in Figs. 7, 8, and 9. We depict the average p^* as a function of N and $W_i^{\max}$ in Fig. 7 and as a function of N and μ in Fig. 8.

Fig. 7 is about here

Fig. 8 is about here

The two figures show that the average p^* tend to decrease in both $W_i^{\max}$ and μ although a few anomalies are observed. It is shown in Fig. 7 for example, that for $N \leq 3$ the average p^* is first increasing when $W_i^{\max}$ is changed from .002 to .004. This could occur because, as the MTS-2 configuration is optimal for such values of $W_i^{\max}$, increasing the target waiting time may result in a lower optimal stock level but may require a higher p^* to keep the system responsive. The decreasing of the average p^* is due to several reasons. As already mentioned earlier, the very loose waiting time constraints and the high production rate would allow the MTO-1 configuration to be optimal. Consequently, since $p = 0$ for this configuration, the overall average p^* will drop. For the problem instances where the ATO configuration is optimal, increasing the production rate and the target waiting time also contribute to the reduced p^* because it would be beneficial to move the postponement point earlier to save the inventory holding cost of the generic component. In Fig. 9 we plot the average p^* as a function of N for each of the three $h_0(p)$ functions. The results show that while there is no significant difference between the linear function and the convex function, the average p^* is higher for the concave function. This observation reveals the potential of considering a late form postponement in the environments where a significant added value is made at the early stage of the production process. In such environments it is possible to make the system become more responsive by having a late form postponement while at the same time taking the advantage of insignificant holding cost increase. In contrast, we do not observe opposite results that motivate earlier postponement when the inventory holding cost follows a convex function. The reason for this could be that applying an early postponement would cause the violation of the service level requirement even though the inventory cost could be reduced.

Fig. 9 is about here

The effect on the threshold premium cost of form postponement

The results regarding the average threshold premium cost $r^*(\%)$ associated with form postponement are summarised in Figs 10, 11, and 12. As expected, the threshold premium is always positive as long as the optimal configuration for the single-stage system is the MTS-1 configuration. If the MTO-1 configuration is feasible then the premium cost is zero, i.e. employing form postponement would not be necessary. It is then clear that the decreasing behaviour shown in Figs 10 and 11 is caused by an increasing number of problem instances for which the MTO-1 configuration is optimal. However, as shown in Fig. 10, the average $r^*(\%)$ is first increasing before decreasing in $W_i^{\max}$. This behaviour becomes clearer when observing Fig. 12 where we see the average $r^*(\%)$ is first increasing in $W_i^{\max}$ before suddenly dropping to zero. For the two-stage system employing form postponement, increasing $W_i^{\max}$ would result in cost reductions obtained from more shifting of the optimal configuration from MTS-2 to ATO and from ATO to MTO-2. In parallel, increasing $W_i^{\max}$ would also give cost reductions in the single-stage (MTS-1) system as a lower stock level is required. The increasing average $r^*(\%)$ occurs because the cost reduction obtained in the two-stage system is higher than that obtained in the single-stage system. The effect of increasing the production rate to the average $r^*(\%)$ as shown in Fig. 12 is not significant until a certain level is reached allowing the MTO-1 system to become optimal. The overall effect of increasing the number of products as shown in Figures 10, 11 and 13 seems to increase the average threshold premium. This is reasonable as the system with form postponement benefits from the risk-pooling effect resulted from holding the inventory of the generic component.

Fig. 10 is about here

Fig. 11 is about here

Fig. 12 is about here

Fig. 13 is about here

The effect on the optimal configuration

The results on how the best configuration is distributed across the problem instances are illustrated in Figs 14-16 (only the 2400 problem instances with the linear holding cost function were included in these figures). Four configurations (MTO-1, MTS-2, ATO, and MTO-2) appear as the possible optimal configuration. However, these results should not be too generalised as we shall show later that the MTS-1 and the MTS-3 configurations could also become optimal under certain circumstances.

Fig. 14 is about here

Fig. 14 depicts the effect of the number of product configurations on the optimal configuration distribution (there are in total 240 instances associated with each value of N). It is shown that the frequency of the MTO-1 and the MTO-2 configurations are not affected. Changing the number of products only affect the distribution of the MTS-2 and the MTS-3 configuration. It is shown that, when there are more product configurations, the frequency of the ATO configuration becomes higher and on the contrary, the frequency of the MTS-2 configuration becomes lower. This is mainly due to the fact that increasing the number of product configurations results in a lower demand rate for each individual product configuration and in higher pooling benefits realised from stocking the generic component thereby allowing more possibilities to employ a MTO mode in the second stage.

Fig. 15 is about here

Fig. 15 shows the effect of the target average waiting time (120 problem instances are associated with each value of $W_i^{\max}$). Only the MTS-2 and ATO configurations are attractive in the situations characterised by very low waiting times ($W_i^{\max} \leq .006$) with the ATO configuration being preferable as the target average waiting time increases. From a certain level of $W_i^{\max}$ (.008 in Figure 15), the MTO-2 configuration starts becoming a possible optimal configuration and its frequency is constant as the target average waiting time increases. It is also shown that, from $W_i^{\max} = 0.012$ onwards, the MTO-1 configuration appears to be a possible optimal configuration and is increasingly dominating the MTS-2 and the ATO configuration as the target average waiting time increases.

The effect of production rate on the optimal configuration distribution is shown in Fig. 16 (200 problem instances are associated with each value of μ). The figure shows that for very low production rate values ($\mu \leq 60$) only the MTS-2, ATO and MTO-2 appear as possible optimal configurations. The MTO-1 configuration becomes feasible when the production rates are higher ($\mu > 60$). What is obvious from this figure is that the frequency of the MTS-2 configuration is decreasing with the production rate except for the four highest production rate values (μ=130, 140, 150 and 160) where the MTS-2 configuration is optimal only for one problem instance with N=1 and $W_i^{\max}$ = .002. It can be shown that only when the production rate is higher ($\mu \geq 170$), the optimal configuration for that particular problem instance is shifted from MTS-2 to ATO. There is no obvious pattern observed on how the frequency of the MTO-2 configuration is affected by changing the production rate. However it is clear that increasing the production rate may create possibilities of shifting the optimal configuration from ATO to MTO-2 as well as from MTO-2 to MTO-1.

Fig. 16 is about here

to see the effect of the inventory holding cost functions we summarise the results in Table 2.

Table 2 is about here

The total frequency of the MTO-1 and MTO-2 configurations are equal for the three functions. An interesting observation is the appearance of the MTS-1 and MTS-3 configurations for the problem instances with the concave holding cost function while both configurations are always sub-optimal for the other two functions. This reveals that when the holding costs of the generic component and the final products do not differ much (as with the concave function), it could be beneficial to stock only the finished products. On the contrary, when the holding cost increases rapidly only at the end-phase of production process it would be more favourable to stock the generic component rather than the final product configurations. This is indicated in Table 2 where the highest frequency of the ATO configuration is found for the problem instances with the convex holding cost function.

6. Conclusions and directions for further research

In this paper we evaluate postponement structures characterised by the positioning of differentiation point and stocking policy. Six classes of manufacturing configurations are identified based on the choice whether or not form postponement is employed and the decision regarding the stocking policy for the final product configurations as well as for the generic component. We developed analytical evaluation methods based on queuing models to assess operational measures for each class of configuration. We developed solution procedures to determine optimal stocking levels and differentiation points minimising the expected total cost that may consist of the inventory holding cost and the amortised cost associated with the employment of form postponement subject to the requirement that the average order response time for each product does not exceed a predetermined threshold level. This allows us to evaluate the relative merits of all manufacturing configurations based on their respective best performances.

The results of our numerical experiment show how different system parameters including the number of product configurations, the production capacity, the maximum order response time and the unit holding cost progression function may affect the preference of a certain class of manufacturing configuration. Our numerical study also reveals important information regarding the choice of early or late differentiation point. It is shown that the high production rates and target waiting times offer the possibility to save the inventory cost of the generic component which contributes to the preference of early differentiation point. Furthermore, the benefits of employing form postponement in terms of relative cost savings are also

examined in this study. The results show that employing form postponement could be beneficial as long as the MTS (instead of MTO) configuration is optimal for the single-stage system. This study also clarifies the concept of form postponement in that delaying the product differentiation in postponement does not necessarily suggest that the differentiation is only processed after customer orders are received. All these outcomes extend the understanding of postponement concept and offer useful guidance in designing the most cost-effective supply chain structure.

This research can be extended in several directions. One possible extension is to more extensively investigate the effect of product variety in the environments characterised by the existence of setup times and/or setup costs associated with changeovers from one product to another. Such an extension would need the use of a different inventory policy (probably involving batching decisions) as opposed to the simple base-stock policy applied in the current model. The study will extend the understanding of the potential of postponement as an important approach dealing with the proliferating product variety. Furthermore, our model can be used as the basis for an extended analysis incorporating both manufacturing and logistics postponement. Clearly, studies seeking to tackle the problem combining all different types of postponement would be valuable to assess the full potential of the postponement concept.

References

Agrawal, M., Kumaresh, T.V. and Mercer, G.A., The false promise of mass customization. *The McKinsey Quarterly*, 2001, 3, 62-71.

Alderson, W., Marketing efficiency and the principle of postponement. *Cost and Profit Outlook*, September 1950.

Arreola-Risa, A. and DeCroix, G.A., Make-to-order versus make-to-stock in a production-inventory system with general production times. *IIE Transactions*, 1998, 30, 705-713.

Aviv, Y. and Federgruen, A., Design for postponement: a comprehensive characterisation of its benefits under unknown demand distributions. *Operations Research*, 2001a, 49, 578-598.

Aviv, Y. and Federgruen, A., Capacitated multi-item inventory systems with random and seasonally demand fluctuating demands: implications for postponement strategies. *Management Science*, 2001b, 47, 512-531.

Bowersox, D.J. and Closs, D.J., *Logistical Management: The Integrated Supply Chain Process*. McGraw-Hill, New York, 1996.

Buzacott, J. and Shanthikumar, J.G., *Stochastic Models of Manufacturing Systems*. Prentice-Hall, Upper Saddle River, 1993.

Federgruen, A. and Katalan, Z., The impact of adding a make-to-order item to a make-to-stock system. *Management Science*, 1999, 45, 980-994.

Feitzinger, E. and Lee, H.L., Mass customization at Hewlett Packard: the power of postponement. *Harvard Business Review*, 1997, 75, 116-121.

Giesberts, P.M.J. and van den Tang, L., Dynamics of the customer order decoupling point: impact on information systems for production control. *Production Planning and Control*, 1992, 3, 300-313.

Gupta, D. and Benjaafar, S., Make-to-order, make-to-stock, or delayed product differentiation? A common framework for modelling and analysis. *IIE Transactions*, 2004, 36, 529-546.

Lee, H.L., Billington, C., and Carter, B., Hewlett Packard gains control of inventory and service through design for localisation. *Interfaces*, 1993, 23, 1-11.

Lee, H.L. and Tang, C.S., Modelling the costs and benefits of delayed product differentiation. *Management Science*, 1997, 43, 40-53.

Lee, Y. and Zipkin, P., Tandem queues with planned inventories. *Operations Research*, 1992, 40, 936-947.

Pagh, J.D. and Cooper, M.C., Supply chain postponement and speculation structures: how to choose the right structure? *Journal of Business Logistics*, 1998, 19, 13-34.

Rajagopalan, S., Make to order or make to stock: model and application. *Management Science*, 2002, 48, 241-256.

Rudberg, M. and Wikner, J., Mass customization in terms of the customer order decoupling point. *Production Planning and Control*, 2004, 15, 445-458.

Sharman, G., The rediscovery of logistics, *Harvard Business Review*, 1994, 62, 71-80.

Swaminathan, J.M. and Tayur, S., Managing broader product lines through form postponement using vanilla boxes. *Management Science*, 1998, 44, S161-S172.

Swaminathan, J.M. and Lee, H., Design for postponement, in *Handbooks in Operations Research and Management Science: Supply Chain Management: Design, Coordination and Operation*, Eds: Graves, S. and de Kok, Elsevier Publishers, 2003, 199-228.

Su, J.C.P., Chang, Y., and Ferguson, M., Evaluation of postponement structures to accommodate mass customization. *Journal of Operations Management*, 2005, 23, 305-318.

Van Hoek, R.I., The rediscovery of postponement: a literature review and directions for research. *Journal of Operations Management*, 2001, 19, 161-184.

Van Houtum, G.J. and Zijm, W.H.M., On the relation between cost and service models for general inventory systems. *Statistica Neerlandica*, 2000, 54, 127-147.

Yang, B. and Burns, N.D., Implications of postponement for the supply chain. *International Journal of Production Research*, 2003, 41, 2075-2090.

Appendix

The formal algorithms for solving the optimization problems are as follows.

A. The single-stage system

For each product i = 1, 2, …, N do:

Step 1: Set the initial solution $S_i = 0$.

Step 2: Calculate $\overline{W}_i(S_i)$. If $\overline{W}_i(S_i) > W_i^{\max}$, go to Step 3. Otherwise stop, $S_i^* = S_i$.

Step 3: Set $S_i = S_i + 1$; go to Step 2.

END

B. The two-stage system

Step 1: Choose an increment factor δ, set $p = \delta$ and $Z^* = \infty$.

Step 2: Determine S_0^* and S_i^* for all i using the algorithms Opt-B1 and Opt-B2 (see below). Calculate the corresponding expected total cost Z. If $Z < Z^*$, set $Z^* = Z$, $S^* = (S_0^*, S_1^*, ..., S_N^*)$, and $p^* = p$.

Step 3: If $p \geq 1 - \delta$ stop. Otherwise set $p = p + \delta$ and go to Step 2.

END

Algorithm Opt-B1

Given that p and S_0 are fixed, for each product i = 1, 2, …, N do the following.

Step 1: Set the initial solution $S_i = 0$.

Step 2: Calculate $\overline{W}_i(p, S_0, S_i)$ using (7) and (9). If $\overline{W}_i(p, S_0, S_i) > W_i^{\max}$, go to Step 3. Otherwise stop; $S_i^*(p, S_0) = S_i$.

Step 3: Set $S_i = S_i + 1$; go to Step 2.

END

Algorithm Opt-B2

Given that p is fixed.

Step 1: Set $S_0 = 0$, $Z^*(p) = \infty$.

Step 2: Calculate $\overline{W}_0^1(p, S_0)$ using (7). If $\overline{W}_0^1(p, S_0) \leq W_i^{\max}$, go to Step 4. Otherwise continue.

Step 3: Set $S_0 = S_0 + 1$; go to Step 2.

Step 4: For each i, obtain $S_i^*(p, S_0)$ using the algorithm Opt-B1.

Step 5: Calculate the corresponding total cost Z. If $Z < Z^*(p)$, set $Z^*(p) = Z$, $S_0^* = S_0$ and $S_i^* = S_i$ for all i.

Step 6: If $\overline{W}_0^i(S_0) \approx 0$ or $h_0(p)\bar{I}_0(S_0) \geq Z^*(p)$ stop. Otherwise set $S_0 = S_0 + 1$ and go to Step 4.

END

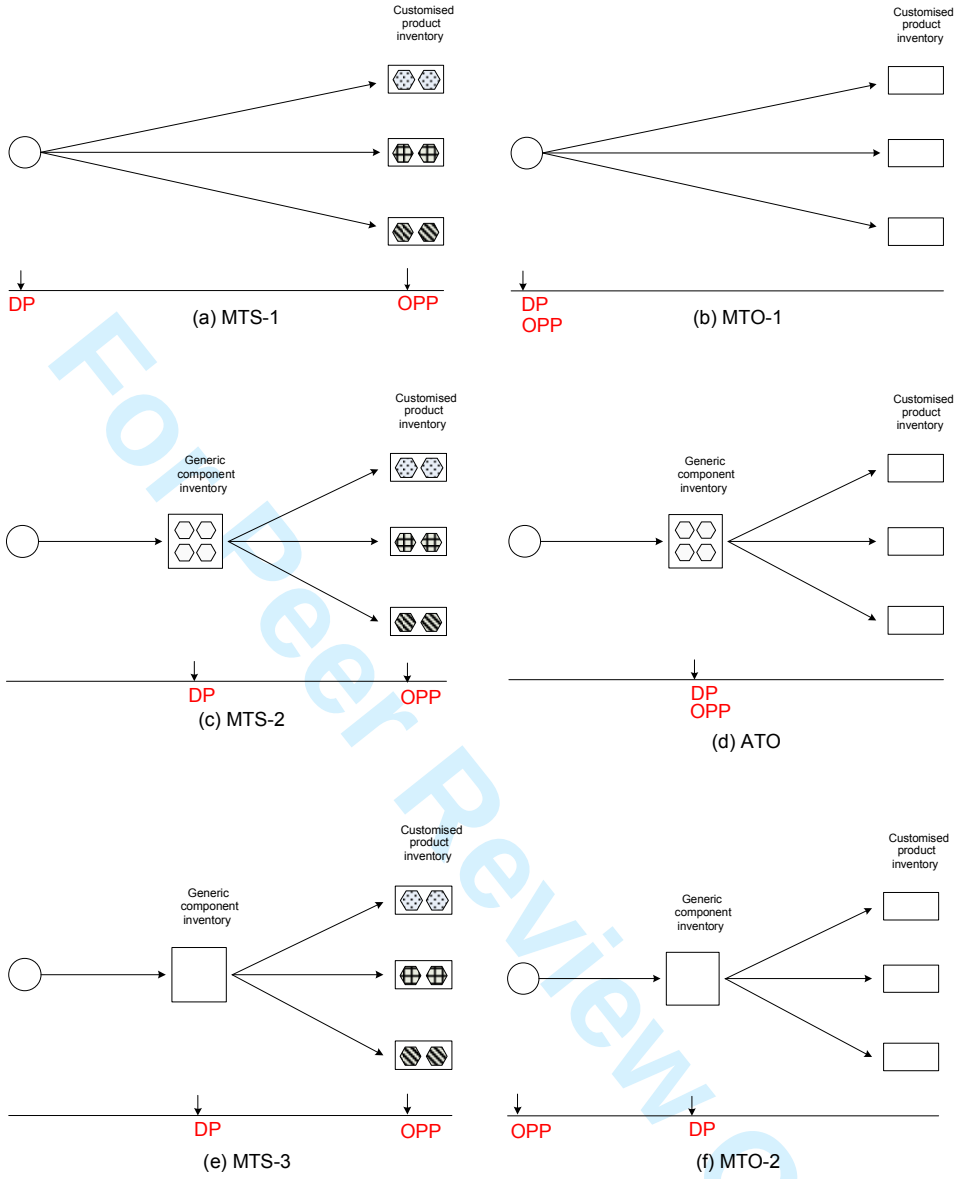

Fig.1. Six configurations based on the position of the differentiation point and stocking policy

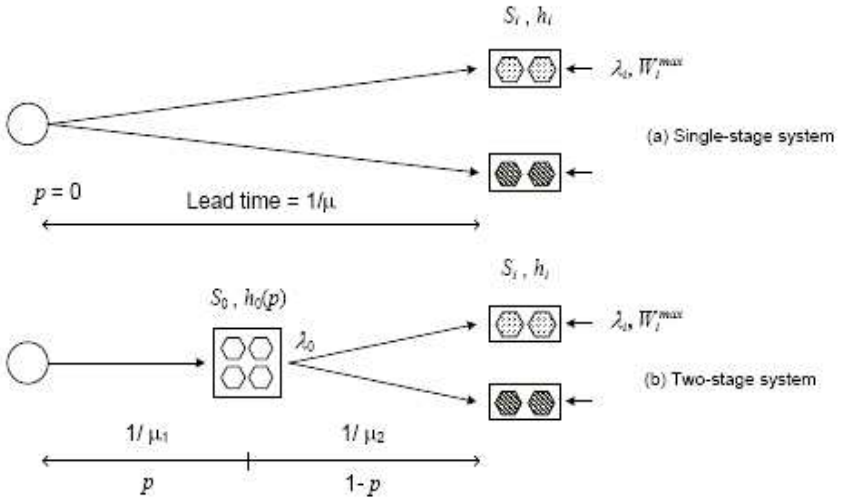

Fig. 2. Notation used to model the single-stage and the two-stage systems

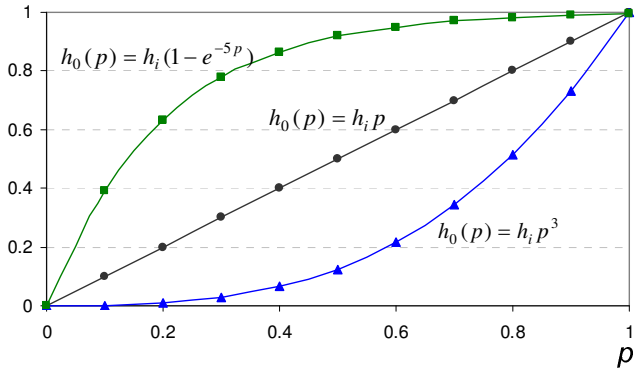

Fig. 3. Three different functions for the inventory holding cost of the generic component

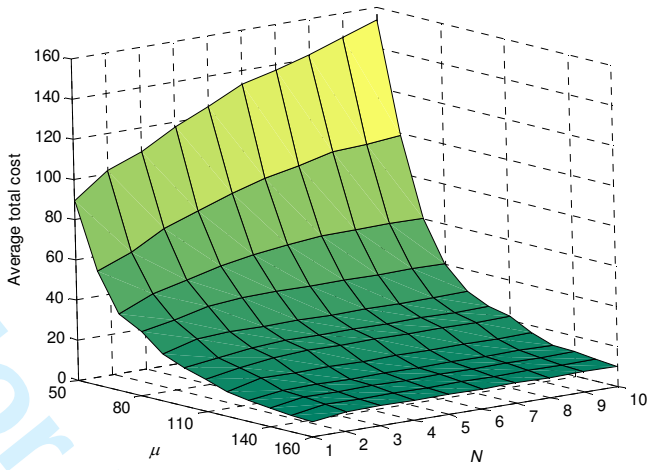

Fig. 4. Average total cost as a function of μ and N

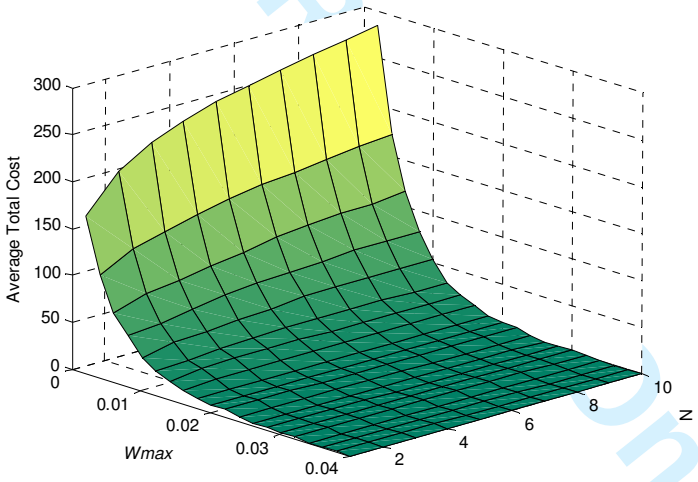

Fig. 5. Average total cost as a function of $W_i^{\max}$ and N

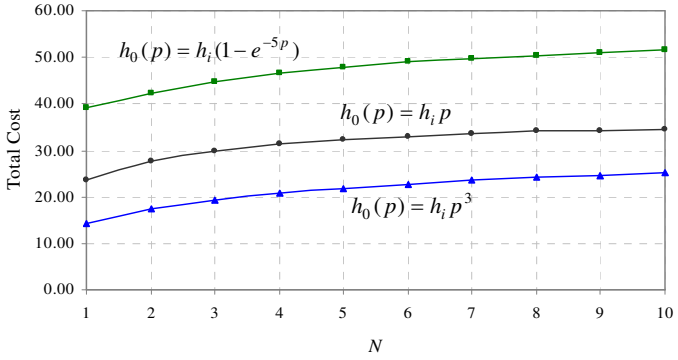

Fig. 6. Average total cost vs N for each of the $h_0(p)$ functions

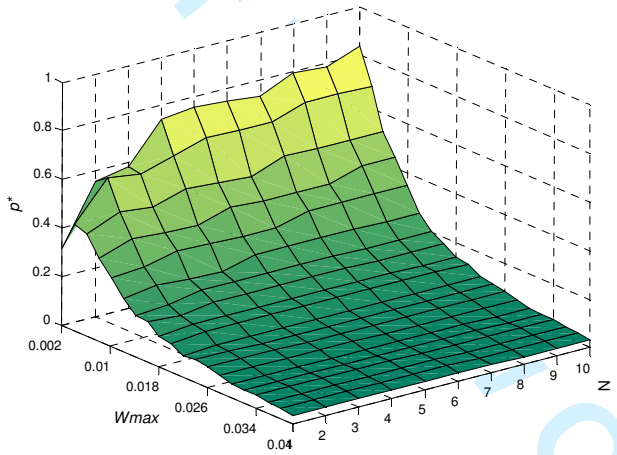

Fig. 7. Average optimal p^* as a function of N and $W_i^{\max}$

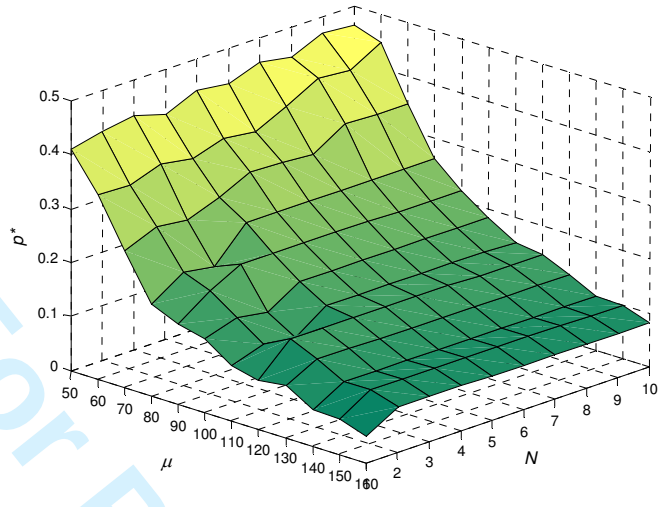

Fig. 8. Average optimal p^* as a function of N and μ

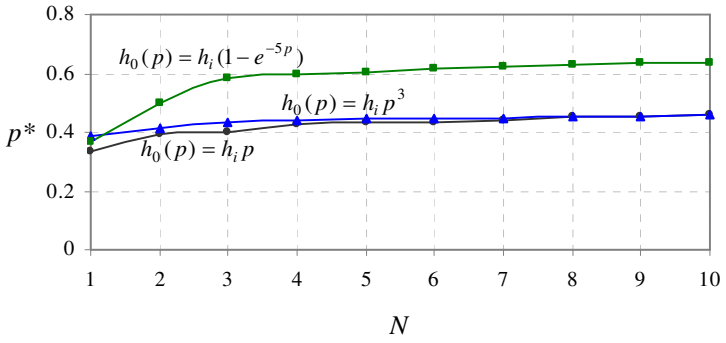

Figure 9. Average p^* as a function of N for each of the three $h_0(p)$ functions

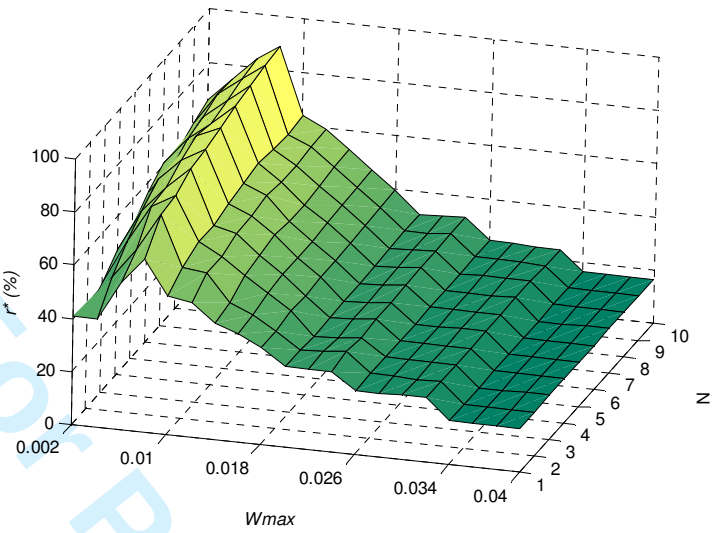

Fig. 10. Average threshold premium $r^*(\%)$ as a function of N and $W_i^{\max}$

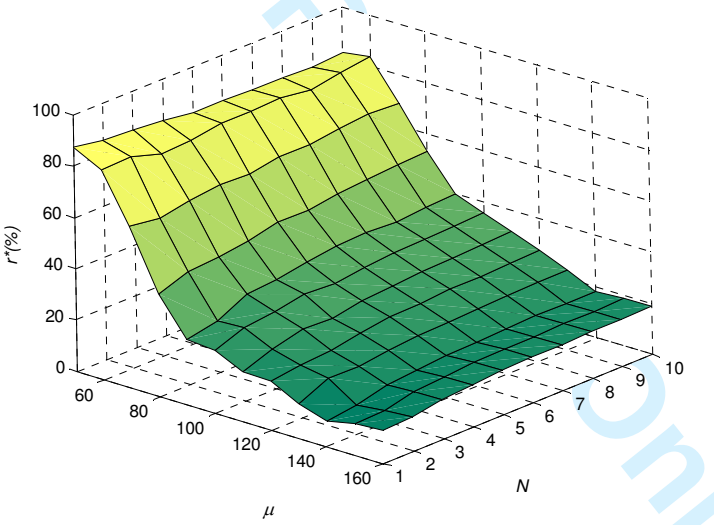

Fig. 11. Average threshold premium $r^*(\%)$ as a function of N and μ

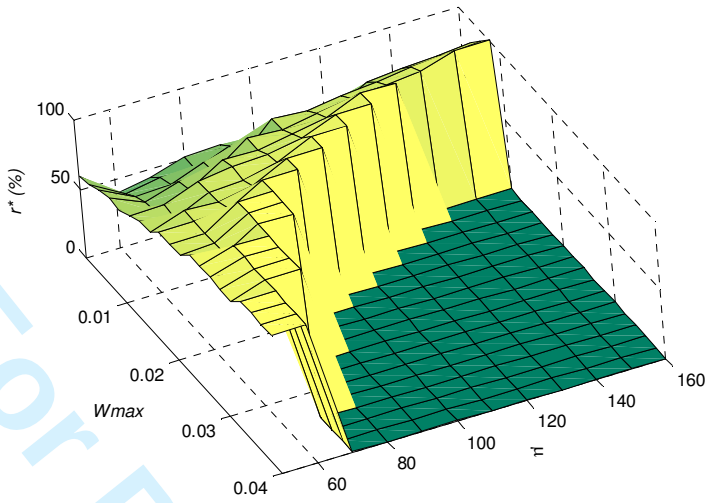

Fig. 12. Average threshold premium $r^*(\%)$ as a function of μ and $W_i^{\max}$

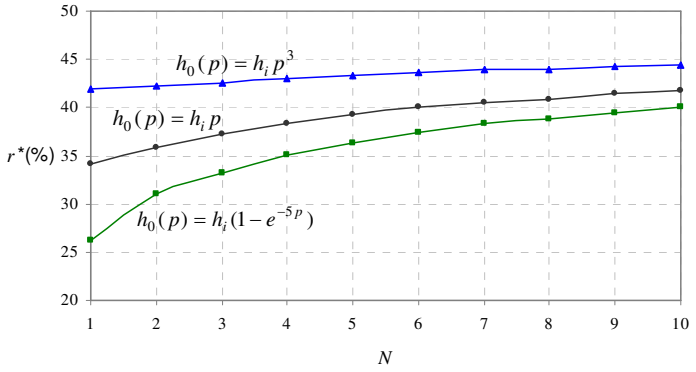

Fig. 13. Average $r^*(\%)$ as a function of N for each of the three $h_0(p)$ functions

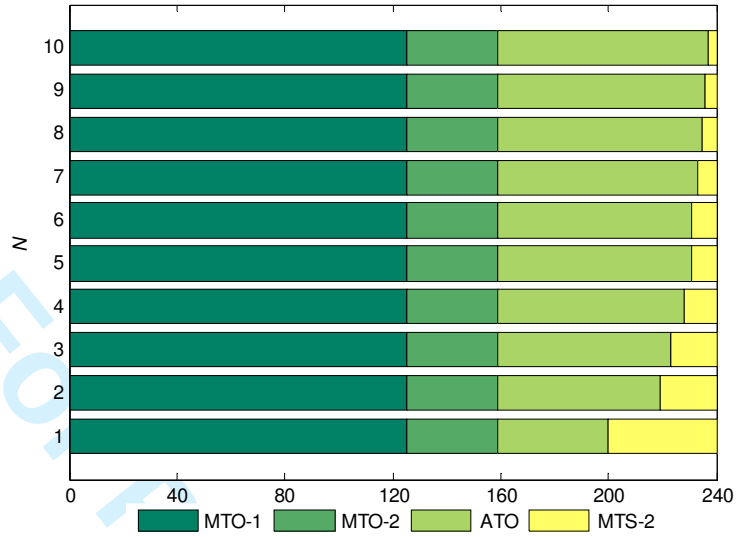

Fig. 14. Optimal configuration distribution as a function of the number of products

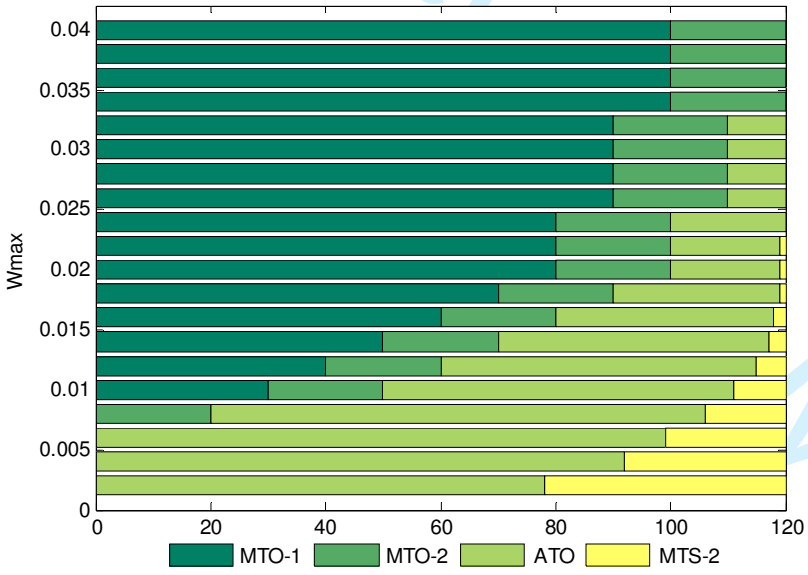

Fig. 15. Optimal configuration distribution as a function of the target waiting time

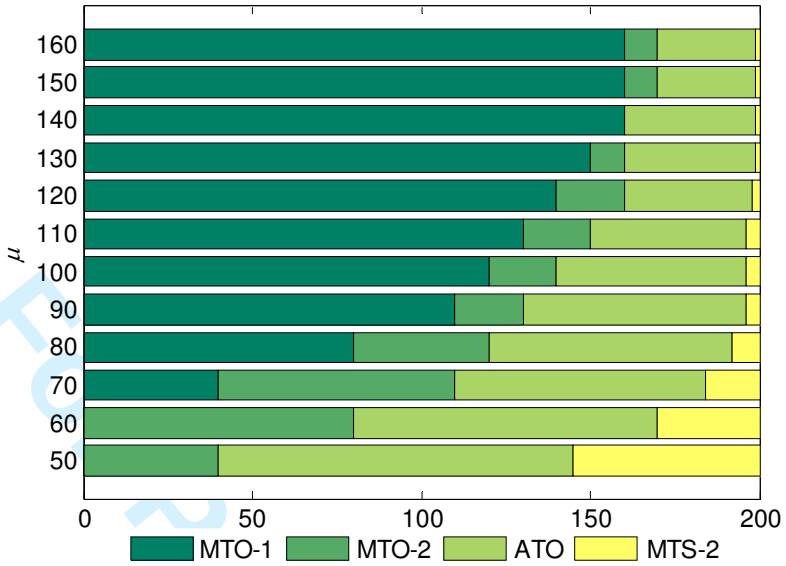

Fig. 16. Optimal configuration distribution as a function of the production rate

Table 1. The parameter values used in the numerical experiment

Parameter	Unit	Number of values	Values
λ_0	/ time unit	1	40
N		10	1, 2, 3, 4, 5, 6, 7, 8, 9, and10
μ	/ time unit	12	50, 60, …, 150, and 160
$W_i^{\max}$	Time unit	20	.002, .004, …, .038, and .04
h_i	$/unit/time unit	1	100
$h_0(p)$	$/unit/time unit	3	See Fig. 2.
r		1	0

Table 2. Frequency of optimal configurations for the three different $h_0(p)$ functions

Configuration	Linear $h_0(p) = h_i p$	Convex $h_0(p) = h_i p^3$	Concave $h_0(p) = h_i(1 - e^{-5p})$
MTS-1	0	0	26 (1.1%)
MTO-1	1250 (52.1%)	1250 (52.1%)	1250 (52.1%)
MTS-2	127 (5.3%)	85 (3.5%)	120 (5.0%)
ATO	683 (28.4%)	724 (30.2%)	658 (27.4%)
MTS-3	0	1	6 (0.25%)
MTO-2	340 (14.2%)	340 (14.2%)	340 (14.2%)