

## *Research Article*

# **Analysis and Optimization of a Combined Make-to-Stock and Make-to-Order Multiproduct Manufacturing System**

**Khaled Hadj Youssef,<sup>1</sup> Christian van Delft,<sup>2</sup> and Yves Dallery<sup>3</sup>**

<sup>1</sup> *Laboratoire Génie Mécanique, École Nationale des Ingénieurs de Monastir, Avenue Ibn El Jazzar, 5019 Monastir, Tunisia*

<sup>2</sup> *HEC School of Management, Paris (GREGHEC) 1, rue de la Libération, 78351 Jouy-en-Josas Cedex, France*

<sup>3</sup> *Laboratoire Génie Industriel, Ecole Centrale Paris, Grande Voie des Vignes, 92295 Chatenay-Malabry Cedex, France*

Correspondence should be addressed to Christian van Delft, vandelft@hec.fr

Received 12 February 2009; Revised 14 August 2009; Accepted 4 September 2009

Recommended by Ömer S. Benli

We consider a single-stage multiproduct manufacturing facility producing several end-products for delivery to customers with a required customer lead-time. The end-products can be split in two classes: few products with high volume demands and a large number of products with low-volume demands. In order to reduce inventory costs, it seems efficient to produce the high-volume products according to an MTS policy and the low volume products according to an MTO policy. The purpose of this paper is to analyze and compare the impact of the scheduling policy on the overall inventory costs, under customer lead-time service level constraints. We consider two policies: the classical FIFO policy and a priority policy (PR) which gives priority to low volume products over high volume products. We show that for some range of parameters, the PR rule can significantly outperform the FIFO rule. In these ranges, the service level constraints are satisfied by the PR rule with much lower inventory costs.

Copyright © 2009 Khaled Hadj Youssef et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## **1. Introduction**

### **1.1. Motivation**

In a marketplace which is increasingly customer oriented, competitive manufacturers have to propose a large variety of products, at low prices, and be able to deliver them to the customers in order to meet the required lead-times. An important (and related) issue is to determine whether the products should be manufactured according to a make-to-order

policy, to a make-to-stock policy, or to some middle ground whereby some products are made to stock and others are made to order. Recall that under make-to-order (MTO) management, a production order is released to the manufacturing facility only after a firm demand has been received. Under make-to-stock (MTS) management, products are manufactured in anticipation of future orders and stored in the finished goods inventory (FGI). Demands are then directly satisfied from the FGI. The advantage of the MTS policy as compared to the MTO policy is its reactivity to customer demands. The drawback is inventory holding costs associated with the FGI. Thus, there is a tradeoff in deciding which policy to use.

Frequently, in multiproduct markets, the customers can be grouped in market segments having specific characteristics, especially concerning the demand variety and/or the required customer lead-time (i.e., the admissible delay between the time an order is placed and the time this order is delivered). It is well known that operations efficiency can be improved by exploiting such market segmentation (see [1, 2]). As a matter of fact, in a context where all products are high-volume low-variety products (standard products) with low customer lead-times, the MTS policy is clearly appropriate. On the other hand, in a context where all products are low volume high variety products with reasonably large customer lead-times, the MTO policy has to be used. Beside these two extreme cases, there are many real systems with a small number of end-products having high volume demands, whereas a fairly large number of end-products have low volume demands. In this situation, it seems appealing to try to produce the high volume products according to an MTS policy and the low volume products according to an MTO policy.

In fact, there are two main issues that need to be investigated, in a combined way, regarding this approach which consists of mixing MTS and MTO policies: the scheduling rule used to process production orders in case of congestion at the manufacturing facility, and the impact of the customer lead-time values.

A simple approach to schedule the manufacturing facility is to use a first-in-first-out (FIFO) scheduling rule whereby all production orders are processed according to their release dates. A potentially interesting alternative is to give priority to production orders corresponding to low-volume high-variety products over production orders corresponding to high volume low variety products. The underlying idea is to give maximal reactivity to low volume products in order to be able to produce them under an MTO management policy, while maintaining reasonable inventories for high volume products.

The second issue to consider in the production-inventory policy is the impact of the customer lead-times, which is significant. For example, large lead-times make an MTO approach possible, while short (or zero) lead-times induce an MTS strategy.

Our study is motivated by many industry cases where the MTS/MTO issue discussed in our paper is encountered. These industry cases relate to situations where the company offers both fairly standardized products that correspond to the need of a large number of customers and specific (customized) products that correspond to the need of a small number of customers. The standardized products that have large volumes would generally be produced according to an MTS policy, while the customized products that have low volumes would be produced according to an MTO policy. In addition, priority in terms of production capacity would be given to the MTO products so that to achieve a low customer lead-time, while the MTS products could be produced with less reactivity without causing damage to the final customers since they are held in stock. This situation can actually be encountered in B2B as well as B2C contexts. A classical example in the B2C context is the automotive industry. Indeed, car manufacturers are often operating under a mixed MTS and MTO modes. The MTS mode corresponds to cars that are produced in anticipation to customer demands and that

are readily available at the car dealers. These cars usually correspond to the variants that correspond to what most customers want in terms of colors, engines, or interior design. On the other hand, the MTO mode corresponds to the manufacturing of a specific car requested by a final customer. The car has all the features that this specific customer wants and the customer is willing to wait some time to get his own customized car. Usually the Assembly Plant will give priority to the MTO cars because it is important for the car manufacturer to be able to deliver the specific car within the customer lead-time, since by not doing so it incurs some financial penalties, which would correspond to a reduced margin for the sale. In B2B contexts, examples include chemical and steel industries. Again, these industries usually have clients that are industries that need fairly standard products in terms of composition. On the other hand, they also have some more specific clients that need products with a customized composition for their own purpose (these products can only be sold to these customers). These industries would again operate under the same MTS/MTO modes with priority given to the MTO products. It should be noted that another reason for giving high priority to customers requesting specific products is that usually these customized products have higher margin than the standard products. It is, therefore, important for the company to serve these customers well in terms of customer lead-times.

The purpose of this paper is to analyze and compare the combined impact of the choice of the scheduling policy and the customer lead-time values on the overall performance of the system. We will provide general theoretical properties as well as managerial insights. To address this issue, we consider a setting in which the modelling assumptions are as simple as possible so as to allow tractability by means of analytical solution techniques. We consider a single-stage manufacturing facility producing multiple heterogeneous products. We assume that these products are divided into two different product families, associated with two market segments. The first family corresponds to a small number of products with a high demand rate. This family is denoted by High Volume (HV). The second family encompasses a large number of products with low demand rate and is denoted by Low Volume (LV). The number of HV products is assumed to be much lower than the number of LV products. We model the demand process for each product by a stochastic Poisson process. The manufacturing facility consists of a single-stage system which is the only constraining resource in terms of production capacity. Moreover, we assume that the manufacturing facility is totally flexible and reliable. The processing times do not depend on the product type and are i.i.d. random variables with exponential distribution. Enough raw materials are assumed to be available.

The system follows a base-stock inventory policy, that is, each product has a target base-stock level and every time a demand for an end-product is received, a production order is generated and sent to the manufacturing facility. Any demand that cannot be satisfied from the FGI at the end of the required customer lead-time is backordered. For a given product, this generic policy can be interpreted as an MTS policy when the base-stock level is positive, and as an MTO policy when the base-stock level is equal to zero.

To compare different ways of running the system, we consider the problem of minimizing the long-run, average inventory holding costs while achieving a service level constraint for each product. The service level constraint, for each product type, is expressed in terms of order lead-time, that is, the time between the instant the order is placed and the instant the finished product is delivered. This order lead-time has to be lower than the required customer lead-time, with a given probability.

By using a fundamental decomposition property, we describe for each scheduling rule a procedure to estimate the optimal base-stock levels and the associated cost. Under FIFO,

we propose an explicit decomposition of this multiple-product system into single-product models. We then provide analytical expressions for the optimal base-stock levels. For the priority rule PR, no closed-form analytical formulas are available to compute the whole system performance. Thus, we propose a simple analytical approximation, providing explicit overall insight.

## **1.2. Related Bibliography**

Performance analysis and optimization of inventory-production systems, under a unique inventory policy, MTO or MTS, have represented an important area for research (see [3–5] for exhaustive references).

The impact of a required customer lead-time in inventory-production systems has been considered in several papers (see Karaesmen et al. [5]). In particular, Buzacott and Shanthikumar [3] analyzed a model of a single-stage make-to-stock queue and investigated how the optimal stock varies as a function of the customer lead-time.

Recently, several papers have studied combined MTS and MTO in multiproduct systems (see Arreola-Risa and Decroix [6] and Federgruen and Katalan [7] for references). Rajagopalan [8] has modeled the MTO/MTS decision and the selection of optimal parameters, under a FIFO scheduling rule and a given customer lead-time, as a mixed-integer nonlinear program. Motivated by a specific industrial case, this author has focused on numerical heuristics, enabling him to solve, in an approximate manner, large-scale problems involving several hundreds of items in presence of setup times and lot size effects.

A similar approach is taken by Sox et al. [9], where optimal base stock levels are computed under a FIFO priority rule. Their model includes a fixed inventory budget constraint and a specific aggregate time window fill rate service level constraint.

Hadj Youssef et al. [4] have modeled the combined MTS and MTO optimization problem in multiproduct systems in the case of service constraints expressed in terms of required average customer lead-time. Their paper is merely focused on the numerical computation of the optimal base-stock levels when considering a preemptive priority scheduling rule for some product families.

Several papers on MTO/MTS systems study the structure and/or the computation of optimal dynamic priority rules (see [4] for extensive references). These papers exploit the theory of optimal control of queuing systems and deal directly with the complexity of the optimal control policies.

## **1.3. Outline and Structure of the Paper**

The outline of this paper and some of the results are as follows: we investigate the structure of the optimal policy (MTO or MTS), in a multiproduct single-stage manufacturing system. The structure of the optimal policy is related to the scheduling rule (FIFO or PR) and to the required customer lead-times. The optimal policy has several intuitive properties, even if the final structure is in general difficult to parameterize efficiently.

Our paper is structured as follows. In Section 2, we present our model and our main assumptions. The fundamental structural properties of the problem are derived in Section 3. It provides the basis of the performance analysis and optimization developed in Sections 4 and 5, leading to a set of managerial insights. In Section 6, we review some interesting extensions and new avenues of research.

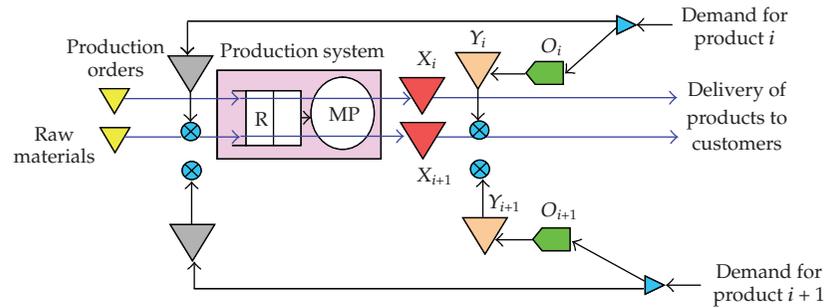


Figure 1: Queuing network model of a multiproduct single-stage BSCS with time lags.

## 2. The Production/Inventory System

### 2.1. General Behavior of the Production/Inventory System

We are considering a single-stage manufacturing facility which produces several types of finished products. This single-stage manufacturing facility is controlled by a Base-Stock Control System (BSCS) (see [3]). In summary, this single-stage BSCS is a control mechanism which consists of releasing raw parts into the manufacturing process each time a new customer demand for finished products is issued. It is assumed that the delivery of parts, corresponding to a given demand for product type- $i$ , has to be done a given time period  $L_i$  after the order is received. This given time period is called the customer lead-time and is thus the admissible delay between the time the order is placed and the time it has to be delivered to the customer. As in [3, Chapter 4.5], we assume that the physical delivery cannot be made to the customer before this time period. If, at the due date, there is no parts in the finished good inventory, demands are then backordered.

In order to model this BSCS in presence of required customer lead-times, we propose a multiproduct queuing model with time lags, which is an extension of the model discussed in [3]. The structure of this basic model is depicted in Figure 1.

This basic model with time lags and several products proceeds as follows. MP represents the considered manufacturing process. It contains parts being processed or waiting to be processed, which are referred to as work-in-process. Let us define the production lead-time as the time elapsed between the release of a raw part in the manufacturing facility (according to the BSCS), and the physical delivery of the corresponding finished part into the FGI.

However,  $O_i$  represents the order time lag process: it contains the list of time lagged orders. Queue  $R$  contains the raw parts waiting for processing on MP. Queue  $X_i$  is the product type- $i$  FGI, that is, the output buffer of the system containing type- $i$  parts that have completed processing and are ready for delivery to customers. Queue  $Y_i$  contains backordered demands (i.e., orders delivered after the requisition date). The sojourn time in  $Y_i$  is called the backorder time for demand of type- $i$  products and is denoted  $D_i$ . Queues  $R$ ,  $X_i$ , and  $Y_i$  are unlimited.

Initially, before any customer demand arrive in the system,  $X_i$  contains a base-stock of  $s_i$  parts, while  $R$ ,  $MP$ ,  $O_i$ , and  $Y_i$  are empty. When a customer demand of type- $i$  product arrives in the system, it triggers a production order of a new part by MP and a demand for the delivery of a type- $i$  finished part to the customer. The new raw part is directly stored in  $R$  while the new demand is stored in  $O_i$  for a deterministic time period equal to the time lag

$L_i$ , after which it will be transferred to the queue  $Y_i$ . This time lag  $L_i$  is the required customer lead-time for product  $i$ . Queues  $X_i$  and  $Y_i$  are linked into a synchronization station: if there are simultaneously a customer demand in  $Y_i$  and an available part in  $X_i$ , this part is immediately delivered to the customer and the demand in  $Y_i$  is satisfied. If there is no available part in  $X_i$ , the customer demand remains waiting in queue  $Y_i$  until the arrival of a new finished product in queue  $X_i$ .

It is worth noting that the time lag  $L_i$  impacts both on the backordered demands queue  $Y_i$  and on the physical inventories  $X_i$ . The global effect of such a time lag is quite intuitive: if the time lag increases, it will progressively become superior to the production lead-time, in such a way that the backordered demands will simultaneously decrease while the finished parts inventory will increase as the parts will be transferred into the physical inventories in advance with respect to the customer due date.

## **2.2. Main Assumptions for the Production/Inventory Model**

The different types of products manufactured in the system are indexed from 1 to  $k$ . The first  $k_{HV}$  indices (i.e.,  $i = 1, \dots, k_{HV}$ ) correspond to HV products while the  $k_{LV}$  remaining indices (i.e.,  $i = k_{HV} + 1, \dots, k_{HV} + k_{LV}$ ) are associated with LV products. We denote by  $\mathcal{C}$ ,  $\mathcal{C}_{HV}$  and  $\mathcal{C}_{LV}$  the sets of indices, respectively, corresponding to all products, to the HV products and to the LV products.

The arrivals of demands occur according to independent Poisson processes, with rate  $\lambda_i$  for product type- $i$ .

Concerning the manufacturing itself, we make the following assumptions. The processing times are exponentially distributed with parameter  $\mu$  and do not depend on the product index. Note that no setup is required to switch from producing one type of product to another, which is the case when products involve similar operations, but supply distinct market segments. Furthermore, the manufacturing process is assumed to be totally reliable.

The base-stocks  $s_i$  are the key parameters to optimize. It should be noted that product type- $i$  is an MTO product if the associated base-stock  $s_i$  is chosen to be equal to zero and an MTS if  $s_i$  is strictly positive.

Furthermore, we consider and compare the performances of two priority rules for production orders waiting for access to the production facility: the well known First-In-First-Out (FIFO) rule and a priority (PR) rule, giving the highest priority to the LV product family over the HV product family. On the other hand, orders which correspond to product types within the same product family (HV or LV) are scheduled among themselves according to the FIFO rule. Furthermore, we assume a preemptive priority rule, that is, if an HV product order is being served when a new order for an LV product enters the system, the current HV product is ejected back in the queue, that is, preempted, and the LV order is produced in priority: this assumption is only required to make the analysis tractable.

## **2.3. The Fundamental Cost/Service Level Formulation and the Associated Notations**

The fundamental optimization problem considered here consists of finding the optimal values for the base-stock levels (under each scheduling rule), which minimize the average holding cost, under a delivery time fill rate constraint for each product type. This constraint corresponds to bounding the probability that an order will be delivered after the required

lead-time. For this formulation, we adopt the following notations:

- $k_{\text{HV}}$ : the number of high volume product types,
- $k_{\text{LV}}$ : the number of low volume product types,
- $k$ : the number of product types (with  $k_{\text{HV}} + k_{\text{LV}} = k$ ),
- $Z$ : the average cost rate for the whole system.

For each product type- $i$ , we define, for  $i \in \mathcal{C}$ ,

- $\lambda_i$ : the average demand rate,
- $h_i$ : the inventory holding cost rate,
- $L_i$ : the required customer lead-time,
- $\gamma_i$ : the effective fill rate, defined as the probability that, for part type- $i$ , the backorder time is equal to zero,
- $\gamma_i^a$ : the required fill rate.

We also introduce

- $1/\mu$ : the average product processing time,
- $\rho$ : the utilization factor of the manufacturing facility, given by  $\rho = \sum_{i \in \mathcal{C}} \lambda_i / \mu$ .

For the HV and LV families, we introduce

the aggregate average demand rates, respectively, given by

$$\lambda_{\text{HV}} = \sum_{i \in \mathcal{C}_{\text{HV}}} \lambda_i, \quad \lambda_{\text{LV}} = \sum_{i \in \mathcal{C}_{\text{LV}}} \lambda_i, \quad (2.1)$$

the aggregate utilization factors, respectively, given by

$$\rho_{\text{HV}} = \sum_{i \in \mathcal{C}_{\text{HV}}} \frac{\lambda_i}{\mu}, \quad \rho_{\text{LV}} = \sum_{i \in \mathcal{C}_{\text{LV}}} \frac{\lambda_i}{\mu}. \quad (2.2)$$

It is also necessary to define

- $s_i$ : the type- $i$  parts base-stock level,

and the following random variables,

- $X_i$ : the number of type- $i$  parts in the finished product inventory,
- $N_i$ : the number of type- $i$  parts in progress in the system,
- $Y_i$ : the number of backordered demands of type- $i$  products,
- $D_i$ : the backorder time for demand of type- $i$  products.

All these random variables are functions of time, depending furthermore on  $s_i$  or other parameters. In the remainder of this paper, when not necessary, we do not explicitly refer to this dependency except for the values of the base-stock levels. For instance, we will use  $X_i(s_i)$

to denote the random variable associated with the number of type- $i$  parts in the finished product inventory, for a base-stock level of  $s_i$  for product  $i$ .

The holding cost/service level formulation for this multiproduct problem can now be expressed as follows:

$$\begin{aligned} \min \quad & Z(s_1, s_2, \dots, s_k) = \sum_{i \in \mathcal{C}} h_i E[X_i(s_1, s_2, \dots, s_k)] \\ \text{s.t.} \quad & \gamma_i(s_1, s_2, \dots, s_k) \geq \gamma_i^a, \\ & s_i \in \mathbb{N}, \quad \forall i \in \mathcal{C}. \end{aligned} \quad (2.3)$$

We denote by  $s_i^*$  the optimal base-stock level for product  $i$  and the associated aggregate optimal cost by

$$Z^* = Z(s_1^*, s_2^*, \dots, s_k^*). \quad (2.4)$$

### 3. Theoretical Analysis for the Considered Problem

#### 3.1. Structural Properties for the Problem

Some important structural properties characterize the problem under consideration. It is worth noting that in fact these properties remain valid even for general arrival and service processes, as it appears in the corresponding proofs.

*Property 1.* For  $i \in \mathcal{C}$ ,

- (1) the random variables  $N_i$  are independent of the base-stock levels  $s_j$  and of the customer lead-times  $L_j$ , for all  $j \in \mathcal{C}$ ,
- (2) the random variables  $X_i, Y_i, D_i$  and the fill rate  $\gamma_i$  are independent of the base-stock levels  $s_j$  and of the lead-times  $L_j$ , for all  $j \in \mathcal{C} \setminus \{i\}$ .

See Appendix A.1 for a sample path proof.

By exploiting this important property, the multiproduct optimization problem (2.3) can be reformulated in a decomposed way as

$$\begin{aligned} \min \quad & Z(s_1, s_2, \dots, s_k) = \sum_{i \in \mathcal{C}} h_i E[X_i(s_i)] \\ \text{s.t.} \quad & \gamma_i(s_i) \geq \gamma_i^a, \quad i \in \mathcal{C}, \\ & s_i \in \mathbb{N}, \quad i \in \mathcal{C}. \end{aligned} \quad (3.1)$$

The impact of the base-stock level  $s_i$  both on the backordered demands queue  $Y_i$  and on the physical inventories  $X_i$  seems quite intuitive: if  $s_i$  increases, the finished parts inventory will simultaneously increase and the backordered demands will decrease. The following property formally establishes this monotone effect of the base-stocks.

*Property 2.* For  $i \in \mathcal{C}$ , one has

$$\begin{aligned}
 E[X_i(s_i + 1)] &> E[X_i(s_i)], & \lim_{s_i \rightarrow \infty} (E[X_i(s_i + 1)] - E[X_i(s_i)]) &= 1, \\
 E[Y_i(s_i + 1)] &< E[Y_i(s_i)], & \lim_{s_i \rightarrow \infty} E[Y_i(s_i)] &= 0, \\
 E[D_i(s_i + 1)] &< E[D_i(s_i)], & \lim_{s_i \rightarrow \infty} E[D_i(s_i)] &= 0, \\
 \gamma_i(s_i + 1) &> \gamma_i(s_i), & \lim_{s_i \rightarrow \infty} \gamma_i(s_i) &= 1.
 \end{aligned} \tag{3.2}$$

See Appendix A.2 for the proof.

**Corollary 3.1.** For every type- $i$  product, the optimal base-stock level  $s_i^*$  can be viewed as the minimum  $s_i$  integer value satisfying the fill rate constraint  $\gamma_i(s_i) \geq \gamma_i^a$ . Furthermore, these  $s_i^*$  are independent of the lead-times  $L_j$ , for all  $j \in \mathcal{C} \setminus \{i\}$ .

*Proof.* This is a direct consequence of the fact that  $E[X_i(s_i)]$  and  $\gamma_i(s_i)$  are increasing functions of  $s_i$  and Properties 1 and 2.  $\square$

Let us denote by  $s_{HV_i}^{*FIFO}$ ,  $s_{LV_i}^{*FIFO}$ ,  $s_{HV_i}^{*PR}$ , and  $s_{LV_i}^{*PR}$  the optimal base-stock levels associated with HV and LV products (and the corresponding customer lead-times) and with the scheduling rules FIFO and PR. The following property formally summarizes two very intuitive properties of these optimal base-stock levels.

*Property 3.* (1) The optimal base-stock levels are ordered as follows:

$$s_{HV_i}^{*PR} \geq s_{HV_i}^{*FIFO}, \quad s_{LV_j}^{*FIFO} \geq s_{LV_j}^{*PR}, \tag{3.3}$$

with  $i \in \mathcal{C}_{HV}$  and  $j \in \mathcal{C}_{LV}$ .

(2) For all product types, the optimal base-stock levels, corresponding to FIFO and PR, are decreasing functions of the corresponding customer lead-times  $L_j$ .

See Appendix A.3 for the proof.

In the remainder of the paper, when necessary for the sake of clarity, we will explicitly express the link between optimal base-stock levels and the customer lead-times by using the notation  $s_i^*(L_i)$ .

### 3.2. Critical Customer Lead-Time and Optimality of Switching from MTO to MTS

Along the same lines as Appendix A.2, it is easily seen that each fill rate  $\gamma_i$  is an increasing function of the associate customer lead-time. As appearing in the following sections, there exist critical required customer lead-time values which are of first importance in the analysis of the considered problem. These values correspond to the particular values for which the fill rate constraints can be satisfied under an MTO policy (i.e., with zero base-stocks). More precisely, if, for a given product type, the required customer lead-time is greater than this critical value, the fill rate constraint will be satisfied even with a zero base-stock. Conversely,

if the required customer lead-time is lower than the critical value, then it will be necessary to introduce a nonzero base-stock to guarantee the fill rate constraint.

*Definition 3.2.* For  $i \in \mathcal{C}_{HV}$  and for  $j \in \mathcal{C}_{LV}$ , let us define  $\tilde{L}_{HV_i}^{FIFO}$ ,  $\tilde{L}_{LV_j}^{FIFO}$ ,  $\tilde{L}_{HV_i}^{PR}$ , and  $\tilde{L}_{LV_j}^{PR}$  as the minimal values of the required customer lead-times such that

under FIFO,

$$s_{HV_i}^{*FIFO}(L_i) = 0 \quad \text{for } L_i \geq \tilde{L}_{HV_i}^{FIFO}, \quad s_{LV_j}^{*FIFO}(L_j) = 0 \quad \text{for } L_j \geq \tilde{L}_{LV_j}^{FIFO}, \quad (3.4)$$

under PR,

$$s_{HV_i}^{*PR}(L_i) = 0 \quad \text{if } L_i \geq \tilde{L}_{HV_i}^{PR}, \quad s_{LV_j}^{*PR}(L_j) = 0 \quad \text{if } L_j \geq \tilde{L}_{LV_j}^{PR}. \quad (3.5)$$

The following property is a key property in this paper, as it underlies the structure of the optimal MTO/MTS policy for the HV and LV classes.

*Property 4.* If we assume that  $\gamma_i^a = \gamma^a$ , for  $i \in \mathcal{C}$ , and that the demand processes are Poisson, then

(1) for  $i \in \mathcal{C}_{HV}$  and for  $j \in \mathcal{C}_{LV}$ , one has

$$\tilde{L}_{HV_i}^{FIFO} = \tilde{L}_{LV_j}^{FIFO} = \tilde{L}^{FIFO}, \quad \tilde{L}_{HV_i}^{PR} = \tilde{L}_{HV}^{PR}, \quad \tilde{L}_{LV_j}^{PR} = \tilde{L}_{LV}^{PR}, \quad (3.6)$$

(2) the three critical values  $\tilde{L}_{HV}^{PR}$ ,  $\tilde{L}^{FIFO}$ , and  $\tilde{L}_{LV}^{PR}$  are ordered as follows:

$$\tilde{L}_{LV}^{PR} < \tilde{L}^{FIFO} < \tilde{L}_{HV}^{PR}. \quad (3.7)$$

*Proof.* See Appendix A.4. □

A remaining difficulty, addressed later in the paper, is the numerical computation of these critical values, which in summary require the numerical solution of the queuing model depicted in Figure 1.

As the HV and LV items represent typically two specific market segments, it is quite natural to consider segment specific lead-time values: we thus assume that the customer lead-time is  $L_{HV}$  for all HV products and  $L_{LV}$  for all LV products. For the remainder of this section, we assume  $\gamma_i^a = \gamma^a$ , for  $i \in \mathcal{C}$ , that is, we assume similar fill rate objectives for the different products. As a consequence, we will eliminate the index “ $i$ ” in the threshold level expressions.

When switching from FIFO to PR, the consequence on the optimal MTO/MTS policy can thus be summarized as in Tables 1 and 2.

For customer lead-time values in ranges 1 and 4, the priority rule has no impact on the optimal policy, which is, respectively, MTS or MTO. Indeed, the customer lead-time in range 1 is so small that even a priority rule does not even induce a sufficient reactivity to produce any

**Table 1:** Optimality of MTO/MTS for the LV family.

$L_{LV}$ values	→				
	0	$\tilde{L}_{LV}^{PR}$	$\tilde{L}_{LV}^{FIFO}$	$\tilde{L}_{LV}^{PR}$	$\infty$
	Range 1	Range 2	Range 3	Range 4	
<b>FIFO</b>	MTS	MTS	MTO		
<b>PR</b>	MTS	MTO	MTO		
Impact of switching from <b>FIFO to PR</b>	No effect	LV products switch from MTS to MTO	No effect		

**Table 2:** Optimality of MTO/MTS for the HV family.

$L_{HV}$ values	→				
	0	$\tilde{L}_{LV}^{PR}$	$\tilde{L}_{LV}^{FIFO}$	$\tilde{L}_{HV}^{PR}$	$\infty$
	Range 1	Range 2	Range 3	Range 4	
<b>FIFO</b>	MTS		MTO	MTO	
<b>PR</b>	MTS		MTS	MTO	
Impact of switching from <b>FIFO to PR</b>	No effect		HV products switch from MTO to MTS	No effect	

item under MTO while still guaranteeing the required customer lead-time. On the contrary, in range 4, the customer lead-time is so high that even under FIFO, it is possible to produce all the items under MTO and simultaneously satisfy the customer lead-time service constraint. In range 2, even if the customer lead-time is still low, a PR scheduling rule induces a sufficient reactivity for LV items, in such a way that they can be produced under MTO (while HV items are MTS). In range 3, the LV customer lead-time is high enough to induce an MTO policy for these items, either under FIFO or PR. For the HV items, with a FIFO rule, the required customer lead-time can be satisfied, while this would not longer be the case with a PR rule, due to the corresponding reactivity loss for the HV class.

It is also important to consider a global cost perspective, usually the most important criterion. The remainder part of the paper focuses on this issue.

## 4. Performance Analysis and Optimization

In this section, we propose procedures in order to compute the optimal base-stock levels, the customer lead-time critical values defining the threshold structure and the associated costs for the considered scheduling rules FIFO and PR. Under FIFO, we derive exact analytical expressions, whereas under PR, as no closed-form formulas are available, we develop a simple analytical approximation.

### 4.1. The FIFO Rule: An Exact Analytical Approach

*Property 5.* Under FIFO, the multiproduct queuing model of the BSCS depicted in Figure 1 can be decomposed into  $k$  independent single-product time-lagged models with a service rate given by

$$\begin{aligned}\hat{\mu}_{LV_i}^{\text{FIFO}} &= \mu - \sum_{j \in \mathcal{C}, j \neq i} \lambda_j, \quad \text{for } i \in \mathcal{C}_{LV}, \\ \hat{\mu}_{HV_i}^{\text{FIFO}} &= \mu - \sum_{j \in \mathcal{C}, j \neq i} \lambda_j, \quad \text{for } i \in \mathcal{C}_{HV}.\end{aligned}\tag{4.1}$$

See Appendix B.1 for the proof.

This property states that, in fact, for each independent single-product time-lagged model  $i$ , the corresponding service rate can be viewed as the initial service rate  $\mu$  minus the workload rate induced by the other product type demand rates. Let us remark that the corresponding utilization factor for product type- $i$  is given directly by

$$\hat{\rho}_i = \frac{\lambda_i}{\hat{\mu}_i}.\tag{4.2}$$

**Corollary 4.1.** *The expression of the fill rate is given by*

$$\gamma_i(s_i) = \Pr\{D_i = 0\} = 1 - \hat{\rho}_i^{s_i} e^{-(\hat{\mu}_i - \lambda_i)L_i}.\tag{4.3}$$

See Appendix B.2 for the proof.

### 4.2. The PR Rule: A Heuristic Approach

Under the PR scheduling rule, the explicit decomposition property exhibited in Section 4.1 no longer holds. We thus propose a new decomposition scheme.

First, we can observe that under PR, the production of LV orders is not affected by HV orders. Along the same lines as in Section 4.1, the LV family can thus be analyzed by decomposition of the associated LV multiproduct problem into  $k_{LV}$  independent single-product subproblems, with equivalent service rates given by

$$\hat{\mu}_{LV_i}^{\text{PR}} = \mu - \sum_{j \in \mathcal{C}_{LV}, j \neq i} \lambda_j, \quad \text{for } i \in \mathcal{C}_{LV}.\tag{4.4}$$

Second, the analysis of the HV family is more complex because the production facility becomes unavailable for HV products each time an LV order is issued. As the demand process for each LV product is a Poisson process, the rate at which the production facility becomes unavailable for the HV family is given by  $\lambda_{LV}$ , which is the aggregate arrival rate for the LV products. Once unavailable, the process remains in this state of unavailability during a random time period which can be viewed as the busy period of a production facility loaded exclusively by the LV family. Due to the assumptions of this paper, this corresponds to the busy period of an M/M/1 queue. As no closed-form formulas exist, we propose a simple heuristic which approximates the random  $HV_i$  item service time by a simple exponentially distributed random variable, with a rate denoted by  $\hat{\mu}_{HV_i}^{PR}$ . In order to be consistent with the global multiproduct model, we choose this rate in such a way that the average sojourn time in this fictitious queue is equal to the average sojourn time for the real system, which via [4] can be computed as

$$W_{HV_i} = \frac{1}{\mu - \lambda_{HV} - \lambda_{LV}(2 - \rho)}. \quad (4.5)$$

The quality of this heuristic approximation has been numerically validated in [10]. Under this exponential approximation, the average sojourn time associated with a production rate  $\hat{\mu}_{HV_i}^{PR}$  is known to be given by

$$\frac{1}{\hat{\mu}_{HV_i}^{PR} - \lambda_{HV_i}}. \quad (4.6)$$

By equating (4.5) and (4.6), we find

$$\hat{\mu}_{HV_i}^{PR} = \mu - \lambda_{HV} - \lambda_{LV}(2 - \rho) + \lambda_{HV_i}, \quad \text{for } i \in \mathcal{C}_{HV}. \quad (4.7)$$

This sojourn time is the same for every HV type item, which stems from the fact that the demand processes are Poisson and that the production rate is the same for all product types.

*Property 6.* Let us denote by  $\hat{\mu}_{LV_i}^{FIFO}$  ( $i \in \mathcal{C}_{LV}$ ) and  $\hat{\mu}_{HV_j}^{FIFO}$  ( $j \in \mathcal{C}_{HV}$ ), respectively, the equivalent production rates of an LV and an HV item under the FIFO rule. We then have

$$\begin{aligned} \hat{\mu}_{LV_i}^{PR} &= \hat{\mu}_{LV_i}^{FIFO} + \lambda_{HV}, \\ \hat{\mu}_{HV_j}^{PR} &= \hat{\mu}_{HV_j}^{FIFO} - \lambda_{LV}(1 - \rho). \end{aligned} \quad (4.8)$$

*Proof.* This is a direct application of (4.1) and (4.4)–(4.7).  $\square$

This property exhibits the following interesting insight concerning the impact of the priority rule when switching from FIFO to PR:  $\lambda_{HV}$  can be interpreted as the reactivity increase for every  $LV_i$  item and  $\lambda_{LV}(1 - \rho)$  as the reactivity decrease for every  $HV_j$  product.

### 4.3. Computation of the Optimal Costs

Through the decomposition property, the aggregate problem (2.3) can be reformulated as a set of single part type problems with time lags. Such single part type models have been extensively studied in [3] and by applying the formulas characterizing the expected inventories  $E[X_i]$ , one finds

$$\begin{aligned} \min \quad & Z_i(s_i) = h_i \left( s_i + \lambda_i L_i - \frac{\hat{\rho}_i}{1 - \hat{\rho}_i} \left( 1 - \hat{\rho}_i^{s_i} e^{-(\hat{\mu}_i - \lambda_i)L_i} \right) \right) \\ \text{s.t.} \quad & \gamma_i(s_i) \geq \gamma_i^a, \\ & s_i \in \mathbb{N}, \quad i \in \mathcal{C}, \end{aligned} \tag{4.9}$$

here  $\hat{\mu}_i$  is calculated from (4.1) or (4.4)–(4.7) depending on the product  $i$ -type (i.e.,  $i \in \mathcal{C}_{\text{HV}}$  or  $i \in \mathcal{C}_{\text{LV}}$ ) and the priority rule (FIFO or PR). The corresponding  $\hat{\rho}_i$  is calculated from (4.2).

Using Corollary 3.1, the optimal base-stock level  $s_i^*$ , corresponding to the optimal solution of each of these optimization problems, can be viewed as the minimum  $s_i$  integer value satisfying the fill rate constraint  $\gamma_i(s_i) \geq \gamma_i^a$ . By Corollary 4.1, a direct solution of the equation  $\gamma_i(s_i) = \gamma_i^a$  amounts to solving

$$1 - \hat{\rho}_i^{s_i} e^{-(\hat{\mu}_i - \lambda_i)L_i} = \gamma_i^a. \tag{4.10}$$

Via some simple mathematical reformulations and due to the integrity constraint,  $s_i^*$  is easily found to be given by

$$s_i^* = \max \left( \left\lceil \frac{\ln(1 - \gamma_i^a)}{\ln \hat{\rho}_i} + \frac{(\hat{\mu}_i - \lambda_i)L_i}{\ln \hat{\rho}_i} \right\rceil, 0 \right). \tag{4.11}$$

The optimal aggregate average cost rate for the whole system, is, therefore, given by

$$Z^* = \sum_{i=1}^k Z_i(s_i^*). \tag{4.12}$$

### 4.4. Computation of the Threshold Structure Values

The following properties formally characterize this threshold structure of the optimal MTO/MTS decision by giving their analytical expression.

*Property 7.* The threshold levels characterizing the optimal MTO/MTS policy can be analytically computed as, for  $j \in C_{HV}$  and  $i \in C_{LV}$ ,

$$\tilde{L}_{LV_i}^{PR} = \frac{-\ln(1 - \gamma_i^a)}{\mu - \lambda_{LV}}, \quad (4.13)$$

$$\tilde{L}_{LV_i}^{FIFO} = \frac{-\ln(1 - \gamma_i^a)}{\mu - \lambda_{LV} - \lambda_{HV}}, \quad \tilde{L}_{HV_j}^{FIFO} = \frac{-\ln(1 - \gamma_j^a)}{\mu - \lambda_{LV} - \lambda_{HV}}, \quad (4.14)$$

and by the exponential approximation, we find

$$\tilde{L}_{HV_j}^{PR} \cong \frac{-\ln(1 - \gamma_j^a)}{\mu - \lambda_{LV} - \lambda_{HV} - \lambda_{LV}(1 - \rho)}. \quad (4.15)$$

*Proof.* See Appendix B.3. □

It is worthwhile to note that, except for the required fill rate  $\gamma_i^a$ , this threshold structure does not depend on the product index. On the other hand, for given aggregate utilization factor  $\rho$  and family demand rate values  $\lambda_{LV}$  and  $\lambda_{HV}$ , the threshold structure does not depend neither on  $k_{LV}$  nor on  $k_{HV}$ .

## 5. Cost Optimal Scheduling Rules: Some Qualitative Insights and General Properties

The present section focuses on the performances of the scheduling rules in terms of global costs. To compare the costs associated to each scheduling rule (FIFO and PR), each using the corresponding optimal base-stock levels, we define the gain  $G$ , or percent cost reduction when PR is used, as

$$G = \frac{Z^{*FIFO} - Z^{*PR}}{Z^{*FIFO}} * 100, \quad (5.1)$$

where  $Z^{*FIFO}$  is the optimal cost for the FIFO rule and  $Z^{*PR}$  the optimal cost for the PR rule, computed by (4.4). A positive gain means that the PR scheduling rule is more efficient than the FIFO rule, while a negative gain means that FIFO outperforms PR. Throughout this section, the theoretical results will be illustrated by means of a numerical example.

### 5.1. Main Intuition

Globally the structure of efficient MTS/MTO decision and efficient scheduling rule selection remains a complex issue. However, some insight can be provided as follows. From an intuitive point of view, two parameters are clearly of first importance and underly the exhibited theoretical properties: the required customer lead-times and the number of LV products (or more precisely the ratio  $k_{LV}/k_{HV}$ ).

As a matter of fact, under the threshold structure described in Section 3.2, the required lead-times determine the optimal MTO/MTS policy as a function of the scheduling rule, and as a consequence, also determines the involved inventory costs.

Secondly, in the two-segment model under consideration, the inventory costs for the HV items and for the LV items have a very different characteristic. The HV cost variation between FIFO and PR is independent of the number of LV items. This is not the case, in general, for the LV items. As the number of LV products is assumed to be large, even slight individual cost variation for each LV item can amount to very high cost variation for the whole LV family. For example, under MTS, the minimal base-stock for each LV item is by definition an integer, namely, 1 unit, even if the LV demand rates are extremely low. As a result, the switching from MTS to MTO (i.e., to zero base-stocks) for LV products has a major impact on inventory costs for sufficiently large  $k_{LV}$  values. Thus, the optimal policy seems to be determined by the LV family characteristics for sufficiently large values of  $k_{LV}/k_{HV}$ . By exploiting this important difference, we will exhibit some fundamental properties.

Before giving these theoretical properties, we provide some insight into the structure of this gain function through a qualitative analysis of a numerical example.

## 5.2. A Numerical Example

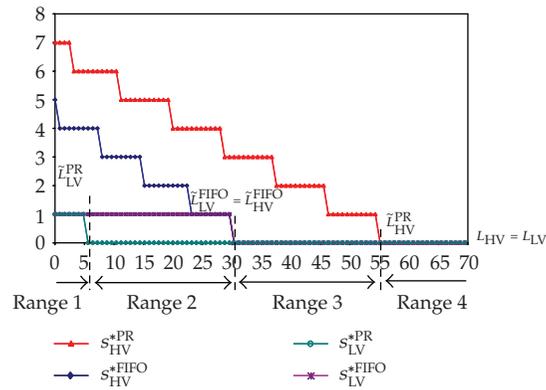
In this example, there are five high volume products with  $\lambda_i = 0.09$  (for  $i = 1, 2, \dots, 5$ ) and 100 low volume products with  $\lambda_i = 0.0045$  (for  $i = 6, 7, \dots, 105$ ). The production rate is  $\mu = 1$ , so that the utilization factor is 90%, equally balanced between the LV and HV products, as  $\rho_{HV} = \rho_{LV} = 0.45$ . This balanced situation illustrates the general theoretical properties established in this paper best. As a matter of fact, when  $\rho_{HV} \gg \rho_{LV}$  or  $\rho_{HV} \ll \rho_{LV}$ , the underlying capacity-sharing priority problem is of minor interest. The holding cost rates are identical, that is,  $h_i = h = 1$ . Furthermore, we assume that the required customer lead-time is the same for all the items, that is, along preceding assumptions, we have  $L_{HV} = L_{LV}$ . The required fill rates are identical and given by  $\gamma_i^a = 0.98$ .

Some intuitive insight can be deduced from the numerical example in Figure 2. Let us first consider range 4 where, as the customer lead-times are very high, all products must be MTO under FIFO and PR. However, due to the base-stock policy used, when these lead-times increase the orders are completed before the requisition time, leading to strictly positive inventory costs. In general in this case, which is in fact of minor interest, FIFO (slightly) outperforms PR.

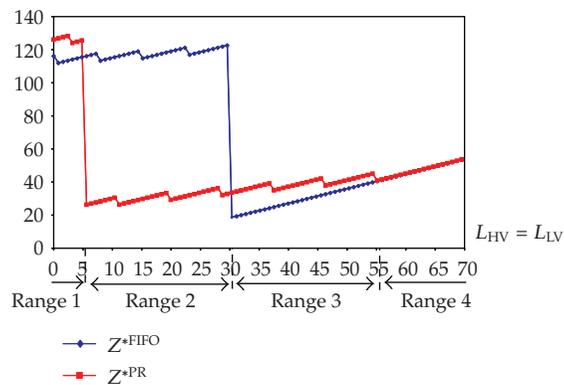
In range 3, the optimal cost associated with the FIFO is lower than the cost with the PR rule. As a matter of fact, switching from FIFO to PR increases the HV base-stocks (as  $s_{HV}^{*PR} > 0$  while  $s_{HV}^{*FIFO} = 0$ ) without decreasing the LV base-stocks as  $s_{LV}^{*FIFO} = s_{LV}^{*PR} = 0$ . The gain  $G$  is thus strictly negative in this range and the FIFO rule outperforms the PR rule.

Let us now consider range 2. In this case, we can see in Figure 2(a) that switching from FIFO to PR results in a decrease of one unit for the LV optimal base-stock levels (i.e.,  $s_{LV}^{*FIFO} = 1$  while  $s_{LV}^{*PR} = 0$ ). As  $k_{LV} = 100$ , this corresponds to an overall decrease of 100 units for the LV products. At the same time, switching from FIFO to PR is associated with an increase of the HV optimal base-stock level. If we consider, as an example, the particular value  $L = 10$ , this increase is equal to 15 (i.e.,  $s_{HV}^{*FIFO} = 3$  and  $s_{HV}^{*PR} = 6$ ) and the resulting gain for the whole system is positive.

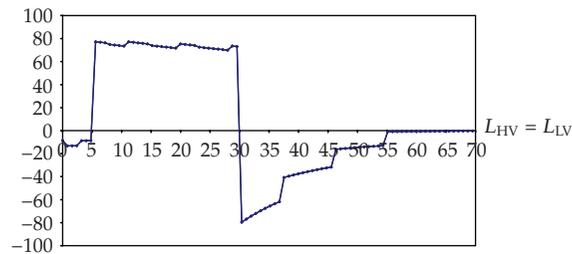
If we consider range 1, we can see that the gain is negative. Consider, as an example, zero lead-times, we can observe that the priority rule has no effect on the optimal base-stock



(a) Optimal base-stock levels



(b) Optimal costs



(c) The gain curve (%)

Figure 2: The numerical example.

levels for the LV products, (i.e.,  $s_{LV}^{*FIFO} = s_{LV}^{*PR} = 1$ ), while the HV optimal base-stock level increases ( $s_{HV}^{*FIFO} = 5$  and  $s_{HV}^{*PR} = 7$ ). In this case, the gain for the whole system is negative.

In summary, we can see that range 2 exhibits a specific effect with respect to the other ranges: for any LV customer lead-time in this range 2, switching from FIFO to PR results in a one unit decrease of the base stock for each LV item. As it is considered here that the number of LV items is huge, this has a major impact. In the other ranges, no such phenomenon can be observed and thus FIFO seems to outperform PR, except maybe for pathological numerical examples.

### 5.3. Fundamental Properties and Heuristic Insights

It is worth noting that the range for which the switching from FIFO to PR has a major impact is the range 2. In the other ranges, the difference between the two scheduling rules is not so significant. We thus propose the following property for range 2.

*Property 8.* For given  $\rho$ ,  $\rho_{LV}$ , and  $\rho_{HV}$  values, when the required LV customer lead-time satisfies

$$L_{LV} \in \left[ \tilde{L}_{LV}^{PR}, \tilde{L}_{LV}^{FIFO} \right] \quad (5.2)$$

for sufficiently large values of  $k_{LV}/k_{HV}$ , PR outperforms FIFO, and under PR, LV items are MTO. Furthermore, the gain converges toward 100% when

$$\frac{k_{LV}}{k_{HV}} \rightarrow \infty. \quad (5.3)$$

*Proof.* See Appendix C. □

This property stems from the fact that when switching from FIFO to PR, the cost increase for HV is independent of  $k_{LV}$ , while the cost decrease for LV is an increasing function of  $k_{LV}$  for large values.

Concerning the other ranges, exact theoretical properties appear to be tedious to prove, even if some strong intuitive ideas underly the structure of the solution. We propose thus in the following, intuitive (and heuristic) explanations for FIFO/PR optimality in other ranges.

First, we can remark that, when the LV demand rates are small, that is, when  $\lambda_i \rightarrow 0$ , for  $i \in \mathcal{C}_{LV}$ , the switching from FIFO to PR effectively reduces the LV base stocks only in range 2, that is, if  $L_{LV} \in [\tilde{L}_{LV}^{PR}, \tilde{L}_{LV}^{FIFO}]$ . For other values of the LV customer lead-times, the switching does not reduce the LV base-stock levels. On the other hand, it will potentially increase the HV base-stocks levels, due to the HV reactivity loss. In particular, this increase can be very important for high utilization rate (i.e., when  $\rho \rightarrow 1$ ) when  $L_{HV_i} \in [0, \tilde{L}_{HV}^{PR}]$ . So, clearly, as far as the base-stock levels are concerned, PR outperforms FIFO only when  $L_{LV} \in [\tilde{L}_{LV}^{PR}, \tilde{L}_{LV}^{FIFO}]$  and FIFO outperforms PR in the other cases.

### 5.4. Extensions

In this section, we use the performance analysis results developed in Section 4 in order to provide a sensitivity analysis with respect to the different parameters of the problem. We still consider the same numerical example as the basic scenario. In particular, we still assume  $L_{LV} = L_{HV}$ .

#### 5.4.1. Impact of the Utilization Factor

Figure 3 depicts the gain curve, as a function of the customer lead-time, for different utilization factor values.

This example shows that when the utilization factor increases, the threshold levels also increase. This result seems quite intuitive. Indeed, when the utilization factor increases, the

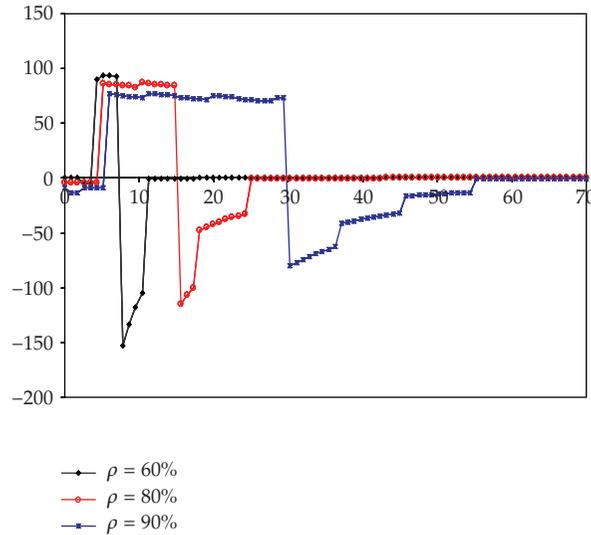


Figure 3: Gain curves for different utilization factor values.

waiting times and cycle times increase, inducing longer delivery lead-times. Optimality of a stockless inventory-production system requires then longer customer lead-times. For the example, we see that for  $\rho = 60\%$ , a PR policy will be optimal when the customer lead-time value is between 3 and 6. However, if  $\rho = 90\%$ , this customer lead-time has to lie between 5 and 30 to induce PR optimality.

It is worth noting that one can show that range 2 increases when  $\rho \rightarrow 1$ . However, the gain appears to be decreasing.

#### 5.4.2. The Case $\rho_{HV} \neq \rho_{LV}$

Up to now, we have assumed in the numerical examples that  $\rho_{HV} = \rho_{LV} = 45\%$ . We consider here a couple of cases:  $\rho_{HV} = 80\%$  and  $\rho_{LV} = 10\%$ , and  $\rho_{HV} = 10\%$  and  $\rho_{LV} = 80\%$ , with the parameters,  $\mu = 1$  and  $k_{HV} = 5$ ,  $k_{LV} = 100$ , with the corresponding demand rate values.

Figures 4 and 5 compare the basic example with these two cases.

These curves show that the threshold levels  $\tilde{L}_{LV}^{FIFO}$  and  $\tilde{L}_{HV}^{FIFO}$  are independent from the  $\rho_{HV}$  and  $\rho_{LV}$  values and in fact depend on the global utilization factor  $\rho$ . On the contrary, the threshold levels  $\tilde{L}_{LV}^{PR}$  and  $\tilde{L}_{HV}^{PR}$  are increasing functions respectively of  $\rho_{LV}$  and  $\rho_{HV}$ . This is a consequence of (4.12) and (4.13).

It is interesting to note that the gain in range 2 increases when  $\rho_{LV}$  decreases. There is a reactivity gain for the HV products if  $\rho_{LV}$  decreases. Thus, the optimal base-stock level increase will be less considerable. Simultaneously, in this range, the gain in terms of inventory costs for the LV family is associated to the switching of the LV base-stock levels from one unit to zero. The overall gain for the LV family remains the same, for a minor loss for the HV family, which explains the result.

A similar explanation holds in range 3 and justifies the decrease of the cost  $Z^{*PR}$  with respect to  $Z^{*FIFO}$  when the utilization factor  $\rho_{LV}$  decreases. This explains why the overall gain is less negative in this range.

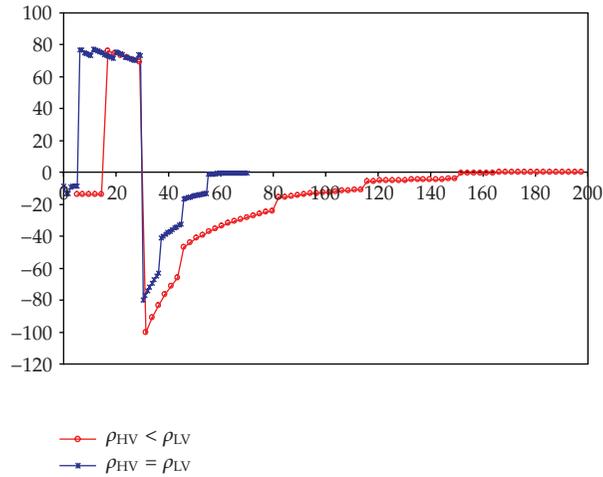


Figure 4: The gain curves if  $\rho_{HV} > \rho_{LV}$ .

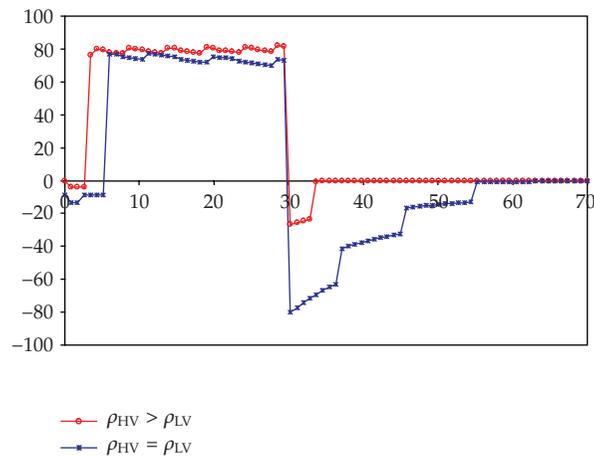


Figure 5: The gain curves if  $\rho_{HV} < \rho_{LV}$ .

#### 5.4.3. The Impact of Holding Costs

First, it can be noted that if the LV holding cost rates are greater than the HV holding cost rates, the gain in range 2 will be more significant. Furthermore, in the model considered here (and in the cited references), the inventory costs are assumed to be proportional to the (value of) inventories. Such models do not include the fixed charges associated to the presence of a strictly positive inventory for a given item. Indeed, having a product that is make-to-stock has an underlying cost, regardless of the value of the corresponding base stock level.

Now, it is easy to incorporate such fixed costs in the comparison described before. It is interesting to note that these fixed costs will give more advantages to a scheduling policy that would have less MTS products. In particular, the superiority of the PR policy in range 2 would become even more important. Indeed, the difference in costs between the PR and the FIFO policy would be increased by the fixed costs times the number of LV products.

### **5.5. Managerial Insight: A Synthesis**

In a multiproduct setting, the structure of efficient MTS/MTO decision and efficient scheduling rule selection is a complex issue. In the two-segment model under consideration here, the number of LV products is assumed to be very large with respect to the number of HV products. As a consequence, even slight individual cost variation for each LV item can amount to very high cost variation for the whole LV family. In particular, under MTS, the minimal base-stock for each LV item is by definition an integer, namely, 1 unit, even if the LV demand rates are extremely low. As a result, the switching from MTS to MTO (i.e., to zero base-stocks) for LV products has a major impact on inventory costs.

#### *Strategic Ranges for the Customer Lead-Times*

Depending on the customer lead-times values, the optimal solutions exhibit different structures. For extreme values (i.e., if the customer lead-times are very high or very low), the inventory policy is fixed (resp., MTO or MTS) and FIFO and PR exhibit similar cost performances. For intermediate values of the customer lead-times, we have two possibilities. For some range, the optimal cost associated with the FIFO is lower than the cost with the PR rule. This arises from the fact that switching from FIFO to PR increases the HV base-stocks without decreasing the LV base-stocks. On the contrary, for some range, switching from FIFO to PR results in a decrease for all the LV optimal base-stock levels, associated with a slight increase of the HV optimal base-stock level. As it is considered here that the number of LV items is huge, this has a major impact and in this particular range PR outperforms FIFO.

#### *Impact of Utilization Factor*

It is a standard result that when the utilization factor grows in a queuing system, the congestion effects become larger and the impact of the scheduling policy is potentially more significant.

#### *Impact of the Holding Costs*

First, as the number of LV products is assumed to be very large with respect to the number of HV products, changes in the LV holding cost rates have a potential impact which is quite larger than changes in the HV holding cost rates. Furthermore, if a fixed charge is associated to the presence of any strictly positive inventory, then a scheduling policy that would have significantly less MTS products will become optimal. The superiority of the PR policy in the critical range would become even more important due to the large number of MTO-LV products.

## **6. Conclusion**

This paper has analyzed the performance of a finite single-stage manufacturing facility producing multiple heterogeneous items that fall into two product families: high volume and low volume products. The production orders are issued according to a base-stock level policy and scheduled following a FIFO rule or a preemptive priority rule (PR), giving the highest priority to low volume items.

We have exhibited the structure of the optimal policy (or the optimal pairs), that is, the optimal flow control rule (MTO or MTS) and the associated scheduling rule (FIFO or PR), as a function of the customer lead-time. We have shown that the range of values of this lead-time can be divided into intervals in which the optimal pair is constant: for each family it is resolved if the optimal approach consists of either introducing high inventory levels or changing the scheduling rule in order to give a higher priority to this family. Several theoretical properties characterizing this structure of the optimal policy have been developed in this paper.

Through numerical examples we have illustrated these results and shown that the potential benefit of an optimal policy appears to be significant for practical problems.

Some interesting extensions of the formalism presented in this paper can be considered. One could consider batch demands, which is a demand corresponding to a (constant or random) number of products of the same type. In such a case, it is expected that the benefits coming from using a priority policy over a FIFO rule would be even higher. One could also consider the introduction of setup costs or changeover times in the formulation. In that case, one expects to get the opposite effect, because, in such a case, small batch sizes induce supplementary costs. It would also be interesting to analyze how the results exhibited here apply to production/inventory systems with advance demand information and timely delivery requirements, as described in [5]. Lastly, in practice, we are often faced to problems which are less structured than the model considered here. For example, the demand rates, the costs, and the service level can be different for each item. Implementation of heuristics (or optimal approach) in order to efficiently split the items into two (or more) priority classes remains a subject of ongoing research and a first working paper about this issue has been published (see [11]).

## Appendices

### A. Structural Properties of the Problem

#### A.1. Proof of Property 1

Let us consider, for  $t \geq 0$ , a general sample path  $P$ , for the considered time-lagged BSCS, with the corresponding variables  $N_{P,i}(t)$ ,  $X_{P,i}(s_i, t)$ ,  $Y_{P,i}(s_i, t)$ , and  $D_{P,i}(s_i, t)$ . For this sample path, let us introduce the following notations:

- (i)  $A_{P,i}(t)$ : the corresponding total number of type- $i$  demands arrived in the time interval  $[0, t]$ , which by definition have to be delivered in the time interval  $[L_i, t+L_i]$ ,
- (ii)  $B_{P,i}(t)$ : the total number of type- $i$  parts produced in the time interval  $[0, t]$ ,
- (iii)  $tA_{P,i,n}$ : the arrival time of the  $n$ th order of type- $i$  products,
- (iv)  $tB_{P,i,n}$ : the completion time of the  $n$ th part of type- $i$  products,
- (v)  $D_{P,i,n}$ : the backorder time for of the  $n$ th demand of type- $i$  products.

By definition, along any sample path  $P$ , we have

$$N_{P,i}(t) = A_{P,i}(t) - B_{P,i}(t). \quad (\text{A.1})$$

Note first that  $A_{P,i}(t)$  exclusively depends on the product type- $i$  arrival process, and not on other parameters. Furthermore, due to the control policy,  $B_{P,i}(t)$  exclusively depends on

the different arrival processes for all the products (which directly trigger the production orders) and on the production process (which realizes these orders). Thus, under the mild assumption of existence of a steady-state pdf for  $N_i(\cdot)$ , first part of Property 1 is established.

Furthermore, by definition, we also have, for  $i \in \mathcal{C}$ ,

$$\begin{aligned} X_{P,i}(s_i, t) &= \text{Max}\{0, [s_i + B_{P,i}(t)] - A_{P,i}(t - L_i)\}, \\ Y_{P,i}(s_i, t) &= \text{Max}\{0, A_{P,i}(t - L_i) - [s_i + B_{P,i}(t)]\}, \\ D_{P,i,n}(s_i) &= \text{Max}\{0, tB_{P,i,n-s_i} - [tA_{P,i,n} + L_i]\} \end{aligned} \quad (\text{A.2})$$

and by the definition of the fill rate, under the assumption of existence of steady-state probability distribution functions for  $X_i(s_i, \cdot)$ ,  $Y_i(s_i, \cdot)$ , and  $D_i(s_i)$ , we directly have the second part of Property 1.

### A.2. Proof of Property 2

Let  $Z_{P,i}(s_i, t)$  be the corresponding surplus level (which can be positive or negative) for type- $i$  items. This variable is defined by

$$Z_{P,i}(s_i, t) = [s_i + B_{P,i}(t)] - A_{P,i}(t - L_i). \quad (\text{A.3})$$

We clearly have the following equations:

$$Z_{P,i}(s_i + 1, t) = Z_{P,i}(s_i, t) + 1, \quad (\text{A.4})$$

$$X_{P,i}(s_i, t) = \text{Max}\{0, Z_{P,i}(s_i, t)\} = \text{Max}\{0, [s_i + B_{P,i}(t)] - A_{P,i}(t - L_i)\}, \quad (\text{A.5})$$

$$Y_{P,i}(s_i, t) = \text{Max}\{0, -Z_{P,i}(s_i, t)\} = \text{Max}\{0, -[s_i + B_{P,i}(t)] + A_{P,i}(t - L_i)\}. \quad (\text{A.6})$$

By the definition (A.5) and under a general ergodicity assumption for the surplus process, we find

$$\begin{aligned} E[X_i(s_i + 1)] &> E[X_i(s_i)], & \lim_{s_i \rightarrow \infty} (E[X_i(s_i + 1)] - E[X_i(s_i)]) &= 1, \\ E[Y_i(s_i + 1)] &< E[Y_i(s_i)], & \lim_{s_i \rightarrow \infty} E[Y_i(s_i)] &= 0. \end{aligned} \quad (\text{A.7})$$

By definition, one has

$$tB_{P,i,n-s_i-1} < tB_{P,i,n-s_i}, \quad (\text{A.8})$$

for  $t > 0$ , we have

$$\begin{aligned} D_{P,i,n}(s_i + 1) &= D_{P,i,n}(s_i) = 0 & \text{if } tB_{P,i,n-s_i} \leq tA_{P,i,n} + L_i, \\ D_{P,i,n}(s_i + 1) &< D_{P,i,n}(s_i) & \text{if } tB_{P,i,n-s_i} > tA_{P,i,n} + L_i. \end{aligned} \quad (\text{A.9})$$

Due to ergodicity property underlying the whole system, by taking the expectation, we find

$$E[D_i(s_i + 1)] < E[D_i(s_i)], \quad \lim_{s_i \rightarrow \infty} E[D_i(s_i)] = 0. \quad (\text{A.10})$$

By the definition of the fill rate  $\gamma_i$ , we directly have

$$\gamma_i(s_i + 1) > \gamma_i(s_i), \quad \lim_{s_i \rightarrow \infty} \gamma_i(s_i) = 1. \quad (\text{A.11})$$

### A.3. Proof of Property 3

Let us introduce the following:

- (i)  $tB_{P,i,n}^{\text{FIFO}}$ : the random completion time of the  $n$ th production order of type- $i$  products under FIFO,
- (ii)  $tB_{P,i,n}^{\text{PR}}$ : the random completion time of the  $n$ th production order of type- $i$  products under PR,
- (iii)  $D_{P,i,n}^{\text{FIFO}}$ : the random backorder time for of the  $n$ th demand of type- $i$  products under FIFO,
- (iv)  $D_{P,i,n}^{\text{PR}}$ : the random backorder time for of the  $n$ th demand of type- $i$  products under PR.

It is easily seen that

$$\begin{aligned} D_{P,i,n}^{\text{PR}} &= \text{Max}\{0, tB_{P,i,n-s_i}^{\text{PR}} - tA_{P,i,n} - L_i\}, \\ D_{P,i,n}^{\text{FIFO}} &= \text{Max}\{0, tB_{P,i,n-s_i}^{\text{FIFO}} - tA_{P,i,n} - L_i\}. \end{aligned} \quad (\text{A.12})$$

Under PR, priority is given to the LV products and, as a consequence, for  $i \in C_{\text{HV}}$ , we have

$$tB_{P,i,n}^{\text{PR}} \geq tB_{P,i,n}^{\text{FIFO}}. \quad (\text{A.13})$$

By combining the two equations of (A.12), we find that, for  $i \in C_{\text{HV}}$ ,

$$D_{P,i,n}^{\text{PR}} \geq D_{P,i,n}^{\text{FIFO}} \quad (\text{A.14})$$

and as a consequence,

$$\gamma_i^{\text{PR}}(s_i) \leq \gamma_i^{\text{FIFO}}(s_i). \quad (\text{A.15})$$

We have shown in Property 2 that the fill rate is an increasing function of  $s_i$ , regardless of the priority rule. It is furthermore shown in Corollary 3.1 that, for any item- $i$ , the optimal

base-stock level can be viewed as the minimum  $s_i$  integer value satisfying the service level constraint:

$$\gamma_i(s_i) \geq \gamma_i^a. \quad (\text{A.16})$$

Thus, we can deduce that

$$s_{HV_i}^{*FIFO} \leq s_{HV_i}^{*PR}. \quad (\text{A.17})$$

On the other hand, under PR, we also have, for  $i \in C_{LV}$ ,

$$tB_{P,i,n}^{PR} \leq tB_{P,i,n}^{FIFO}. \quad (\text{A.18})$$

Proceeding as above, we easily find

$$s_{LV_i}^{*FIFO} \geq s_{LV_i}^{*PR}. \quad (\text{A.19})$$

#### **A.4. Proof of Property 4**

By Property 1, the fill rates  $\gamma_i$  are independent of the base-stock levels  $s_j$  and of the lead-times  $L_j$ , for all  $j \neq i$ . First, let us remark that under MTO, that is, with zero base-stocks, the fill rates can be rewritten as

$$\gamma_i(0) = \Pr\{W_i \leq L_i\}, \quad (\text{A.20})$$

with  $W_i$  defined as the sojourn time in the system.

Clearly, under FIFO and under the assumptions concerning the arrival and service processes,  $W_i$  corresponds to the sojourn time in a  $k$ -multiclass M/M/1 queue with equal service rates. It is a standard result that the pdf's of  $W_i$ , for any product  $i \in C$ , are independent of the product type. In this case, the  $W_i$  pdf is the same for the  $k$  items. Clearly, under PR, the result still holds for the LV product types. However, as explained in Section 4.2, the HV product dynamics can be associated to a  $k_{HV}$ -multiclass M/G/1 queue with equal service rates. It is also a standard result that the pdf's of  $W_i$ , for any product  $i \in C_{HV}$ , are independent of the product type. In this case, to each class of products corresponds thus a unique  $W_i$  pdf.

Now, the ordering property stems from (A.15) and the fact that the fill rate is a increasing function of the customer lead-time.

## **B. Performances Analysis**

### **B.1. Proof of Property 5**

By definition of the BSCS model, the dynamics of order arrival processes and of the order production process are independent of the time-lags structure. In such a way that the state variables  $N_i$  (and in fact the associate waiting or sojourn times) can be characterized as in [4], where a model without time lags has been completely studied. In particular, it is shown

that a decomposition property holds under which the state variables related to each product families can be separately analyzed. The probability distribution of the state variables  $N_i$  is explicitly given by

$$\text{Prob}\{N_i = n\} = (1 - \hat{\rho}_i) \hat{\rho}_i^n \quad \text{for } n = 0, 1, 2, \dots, \quad (\text{B.1})$$

with  $\hat{\rho}_i = \lambda_i / \hat{\mu}_i$  and  $\hat{\mu}_i = \mu - \sum_{j \neq i} \lambda_j$ . Now, as for each product type- $i$ , the corresponding variables  $X_i$ ,  $Y_i$ ,  $D_i$  and the fill rate  $\gamma_i$  depend on the variable  $N_i$  (and in fact of the associated waiting or sojourn times) and exclusively of the parameters  $s_i$  and  $L_i$ , the multiproduct time-lag queuing model can be decomposed into independent single product models.

### B.2. Fill Rate Expression in the Multiproduct Case

In [3, page 106], it is proved that one has

$$\Pr\{\tilde{D}_i > d\} = [\hat{\rho}_i^{s_i}] e^{-(\hat{\mu}_i - \lambda_i)d}, \quad (\text{B.2})$$

where  $\tilde{D}_i$  is the backordered time for a zero-time lag model (i.e., if  $L_i = 0$ ). Now, let us define

- (i)  $\tilde{t}A_{i,n}$ : the random arrival time of the  $n$ th order of type- $i$  products (if  $L_i = 0$ ),
- (ii)  $\tilde{t}B_{i,n}$ : the random completion time of the  $n$ th part of type- $i$  products (if  $L_i = 0$ ),
- (iii)  $\tilde{D}_{i,n}$ : the random backorder time for of the  $n$ th demand of type- $i$  products (if  $L_i = 0$ ).

By definition, one has  $\tilde{D}_{i,n} = \text{Max}\{0, \tilde{t}B_{i,n-s_i} - \tilde{t}A_{i,n}\}$  and  $\Pr\{\tilde{D}_{i,n} > d\} = \Pr\{\tilde{t}B_{i,n-s_i} - \tilde{t}A_{i,n} > d\}$ . As  $tA_{i,n}$  and  $tB_{i,n}$  are independent from the time lag  $L_i$ , one has

$$\begin{aligned} \Pr\{D_{i,n} > d\} &= \Pr\{\tilde{t}B_{i,n-s_i} - \tilde{t}A_{i,n} - L_i > d\} = \Pr\{\tilde{D}_{i,n} > d + L_i\}, \\ \Pr\{D_i > d\} &= [\hat{\rho}_i^{s_i}] e^{-(\hat{\mu}_i - \lambda_i)(L_i + d)}. \end{aligned} \quad (\text{B.3})$$

### B.3. Proof of Property 7

The expression of the optimal base-stock levels for the different product types is given by (4.10)

$$s_i^* = \text{MAX}\left(\left[\frac{\ln(1 - \gamma_i^a)}{\ln \hat{\rho}_i} + \frac{(\hat{\mu}_i - \lambda_i)L_i}{\ln \hat{\rho}_i}\right], 0\right). \quad (\text{B.4})$$

The threshold  $L_i$  values corresponding to the priority rules (FIFO or PR) and to the product class type (LV and HV) are computed by solving  $s_i^* = 0$  with respect to the customer lead-time  $L_i$ . So, one has just to solve

$$\left[\frac{\ln(1 - \gamma_i^a)}{\ln \hat{\rho}_i} + \frac{(\hat{\mu}_i - \lambda_i)L_i}{\ln \hat{\rho}_i}\right] = 0, \quad (\text{B.5})$$

with the  $\hat{\rho}_i$  value corresponding to the considered priority rule (FIFO or PR) and the considered product class (LV or HV). The different  $\hat{\rho}_i$  values are given by (4.1), (4.4), and (4.7).

### C. Proof of Property 8

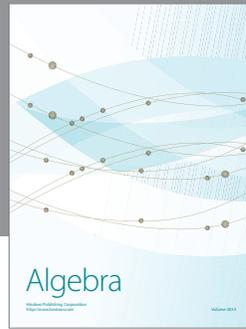
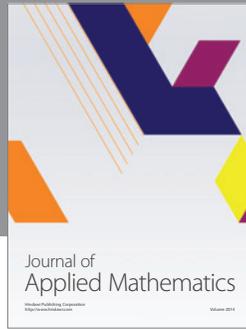
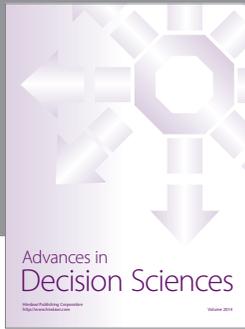
When  $L_{LV} \in [\tilde{L}_{LV}^{PR}, \tilde{L}_{LV}^{FIFO}]$ , by definition of the threshold levels, the LV optimal base-stock levels satisfy

$$s_{LV_i}^{*FIFO} > s_{LV_i}^{*PR} = 0. \quad (C.1)$$

In such a way that the cost decrease for LV is an increasing function of  $k_{LV}$ . Furthermore, when switching from FIFO to PR, the cost increase for HV is positive, independent of  $k_{LV}$  and finite (under some mild ergodicity assumption). We thus have that PR outperforms FIFO, and under PR, LV items are MTO. Furthermore, the gain converges toward 100% when  $k_{LV}/k_{HV}$  increases.

### References

- [1] F. Chen, "Market segmentation, advanced demand information, and supply chain performance," *Manufacturing and Service Operations Management*, vol. 3, no. 1, pp. 53–67, 2001.
- [2] M. M. Davis, N. J. Aquilano, and R. B. Chase, *Fundamentals of Operations Management*, McGraw-Hill, New York, NY, USA, 2002.
- [3] J. A. Buzacott and J. G. Shanthikumar, *Stochastic Models of Manufacturing Systems*, Prentice-Hall, Upper Saddle River, NJ, USA, 1993.
- [4] K. Hadj Youssef, Ch. Van Delft, and Y. Dallery, "Efficient scheduling rules in a combined make-to-stock and make-to-order manufacturing system," *Annals of Operations Research*, vol. 126, pp. 103–134, 2004.
- [5] F. Karaesmen, J. A. Buzacott, and Y. Dallery, "Integrating advance order information in make-to-stock production systems," *IIE Transactions*, vol. 34, no. 8, pp. 649–662, 2002.
- [6] A. Arreola-Risa and G. A. Decroix, "Make-to-order versus make-to-stock in a production-inventory system with general production times," *IIE Transactions*, vol. 30, no. 8, pp. 705–713, 1998.
- [7] A. Federgruen and Z. Katalan, "The impact of adding a MTO item to a MTS production system," *Management Sciences*, vol. 45, no. 7, pp. 980–994, 1999.
- [8] S. Rajagopalan, "Make to order or make to stock: model and application," *Management Science*, vol. 48, no. 2, pp. 241–256, 2002.
- [9] C. R. Sox, L. J. Thomas, and J. O. McClain, "Coordinating production and inventory to improve service," *Management Science*, vol. 43, no. 9, pp. 1189–1197, 1997.
- [10] K. Hadj Youssef, *Pilotage de systèmes de production à flux mixte: production à la commande et production par anticipation*, Ph.D. thesis, Ecole Centrale Paris, Paris, France, 2003.
- [11] K. Hadj Youssef, Ch. van Delft, and Y. Dallery, "Priority allocation decisions in large scale MTO/MTS-multi-product manufacturing systems," Tech. Rep. HEC 905/2008, 2008.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

