# Optimal Production and Admission Policies in Make-to-Stock/Make-to-Order Manufacturing Systems

June 8, 2009

Seyed Iravani[1], Tieming Liu[2], David Simchi-Levi[3]

**Abstract**

In this paper, we address production and admission control policies in manufacturing systems that produce two types of products. One type consists of identical items that are produced to stock, while the other has varying features and is produced to order. This general version is motivated by applications from various industries, in particular, the automobile industry and PC industry. Two models are studied. In the first model, the make-to-stock product has higher priority than the make-to-order product, and in the second model the manufacturer gives higher priority to the make-to-order product over the make-to-stock product. We characterize the optimal production and admission policies with simple structures, and using a computational analysis, we provide insights into the benefits of the new policies. We also investigate the impact of production capacity, cost structure, demand structure, and variabilities in demand and production process on system performance.

## 1 Introduction

In recent years, many retail and manufacturing companies have started exploring innovative revenue management techniques in an effort to improve their operations and ultimately the bottom line. Manufacturing systems that can produce different types of products have raised more research interest in recent years as many firms begin to practice market segmentation by providing multiple types of products to customers, and then differentiating customers according to their choices.

Customized production is a strong trend in the manufacturing industry. For example, in the computer industry, many companies allow customers to configure their products, and companies use the make-to-order (MTO) mode to manage the production. The MTO production not only gives more satisfaction to customers, but it also helps manufacturers eliminate finished goods inventory. However, the MTO environment suggests important challenges associated with matching fixed production capacity with highly variable demand. Specifically, an MTO system implies periods where the facility is idle and other times in which a large number of orders are awaiting production.

A good example of companies that aggressively use both customized production and customized pricing strategies is Dell Computers. Dell provides very high flexibility in configurations for customers who are willing to pay high prices. At the same time, Dell also frequently provides promotions for some low-end products to attract more customers, and for these products, customers usually have very little flexibility on

[1]Department of Industrial Engineering and Management Science, Northwestern University.
[2]School of Industrial Engineering and Management, Oklahoma State University.
[3]Department of Civil and Environmental Engineering and Engineering Systems Division, MIT.

configurations. To satisfy these demands, Dell also produces some standard products to stock, the make-to-stock (MTS) environment. This gives raise to a combination of a make-to-stock/make-to-order environment that allows Dell to better manage their production capacity and increase expected profit.

The application of make-to-stock/make-to-order manufacturing systems is also important among part suppliers who face demands from both original equipment manufacturers (OEMs) and the so-called aftermarket. For example, in the automobile industry, a part supplier sells its products to automotive assembly plants for installation into new vehicles, as well as repair shops for replacement in old vehicles (Carr and Duenyas [3]). OEM and aftermarket demands are both important to the part supplier. OEM demands guarantee high utilization of the production capacity, while aftermarket demands bring high profit margins to the supplier. OEM sales are based on long-term contracts, and they are produced under the make-to-stock mode. In contrast, aftermarket items are produced under the make-to-order mode due to their large variety. Note that, as opposed to the Dell example, in this case the high priority products are produced to stock. All these developments call for models that integrate production, sequencing and admission decisions for a hybrid production system with both make-to-stock and make-to-order products.

There is a significant amount of literature addressing multi-product production scheduling problems in either a make-to-stock or a make-to-order manufacturing system. Problems with or without set-ups associated with switching from producing one product to another have both been extensively studied. We refer the reader to Graves [12], Zheng and Zipkin [40], Pena-Perez and Zipkin [28], Wein [38], Ha [17], and De Vericourt et al. [6] for scheduling policies in make-to-stock environments, and to Varaiya et al. [36], Gittins [10], Duenyas and Van Oyen [9], Reiman and Wein [31] and the references therein for scheduling policies in make-to-order environments.

Single-product models with multi-class customers in a make-to-stock manufacturing system, the so-called stock rationing problems, have been studied in various contexts since the late 1960's. Some of the examples of stock rationing in uncapacitated systems are Topkis [35], Nahmias and Demmy [26], Cohen et al. [5], and Melchiors et al. [23]. For stock rationing in capacitated systems see Ha [15, 16, 18], de Vericourt et al. [7], and Huang and Iravani [19], Iravani et al. [21, 20], Benjaafar et al. [1]. Extensive reviews for the stock rationing problems can be found in Kleijn and Dekker [22].

All the papers above focused on the dynamic production and sequencing problems for either make-to-order systems or make-to-stock systems; however, none of them considered a hybrid manufacturing system, and admission control was rarely studied. Carr and Duenyas [3] considered both the production and admission decisions for a make-to-stock/make-to-order system, where the make-to-stock orders have higher priorities. They found an optimal policy characterized by monotonic switching curves. Gupta and Wang [13] study a MTS/MTO problem in a periodic-review model with stochastic demands and deterministic production capacities. Leadtime is quoted in their model. Due to the complexity and the large state space, they focused on the case with the leadtime of one period, and they developed a three-dimensional lookup table to find the optimal decisions. For models with longer leadtimes, they acknowledge that "even if one could compute and store look-up tables, the implementation of the resulting policy can be problematic." Dobson and Yano [8] consider product offering, pricing, and MTO and MTS decisions with linear deterministic demand functions in strategic levels, while production and admission controls are not their focus. Youssef, van Delft, and Dallery [39] examine the manufacturing scheduling problem in an MTS/MTO system under two static policies: the classical FIFO rule and a priority policy. In the latter, preemptive priority is given to low volume (make-to-order) products, and approximations are used to obtain analytical and numerical results. They consider the base-stock threshold, but rationing, admission, and contingent outsourcing are not their focus.

In the literature on inventory systems (including the above papers, except for [3, 13]), it is commonly assumed that unsatisfied demands are either fully backordered or lost. A relatively small amount of papers considered partial backordering, where in case of shortage, arriving customers may be either backordered or lost, each with a certain probability that is an outcome of the customer behavior (see Montgomery [25], Moinzadeh [24], Smeitink [33], and Nahmias and Smith [27]). In a few recent papers, partial backordering became part of the seller's decision. The seller determines the amount of unsatisfied demands to be backordered or rejected (see Rabinowitz [32], Benjaafar et. al. [1], and Chen and Kulkarni [4]). However, the admission decision increases the complexity of the model, and the current optimal policies in the literature are characterized by state-dependent non-linear switching curves. The complexity of the non-linear structure makes the optimal policies difficult to implement in the real world.

In this paper, we study policies to coordinate production, sequencing and admission controls for two types of manufacturing systems both with make-to-stock products and make-to-order products. Our paper is related to Carr and Duenyas [3], but it is different from [3] in the following sense: (*i*) Our paper studies two types of manufacturing systems: in the first case, MTS product has higher priority than MTO product, and in the second case, MTS product has lower priority; While Carr and Duenyas only studied the first case. (*ii*) In our paper, unsatisfied high-priority demands are fully backlogged; while in [3], when the high-priority product is out of stock, the demand is lost. Our backlogging assumption is reasonable since it is often difficult to find a product matching all the required features from other suppliers immediately. (*iii*) In our paper, the optimal policies are characterized by a linear or near-linear structure, while in [3], the threshold levels do not have such a simple structure. (*iv*) We also examine the possibility of outsourcing in our model, which is not considered in [3].

In the first model of this paper, we consider a manufacturing system with both high-priority MTS (e.g., OEM) and low-priority MTO (e.g., aftermarket) demands. High-priority customers order identical items from the manufacturer, and these orders cannot be rejected. If the product is out of stock, high-priority orders will still be accepted and backlogging cost will be incurred until they are satisfied. Demands for low-priority, customized products can be rejected if the manufacturer does not have enough production capacity, or otherwise accepted. We characterize the structure of the optimal production and admission control policies with a near-linear structure, and we extend our results for the case where the manufacturer can outsource low-priority orders.

In the second model, we examine the production and admission controls in a manufacturing system with high-priority MTO (customized) and low-priority MTS (pre-configured) products. MTO products provide high configuration flexibilities to customers, and thus are sold at higher prices, and MTS products are sold at lower prices for promotion. Demands for high-priority MTO products will all be accepted, while demands for low-priority MTS products could be satisfied, backlogged, or rejected. We characterize the optimal policy with linear threshold levels. We also extend our results to systems in which the low-priority MTS orders can be outsourced.

The paper is organized as follows. In Section 2, we study the model with both OEM and aftermarket demands, and we extend it to cases in which the manufacturer has the option of outsourcing the aftermarket orders, as well as to cases with multiple types of aftermarket products. In Section 3, we investigate the model with both customized and pre-configured products, and extend the results to systems with outsourcing option for pre-configured products. In Section 4, we use computational analysis to obtain insights into the benefits of the new policies, and the impact of production capacity, demand structure and price structure on system performance. In Section 5, we study the effects of variability in demand and production process using a system with Erlang distributed demand interarrival and production times. Finally in Section 6, we summarize and suggest future research directions.

# 2 Model with OEM and Aftermarket Products

## 2.1 Problem Formulation

We consider a manufacturing system with two types of products in an infinite horizon. Type 1 is for high-priority orders and has standard configuration, so it is produced to stock. Demands for type 1 are satisfied if inventory is available, or are fully backlogged with high backlogging cost $b_1$ per item per unit time if type 1 is not available in inventory. Type 2 has varying features and is produced to order. Demands for type 2 can be either accepted or rejected. As in literature on MTO systems (see [3] and [39]), we consider backlogging cost $b_2$ per unit time for accepted orders of type 2. Backlogging cost serves as a proxy for the manufacturer's cost due to having delays in production of type 2 orders and thus becoming uncompetitive in that market, loss of good will, as well as extra shipping and handling cost due to the tardiness. As in literature on MTO systems (see [3] and [39]), we assume that the backlogging cost for the MTO product is charged since the order is accepted.[4] We assume $b_2 < b_1$, since suppliers usually have long-term contracts with OEMs, and delayed fulfillments are heavily penalized. If a demand for type 2 is rejected, a rejection penalty, $r_2$, associated with lost sales and loss of goodwill, is incurred. The contribution margin for product $i$ is $p_i$, and the inventory holding cost for product 1 is $h$ per unit time.

We assume that customers for each type of product arrive according to a Poisson Process, and we let $\lambda_i$ be the demand arrival rate of Class $i$, for $i = 1, 2$. Also, we assume that the production time of each type of product follows the same exponential distribution, with production rate $\mu$. The same-production-time assumption is reasonable if the production process is mainly assembly operations, where the time difference between assembling a standard product (e.g., a Dell Inspiron 700m with a 1G memory chip) and assembling a customized product (e.g., a Dell Inspiron 700m with a 2G memory chip) is insignificant. In Section 5, we relax this assumption and study systems with different production rates. We further assume that preemptions are allowed, and no set-up time is needed when the manufacturing system switches from one type of job to the other. This assumption is also reasonable for assembly production systems where setup times are negligible compared to production times.

The assumption of Poisson arrival demand and exponential production times is what allows us to formulate the problem and characterize the structure of the optimal policy. In Section 5, we extend our analysis to Erlang distributed interarrival and production times and examine the impact of variability in demand and production process on system performance. We show that most of the insights we obtained with our exponential model are not influenced by the assumption on demand and production processes.

Since the system cost (e.g., backlogging or holding cost) is the same for the same type of product, it is clear that there is no need for rationing among orders for the same type of product. Therefore, the system state can be described by a vector of two variables, $\mathbf{y}(t) = (y_1(t), y_2(t))$, where $y_i(t)$ is the net inventory level of product $i$ at time $t$, with $y_1(t) \in Z$ and $y_2(t) \in Z^-$, where $Z$ is the set of all positive and negative integer numbers, while $Z^-$ only includes non-positive integer numbers. We use $y_1^+(t) = \max\{0, y_1(t)\}$ to

[4] Another case is when the backlogging penalties are incurred only when orders spend more than a predetermined length of time. The analysis of this case is very complex, since in addition to the number of orders in the system, the state of the system should also include how many time periods each order has spent in the system.

show the amount of inventory of type 1 at time $t$, and $y_1^-(t) = \max\{0, -y_1(t)\}$ to show the number of waiting customers at time $t$. Similarly, $-y_2(t)$ is the number of waiting customers for product 2 at time $t$. The system state space is $\Omega = Z \times Z^-$.

In state $\mathbf{y} = (y_1, y_2)$, the system incurs a cost at rate,

$$c(\mathbf{y}) = -hy_1^+ - b_1 y_1^- + b_2 y_2.$$

Let $\alpha$ be the time discount rate, let $N_i^a(t)$ be the number of accepted orders over interval $[0, t]$ for product $i$ $(i = 1, 2)$, and let $N_2^r(t)$ be the number of rejected orders for type 2 over the same period of time. We seek an optimal control policy $\pi$ so as to maximize either the discounted system profit over an infinite horizon,

$$\max_{\pi} \quad J^{\pi}(\mathbf{y}(0)) = \mathrm{E}^{\pi}_{\mathbf{y}(0)} \left[ \sum_{i=1}^{2} \int_0^{\infty} e^{-\alpha t} p_i dN_i^a(t) - \int_0^{\infty} e^{-\alpha t} r_2 dN_2^r(t) + \int_0^{\infty} e^{-\alpha t} c(\mathbf{y}(t)) dt \right], \tag{1}$$

or the average profit over an infinite horizon,

$$\max_{\pi} \quad J_a^{\pi} = \lim_{T \to \infty} \frac{1}{T} \mathrm{E}^{\pi} \left[ \sum_{i=1}^{2} p_i N_i^a(T) - r_2 N_2^r(T) + \int_0^T c(\mathbf{y}(t)) dt \right]. \tag{2}$$

In (1), $J_{\pi}(\mathbf{y}(0))$ is the expected profit function under policy $\pi$ starting from initial state $\mathbf{y}(0) = (y_1(0), y_2(0))$. In the rest of the paper we will mainly focus on the discount-profit problem. However, as shown in de Vericourt, Karaesmen and Dallery [7], the theoretical results for the discounted-profit model also apply to the average-profit problem.

Without loss of generality, we redefine the time scale such that $\alpha + \mu + \lambda_1 + \lambda_2 = 1$. According to Bertsekas [2], the optimality equation $J^*(y_1, y_2)$ under the time-discount criterion satisfies the following Bellman's equation:

$$J(y_1, y_2) = c(y_1, y_2) + \mu H_0 J(y_1, y_2) + \lambda_1 H_1 J(y_1, y_2) + \lambda_2 H_2 J(y_1, y_2) := HJ(y_1, y_2) \tag{3}$$

where $H_0$, $H_1$, and $H_2$ are functions defined by,

$$\begin{aligned} H_0 J(y_1, y_2) &= \max\Big\{ J(y_1, y_2), J(y_1 + 1, y_2), J(y_1, y_2 + 1 | y_2 < 0) \Big\} \\ H_1 J(y_1, y_2) &= J(y_1 - 1, y_2) + p_1 \\ H_2 J(y_1, y_2) &= \max\Big\{ J(y_1, y_2 - 1) + p_2, J(y_1, y_2) - r_2 \Big\}. \end{aligned}$$

$H_0$ corresponds to the production decision: the manufacturer can choose to either produce or stop production. $J(\ T \mid C\ )$ is a conditional value function, which indicates that the transition to state $T$ is only a valid option when condition $C$ holds. Therefore, the third term means producing product 2 when there are backorders for it. $H_1$ indicates that the demands for type 1 will always be accepted. $H_2$ is associated with the admission control for an arriving order for type 2. The manufacturer can either accept (backlog) or reject the order. We denote the overall Bellman function as $HJ(y_1, y_2)$.

The optimality equation under the average-profit criterion is:

$$J(y_1, y_2) + g = c(y_1, y_2) + \mu H_0 J(y_1, y_2) + \lambda_1 H_1 J(y_1, y_2) + \lambda_2 H_2 J(y_1, y_2) \quad (4)$$

where $g$ is the optimal average profit per unit time.

## 2.2 The Optimal Policy

We investigate the structure of the optimal policy following the approach of Ha [16] and De Vericourt, Karaesmen and Dallery [7]. We first define a set of optimality conditions and decision rules, and then show that the optimal expected profit function, $J(\mathbf{y})$, satisfies the conditions.

For any function $f$ defined on $\Omega$, let $\Delta_1 f(y_1, y_2) = f(y_1 + 1, y_2) - f(y_1, y_2)$, $\Delta_2 f(y_1, y_2) = f(y_1, y_2 + 1) - f(y_1, y_2)$, and $\Delta_{12} f(\mathbf{y}) = f(y_1 + 1, y_2) - f(y_1, y_2 + 1)$. We define the set of functions as $\mathcal{C}1$, such that if $f(y_1, y_2) \in \mathcal{C}1$, then,

**Condition C1:** For $(y_1, y_2) \in \Omega$,

- C.1.1: $\Delta_i f(y_1, y_2) \geq 0$, if $y_i < 0$, $i = 1, 2$;
- C.1.2: $\Delta_{12} f(y_1, y_2) \geq 0$, if $y_1 < 0$.
- C.1.3: $\Delta_1 f(y_1, y_2)$ and $\Delta_2 f(y_1, y_2)$ are non-increasing in $y_1$ and $y_2$;
- C.1.4: $\Delta_{12} f(y_1, y_2)$ is non-increasing in $y_1$ and non-decreasing in $y_2$;

To have some intuition about the above conditions, we apply the condition set $\mathcal{C}1$ to the expected profit function $J(y_1, y_2)$. Condition C.1.1 implies that, if there are waiting customers for a product, it is better to produce the product rather than idle the facility. Condition C.1.2 implies that, if there are waiting customers for both products, type 1 has higher priority than type 2. Conditions C.1.3 and C.1.4 are the concavity and submodularity conditions.

We first show that any functions satisfying the above conditions have the following properties.

**Property 1** *If $f(y_1, y_2) \in \mathcal{C}1$, then*

*(i)* $\Delta_1 f(y_1, y_2) \geq 0$ *for* $y_1 < S^f$, *where* $S^f = \min\{z | \Delta_1 f(z, 0) < 0\}$;

*(ii)* $\Delta_{12} f(y_1, y_2) \geq 0$ *for* $y_2 < 0$, $y_1 < Q^f(y_2)$, *where* $Q^f(y_2) = \min\{z | \Delta_{12} f(z, y_2) < 0\}$;

*(iii)* $\Delta_2 f(y_1, y_2 - 1) \leq p_2 + r_2$ *for* $y_2 > A^f(y_1)$, *where* $A^f(y_1) = \max\{z | \Delta_2 f(y_1, z - 1) > p_2 + r_2\}$;

*(iv)* $0 \leq A^f(y_1) - A^f(y_1 + 1) \leq 1$,

*(v)* $S^f \geq Q^f(-1) \geq Q^f(y_2)$, *for* $y_2 < -1$.

Please refer to On-line Appendix A for the proof for all the propositions, lemmas, and theorems in Section 2.1. Property (i) of the proposition implies that for threshold level $S^f$, if $y_1 < S^f$, then producing type 1 is better than idling the facility. Property (ii) implies that type 1 has higher priority if $y_1 < Q^f(y_2)$), and type 2 has higher priority otherwise. Property (iii) implies that for threshold level $A^f(y_1)$, if $y_2 > A^f(y_1)$, then accepting a type 2 order is better than rejecting. Property (iv) implies that the admission threshold level $A^f(y_1)$ goes up either vertically (when $A^f(y_1) - A^f(y_1+1) = 0$) or in a 45° line (when $A^f(y_1) - A^f(y_1+1) = 1$). Property (v) implies that the production threshold level $S^f$ is no lower than the rationing threshold level $R^f$.

In the following, we show that the optimal profit function $J(y_1, y_2)$ satisfies all the conditions in set $\mathcal{C}1$, i.e., $J(y_1, y_2) \in \mathcal{C}1$. Lemma 2 indicates that the structure of function $f$ in $\mathcal{C}1$ is preserved under function $H$.

**Lemma 1** *If $f(y_1, y_2) \in \mathcal{C}1$, then $Hf(y_1, y_2) \in \mathcal{C}1$.*

In the next proposition we use the lemma to show that there exists an optimal policy satisfying all the conditions in $\mathcal{C}1$. It will be a switching-curve policy characterized by a base-stock level ($S$), a rationing threshold curve ($Q(y_2)$), and an admission threshold curve ($A(y_1)$).

**Proposition 1** *The optimal policy is characterized by three thresholds: the base-stock level, $S$, the rationing level, $Q(y_2)$ , and the admission level, $A(y_1)$, such that at state $(y_1, y_2)$,*

- *Production control: when there are no orders in the system (i.e., $y_2 = 0$), it is optimal to produce type 1 if $y_1 < S$, and to stop production if $y_1 \geq S$.*
- *Rationing control: when there are orders in the system (i.e., $y_2 \neq 0$), it is optimal to produce type 1 if $y_1 < Q(y_2)$, and to produce type 2 if $y_1 \geq Q(y_2)$ and $y_2 < 0$, where $Q(y_2)$ decreases as $y_2$ decreases (the number of backorders increases).*
- *Admission control: when $y_1 < R$, it is optimal to accept an arriving demand for type 2 if $y_2 > A(y_1)$, and to reject otherwise, where $A(y_1)$ decreases as $y_1$ increases.*

The above policy is similar to the switching-curve policy in Carr and Duenyas [3], and thus not surprising. (The proof for the above policy is omitted because later we will prove a more restrictive policy using the same approach. We denote $f^*(y_1, y_2)$ as the optimal function that satisfies equation (3).

In the following, we will show that the optimal policy could be characterized more specifically by a near-linear structure, which is a special case of the switching-curve structure. For this purpose, we introduce $\mathcal{C}2$, a set of functions that satisfy all the conditions in $\mathcal{C}1$ and the following additional conditions. Denote $R^f = Q^f(-1)$ and $B^f = A^f(0)$, we define $\mathcal{C}2$, such that, if $f(y_1, y_2) \in \mathcal{C}2 \subset \mathcal{C}1$, then,

**Condition C2:** For $y_1 < R^f$ and $y_2 < 0$,

- C.2.1: $\Delta_{12} f(y_1, y_2)$ is independent of $y_2$;
- C.2.2: $\Delta_{12} f(y_1, y_2) \geq 0$;
- C.2.3: $\Delta_{12} f(y_1, y_2) \geq \Delta_{12} H f(y_1, y_2) \geq \Delta_{12} f^*(y_1, y_2)$.
- C.2.4: $\Delta_2 f(y_1, y_2)$ is independent of $y_1$ and $y_2$ as long as $y_1 + y_2$ is fixed;

Condition C.2.1 implies that the marginal difference between producing type 1 or type 2 is independent of the number of backorders for type 2, when $y_1 < R^f$. Condition C.2.2 implies that type 1 has higher priority if the inventory of product 1 is less than a threshold level $R^f$ (i.e., if $y_1 < R^f$), and type 2 has higher priority otherwise. Condition C.2.3 implies that the value of $\Delta_{12} f(y_1, y_2)$ monotonically decreases during the value iteration process (until it converges). The sign of $p_2 + r_2 - \Delta_2 J(y_1, y_2 - 1)$ determines whether to reject an order for type 2. Condition C.2.4 suggests that admission decisions depend on the total inventory level, but not on the inventory level of each product.

We next show that under the above additional conditions, the threshold levels have the following additional properties.

**Property 2** *If $f(y_1, y_2) \in \mathcal{C}2 \subset \mathcal{C}1$, then*

*(vi)* $\Delta_2 f(y_1, y_2 - 1) \leq p_2 + r_2$ *for* $y_1 < R^f$ *and* $y_1 + y_2 > B^f$.

*(vii)* $y_1 + A^f(y_1) = B^f$ *when* $y_1 < R^f$,

*(viii)* $R^{Hf} \leq R^f$.

Property (vi) implies that accepting a new order for type 2 is profitable, as long as the the total inventory of products 1 and 2 are larger than a threshold level $B^f$ (i.e., $y_1 + y_2 > B^f$). Property (vii) implies that the admission threshold $A^f(y_1)$ presents a 45° line when $y_1 < R^f$. Property (viii) of the proposition suggests that the rationing threshold level monotonically decreases during the value iteration process (until it converges).

In the following, we show that the optimal profit function $J(y_1, y_2)$ satisfies all the conditions in set $\mathcal{C}2$, i.e., $J(y_1, y_2) \in \mathcal{C}2$. Lemma 2 indicates that the structure of function $f$ in $\mathcal{C}$ is preserved under function $H$.

**Lemma 2** *If $f(y_1, y_2) \in \mathcal{C}2 \subset \mathcal{C}1$, then $Hf(y_1, y_2) \in \mathcal{C}2 \subset \mathcal{C}1$.*

In the next theorem we use the lemma to show that a modified base-stock policy is optimal.

**Theorem 1** *The optimal policy is characterized by three parameters: the base-stock level, $S$, the rationing level, $R$ ($R \leq S$), and the admission level, $B$, such that at state $(y_1, y_2)$,*

- *Production control: when there are no orders in the system (i.e., $y_2 = 0$), it is optimal to produce type 1 if $y_1 < S$, and to stop production if $y_1 \geq S$.*

- *Rationing control: when there are orders in the system (i.e.,* $y_2 \neq 0$*), it is optimal to produce type 1 if* $y_1 < R$*, and to produce type 2 if* $y_1 \geq R$ *and* $y_2 < 0$*.*

- *Admission control: when* $y_1 < R$*, it is optimal to accept an arriving demand for type 2 if* $y_1 + y_2 > B$*, and to reject otherwise; when* $y_1 \geq R$*, it is optimal to accept an arriving demand for type 2 if* $y_2 > A(y_1)$*, and to reject otherwise.*

The optimal policy is illustrated in Figure 1-Left. Under the optimal policy, the manufacturer will stop production if the inventory level of type 1 is greater or equal to $S$ and there are no waiting orders of type 2. When the inventory level of type 1 is higher than $R$ and there are orders for type 2 in the system, the manufacturer will give higher priority to type 2 orders. If the inventory level of type 1 is less than $R$ or when there are no orders for type 2, the manufacturer will produce type 1 to increase its inventory level. When the inventory level of type 1 is lower than $R$, the manufacturer will accept an arriving order for type 2 if the total net inventory $y_1 + y_2$ is higher than $B$ or reject it otherwise; When $y_1 \geq R$, the manufacturer will accept a type 2 order if $y_2 > A(y_1)$.

The optimal policy has a near-linear structure. The rationing threshold is a horizontal line, and the admission threshold below $y_1 = R$ represents a 45° line. Only the admission threshold above $R$ may not be linear. If we use a 45° to approximate the admission threshold, as shown by the dotted line in Figure 1-Left, we come to a linear structured policy. We refer to this linear heuristic policy as the $(S, R, B)$ policy. Due to its simple structure, the computation complexity of the $(S, R, B)$ policy is much lower than that of the optimal policy. As we will later show in Section 4.1, the performance of the $(S, R, B)$ policy is very close to that of the optimal policy.

**Remark:** The near-linear structure of the optimal threshold policy can be extended to problems with a single MTS product and *multiple* MTO products, using the same approach in Carr and Duenyas [3], when all MTO products have the same production rate, the same backlogging penalty, but different profits and rejection penalties. We present the extension in the On-line Appendix B.

## 2.3 Expected Due Date Quoting

Many manufacturers quote expected due dates (i.e., leadtime) to their customers. For example, Dell Computers quote *Preliminary Ship Date*, and has the following statement on its website: "The preliminary ship date is not intended to provide you with an actual estimated ship date. The preliminary ship date represents the estimated time it takes to process your order and custom build your computer." In this section we present the expected due date for both the OEM and the aftermarket products.

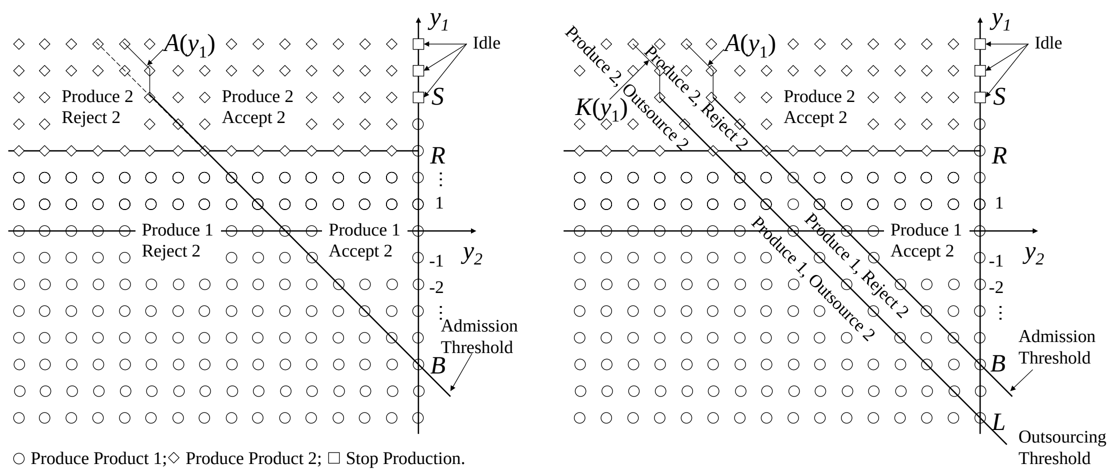


Figure 1: *Left:* The Optimal and the Linear *(S, R, B)* Heuristic Policy; *Right:* The Optimal Policy with Contingent Outsourcing.

**Lemma 3** *If an OEM order is not satisfied immediately, its expected due date (i.e., leadtime) is* $T_1 = (y_1^- + 1) \times \frac{1}{\mu}$.

**Lemma 4** *When* $y_1 < R$*, the expected leadtime for an arriving aftermarket order is* $T_2 = \frac{1}{\mu - \lambda_1}(R - y_1 + y_2^- + 1)$.

When $y_1 \geq R$, it is difficult to obtain a simple expression to calculate the expected leadtime for an arriving aftermarket order. We present an algorithm in On-line Appendix A for this purpose.

## 2.4 Contingent Outsourcing for Aftermarket Orders

In this section, we consider a manufacturer who has the option to outsource the production of some orders to other suppliers, and the manufacturer pays by the number of units outsourced. In this model, we assume that the manufacturer only outsources the production of aftermarket products but not the OEM orders. Such an assumption is reasonable for manufacturers who want to provide consistent quality for OEM products. The quality of outsourced products is not guaranteed, and it may hurt the relationship between the manufacture and the OEM, if the quality of the outsourced product is low. In some industries, outsourcing may simply be prohibited by the contract between a manufacturer and an OEM customer.

We focus on the case where contingent outsourcing of an order for product 2 incurs a cost of $l_2$, higher than the rejection penalty $r_2$. (The case with $l_2 \leq r_2$ is trivial, as the manufacturer will never reject any demand – s/he will accept and then outsource the production immediately.) As a result, when an order for type 2 arrives, the outsourcing option $(J(y_1, y_2) - l_2)$ is always dominated by the rejection option $(J(y_1, y_2) - r_2)$.

In other words, since the outsourcing cost is higher than the rejection penalty, an arriving type 2 order will be rejected directly (rather than being outsourced) if the production capacity is not enough. So we do not need to consider outsourcing in our $H_2$ function. Similarly, we do not need to consider outsourcing in our $H_0$ function either. The outsourcing option will be adopted only when the inventory level is low. If outsourcing is not adopted before the production completion (otherwise the pre-event state would be transient since the outsourcing decision can be made anytime), outsourcing would not be profitable after the completion either. Upon the completion of a job, the inventory level is just increased by one unit. Therefore, the manufacturer only needs to decide whether to outsource an aftermarket order at the event when an OEM order arrives. To incorporate the new outsourcing feature, we change function $H_1$ as

$$H_1 J(y_1, y_2) = \max\Big\{J(y_1 - 1, y_2) + p_1, J(y_1 - 1, y_2 + 1) + p_1 - p_2 - l_2\Big\}.$$

To prove our main result in Theorem 2, we need to add a few more properties.

**Property 3** *If $f(y_1, y_2) \in \mathcal{C}2 \subset \mathcal{C}1$, then*

*(ix)* $\Delta_2 f(y_1, y_2 - 1) \leq p_2 + l_2$ *for* $y_2 > K^f(y_1)$*, where* $K^f(y_1) = \max\{z | \Delta_2 f(y_1, z - 1) > p_2 + l_2\}$*,*

*(x)* $K^f(y_1) \leq A^f(y_1)$*,*

*(xi)* $0 \leq K^f(y_1) - K^f(y_1 + 1) \leq 1$*,*

*(xii)* $y_1 + K^f(y_1) = L^f$ *when* $y_1 < R^f$*, where* $L^f = K^f(0)$*,*

*(xiii)* $L^f \leq B^f$ *when* $y_1 < R^f$*.*

Property (ix) implies that for threshold level $K^f(y_1)$, if $y_2 > K^f(y_1)$, then keeping a type 2 order is better than outsourcing it. Property (x) of the proposition implies that the outsourcing threshold $K^f(y_1)$ cannot be higher than the admission threshold $A^f(y_1)$. Property (xi) implies that the outsourcing threshold $K^f(y_1)$ goes up either vertically or in a 45° line. Property (xii) concludes that when $y_1 < R^f$, the outsourcing threshold $K^f(y_1)$ presents a 45° line, and $y_1 + K^f(y_1) = L^f$. When $y_1 < R^f$, outsourcing an order for product 2 is profitable when $y_1 < R^f$ and $y_1 + y_2 \leq L^f$. Finally, Property (xiii) is a special case of the result in Property (x). It shows that when $y_1 < R^f$, outsourcing threshold $L^f$ cannot be higher than the admission threshold $B^f$ (both present linear functions now).

We show in the following lemma that, for systems with contigent outsourcing, the optimality conditions in sets $\mathcal{C}1$ and $\mathcal{C}2$ are preserved by the modified functions.

**Lemma 5** *If $f(y_1, y_2) \in \mathcal{C}2 \subset \mathcal{C}1$, then $Hf(y_1, y_2) \in \mathcal{C}2 \subset \mathcal{C}1$.*

The following theorem characterizes the structure of the optimal policy for the two-product problem with outsourcing options:

**Theorem 2** *The optimal policy is characterized by four parameters: the base-stock level, $S$, the rationing level, $R$, the admission level, $B$, and the outsourcing level, $L$ ($L \leq B$), such that at state $(y_1, y_2)$,*

- *Production control: when there are no orders in the system (i.e., $y_2 = 0$), it is optimal to produce type 1 if $y_1 < S$, and to stop production if $y_1 \geq S$.*
- *Rationing control: when there are orders in the system (i.e., $y_2 \neq 0$), it is optimal to produce type 1 if $y_1 < R$, and to produce type 2 if $y_1 \geq R$ and $y_2 < 0$.*
- *Admission control: when $y_1 < R$, it is optimal to accept an arriving demand for type 2 if $y_1 + y_2 > B$, and to reject otherwise; when $y_1 \geq R$, it is optimal to accept an arriving demand for type 2 if $y_2 > A(y_1)$, and to reject otherwise.*
- *Outsourcing control: when $y_1 < R$, it is optimal to outsource an order for type 2 upon arrival of an order for type 1, if $y_1 + y_2 \leq L$; when $y_1 \geq R$, it is optimal to outsource an order for type 2 if $y_2 \leq K(y_1)$.*

The optimal policy for a two-product problem with outsourcing options is described in Figure 1-Right. Under the optimal policy, when $y_1 < R$, if the total inventory $y_1 + y_2$ is lower or equal to $L$, the manufacturer will outsource an order for type 2 when an order for type 1 arrives; when $y_1 \geq R$, the manufacturer will outsource an order for type 2 if $y_2 \leq K(y_1)$. Except for the admission and outsourcing thresholds above $R$, the optimal policy has linear thresholds anywhere else.

# 3 Model with Customized and Pre-configured Products

## 3.1 Problem Formulation

In this section, we consider another type of MTS/MTO manufacturing system, which pervades in companies facing consumers directly. For example, Dell provides very high flexibility for customers to customize their products, and at the same time, it also frequently provides promotions for some low-end, pre-configured products. The promotional products not only enlarge the company's market share, but also allows the manufacturer to better manage the production capacity.

In this model, type 1 product is produced to stock and for promotion; type 2 is the regular product for customization, and hence it is produced to order. Demand for type 1 is satisfied if inventory is available, and is backlogged, or rejected otherwise depending on the manufacturer's decision. We consider backlogging cost $b_1$ for the backlogged orders of type 1. Demand for type 2 provides higher profit margin to the manufacturer, and we assume they are fully accepted. We consider unit backlogging cost $b_2$ per unit time for the orders of type 2 in the MTO system, and we assume $b_1 \leq b_2$. This is because customers who buy promotional products are usually sensitive to the price, but not so sensitive to the leadtime as other customers. We continue with the notation and the assumptions in Section 2.

In the following, we use an approach parallel to Section 2 to analyze a system with both promotional and customized products. In state $\mathbf{y} = (y_1, y_2)$, the system incurs a cost at rate

$$c(\mathbf{y}) = -hy_1^+ - b_1 y_1^- - b_2 y_2^- .$$

We seek an optimal control policy $\pi$ so as to maximize either the discounted profit over an infinite horizon,

$$\max_{\pi} \quad J^{\pi}(\mathbf{y}(0)) = E^{\pi}_{\mathbf{y}(0)} \left[ \sum_{i=1}^{2} \int_0^{\infty} e^{-\alpha t} p_i dN_i^a(t) - \int_0^{\infty} e^{-\alpha t} r_1 dN_1^r(t) + \int_0^{\infty} e^{-\alpha t} c(\mathbf{y}(t)) dt \right], \tag{5}$$

or the average profit over an infinite horizon,

$$\max_{\pi} \quad J_a^{\pi} = \lim_{T \to \infty} E^{\pi} \left[ \sum_{i=1}^{2} p_i N_i^a(T) - r_1 dN_1^r(T) + \int_0^T c(\mathbf{y}(t)) dt \right]. \tag{6}$$

In the rest, we will mainly focus on the discounted-profit problem. The theoretical results for the discounted-profit problem also hold for the average-profit problem.

Redefining the time scale such that $\alpha + \mu + \lambda_1 + \lambda_2 = 1$, the optimality equation $J^*(y_1, y_2)$ under the time-discount criterion satisfies the following Bellman's equation:

$$J(y_1, y_2) = c(y_1) + \mu H_0 J(y_1, y_2) + \lambda_1 H_1 J(y_1, y_2) + \lambda_2 H_2 J(y_1, y_2) := HJ(y_1, y_2), \tag{7}$$

where $H_i$, $i = 0, 1, 2$, are functions defined in $\Omega$,

$$\begin{aligned} H_0 J(y_1, y_2) &= \max \Big\{ J(y_1, y_2), J(y_1 + 1, y_2), J(y_1, y_2 + 1 | y_2 < 0) \Big\}, \\ H_1 J(y_1, y_2) &= \max \Big\{ J(y_1, y_2) - r_1, J(y_1 - 1, y_2) + p_1 \Big\}, \\ H_2 J(y_1, y_2) &= J(y_1, y_2 - 1). \end{aligned}$$

$H_0$ corresponds to the production decision. The manufacturer can choose to either produce or stop production. Particularly, in the third term, condition $y_2 < 0$ implies that make-to-order products can only be produced when there are waiting customers. $H_1$ is associated with the admission control for an arriving demand for type 1 product. The manufacturer will satisfy the demand if inventory is available, or choose to backlog or to reject otherwise. Function $H_2$ indicates that demand for type 2 will always be accepted.

The optimality equation under the average-profit criterion is:

$$J(y_1, y_2) + g = c(y_1) + \mu H_0 J(y_1, y_2) + \lambda_1 H_1 J(y_1, y_2) + \lambda_2 H_2 J(y_1, y_2), \tag{8}$$

where $g$ is the optimal average profit per unit time.

## 3.2 The Optimal Policy

We use the same approach as in Section 2.2 to investigate the structure of the optimal policy for this model. Here we only present the main results. The supporting analytical results are provided in On-line Appendix D.

**Theorem 3** *The optimal policy is characterized by two parameters: the base-stock level, $S$, and the admission level, $B$ ($B \leq 0$), such that at state $(y_1, y_2)$,*

- *Production control: if there are orders for type 2 (i.e., $y_2 \neq 0$), it is optimal to produce type 2; otherwise, it is optimal to produce type 1 if $y_1 < S$, and to stop production if $y_1 \geq S$.*
- *Admission control: it is optimal to satisfy an arriving demand for type 1 if $y_1 > 0$, to reject if $y_1 + y_2 \leq B$, and to backlog otherwise.*

The optimal policy is described in Figure 2-Left. Under the optimal policy, the manufacturer will stop production if the inventory level of type 1 reaches $S$, and keep producing otherwise. When there is no inventory available, the manufacturer will backlog an arriving demand for type 1 if the net total inventory, $y_1 + y_2$, is higher than $B$, or reject otherwise. We show in Proposition 5 of On-line Appendix D that the threshold level $B$ is non-positive. Therefore, it is always optimal to satisfy promotional demands when the inventory is available. In this policy, type 2 orders always have higher priority and cannot be rejected. We refer to this optimal policy as the $(S, B)$ policy.

**Remark:** Note that the optimal policy for product 2 is not affected by the system state. Orders for type 2 will always be accepted, and they always have higher priorities than type 1. An arriving demand for type 2 will receive immediate service, if there are no other orders in the system for type 2; otherwise, it will wait until the manufacturer has satisfied the type 2 orders that arrived earlier. The expected profit generated by type 2 product is determined by the demand process itself. The manufacturer only maximizes the "extra" profit generated by the promotional product. Therefore, the two-product problem could be seen as a problem with a single MTS product, but under the impact of the higher-priority product. For the same reason, this problem can be easily extended with $n-1$ types of MTO products whose holding costs, $b_2, b_3, \ldots, b_n$ are all higher than $b_1$. Orders for different types of MTO products are sequenced according to their backlogging penalties, and their impacts on the MTS product are independent of their types.

Similar to our model with OEM and aftermarket products, here we present the expected due date quoting for type 1 and type 2 orders in Lemmas 6 and 7.

**Lemma 6** *The expected due date (i.e., leadtime) for a customized order is $T_2 = (y_2^- + 1) \times \frac{1}{\mu}$.*

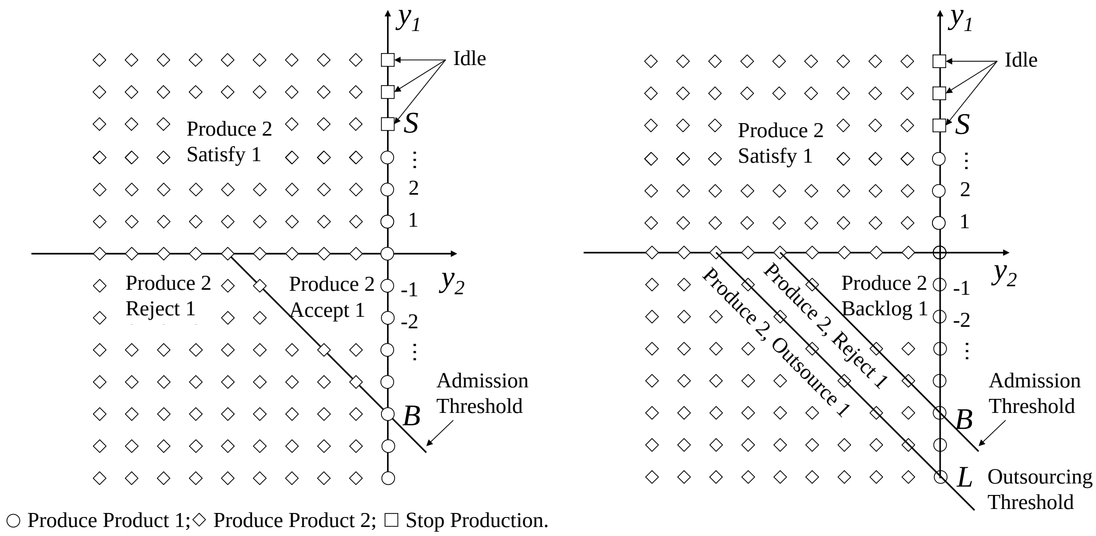


Figure 2: *Left:* The Optimal $(S, B)$ Policy; *Right:* The Optimal $(S, B, L)$ Policy with Contingent Outsourcing.

**Lemma 7** *If an order for the prefigured product cannot be satisfied immediately, the expected due date (i.e., leadtime) is* $T_1 = \frac{1}{\mu - \lambda_2}(y_2^- + y_1^- + 1)$.

## 3.3 Contingent Outsourcing for Promotional Orders

In this section, we assume the manufacturer can outsource an order for promotional product with a cost of $l_1$, $l_1 \geq r_1$. The manufacturer does not outsource orders for customized products because s/he wants to provide guaranteed quality to the primary customers. For the same reason as discussed in Section 2.4, in this case, the manufacturer needs to decide whether to outsource an order for type 1 only when an order for product 2 arrives. To adjust the model to this situation, we first change function $H_2$ as follows

$$H_2 J_1(y_1, y_2) = \max\Big\{J_1(y_1, y_2 - 1), J_1(y_1 + 1, y_2 - 1) - p_1 - l_1\Big\}.$$

We use the same approach in Section 2.4 to investigate the structure of the optimal policy for this model. Here we only present the main results. The supporting analytical details are provided in On-line Appendix E.

**Theorem 4** *The optimal policy for the problem with contingent outsourcing option for the make-to-stock orders is characterized by three parameters: the base-stock level, $S$, the admission level, $B$, and the outsourcing level, $L$ ($L \leq B$), such that at state $(y_1, y_2)$,*

- *Production control: if there are orders for type 2 (i.e., $y_2 \neq 0$), it is optimal to produce type 2; otherwise, it is optimal to produce type 1 if $y_1 < S$, and to stop production if $y_1 \geq S$.*

- *Admission control: it is optimal to satisfy an arriving demand for type 1 if $y_1 > 0$, to reject if $y_1 + y_2 \leq B$, and to backlog otherwise.*
- *Outsourcing control: it is optimal to outsource an order for type 1 after accepting an order of type 2 product, if $y_1 + y_2 \leq L$.*

The optimal policy is described in Figure 2-Right. Under the optimal policy, if the net total inventory level, $y_1 + y_2$, is lower or equal to $L$, the manufacturer will outsource an order for type 1 when an order for type 2 arrives. We refer to this optimal policy as the $(S, B, L)$ policy.

# 4 Computational Analysis

In this section we report the results of the computational study that we performed based on the model with OEM and aftermarket products. The computational analysis based on the model with customized and pre-configured products presents similar properties, and is therefore omitted. The goal of our numerical analysis is to investigate the following questions:

1. How much profit will a manufacturer lose if s/he uses the linear-structured $(S, R, B)$ policy instead of the optimal policy?
2. Outsourcing orders prevents large backlogging costs for the manufacturer. However, it also decreases customer satisfaction. How much will the profit decrease if the manufacturer does not utilize the flexibility of outsourcing?
3. What is the profit improvement of the $(S, R, B)$ policy relative to other commonly used policies? Under what circumstances the improvement is or is not very significant?
4. When the objective is to maximize profit, it is sometimes inevitable to sacrifice the service level of some customers. Therefore, an important question is: What is the impact of the $(S, R, B)$ policy on the service level of both the aftermarket and the OEM customers?

Our numerical study consists of 320 cases generated by varying the following parameters:

- $\rho = (\lambda_1 + \lambda_2)/\mu$: This parameter is an indication of the manufacturer's relative capacity compared to the market size. We refer to $\rho$ as *potential facility utilization* in the rest of the paper. Please notice that $\rho$ is not the real facility utilization since aftermarket demands may be rejected. We considered five values for $\rho$, specifically, $\rho \in \{0.6, 0.8, 1.0, 1.2, 1.4\}$.
- $\lambda_1/(\lambda_1 + \lambda_2)$: This demand ratio represents the size of the OEM demand compared to the aftermarket demand. In our numerical study we considered four cases, namely $\lambda_1/(\lambda_1 + \lambda_2) \in \{0, 0.3, 0.7, 1.0\}$.

- $p_2/p_1$: The price ratio represents the price difference between the two products. Since the price of aftermarket product is higher than the price offered to the OEM, we consider four values for $p_2/p_1$ that are all greater than one, namely $p_2/p_1 \in \{1.0, 1.3, 1.6, 2.0\}$. We did not consider ratios grater than 2, since it is very uncommon in practice to have cases in which an item is sold twice of its price for OEM.
- $b_2/b_1$: The penalty ratio represents the backlogging penalty difference between the two products. In the numerical study we assume the backlogging cost of each product is proportional to its price, and we consider four scenarios: $(b_1 = 20\%p_1, b_2 = 5\%p_2)$, $(b_1 = 40\%p_1, b_2 = 5\%p_2)$, $(b_1 = 100\%p_1, b_2 = 5\%p_2)$, and $(b_1 = 200\%p_1, b_2 = 5\%p_2)$. Note that, although $p_1 < p_2$, in all the scenarios, backlogging of type 1 customers is costlier than that of type 2 customers, which follows the assumption for the model with OEM and aftermarket products.

In order to better present the effects of the parameters on the system performance (and omit the effects of discount factor), our numerical study uses total expected profit per unit time as the performance measure.

## 4.1 $(S, R, B)$ Policy Versus Optimal Policy

To evaluate the performance of the $(S, R, B)$ policy, we first compare it with the optimal policy in our numerical study. To obtain the expected profit under the $(S, R, B)$ policy, we revise the admission decision in the MDP model to follow the linear threshold level $B$. This changes our MDP model from an optimization model to a performance evaluation model that obtains the optimal expected profits for the $(S, R, B)$ policy. The thresholds $(S, R, B)$ are found by searching. (A heuristic algorithm that can quickly approximate thresholds $(S, R, B)$ is also presented in On-line Appendix F.)

By examining the 320 cases in our numerical study, we found that, on average, the $(S, R, B)$ policy results in 0.6% less profit than the optimal policy. The maximum difference that we observed was 2.1%. Besides the similarities between the two policies, another key reason that the $(S, R, B)$ policy performs so close to the optimal policy is that the probability that the system visits a state in the upper left area (where $y_1 \geq R$ and $y_2 \leq B - R$ ) in Figure 1 is very small. Intuitively, on one hand, when the production capacity is tight, the manufacturer may have low inventory or high backorders for both products most of the time, and the system stays in the lower area (where $y_1 < R$) most of the time; On the other hand, when the production capacity is sufficient, the manufacturer may have high inventory level for product 1 and very few backorders for product 2, and the probability that the system stays in the upper right area (where $y_1 \geq R$ and $y_2 > B - R$) is high. So although the admission threshold may not be linear when $y_1 \geq R$, the frequency that we need to refer to it is very small.

The close performance of the linear $(S, R, B)$ policy to that of the optimal policy is an indication that changing the near-linear structured optimal policy to a linear heuristic does not have a significant impact on the expected profit. The $(S, R, B)$ policy approximates the optimal policy very well.

## 4.2 Effectiveness of Contingent Outsourcing

We also studied the performance of the outsourcing policy analyzed in Section 2.4. Interestingly, we found that in all the scenarios in our numerical study, outsourcing provides very slight profit increase (i.e., less than 0.8% in maximum), compared with the $(S, R, B)$ policy in which outsourcing is not applied. This is quite intuitive in the sense that under the $(S, R, B)$ policy, an aftermarket demand will be accepted only when system has enough capacity, so the probability that an accepted order is later be outsourced is very small. Therefore, for a manufacturer with the flexibility to accept or reject an order, having additional flexibility of outsourcing an order later does not have much value.

## 4.3 $(S, R, B)$ Policy Versus Simple Base-Stock Policy

Because of its non-linear component, the optimal policy is not easy to implement in practice. Fortunately, the $(S, R, B)$ policy has a simple structure and well approximates the optimal policy. In this section we compare the performance of the $(S, R, B)$ policy with a policy that is commonly used in practice (i.e., the base-stock policy). The objective is to investigate how much the expected profit increases if a manufacturing system switches from other commonly used policies to the $(S, R, B)$ policy.

The simple base-stock policy is often used in practice, and it is characterized by two threshold levels: base-stock level, $S'$, for type 1 products, and admission level, $B'$, for type 2 products. Under this policy, the system produces type 1 products as long as the inventory is less than $S'$. When the inventory of type 1 product reaches $S'$, the system produces type 2 products if there is an order of type 2 in the system, or idles otherwise. The orders for type 2 product will only be accepted if the number of orders of type 2 in the system is less than $B'$. We call this policy the $(S', B')$ policy. To obtain the optimal values for $S'$ and $B'$, we searched all possible combinations for values $S'$ and $B'$ and obtained $(S'^*, B'^*)$ that results in the maximum expected profit.

### 4.3.1 Impact on Expected Profit

Denoting $J_{SB}$ as the system's expected profit per unit time under the optimal $(S', B')$ policy, we evaluate the profit improvement under the $(S, R, B)$ policy using the following measure:

$$\text{Profit Potential}_{SRB} = \frac{J_{SRB} - J_{SB}}{J_{SB}} \times 100\%, \tag{9}$$

where $J_{SRB}$ is the expected profit per unit time under the $(S, R, B)$ policy.

Based on our numerical study, we found that, on average, the profit improvement obtained by using the $(S, R, B)$ policy is 8%. The maximum profit improvement can be up to 40%.

We also examined the effects of price ratio and potential facility utilization on the profit improvement. Figure 3-Left, which is for one set of examples among several that we studied, depicts the typical behavior of the system. In Figure 3-Left, we fix $\lambda_1/(\lambda_1+\lambda_2)=0.7$ and $b_1=20\%p_1$, and we change the price ratio and the potential facility utilization rate. As the figure shows, when the price ratio increases, implementing the $(S,R,B)$ policy results in more profit improvement. The reason is as follows. Both the $(S,R,B)$ and the base-stock policies do not reject the type 1 orders, and thus in the long-run, the number of satisfied type 1 orders is the same under both policies. However, the number of type 2 orders satisfied under these two polices are different. The $(S,R,B)$ policy satisfies more type 2 orders than the base-stock policy, because as shown in Section 2.2, the $(S,R,B)$ policy is almost as efficient as the optimal policy (which has the highest number of accepted orders of type 1). As a result, the gap between the performance of these two policies (i.e., the profit potential) increases as the orders of type 2 become more profitable (i.e., the price ration increases).

The figure also suggests that, when the capacity is sufficient ($\rho \leq 1.0$), profit potential increases as $\rho$ increases. This is true since as potential facility utilization rate increases, i.e., capacity becomes tighter, it is more and more important to allocate capacity effectively between the two classes of products, which is exactly what the $(S,R,B)$ policy achieves. However, when $\rho > 1.0$ (please refer to Figure 8-Left in On-line Appendix H), the expected profit under both the $(S,R,B)$ policy and the simple base-stock policy turns negative as $\rho$ further increases, and the profit difference between the two policies becomes smaller and finally approaches zero. In this case, since the production capacity is very tight, both policies behave the same, keeping the facility running almost all the time and rejecting aftermarket demand if capacity is not enough. So manufacturers benefit most from implementing the $(S,R,B)$ policy when the capacity is neither too tight nor too loose.

We also examined the effects of backlogging penalty ratio for OEM products. For this purpose, we keep $b_2=5\%$ unchanged, and increase $b_1$ from 20% to 40%, 100%, and 200%. We find that the profit potential of the $(S,R,B)$ policy decreases and finally approaches zero. Intuitively, as the OEM backlogging penalty increases, giving priority to OEM product becomes more critical, and thus any policy such as our base-stock policy that gives higher priority to the OEM also performs relatively well. Therefore, the gap between the performance of the base-stock policy and the $(S,R,B)$ policy (i.e., the profit potential) decreases as the OEM backlogging penalty increases.

Figure 3-Right shows the typical system behavior of how changes in demand ratio and potential facility utilization affect the profit improvement. Figure 3-Right corresponds to one set of our numerical study in which $p_2/p_1=1.6$ and $b_1=20\%p_1$. As the figure suggests, when the demand ratio, $\frac{\lambda_1}{\lambda_1+\lambda_2}$, equals 0 or 1,

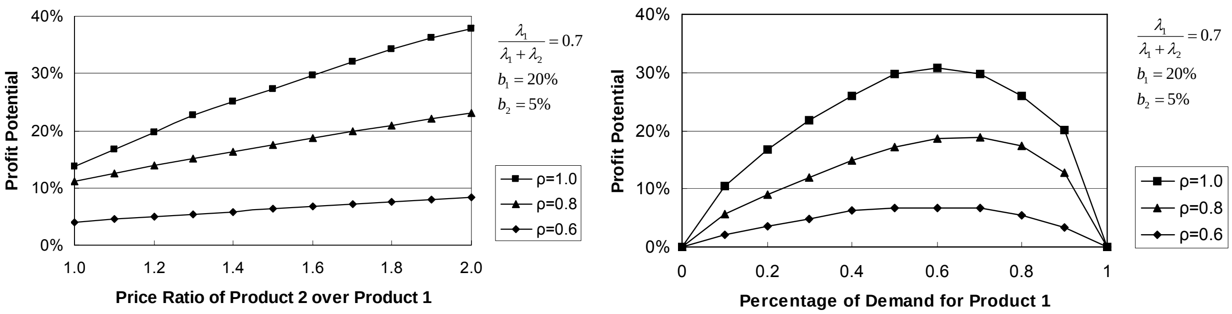


Figure 3: *Left:* Impact of Price Ratio $(p_2/p_1)$; *Right:* Impact of Demand Ratio $\lambda_1/(\lambda_1+\lambda_2)$.

there is only one type of products in the system, and therefore there is no difference between the $(S, R, B)$ policy and the simple base-stock policy. However, as the demand ratio increases from 0 to 1, the relative performance of the $(S, R, B)$ policy increases and then decreases. The reason is the same as what was described for Figure 3-Left. When the two products are intensively competing for the capacity, the benefit of the $(S, R, B)$ policy becomes prominent, because the $(S, R, B)$ policy can manage the production and the inventory more effectively than the simple base-stock policy.

Figure 3-Right also depicts that, when $\rho \leq 1.0$, the profit improvement of the $(S, R, B)$ policy increases as the total demand relative to the production capacity increases. However, when $\rho > 1.0$ (please refer to Figure 8-Right in On-line Appendix H), the expected profit under both the $(S, R, B)$ policy and the simple base-stock policy turns negative as $\rho$ further increases, and the profit difference between the two policies becomes smaller and finally approaches zero. The reason is the same as for Figure 3-Left.

### 4.3.2 Impact on Service Levels

When the objective is to maximize profit, it is sometimes inevitable to sacrifice the service level of some customers. Therefore, it is important to understand what the impact of the $(S, R, B)$ policy is on the service levels of both the aftermarket and the OEM customers. Thus, in this section we study the impact of the $(S, R, B)$ policy on customers' service levels.

We measure the service level for aftermarket by the fraction of aftermarket orders that are accepted by the manufacturer. On the other hand, we measure the service level for OEM demand by the fraction of OEM orders that are immediately satisfied from inventory.

Figure 4-Left shows the service levels for aftermarket demands under the $(S, R, B)$ policy and the simple $(S', B')$ policy, respectively. As the figure shows, when the price ratio increases, the service level for aftermarket demands increase slightly under both policies. This is very intuitive, since the manufacturer will

accept more aftermarket demands when they contribute more profits.

Figure 4-Left also shows that when the production capacity is sufficient ($\rho = 0.8$), both policies maintain high service levels for aftermarket demands, and the service level under the $(S, R, B)$ policy is slightly higher than that under the simple base-stock policy. As capacity becomes tight ($\rho = 1.0$), the $(S, R, B)$ policy provides higher service level than the simple base-stock policy. Since the $(S, R, B)$ policy can manage the limited capacity more effectively than the simple policy, more aftermarket demand could be satisfied under the $(S, R, B)$ policy. However, as $\rho$ further increases ($\rho = 1.2$), the difference in service levels of the two policies becomes smaller and finally approaches zero. In this case, since the production capacity is very tight, both policies behave similarly, keeping the facility running almost all the time and rejecting aftermarket demands if the number of orders exceeds the threshold.

Figure 4-Right illustrates the service level for OEM demand, i.e., the percentage of OEM orders that are immediately satisfied upon their arrivals. When the price ratio increases, the service level for OEM demand under both the $(S, R, B)$ policy and the simple base-stock policy decreases.

In Figure 4-Right, we find that the service level for OEM is lower under the $(S, R, B)$ policy than under the simple base-stock policy. This can be explained in two ways. One reason is that the $(S, R, B)$ policy sacrifices the service level for OEM demands in order to accept more demands from the aftermarket (which has a higher profit margin) to improve the overall profit. Another reason is that the service level for OEM demand is mainly determined by $R$ in the $(S, R, B)$ policy, and by $S'$ in the simple policy. In the numerical results, we find $S' \geq R$, so the the service level for OEM demand is higher under the simple policy. The reason for $S' \geq R$ is also intuitive. In the simple policy, $S'$ acts as both the base-stock level for product 1 and the rationing level for product 2, so the value of $S'$ is usually between $S$ and $R$.

Figure 4-Right also shows that when the production capacity is sufficient ($\rho = 0.8$), both policies maintain high service levels for OEM, and the service level difference between the two policies is very small. As capacity becomes tight ($\rho = 1.0$), the service levels of both policies decrease, and the difference between them increases. As $\rho$ further increases ($\rho = 1.2$), the service level difference between the two policies becomes smaller, which could be explained similarly as for Figure 4-Left.

#### 4.3.3 Impact on Expected Leadtimes

The service levels for both types of demands can also be measured by the expected leadtime, which directly depends on the average number of orders in the system upon arrival of an order (see Lemmas 3 and 4). In this section, we compare the average number of orders in the system under both the $(S, R, B)$ and the simple base-stock policies.

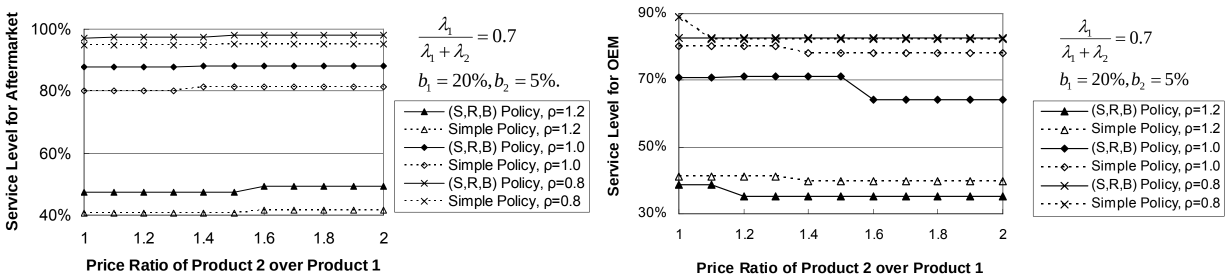


Figure 4: *Left*: Service Level for Aftermarket Demands under the $(S, R, B)$ Policy and the $(S', B')$ Policy; *Right*: OEM Service Level under the $(S, R, B)$ Policy and the $(S', B')$ Policy

Figure 5-Left illustrates that the average number of aftermarket orders under the $(S, R, B)$ policy is higher than that under the simple base-stock policy, which is mainly due to more demands being accepted under the $(S, R, B)$ policy. It, together with the result in Figure 5-Right, implies that an incoming aftermarket demand, if accepted, may expect a longer waiting time (leadtime) under the $(S, R, B)$ policy than under the simple base-stock policy.

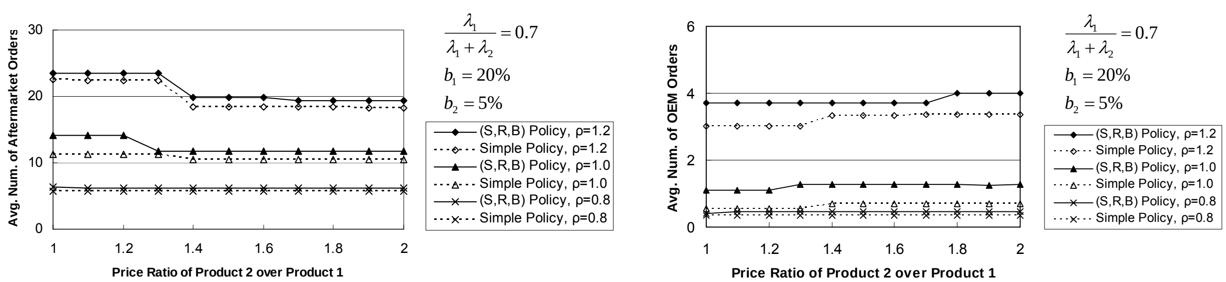


Figure 5: *Left*: Average Number of Aftermarket Orders under the $(S, R, B)$ Policy and the $(S', B')$ Policy; *Right*: Average Number of OEM Orders under the $(S, R, B)$ Policy and the $(S', B')$ Policy.

Figure 5-Right also illustrates that the average number of OEM orders under the $(S, R, B)$ policy is higher than that under the simple base-stock policy, which is mainly due to less OEM demands being immediately satisfied under the $(S, R, B)$ policy. It implies that an incoming OEM demand, if not immediately satisfied, may expect a longer waiting time (leadtime) under the $(S, R, B)$ policy than under the simple base-stock policy.

# 5 Extension to Erlang Model

In this section we extend the model of OEM and aftermarket products to cases with Erlang distributed interarrival and production times. Erlang distribution is a flexible distribution that can cover a wide range of processes in practice. It also allows us to use *Method of Stages* (see Gross and Harris [14]) to develop an MDP model and obtain the optimal expected profit. The objective of the extension is to answer the following questions:

1. Will the near-linear structure of the optimal policy still hold with Erlang distributed interarrival and production times? If not, how much profit will a manufacturer lose if s/he uses linear threshold structure instead of the optimal policy with non-linear structure?

2. How does the difference in production rates of the two product affect the manufacturer's inventory control policy as well as the expected profit?

3. How does the variability in demand and production processes influence the manufacturer's inventory control policy as well as the expected profit?

In the extended model, we assume the demand interarrival time for product $i$ follows an *Erlang*-$K_i$ distribution, and the production time of a type $i$ product is an *Erlang*-$m_i$ random variable. The phase arrival rate in the demand process for product $i$ is $K_i\lambda_i$, so the demand arrival rate for product $i$ is still $\lambda_i$. The time to finish one phase of product $i$ follows an exponential distribution with mean $1/\gamma_i$. Thus, the average time to produce a type $i$ product is $1/\mu_i = m_i/\gamma_i$, and the weighted average production time $1/\bar{\mu}$ is calculated as

$$\frac{1}{\bar{\mu}} = \frac{\lambda_1}{\lambda_1+\lambda_2}\Big(\frac{m_1}{\gamma_1}\Big) + \frac{\lambda_2}{\lambda_1+\lambda_2}\Big(\frac{m_2}{\gamma_2}\Big). \tag{10}$$

Let $s_i$ be the number of production phases completed for an unfinished job of product $i$. The problem can be formulated in terms of an aggregate state variable, i.e., the work storage level, $x_i = s_i + K_i y_i$, which links inventory and partially completed production [18]. We describe the system state with a tuple of four variables, $(x_1, x_2, k_1, k_2)$, where $x_1 \in Z$ and $x_2 \in Z^-$ are the work storage level of each product, and $k_i \in \{1, 2, \ldots, K_i\}$ is the number of phases until the arrival of the next type $i$ demand. Note that, having product $i's$ work storage level $x_i$, it is easy to calculate its inventory level[5], $y_i = \lfloor \frac{x_i}{m_i} \rfloor$, and the number of finished phases of a job under production, $s_i = x_i - m_i\lfloor \frac{x_i}{m_i} \rfloor$.

The Erlang model is more complicated than the exponential model, and it is difficult to obtain analytical properties (Please refer to On-line Appendix I for the MDP formulations). In the following, we therefore perform an extensive numerical study to gain some insights into the system performance. We test the same 320 cases defined in Section 4, and for each case, we consider 50 combinations of $(K_1, K_2)$ and $(m_1, m_2)$, by fixing one pair at $(1, 1)$ and letting each parameter of the other pair take a value out of $\{1, 2, 3, 4, 5\}$. This results in an extensive set of examples consisting of 16,000 ($= 320 \times (25 + 25)$) cases. We compare the results

[5] Note that in our model, the inventory can be positive or negative.

of the above Erlang model and the exponential model developed in Section 2.1. To make the two models comparable, we let the average production times in the two models to be equal to each other.

## 5.1 Impact on the Structure of the Optimal Policy

Our numerical results show that the optimal policy for the Erlang model does not always follow the same structure as that for exponential model. Figure 6-Left shows an example of the optimal policy. The horizontal and vertical axis in that figure correspond to $x_2$ and $x_1$, respectively, which are the work storage levels in the system. As the figure shows, the optimal policy has a stepwise function and does not follow the linear structure similar to the exponential case.

Note that, since production rates are different for both products, completing one stage of product 1 takes $1/\gamma_1$ in average, while completing one stage of product 2 takes $1/\gamma_2$. Thus, the average amount of completed work in the system for product 1 is $x_1/\gamma_1$, and for product 2 is $x_2/\gamma_2$. So, one can use *average work-time-storage levels* to represent the optimal policy, which is done in Figure 6-Middle. Figure 6-Middle shows the optimal policy in 6-Left, but in terms of average work-time-storage levels in the system. The vertical and horizontal axis, $x_1/\gamma_1$, and $x_2/\gamma_2$, respectively, are measured in time.

As the figure shows, the optimal threshold levels are still not linear, but they fluctuate around some straight lines. For example, the admission control threshold, although a non-linear function, looks fluctuating around a 45° line. We examine several different examples and we observed the same phenomenon, i.e., the optimal rationing thresholds fluctuate around a horizontal line, and the optimal admission thresholds fluctuate around a 45° line. This suggests that the linear structure in the exponential model, if translated into average work-time-storage levels in the system, may be a good heuristic policy for systems with non-exponential (i.e., Erlang) production and demand interarrival times.

To examine this, we compared the expected profit under the optimal policy for the Erlang model, with a heuristic policy, called *Linear Work-Time-Storage* $(S_L, R_L, B_L)$ *Heuristic Policy.* This policy has a similar linear structure as the $(S, R, B)$ policy in our exponential model. The only difference is that threshold levels $S_L$, $R_L$ and $B_L$ are measured in average work-time-storage level instead inventory level. An example of the structure of this policy is shown in Figure 6-Right. Under this policy, for example, an arriving order is accepted at state $(x_1, x_2)$ if the total average work-time-storage $(x_1/\gamma_1) + (x_2/\gamma_2)$ is greater than $B_L$. To find the optimal $(S_L, R_L, B_L)$ thresholds, we search for the values of $S_L$, $R_L$, and $B_L$ that maximize the expected profit.

To evaluate the performance of the Linear Work-Time-Storage Heuristic Policy, we compared its expected profit with that under the optimal policy for the Erlang model in an extensive numerical study. To obtain the

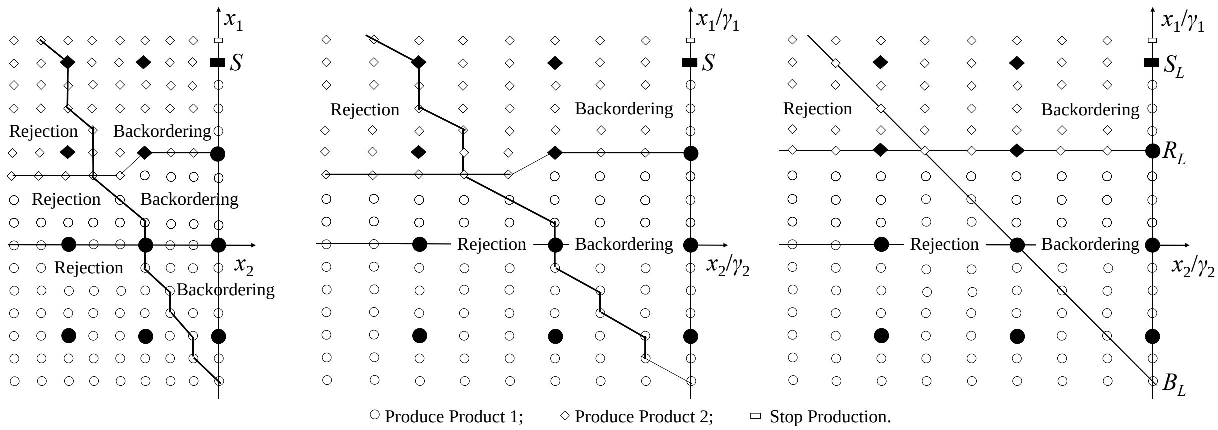


Figure 6: *Left*: Optimal Policy in Work-Storage Levels; *Middle*: Optimal Policy in Work-Time-Storage Levels; *Right*: Linear Work-Time-Storage Heuristic Policy. In all figures, filled points correspond to system states with no uncompleted jobs, i.e., $x_i/m_i \in Z$.

expected cost under our heuristic policies, we revise the Erlang model to follow the production, rationing, and admission decisions according to our heuristic policies. This changes our MDP model from an optimization model to a performance evaluation model that obtains the optimal expected profits for our heuristics.

By examining the 16,000 cases in our numerical study, we found that, on average, the linear work-time-storage heuristic policy results in 2.0% less profit than the optimal policy. The maximum difference that we observed was 4.4%. The very close performance of the linear work-time-storage heuristic policy to the optimal policy is an indication that changing the nonlinear structure of the optimal policy, to a linear structure (as in exponential model) does not have a significant impact on the expected profit.

## 5.2 Impacts of Production Rates

To examine the impact of the production rate of each product, we fix the values of $\lambda_i$, $K_i$, and $m_i$, for $i = 1, 2$, and change the value of $\gamma_i$. We choose the values for $\gamma_1$ and $\gamma_2$ to make sure the weighted average production time is the same in all the cases (please refer to Equation (10)).

Compared to reducing average production time of the aftermarket products, we find that in most cases of our numerical study, reducing the average production time for the OEM product is more profitable (unless when $\lambda_2/(\lambda_1 + \lambda_2)$ is close to 1, which means the amount of OEM demand is insignificant compare to aftermarket demand). Intuitively, this could be explained by the following three reasons: (i) The backlogging cost for type 1 orders is much higher than that for type 2 orders, and increasing the production rate for the OEM product can save the manufacturer's backlogging cost; (ii) Since product 1 has higher priority than product 2, there is a downstream impact of product 1 on product 2's production. Type 2 orders can only be processed after all type 1 orders are satisfied and rationing level has been built up. So reducing product 1's production time not only reduce the average backlogging time of type 1 orders, but can also reduce the holding time for type 2 orders. However, in the other direction, reducing product 2's production time has little impact on product 1. (iii) In our numerical analysis, we also find that increasing the production rate for the OEM product results in smaller base-stock and rationing threshold levels. This is because the manufacture tends to keep less safety stock for product 1, as the product could be produced in a faster rate. Smaller base-stock and rationing threshold levels save the manufacturer the inventory holding cost for the OEM product. At the same time, smaller rationing threshold level also allows the manufacturer to process aftermarket orders earlier, as well as to accept more aftermarket demands, which increases the manufacturer's profit.

## 5.3 Impacts of Demand and Production Variabilities

To examine the impacts of the variabilities of demand and production times separately, we study two scenarios. In the first scenario, we fix the values of $\gamma_i$, $m_i$, and $\lambda_i$, for $i = 1, 2$, and change the values of $K_1$ and $K_2$ to get different demand variabilities. In the second scenario, we fix the value of $\lambda_i$, $K_i$, and $\mu_i$, $i = 1, 2$, and change the values of $m_i$ to evaluate the impact of the variability of production times. The values of $\gamma_1$ and $\gamma_2$ are adjusted at the same time in order to keep the average processing time of each product, $\mu_i = \frac{m_i}{\gamma_i}$ unchanged.

In the first scenario, we find that the manufacturer's expected profit increases as the demand arrival process becomes less variable. As the system becomes more deterministic and more predictable, the manufacturer accepts more aftermarket demands without sacrificing the service level for either product.

Although reducing the demand variability for either product improves the system profit, however, we find that reducing the variability of OEM demand is more effective in most cases (unless when $\lambda_2/(\lambda_1 + \lambda_2)$ is close to 1). Intuitively, this is caused by the downstream impact of product 1 on product 2's production. Reducing the variability of OEM demand results in lowered base-stock and rationing threshold levels, because the manufacture tends to keep less safety stock for product 1, as demand arrival process becomes more predictable. Smaller base-stock and rationing threshold levels save the manufacturer the inventory holding cost for the OEM product. At the same time, smaller rationing threshold level also allows the manufacturer to process aftermarket orders earlier and accept more aftermarket demands.

In the second scenario, we find that the manufacturer's expected profit increases as the production time of either product becomes less variable. We also find that reducing the variability in production time of the OEM products is more beneficial compared to reducing the variability in the production time of the aftermarket products. Both results could be explained similarly based on the smaller base-stock and rationing levels for the OEM product.

# 6 Conclusions

In this paper, we studied the production and inventory policies in two different make-to-stock/make-to-order manufacturing environments. For the model with both OEM and aftermarket demands, we show that the optimal policy has a near-linear structure characterized by three parameters, the base-stock level, the rationing level, and the admission level. We extend the optimal policy to problems in which aftermarket orders can be outsourced, as well as to problems with multiple types of aftermarket products. Numerical study shows that a linear structured heuristic approximates the optimal policy very well. For the model where the manufacturer offers both customized and pre-configured (promotional) products, we show that the optimal policy has a linear structure characterized by two parameters, the base-stock level and the admission level. We also extend the optimal policy to problems with outsourcing option for pre-configured products.

After conducting an extensive computational analysis, we observed the following:

1. Implementing the $(S, R, B)$ policy can result in, on average, 8% (and sometimes up to 40%) more profit than that under the commonly used base-stock policy. This difference in profit is high when production capacity is tight, when the backlogging cost for MTS product and the backlogging cost for MTO product are close, or when demand arrival rates in the two classes are close.

2. When the manufacturer has the option of rejecting low priority orders, having the flexibility of outsourcing an accepted low-priority order does not provide a significant increase in profit.

3. The $(S, R, B)$ policy improves the profit by allocating the production capacity more effectively, and therefore having the capability to accept more low-priority demands. As a tradeoff, when the production capacity is very tight, the service level for the high-priority demands under the optimal policy is lower than that under the simple base-stock policy.

4. The linear work-time-storage policy works as a good heuristic for problems with non-exponential inter-arrival and production times. Our numerical analysis shows that the expected profit under the linear work-time-storage heuristic, on average, is only 2.0% lower than that under the optimal solution. The maximum difference that we observe was 4.4%. The good performance and simple structure of our linear inventory heuristic makes it an attractive policy to use in practice.

5. The production time and variability of the high-priority product has more impact on the system performance than those of the low-priority one, because of the down stream effect of the high-priority product.

In this paper we investigated the capacity allocation in a MTS/MTO system. An important future research direction is the joint pricing and capacity allocation decisions for multiple products. So far, academic literature on joint pricing and allocation decisions has confined itself mainly to single-product and single-class-customer problems, which are not applicable to the current practice of market segmentation in the manufacturing industry. The complexity of the joint pricing and allocation problem among multiple products increases dramatically due to the fact of demand diversions among different products, i.e., the demands for one product not only depends on their own price, but may also be influenced by the prices of other products. Another issue complicating the problem is that different products may compete for same resources, i.e., the production capacity and components.

# References


[1] Benjaafar, S., M. ElHafsi and T. Huang. Optimal control of a production-inventory system with both backorders and lost sales. Working Paper. University of Minnesota.

[2] Bertsekas, D. 1995. *Dynamic Programming and Optimal Control*, Volume Two, Athena Scientific.

[3] Carr, S. and I. Duenyas. Optimal admission control and sequencing in a make-to-stock/make-to-order production system. *Operations Research*, 48, pp. 709-720.

[4] Chen F. and V. G. Kulkarni. 2007. Individual, class-based, and social optimal admission policies in two-priority queues. *Stochastic Models*, 23, pp 97127.

[5] Cohen, M.A., P.R. Kleindorfer and H.L. Lee. 1988. Service constrained (s,S) inventory systems with priority demand classes and lost sales. *Management Science*, Vol. 34, pp. 482-499.

[6] de Vericourt F., F. Karaesmen and Y. Dallery. 2000. Dynamic scheduling in a make-to-stock system: s partial characterization of optimal policies. *Operations Research*, 48, pp 811-819.

[7] de Vericourt F., F. Karaesmen and Y. Dallery. 2002. Optimal stock allocation for a capacitated supply system. *Management Science*, 48, pp 1486-1501.

[8] Dobson, G. and C.A. Yano. 2001. Production offering, pricing, and make-to-stock/make-to-order decisions with shared capacity. *Production and Operations Management*, 11, pp 293-312.

[9] Duenyas, I. and M. Van Oyen. 1996. Heuristic scheduling problem with random demand. *Management Science.* Vol. 42, pp. 814-829.

[10] Gittins, J.C., *Multi-armed Bandit Allocation Indices*, Wiley, New York, 1989.

[11] Glasserman, P., and D. Yao. 1994. Monotone optimal control for permutable GSMPs. *Mathematics of Operations Research*, 19, pp 449-476.

[12] Graves, S.C. 1980. The multi-product production cycling problem. *AIIE Trans.* 12, pp 233-240.

[13] Gupta, D. and L. Wang. 2007. Capacity Management for Contract Manufacturing. *Operations Research*, 55(2), pp. 367377.

[14] Gross D. and C.M. Harris. 1985. *Fundamentals of Queueing Theory*, John Wiley & Sons.

[15] Ha, A. 1997. Inventory rationing in a make-to-stock production system with several demand classes and lost sales. *Management Science*, 43, pp 1093-1103.

[16] Ha, A. 1997. Stock-rationing policy for a make-to-stock production system with two priority classes and backordering. *Naval Research Logistics*, 44, pp 458-472.

[17] Ha, A. 1997. Optimal dynamic scheduling policy for a make-to-stock production system. Operations Research, 45, pp 42-53.

[18] Ha, A. 2000. Stock-rationing in an M/Ek/1 make-to-stock queue. *Management Science*, 46, pp. 77-87.

[19] Huang, B., and S.M.R. Iravani. 2008. Technical Note - A make-to-stock system with multiple customer classes and batch ordering. *Operations Reserach*, 56, 1312-1320.

[20] Iravani, S.M.R., B. Kolfal, M.P. Van Oyen. Technical Note - Inventory and Process Flexibility in a Make-to-Stock Systems with Product Substitution. Working paper. Northwestern University.

[21] Iravani, S.M.R., T. Liu, K. L. Luangkesorn, and D. Simchi-Levi. 2007. A produce-to-stock system with advance demand information and secondary customers. *Naval Reserach Logistic*, 54, 331-345.

[22] Kleijn, M. and R. Dekker. 1998. An overview of inventrory systems with several demand classes. Report EI9838/A, Econometric Institute, Erasmus University Rotterdam, The Netherlands.

[23] Melchiors, P., R. Dekker and M.J. Kleijn. 2000. Inventory rationing in an (s, Q) inventory model with lost sales and two demand classes. *Journal of the Operational Research Society*, 51, pp 111-122.

[24] Moinzadeh, K. 1989. Operating characteristics for the $(S-1, S)$ inventory system with partial backorders and constant resupply time. *Management Science*, 35, 472-477.

[25] Montgomery, D. C., M. S. Bazaraa and A. K. Keswani. 1981. Inventory models with a mixture of backorders and lost sales. *Naval Research Logistics*, 20, pp. 255-263.

[26] Nahmias, S. and W.S. Demmy. 1981. Operating characteristics of an inventory system with rationing. *Management Science*, 27, pp. 1236-1245.

[27] Nahmias, S. and S. A. Smith. 1994. Optimizing inventory levels in a two-echelon retailer system with partial lost sales. *Management Science*, 40, pp. 582-596.

[28] Pena-Perez, A. and P. Zipkin. 1997. Dynamic scheduling rules for a multi-product make-to-stock queue. *Operations Research*, 45, pp 919-930.

[29] Porteus, E. 1982. Conditions for characterizing the structure of optimal strategies in infnite-horizon dynamic programming. Journal of Optimization Theory and Applications, Vol. 36, pp. 419-431.

[30] Puterman M.L. 1994. *Markov Decision Processes*. Wiley, New York.

[31] Reiman, M. I. and L. M. Wein. 1998. Dynamic scheduling of a two-class queue with setups. *Operations Research*, 46, pp. 532-547.

[32] Rabinowitz, G. A., A. Mehrez, C. Chull and B. E. Patuwo. 1995. A partial backorder control for continuous review $(r, Q)$ inventory system with poisson demand and constant leadtime. *Computers and Operations Research*, 22, pp. 689-700.

[33] Smeitink, E. 1990. A note on the operating characteristics of the $(S-1, S)$ inventory systems with partial backorders and constant resupply times. *Management Science*, 36, pp. 1413-1414.

[34] Sennott, L.I. 1999. Stochastic Dynamic Programming and the Control of Queueing Systems. Wiley.

[35] Topkis D.M. 1968. Optimal ordering and rationing policies in a nonstationary dynamic inventory model with n demand classes. *Management Science*, 15, pp. 160-176.

[36] Varaiya, P., J. Walrand and C. Buyukkoc. 1985. Extensions of the multi-armed brandit problem. *IEEE Transactions Automatic Control*, Vol. AC-30, pp.426-439.

[37] Veatch, M. and L.M. Wein. 1996. Scheduling a make-to-stock queue: index policies and hedging points. *Operations Research*, 44, pp 634-647.

[38] Wein L. M.. 1992. Dynamic scheduling of a multiclass make-to-stock queue. *Operations Research*, 40, pp. 724-735.

[39] Youssef, K.H., C. van Delft, and Y. Dallery. 2004. Efficient scheduling rules in a combined make-to-stock and make-to-order manufacturing system. *Annals of Operations Research*, 126, pp 103-134.

[40] Zheng, Y. and P. Zipkin. 1990. Aqueueing model to analyze the value of centralized information. *Operations Research*, 38, pp 296-307.

# ON-LINE APPENDIX A
## Proof of Analytical Results

**PROOF OF PROPOSITION 1**

**Property (i):** Property (i) is a direct result of Condition C.1.3. Since $\Delta_1 f(y_1, y_2)$ is non-increasing in $y_1$ and $y_2$, there exists a threshold level, $S^f = \min\{z|\Delta_1 f(z, 0) < 0\}$, such that $\Delta_1 f(y_1, y_2) \geq 0$ for $y_1 < S^f$ (remember $y_2$ is non-positive).

**Property (ii):** Property (ii) is a direct result of Condition C.1.4. Since $\Delta_{12} f(y_1, y_2)$ is non-increasing in $y_1$, for a given value of $y_2$, there exists a threshold level, $Q^f(y_2) = \min\{z|\Delta_{12} f(z, y_2) < 0\}$, such that $\Delta_{12} f(y_1, y_2) \geq 0$ for $y_1 < R^f(y_2)$.

**Property (iii):** Property (iii) is a direct result of Condition C.1.3. Since $\Delta_2 f(y_1, y_2)$ is non-increasing in $y_2$, for a given value of $y_1$, there exists a threshold level, $A^f(y_1) = \max\{z|\Delta_2 f(y_1, z-1) > p_2 + r_2\}$, such that $\Delta_2 f(y_1, y_2) \leq p_2 + r_2$ for $y_2 > A^f(y_1)$.

**Property (iv) Left:** By Condition C.1.3, $\Delta_2 f(y_1, y_2) \geq \Delta_2 f(y_1 + 1, y_2)$. Therefore $A^f(y_1) - A^f(y_1 + 1) \geq 0$.

**Property (iv) Right:** By Condition C.1.4, we have,

$$\begin{array}{rrcll} & \Delta_{12} f(y_1, y_2 - 1) & \geq & \Delta_{12} f(y_1, y_2 - 2) & (C.1.4) \\ \Rightarrow & f(y_1 + 1, y_2 - 1) - f(y_1, y_2) & \geq & f(y_1 + 1, y_2 - 2) - f(y_1, y_2 - 1) & (Definition) \\ \Rightarrow & f(y_1 + 1, y_2 - 1) - f(y_1 + 1, y_2 - 2) & \geq & f(y_1, y_2) - f(y_1, y_2 - 1) & (Rearrangement) \\ \Rightarrow & \Delta_2 f(y_1 + 1, y_2 - 2) & \geq & \Delta_2 f(y_1, y_2 - 1) & (Definition) \end{array}$$

Suppose $y_2^A = A^f(y_1)$, then by Property (iii), $\Delta_2 f(y_1, y_2^A - 1) > p_2 + r_2$. With the above inequality, we have $\Delta_2 f(y_1 + 1, y_2 - 2) > p_2 + r_2$. By C.1.3, $\Delta_2 f(y_1, y_2)$ is non-increasing in $y_2^A$, so $y_2^A - 1 \leq A^f(y_1 + 1)$. Therefore $A^f(y_1) - A^f(y_1 + 1) \leq 1$.

**Property (v) Left:** By contradiction, assume $Q^f(-1) > S^f$. By Condition C.2.2, we will have $\Delta_{12} f(S^f, -1) \geq 0$. Notice that $\Delta_{12} f(S^f, -1) = f(S^f + 1, -1) - f(S^f, 0) + f(S^f + 1, 0) - f(S^f + 1, 0) = \Delta_1 f(S^f, 0) - \Delta_2 f(S^f + 1, -1)$, so we get $\Delta_1 f(S^f, 0) \geq \Delta_2 f(S^f + 1, -1)$, which cannot be true since we know that $\Delta_1 f(S^f, 0) < 0$ (C.2.2) and $\Delta_2 f(S^f + 1, -1) \geq 0$ (C.1.1).

**Property (v) Right:** By Condition C.1.4, $\Delta_{12} f(y_1, y_2)$ is non-decreasing in $y_2$, and thus, $Q^f(-1) \geq Q^f(y_2)$, for $y_2 < -1$. ■

**PROOF OF LEMMA 1**

We prove that function $H$ preserves each condition in $\mathcal{C}1$. In the following, we denote $\mathbf{y} = (y_1, y_2)$, $\mathbf{e_0} = (0, 0)$, $\mathbf{e_1} = (1, 0)$, and $\mathbf{e_2} = (0, 1)$.

**Proof for Condition C.1.1:** We prove that $H_k f$ preserves Condition C.1.1, i.e., $\Delta_i f(\mathbf{y}) \geq 0$, for $y_i < 0$, and $i = 1, 2$. Since

$$\Delta_i H f(\mathbf{y}) = \Delta_i c(\mathbf{y}) + \Delta_i H_0 f(\mathbf{y}) + \Delta_i H_1 f(\mathbf{y}) + \Delta_i H_2 f(\mathbf{y}),$$

in order to prove that $\Delta_i H f(\mathbf{y}) \geq 0$ for $y_i < 0$, we show that all terms on the right hand side of the above equation are non-negative for $y_i < 0$ and $i = 1, 2$.

- For $H_0 f$, note that $H_0 f(\mathbf{y} + \mathbf{e_i})$ can be written as $H_0 f(\mathbf{y} + \mathbf{e_i}) = f(\mathbf{y} + \mathbf{e_i} + \mathbf{e_p})$, where $\mathbf{e_p}$ is one of the following three cases: $\mathbf{e_0}$, $\mathbf{e_1}$, or $\mathbf{e_2}$. Similarly, $H_0 f(\mathbf{y}) = f(\mathbf{y} + \mathbf{e_q})$, where $\mathbf{e_q}$ can be $\mathbf{e_0}$, $\mathbf{e_1}$, or $\mathbf{e_2}$. Then we have

$$\begin{array}{rcll} \Delta_i H_0 f(\mathbf{y}) & = & H_0 f(\mathbf{y} + \mathbf{e_i}) - H_0 f(\mathbf{y}) & (Definition) \\ & = & f(\mathbf{y} + \mathbf{e_i} + \mathbf{e_p}) - f(\mathbf{y} + \mathbf{e_q}) & (Above) \\ & \geq & f(\mathbf{y} + \mathbf{e_i} + \mathbf{e_q}) - f(\mathbf{y} + \mathbf{e_q}) & (Max) \end{array}$$

$$
\begin{array}{lll}
= & \Delta_i f(\mathbf{y}+\mathbf{e_q}) & (Definition) \\
\geq & 0 & (C.1.1)
\end{array}
$$

- For $H_1 f$, we have $H_1 f(\mathbf{y}+\mathbf{e_i}) = f(x+\mathbf{e_i}-\mathbf{e_1})+p_1$ and $H_1 f(\mathbf{y}) = f(\mathbf{y}-\mathbf{e_1})+p_1$. Then we will have

$$
\Delta_i H_1 f(\mathbf{y}) = H_1 f(\mathbf{y}+\mathbf{e_i}) - H_1 f(\mathbf{y}) = f(\mathbf{y}+\mathbf{e_i}-\mathbf{e_1})+p_1 - f(\mathbf{y}-\mathbf{e_1}) - p_1 = \Delta_i f(\mathbf{y}-\mathbf{e_1}) \geq 0.
$$

- For $H_2 f$, using a similar approach as in the first case, we can write $H_2 f(\mathbf{y}+\mathbf{e_i}) = f(\mathbf{y}+\mathbf{e_i}-\mathbf{e_p})+g_2(p)$, and $H_2 f(\mathbf{y}) = f(\mathbf{y}-\mathbf{e_q})+g_2(q)$ with $\mathbf{e_p}$ and $\mathbf{e_q}$ equal to either $\mathbf{e_0}$ or $\mathbf{e_2}$, and $g_2(0) = -r_2$ and $g_2(2) = p_2$. We have

$$
\begin{array}{rcll}
\Delta_i H_2 f(\mathbf{y}) & = & H_2 f(\mathbf{y}+\mathbf{e_i}) - H_2 f(\mathbf{y}) & (Definition) \\
& = & f(\mathbf{y}+\mathbf{e_i}-\mathbf{e_p}) + g_2(p) - f(\mathbf{y}-\mathbf{e_q}) - g_2(q) & (Above) \\
& \geq & f(\mathbf{y}+\mathbf{e_i}-\mathbf{e_q}) + g_2(q) - f(\mathbf{y}-\mathbf{e_q}) - g_2(q) & (Max) \\
& = & \Delta_i f(\mathbf{y}-\mathbf{e_q}) & (Definition) \\
& \geq & 0 & (C.1.1)
\end{array}
$$

Thus, we have $\Delta_i H_k f(\mathbf{y}) \geq 0$ for $k = 0, 1, 2$. On the other hand,

$$
\Delta_i c(\mathbf{y}) = c(\mathbf{y}+\mathbf{e_i}) - c(\mathbf{y}) = \begin{cases} b_1(y_1+1) + b_2 y_2 - b_1 y_1 - b_2 y_2 = b_1 \geq 0 & if \quad y_1 < 0,\ y_2 \leq 0,\ i = 1 \\ b_1 y_1 + b_2(y_2+1) - b_1 y_1 - b_2 y_2 = b_2 \geq 0 & if \quad y_1 < 0,\ y_2 < 0,\ i = 2 \\ -h y_1 + b_2(y_2+1) + h y_1 - b_2 y_2 = b_2 \geq 0 & if \quad y_1 \geq 0,\ y_2 < 0,\ i = 2. \end{cases}
$$

so $\Delta_i c(\mathbf{y}) = b_i \geq 0$ if $y_i < 0$. Therefore, we have $\Delta_i H f(\mathbf{y}) = \Delta_i c(\mathbf{y}) + \Delta_i H_0 f(\mathbf{y}) + \Delta_i H_1 f(\mathbf{y}) + \Delta_i H_2 f(\mathbf{y}) \geq 0$. This completes the proof for Condition C.1.1.

**Proof for Condition C.1.2:** The proof is similar to the proof for Condition C.1.1, and is therefore omitted.

**Proof for Condition C.1.3:** We must prove that $\Delta_1 H_k f(\mathbf{y})$, $\Delta_1 c(\mathbf{y})$, $\Delta_2 H_k f(\mathbf{y})$, and $\Delta_2 c(\mathbf{y})$ are non-increasing in $y_1$ and $y_2$.

**Proof for $\Delta_1 H_0 f(\mathbf{y})$:** First we prove that $\Delta_1 H_0 f(\mathbf{y})$ is non-increasing in $y_1$ and $y_2$. By definition $\Delta_1 H_0 f(\mathbf{y}) = H_0 f(\mathbf{y}+\mathbf{e_1}) - H_0 f(\mathbf{y})$, so we have,

$$
\Delta_1 H_0 f(\mathbf{y}) = \begin{cases}
f(\mathbf{y}+2\mathbf{e_1}) - f(\mathbf{y}+\mathbf{e_1}) = \Delta_1 f(\mathbf{y}+\mathbf{e_1}) & if \quad y_1 < Q^f(y_2) - 1 \\
f(\mathbf{y}+2\mathbf{e_1}) - f(\mathbf{y}+\mathbf{e_1}) = \Delta_1 f(\mathbf{y}+\mathbf{e_1}) & if \quad y_1 = Q^f(y_2) - 1, \quad y_2 = 0 \\
f(\mathbf{y}+\mathbf{e_1}+\mathbf{e_2}) - f(\mathbf{y}+\mathbf{e_1}) = \Delta_2 f(\mathbf{y}+\mathbf{e_1}) & if \quad y_1 = Q^f(y_2) - 1, \quad y_2 < 0 \\
f(\mathbf{y}+2\mathbf{e_1}) - f(\mathbf{y}+\mathbf{e_1}) = \Delta_1 f(\mathbf{y}+\mathbf{e_1}) & if \quad Q^f(y_2) \leq y_1 < S^f - 1, \quad y_2 = 0 \\
f(\mathbf{y}+\mathbf{e_1}+\mathbf{e_2}) - f(\mathbf{y}+\mathbf{e_2}) = \Delta_1 f(\mathbf{y}+\mathbf{e_2}) & if \quad Q^f(y_2) \leq y_1 < S^f - 1, \quad y_2 < 0 \\
f(\mathbf{y}+\mathbf{e_1}) - f(\mathbf{y}+\mathbf{e_1}) = 0 & if \quad y_1 = S^f - 1, \quad y_2 = 0 \\
f(\mathbf{y}+\mathbf{e_1}+\mathbf{e_2}) - f(\mathbf{y}+\mathbf{e_2}) = \Delta_1 f(\mathbf{y}+\mathbf{e_2}) & if \quad y_1 = S^f - 1, \quad y_2 < 0 \\
f(\mathbf{y}+\mathbf{e_1}) - f(\mathbf{y}) = \Delta_1 f(\mathbf{y}) & if \quad y_1 \geq S^f, \quad y_2 = 0 \\
f(\mathbf{y}+\mathbf{e_1}+\mathbf{e_2}) - f(\mathbf{y}+\mathbf{e_2}) = \Delta_1 f(\mathbf{y}+\mathbf{e_2}) & if \quad y_1 \geq S^f, \quad y_2 < 0
\end{cases}
$$

By C.1.3, $\Delta_1 f(\mathbf{y})$ and $\Delta_2 f(\mathbf{y})$ are both non-increasing in $y_1$ and $y_2$, so $\Delta_1 H_0(\mathbf{y})$ is non-increasing in $y_1$ and $y_2$ *within* each of the nine sub-condition intervals listed above. In the following, we will show that $\Delta_1 H_0(\mathbf{y})$ is non-increasing in $y_1$ and $y_2$ *across* any two adjacent intervals.

Here we only show that $\Delta_1 H_0 f(\mathbf{y})$ is non-increasing in $y_1$ and $y_2$ across the second and the third intervals (when $y_1 = Q^f(-1) - 1$ and $y_2 = 0$). The proofs for other intervals are similar and are therefore omitted.

Let us denote $\mathbf{y^{Q_0}} = (Q^f(-1) - 1, 0)$. By the results above, we have

$$
\begin{array}{rcl}
\Delta_1 H_0 f(\mathbf{y^{Q_0}}+\mathbf{e_1}) & = & \Delta_1 f(\mathbf{y^{Q_0}}+2\mathbf{e_1}) \\
\Delta_1 H_0 f(\mathbf{y^{Q_0}}) & = & \Delta_1 f(\mathbf{y^{Q_0}}+\mathbf{e_1}) \\
\Delta_1 H_0 f(\mathbf{y^{Q_0}}-\mathbf{e_1}) & = & \Delta_1 f(\mathbf{y^{Q_0}}).
\end{array}
$$

On the other hand, by C.1.3, we have

$$
\Delta_1 f(\mathbf{y^{Q_0}}+2\mathbf{e_1}) \leq \Delta_1 f(\mathbf{y^{Q_0}}+\mathbf{e_1}) \leq \Delta_1 f(\mathbf{y^{Q_0}}),
$$

so $\Delta_1 H_0 f(\mathbf{y})$ is non-increasing in $y_1$ across the adjacent intervals at $\mathbf{y^{Q_0}}$.

To show that $\Delta_1 H_0 f(\mathbf{y})$ is also non-increasing in $y_2$ across the adjacent intervals at $\mathbf{y^{Q_0}}$, note that

$$\begin{aligned} \Delta_1 H_0 f(\mathbf{y^{Q_0}} - \mathbf{e_2}) &= \Delta_2 f(\mathbf{y^{Q_0}} - \mathbf{e_2} + \mathbf{e_1}) \\ \Delta_1 H_0 f(\mathbf{y^{Q_0}}) &= \Delta_2 f(\mathbf{y^{Q_0}} + \mathbf{e_1}) \end{aligned}$$

On the other hand, according to the definition of $Q^f(-1)$, $\Delta_{12} f(\mathbf{y^{Q_0}} + \mathbf{e_1}) < 0$, which implies $\Delta_2 f(\mathbf{y^{Q_0}} + \mathbf{e_1}) > \Delta_1 f(\mathbf{y^{Q_0}} + \mathbf{e_1})$. And by C.1.3, we have

$$\Delta_2 f(\mathbf{y^{Q_0}} - \mathbf{e_2} + \mathbf{e_1}) \geq \Delta_2 f(\mathbf{y^{Q_0}} + \mathbf{e_1}) > \Delta_1 f(\mathbf{y^{Q_0}} + \mathbf{e_1}),$$

so $\Delta_1 H_0 f(\mathbf{y})$ is non-increasing in $y_2$ across the adjacent intervals at $\mathbf{y^{Q_0}}$.

Using the same approach, it is easy to show that $\Delta_1 H_0 f(\mathbf{y})$ is non-increasing in $y_1$ and $y_2$ across other adjacent intervals.

**Proof for $\Delta_1 H_1 f(\mathbf{y})$:** Now let us prove that $\Delta_1 H_1 f(\mathbf{y})$ is non-increasing in $y_1$ and $y_2$. We know,

$$\Delta_1 H_1 f(\mathbf{y}) = f(\mathbf{y} + \mathbf{e_1} - \mathbf{e_1}) + p_1 - f(\mathbf{y} - \mathbf{e_1}) - p_1 = \Delta_1 f(\mathbf{y} - \mathbf{e_1}).$$

By Condition C.1.3, $\Delta_1 f(\mathbf{y})$ is non-increasing in $y_1$ and $y_2$. So $\Delta_1 H_1 f(\mathbf{y})$ is also non-increasing in $y_1$ and $y_2$.

**Proof for $\Delta_1 H_2 f(\mathbf{y})$:** Now we prove that $\Delta_1 H_2 f(\mathbf{y})$ is non-increasing in $y_1$ and $y_2$. By definition, $\Delta_1 H_2 f(\mathbf{y}) = H_2(\mathbf{y} + \mathbf{e_1}) - H_2(\mathbf{y})$, so we have

$$\Delta_1 H_2 f(\mathbf{y}) = \begin{cases} f(\mathbf{y} + \mathbf{e_1} - \mathbf{e_2}) + p_2 - f(\mathbf{y} - \mathbf{e_2}) - p_2 = \Delta_1 f(\mathbf{y} - \mathbf{e_2}) & if \quad y_2 > A^f(y_1) \\ f(\mathbf{y} + \mathbf{e_1} - \mathbf{e_2}) + p_2 - f(\mathbf{y}) + r_2 \qquad = \Delta_{12} f(\mathbf{y} - \mathbf{e_2}) + p_2 + r_2 & if \quad A^f(y_1) = y_2 > A^f(y_1 + 1) \\ f(\mathbf{y} + \mathbf{e_1}) - r_2 - f(\mathbf{y}) + r_2 \qquad = \Delta_1 f(\mathbf{y}) & if \quad y_2 \leq A^f(y_1 + 1) \end{cases} \tag{11}$$

Using Condition C.1.3, it is easy to show that $\Delta_1 H_2 f(\mathbf{y})$ is non-increasing in $y_1$ and $y_2$ in the first and the last cases. In the following, we show that it is also true in the second case. Let us denote $\mathbf{y^A} = (y_1, y_2)$, such that $A^f(y_1) = y_2 > A^f(y_1 + 1)$. In the following, we will compare $\Delta_1 H_2 f(\mathbf{y^A})$ with $\Delta_1 H_2 f(\mathbf{y^A} + \mathbf{e_1})$, $\Delta_1 H_2 f(\mathbf{y^A} + \mathbf{e_2})$, $\Delta_1 H_2 f(\mathbf{y^A} - \mathbf{e_1})$ and $\Delta_1 H_2 f(\mathbf{y^A} - \mathbf{e_2})$, one by one.

$$\begin{aligned} \Delta_1 H_2 f(\mathbf{y^A}) - \Delta_1 H_2 f(\mathbf{y^A} + \mathbf{e_1}) &= \Delta_{12} f(\mathbf{y^A} - \mathbf{e_2}) + p_2 + r_2 - \Delta_1 f(\mathbf{y^A} - \mathbf{e_2} + \mathbf{e_1}) && (Equation(11)) \\ &\geq \Delta_{12} f(\mathbf{y^A} - \mathbf{e_2}) + \Delta_2 f(\mathbf{y^A} - \mathbf{e_2} + \mathbf{e_1}) - \Delta_1 f(\mathbf{y^A} - \mathbf{e_2} + \mathbf{e_1}) && (Property(iii)) \\ &= \Delta_{12} f(\mathbf{y^A} - \mathbf{e_2}) - \Delta_{12} f(\mathbf{y^A} - \mathbf{e_2} + \mathbf{e_1}) && (\Delta_{12} f = \Delta_1 f - \Delta_2 f) \\ &\geq 0 && (C.1.4) \end{aligned}$$

$$\begin{aligned} \Delta_1 H_2 f(\mathbf{y^A}) - \Delta_1 H_2 f(\mathbf{y^A} - \mathbf{e_1}) &= \Delta_{12} f(\mathbf{y^A} - \mathbf{e_2}) + p_2 + r_2 - \Delta_1 f(\mathbf{y^A} - \mathbf{e_1}) && (Equation(11)) \\ &\leq \Delta_{12} f(\mathbf{y^A} - \mathbf{e_2}) + \Delta_2 f(\mathbf{y^A} - \mathbf{e_2}) - \Delta_1 f(\mathbf{y^A} - \mathbf{e_1}) && (C.2.3) \\ &= \Delta_1 f(\mathbf{y^A} - \mathbf{e_2}) - \Delta_1 f(\mathbf{y^A} - \mathbf{e_1}) && (\Delta_{12} f = \Delta_1 f - \Delta_2 f) \\ &= f(\mathbf{y^A} - \mathbf{e_2} + \mathbf{e_1}) - f(\mathbf{y^A} - \mathbf{e_2}) - f(\mathbf{y^A}) + f(\mathbf{y^A} - \mathbf{e_1}) && (Definition) \\ &= \Delta_{12} f(\mathbf{y^A} - \mathbf{e_2}) - \Delta_{12} f(\mathbf{y^A} - \mathbf{e_2} - \mathbf{e_1}) && (Definition) \\ &\leq 0 && (C.1.4) \end{aligned}$$

Thus, $\Delta_1 H_2 f(\mathbf{y^A})$ is non-increasing in $y_1$ across the adjacent intervals at $\mathbf{y^A}$. We now compare $\Delta_1 H_2 f(\mathbf{y^A})$ with $\Delta_1 H_2 f(\mathbf{y^A} + \mathbf{e_2})$ and $\Delta_1 H_2 f(\mathbf{y^A} - \mathbf{e_2})$ to show that $\Delta_1 H_2 f(\mathbf{y^A})$ is also non-increasing in $y_2$ across the adjacent intervals.

$$\begin{aligned} \Delta_1 H_2 f(\mathbf{y^A}) - \Delta_1 H_2 f(\mathbf{y^A} + \mathbf{e_2}) &= f(\mathbf{y^A} + \mathbf{e_1} - \mathbf{e_2}) - f(\mathbf{y^A}) + p_2 + r_2 - f(\mathbf{y^A} + \mathbf{e_1}) + f(\mathbf{y^A}) && (Equation(11)) \\ &= -\Delta_2 f(\mathbf{y^A} + \mathbf{e_1} - \mathbf{e_2}) + p_2 + r_2 \\ &\geq 0 && (Property(iii)) \end{aligned}$$

$$\begin{aligned}\Delta_1 H_2 f(\mathbf{y^A}) - \Delta_1 H_2 f(\mathbf{y^A}-\mathbf{e_2}) &= \Delta_{12} f(\mathbf{y^A}-\mathbf{e_2}) + p_2 + r_2 - \Delta_1 f(\mathbf{y^A}-\mathbf{e_2}) && (Equation(11))\\ &= p_2 + r_2 - \Delta_2 f(\mathbf{y^A}-\mathbf{e_2}) && (Definition)\\ &\leq 0 && (C.2.3)\end{aligned}$$

Therefore, $\Delta_1 H_2 f(\mathbf{y^A})$ is also non-increasing in $y_2$ across the adjacent intervals at $\mathbf{y^A}$.

**Proof for $\Delta_1 c(\mathbf{y})$:** We now show that $\Delta_1 c(\mathbf{y})$ is non-increasing in $y_1$, $y_2$. Note that,

$$\Delta_1 c(\mathbf{y}) = c(\mathbf{y}+\mathbf{e_1}) - c(\mathbf{y}) = \begin{cases} b_1(y_1+1) + b_2y_2 - b_1y_1 - b_2y_2 = b_1 \geq 0 & if \quad y_1 < 0 \\ -h(y_1+1) + b_2y_2 + hy_1 - b_2y_2 = -h \leq 0 & if \quad y_1 \geq 0 \end{cases}$$

which is is non-increasing in $y_1$, $y_2$.

In conclusion, since $\Delta_1 H_0 f(\mathbf{y})$, $\Delta_1 H_1 f(\mathbf{y})$, $\Delta_1 H_2 f(\mathbf{y})$, and $\Delta_1 c(\mathbf{y})$ are all non-increasing in $y_1$, $y_2$, then $\Delta_1 H f(\mathbf{y})$ is non-increasing in $y_1$, $y_2$. This completes the proof for $\Delta_1 H f(\mathbf{y})$.

**Proof for $\Delta_2 H_0 f(\mathbf{y})$:** Next, we prove that $\Delta_2 H_k f(\mathbf{y})$ is non-increasing in $y_1$ and $y_2$. By definition, $\Delta_2 H_k f(\mathbf{y}) = H_k f(\mathbf{y}+\mathbf{e_2}) - H_k f(\mathbf{y})$, so we have,

$$\Delta_2 H_0 f(\mathbf{y}) = \begin{cases} f(\mathbf{y}+\mathbf{e_2}+\mathbf{e_1}) - f(\mathbf{y}+\mathbf{e_1}) = \Delta_2 f(\mathbf{y}+\mathbf{e_1}) & if \quad y_1 < Q^f(y_2) \\ f(\mathbf{y}+2\mathbf{e_2}) - f(\mathbf{y}+\mathbf{e_2}) \quad = \Delta_2 f(\mathbf{y}+\mathbf{e_2}) & if \quad y_1 \geq Q^f(y_2), \quad y_2 < -1 \\ f(\mathbf{y}+\mathbf{e_2}+\mathbf{e_1}) - f(\mathbf{y}+\mathbf{e_2}) = \Delta_1 f(\mathbf{y}+\mathbf{e_2}) & if \quad S^f > y_1 \geq Q^f(y_2), \quad y_2 = -1 \\ f(\mathbf{y}+\mathbf{e_2}) - f(\mathbf{y}+\mathbf{e_2}) \quad = 0 & if \quad y_1 \geq S^f, \quad y_2 = -1 \end{cases}$$

Using the same approach as in the proof for $\Delta_1 H_0 f(\mathbf{y})$, it is easy to show that $\Delta_2 H_0(\mathbf{y})$ is non-increasing in $y_1$ and $y_2$.

**Proof for $\Delta_2 H_1 f(\mathbf{y})$:** Now we prove that $\Delta_2 H_1 f(\mathbf{y})$ is non-increasing in $y_1$ and $y_2$, we have

$$\Delta_2 H_1 f(\mathbf{y}) = f(\mathbf{y}+\mathbf{e_2}-\mathbf{e_1}) + p_1 - f(\mathbf{y}-\mathbf{e_1}) - p_1 = \Delta_2 f(\mathbf{y}-\mathbf{e_1}).$$

By Condition C.1.3, we know that $\Delta_2 f(\mathbf{y})$ is non-increasing in $y_1$ and $y_2$. Thus, $\Delta_2 H_1 f(\mathbf{y})$ is also non-increasing in $y_1$ and $y_2$.

**Proof for $\Delta_2 H_2 f(\mathbf{y})$:** Now we prove that $\Delta_2 H_2 f(\mathbf{y})$ is non-increasing in $y_1$ and $y_2$, we have

$$\Delta_2 H_2 f(\mathbf{y}) = \begin{cases} f(\mathbf{y}+\mathbf{e_2}-\mathbf{e_2}) + p_2 - f(\mathbf{y}-\mathbf{e_2}) - p_2 = \Delta_2 f(\mathbf{y}-\mathbf{e_2}) & if \quad y_2 > A^f(y_1) \\ f(\mathbf{y}+\mathbf{e_2}-\mathbf{e_2}) + p_2 - f(\mathbf{y}) + r_2 \quad = p_2 + r_2 & if \quad y_2 = A^f(y_1) \\ f(\mathbf{y}+\mathbf{e_2}) - r_2 - f(\mathbf{y}) + r_2 \quad = \Delta_2 f(\mathbf{y}) & if \quad y_2 < A^f(y_1) \end{cases}$$

According to Condition C.1.3, we know that $\Delta_2 f(\mathbf{y})$ is non-increasing in $y_1$ and $y_2$. Therefore, $\Delta_2 H_2(\mathbf{y})$ is non-increasing in $y_1$ and $y_2$ when $y_2 > A^f(y_1)$ or when $y_2 < A^f(y_1)$. Using the same approach as in the proof of $\Delta_1 H_2 f(\mathbf{y})$, it is easy to prove that $\Delta_2 H_2(\mathbf{y})$ is non-increasing in $y_1$ and $y_2$ when $y_2 = A^f(y_1)$.

**Proof for $\Delta_2 c(\mathbf{y})$:** Finally, note that $\Delta_2 c(\mathbf{y}) = c(\mathbf{y}+\mathbf{e_2}) - c(\mathbf{y}) = b_2$ is independent of $y_1$ and $y_2$, and therefore non-increasing in $y_1$ and $y_2$. Since $\Delta_2 H_0 f(\mathbf{y})$, $\Delta_2 H_1 f(\mathbf{y})$, $\Delta_2 H_2 f(\mathbf{y})$, and $\Delta_2 c(\mathbf{y})$ are all non-increasing in $y_1$, $y_2$, then $\Delta_2 H f(\mathbf{y})$ is non-increasing in $y_1$, $y_2$. This completes the proof for $\Delta_2 H f(\mathbf{y})$, and therefore the proof for Condition C.1.3.

**Proof for Condition C.1.4:**

**Proof for $\Delta_{12} H_0 f(\mathbf{y})$:** We prove that $\Delta_{12} H_k f(\mathbf{y})$ is non-increasing in $y_1$ and non-decreasing in $y_2$. By definition, $\Delta_{12} H_k f(\mathbf{y}) = H_k f(\mathbf{y}+\mathbf{e_1}) - H_k f(\mathbf{y}+\mathbf{e_2})$, so we have,

$$\Delta_{12} H_0 f(\mathbf{y}) = \begin{cases} f(\mathbf{y}+\mathbf{e_1}+\mathbf{e_1}) - f(\mathbf{y}+\mathbf{e_2}+\mathbf{e_1}) = \Delta_{12} f(\mathbf{y}+\mathbf{e_1}) & if \quad y_1 < Q^f(y_2) - 1 \\ f(\mathbf{y}+\mathbf{e_1}+\mathbf{e_2}) - f(\mathbf{y}+\mathbf{e_2}+\mathbf{e_1}) = 0 & if \quad y_1 = Q^f(y_2) - 1 \\ f(\mathbf{y}+\mathbf{e_1}+\mathbf{e_2}) - f(\mathbf{y}+\mathbf{e_2}+\mathbf{e_2}) = \Delta_{12} f(\mathbf{y}+\mathbf{e_2}) & if \quad y_1 \geq Q^f(y_2), \quad y_2 < -1 \\ f(\mathbf{y}+\mathbf{e_1}+\mathbf{e_2}) - f(\mathbf{y}+\mathbf{e_2}+\mathbf{e_1}) = 0 & if \quad S^f > y_1 \geq Q^f(y_2), \quad y_2 = -1 \\ f(\mathbf{y}+\mathbf{e_1}+\mathbf{e_2}) - f(\mathbf{y}+\mathbf{e_2}) \quad = \Delta_1 f(\mathbf{y}+\mathbf{e_2}) & if \quad y_1 \geq S^f, \quad y_2 = -1 \end{cases}$$

$$\Delta_{12}H_1f(\mathbf{y}) \quad = \quad f(\mathbf{y}+\mathbf{e_1}-\mathbf{e_1})+p_1-f(\mathbf{y}+\mathbf{e_2}-\mathbf{e_1})-p_1=\Delta_{12}f(\mathbf{y}-\mathbf{e_1})$$

$$\Delta_{12}H_2f(\mathbf{y}) \quad = \quad \begin{cases} f(\mathbf{y}+\mathbf{e_1}-\mathbf{e_2})+p_2-f(\mathbf{y}+\mathbf{e_2}-\mathbf{e_2})-p_2=\Delta_{12}f(\mathbf{y}-\mathbf{e_2}) & if \quad y_2 > A^f(y_1+1) \\ f(\mathbf{y}+\mathbf{e_1})-r_2-f(\mathbf{y}+\mathbf{e_2}-\mathbf{e_2})-p_2 \qquad =\Delta_1 f(\mathbf{y})-p_2-r_2 & if \quad A^f(y_1)-1<y_2=A^f(y_1+1) \\ f(\mathbf{y}+\mathbf{e_1})-r_2-f(\mathbf{y}+\mathbf{e_2})+r_2 \qquad =\Delta_{12}f(\mathbf{y}) & if \quad y_2 \le A^f(y_1)-1 \end{cases}$$

Using the same approach as in the proof for C.1.3, it is easy to show that $\Delta_{12}H_k(\mathbf{y})$ is non-increasing in $y_1$ and non-decreasing in $y_2$.

Note that,

$$\Delta_{12}c(\mathbf{y}) = \begin{cases} b_1-b_2 & if \quad y_1<0 \\ -h_1-b_2 & if \quad y_1 \ge 0 \end{cases}$$

which is also non-increasing in $y_1$ and non-decreasing in $y_2$. So in summary, $\Delta_{12}Hf(\mathbf{y})$ is non-increasing in $y_1$ and non-decreasing in $y_2$. ■

**PROOF OF PROPOSITION 2**

**Property (vi):** By Condition C.2.4, $\Delta_2 f(y_1,y_2)$ only depends on $y_1+y_2$ when $y_1<R^f$, so $\Delta_2 f(y_1,y_2) \le p_2+r_2$ for $y_1+y_2>B^f$, where $B^f=A^f(0)=\min\{z|\Delta_2 f(0,z-1)>p_2+r_2\}$.

**Property (vii):** By Condition C.2.4, $\Delta_2 f(y_1,y_2)$ is independent of $y_1$ and $y_2$ if $y_1+y_2$ is fixed. Comparing the definitions of $A^f(y_1)$ and $B^f$, we have $0+B^f-1=y_1+A^f(y_1)-1$, and thus $B^f=y_1+A^f(y_1)$, when $y_1<R^f$.

**Property (viii):** By Condition C.2.3, $\Delta_{12}Hf(y_1,y_2) \le \Delta_{12}f(y_1,y_2)$ when $y_1<R^f$, and therefore $R^{Hf} \le R^f$. ■

**PROOF OF LEMMA 2**

We prove that function $H$ preserves each condition in $\mathcal{C}2$.

**Proof for Condition C.2.1:** We prove that $\Delta_{12}H_kf(\mathbf{y})$ and $\Delta_{12}c(\mathbf{y})$ are independent of $y_2$ when $y_1<R^{Hf}$. By definition, $\Delta_{12}H_kf(\mathbf{y})=H_kf(\mathbf{y}+\mathbf{e_1})-H_kf(\mathbf{y}+\mathbf{e_2})$, so we have,

$$\Delta_{12}H_0f(\mathbf{y}) \quad = \quad \begin{cases} f(\mathbf{y}+\mathbf{e_1}+\mathbf{e_2})-f(\mathbf{y}+\mathbf{e_2}+\mathbf{e_1})=0 & if \quad y_1=R^f-1 \\ f(\mathbf{y}+\mathbf{e_1}+\mathbf{e_1})-f(\mathbf{y}+\mathbf{e_2}+\mathbf{e_1})=\Delta_{12}f(\mathbf{y}+\mathbf{e_1}) & if \quad y_1<R^f-1 \end{cases}$$

$$\Delta_{12}H_1f(\mathbf{y}) \quad = \quad f(\mathbf{y}+\mathbf{e_1}-\mathbf{e_1})+p_1-f(\mathbf{y}+\mathbf{e_2}-\mathbf{e_1})-p_1=\Delta_{12}f(\mathbf{y}-\mathbf{e_1})$$

$$\Delta_{12}H_2f(\mathbf{y}) \quad = \quad \begin{cases} f(\mathbf{y}+\mathbf{e_1}-\mathbf{e_2})+p_2-f(\mathbf{y}+\mathbf{e_2}-\mathbf{e_2})-p_2=\Delta_{12}f(\mathbf{y}-\mathbf{e_2}) & if \quad y_1<R^f, \quad y_1+y_2 \ge B^f \\ f(\mathbf{y}+\mathbf{e_1})-r_2-f(\mathbf{y}+\mathbf{e_2})+r_2 \qquad =\Delta_{12}f(\mathbf{y}) & if \quad y_1<R^f, \quad y_1+y_2<B^f \end{cases}$$

Using the same argument as in the proof for C.1.3, it is easy to show that $\Delta_{12}H_kf(\mathbf{y})$ is independent of $y_2$. Note that, $\Delta_{12}c(\mathbf{y})$ is also independent of $y_2$. By Property (iv) in Proposition 2, $R^{Hf} \le R^f$, therefore, $\Delta_{12}Hf(\mathbf{y})$ is independent of $y_2$ when $y_1<R^{Hf}$.

**Proof for Condition C.2.2:** Condition C.2.2 is a direct result of Conditions C.1.4 and C.2.1. Since $\Delta_{12}f(y_1,y_2)$ is non-increasing in $y_1$ and independent of $y_2$, $\Delta_{12}Hf(y_1,y_2)=\Delta_{12}Hf(y_1,-1) \ge 0$ for $y_1<R^{Hf}$.

**Proof for Condition C.2.3:** Note that $\Delta_{12}H[Hf(\mathbf{y})]=\Delta_{12}c(y_1,y_2)+\mu\Delta_{12}H_0[Hf(y_1,y_2)]+\lambda_1\Delta_{12}H_1[Hf(y_1,y_2)]+\lambda_2\Delta_{12}H_2[Hf(y_1,y_2)]$, we first prove $\Delta_{12}H[Hf(\mathbf{y})] \le \Delta_{12}Hf(\mathbf{y})$ by showing that $\Delta_{12}H_k[Hf(\mathbf{y})] \le \Delta_{12}H_kf(\mathbf{y})$ when $y_1<R^{Hf}$. By definition, $\Delta_{12}H_k[Hf(\mathbf{y})]=H_k[Hf(\mathbf{y}+\mathbf{e_1})]-H_k[Hf(\mathbf{y}+\mathbf{e_2})]$, so we have, so we have,

$$\Delta_{12}H_0[Hf(\mathbf{y})] \quad = \quad \begin{cases} Hf(\mathbf{y}+\mathbf{e_1}+\mathbf{e_2})-Hf(\mathbf{y}+\mathbf{e_2}+\mathbf{e_1})=0 & if \quad y_1=R^{Hf}-1 \\ Hf(\mathbf{y}+\mathbf{e_1}+\mathbf{e_1})-Hf(\mathbf{y}+\mathbf{e_2}+\mathbf{e_1})=\Delta_{12}Hf(\mathbf{y}+\mathbf{e_1}) & if \quad y_1<R^{Hf}-1 \end{cases}$$

$$\Delta_{12}H_1[Hf(\mathbf{y})] \quad = \quad Hf(\mathbf{y}+\mathbf{e_1}-\mathbf{e_1})+p_1-Hf(\mathbf{y}+\mathbf{e_2}-\mathbf{e_1})-p_1=\Delta_{12}Hf(\mathbf{y}-\mathbf{e_1})$$

$$\Delta_{12}H_2[Hf(\mathbf{y})] \quad = \quad \begin{cases} Hf(\mathbf{y}+\mathbf{e_1}-\mathbf{e_2})+p_2-Hf(\mathbf{y}+\mathbf{e_2}-\mathbf{e_2})-p_2=\Delta_{12}Hf(\mathbf{y}-\mathbf{e_2}) & if \quad y_1<R^{Hf}, \quad y_1+y_2\geq B^{Hf} \\ Hf(\mathbf{y}+\mathbf{e_1})-r_2-Hf(\mathbf{y}+\mathbf{e_2})+r_2 \qquad\qquad\quad =\Delta_{12}Hf(\mathbf{y}) & if \quad y_1<R^{Hf}, \quad y_1+y_2<B^{Hf} \end{cases}$$

By Condition C.2.3, $\Delta_{12}Hf(\mathbf{y})\leq\Delta_{12}f(\mathbf{y})$. Comparing the above results with those shown in the proof for Condition C.2.1, we have $\Delta_{12}H_k[Hf(\mathbf{y})]\leq\Delta_{12}H_kf(\mathbf{y})$ when $y_1<R^{Hf}$.

Note that $\Delta_{12}f^*(\mathbf{y})=\Delta_{12}Hf^*(\mathbf{y})=\Delta_{12}c(y_1,y_2)+\mu\Delta_{12}H_0f^*(y_1,y_2)+\lambda_1\Delta_{12}H_1f^*(y_1,y_2)+\lambda_2\Delta_{12}H_2f^*(y_1,y_2)$, using a similar argument as above, we have $\Delta_{12}H[Hf(\mathbf{y})]\leq\Delta_{12}f^*(\mathbf{y})$ when $y_1<R^{Hf}$.

**Proof for Condition C.2.4:** Condition C.2.4 can be derived from Condition C.2.1. Since $\Delta_{12}f(y_1,y_2)$ is independent of $y_2$ when $y_1<R^f$, we have

$$\begin{aligned} \Delta_{12}f(y_1-1,y_2) &= \Delta_{12}f(y_1-1,y_2-1) && (C.2.1) \\ \Rightarrow \quad f(y_1,y_2)-f(y_1-1,y_2+1) &= f(y_1,y_2-1)-f(y_1-1,y_2) && (Definition) \\ \Rightarrow \quad f(y_1,y_2)-f(y_1,y_2-1) &= f(y_1-1,y_2+1)-f(y_1-1,y_2) && (Rearrangement) \\ \Rightarrow \quad \Delta_2f(y_1,y_2-1) &= \Delta_2f(y_1-1,y_2) && (Definition) \end{aligned}$$

Thus, in general we have $\Delta_2f(y_1,y_2)=\Delta_2f(y_1\mp1,y_2\pm1)$, which directly implies Condition C.2.4 (Note that Condition C.2.4 can be obtained by adding 1 to one element and deducing 1 form the other element of vector $\mathbf{y}$ in the above result.) ∎

## PROOF OF THEOREM 1

We first prove the existence of an optimal stationary policy by following the approaches in [15] and [16]. For this purpose, we need to show: $(i)$ The set of structured functions $\mathcal{C}$ is complete, and $(ii)$ $J(y_1,y_2)\in\mathcal{C}$.

Function $c^{'}(y_1,y_2)=-h(y_1-N)^+-b_1(y_1-N)^-+b_2y_2$, where $N$ could be any integer number, satisfies all the conditions in $\mathcal{C}1$ and $\mathcal{C}2$, except for C.2.3. To make sure that C.2.3 is preserved in each iteration, we let $N$ take a "sufficiently" large number, i.e., a number that is guaranteed to be larger than $R$. For example, $S$, whose existence is secured in Proposition 1 is larger than $R$ by Properties $(v)$ and $(viii)$. The value iteration process starts with $J(y_1,y_2)=c^{'}(y_1,y_2)$. By Lemma 2, the operator $H=c+\mu H_0+\lambda_1H_1+\lambda_2H_2$ preserves the structural properties from Condition C.1.1 to C.2.4. Because the limit of any converging sequences of functions in $\mathcal{C}1$ and $\mathcal{C}2$ will be in $\mathcal{C}1$ and $\mathcal{C}2$ as well, the set of structured functions $\mathcal{C}1$ and $\mathcal{C}2$ is complete. On the other hand, since $c^{'}(y_1,y_2)\in\mathcal{C}2\subset\mathcal{C}1$, Lemmas 1 and 2 imply that $J(y_1,y_2)\in\mathcal{C}2\subset\mathcal{C}1$. So the optimal expected profit function $J$ is structured and satisfies all conditions in $\mathcal{C}1$ and $\mathcal{C}2$. Hence, the existence of an optimal stationary policy under the discounted-profit criterion follows from the corollaries of Theorem 5.1 of [29].

The existence and convergence of the optimal profit functions and the optimal stationary policies under the average-profit criterion are more complicated since there are infinite number of states. Moreover, the cost in a period can be unbounded from above when $y_1\to\pm\infty$ or $y_2\to-\infty$. We will show that our MDP model satisfies the three conditions (SEN 1), (SEN2) and (SEN3) in Sennott ( 1999, pp 132) so that the average-profit exists by letting $\alpha\to0^+$.

Note that our MDP model is *unichain* because all the states lead to a single recurrent class for any deterministic stationary policy (see Puterman [30] (p.348)). Let $\pi_I$ be a policy in which the manufacturer *only* produces type 1 if there are orders for it, and *never* accepts demand for type 2 product, i.e., $S=R=B=0$. It is not hard to show that $\pi_I$ induces a positive recurrent class and $\pi_I$ is a *z standard policy* defined in Sennott (1999, p. 43) with a distinguished single state $(0,0)$. By Proposition 7.5.3 and 7.2.4 in Sennott (1999), (SEN1 ) and (SEN2) hold for every state in $\Omega$.

Now we study the one-step expected profit function (i.e., expected profit during a small time interval $\delta t$) under policy $\pi_I$: $V_{\pi_I}(y_1,y_2)=-(b_1y_1^-+b_2y_2^-)\delta t+p_1\lambda_1\delta t-r_2\lambda_2\delta t$. $V_{\pi_I}(y_1,y_2)$ is bounded in both sides, $-(b_1y_1^-+b_2y_2^-+r_2\lambda_2)\delta t\leq V_{\pi_I}(y_1,y_2)\leq p_1\lambda_1\delta t$. As a result, the total discounted-profit under the optimal policy starting at state $(y_1,y_2)$ is bounded, $-\frac{1}{1-e^{-\alpha\delta t}}(b_1y_1^-+b_2y_2^-+r_2)\delta t\leq J(y_1,y_2)\leq\frac{1}{1-e^{-\alpha\delta t}}p_1\lambda_1\delta t$.

Let $M$ be the optimal total discounted-profit associated with the distinguished set of states $(0,0)$, then we get $-\frac{1}{1-e^{-\alpha\delta t}}(b_1y_1^-+b_2y_2^-+r_2\lambda_2)\delta t-M\leq J(y_1,y_2)-J(0,0)\leq\frac{1}{1-e^{-\alpha\delta t}}p_1\lambda_1\delta t-M$, for all $(y_1,y_2)\in\Omega$. Thus, (SEN3) holds and the average profit exists by Theorem 7.2.3 in Sennott (1999). By the same theorem, any limit point of the optimal stationary policies under total discount-profit criterion is average-profit optimal. The limit point can be obtained by appropriately choosing a sequence $\alpha \to 0^+$.

**Part 1:** From equation (3), we can see that when there is no order for any product, the optimal production control only depends on the sign of $\Delta_1 J(y_1, y_2)$: to produce if the sign is positive, and to stop otherwise. By Condition C.1.3 and Property (i), it is optimal to produce if $y_1 < S$.

**Part 2:** When there are orders for both products, by Condition C.1.1 and Part (i), it is always optimal to produce, and by Condition C.1.2 and Part (ii), the manufacturer will always give priority to type 1. If there are only orders for type 2, by Conditions C.2.1, C.2.2, and Part (ii), the manufacturer will produce type 2 if $y_1 \geq R$, or otherwise produce type 1 to build up the inventory.

**Part 3:** The admission control for an arriving order for type 2 depends on the sign of $p_2 + r_2 - \Delta_2 J(y_1, y_2 - 1)$: satisfy or backlog the demand if the sign is positive, reject otherwise. When $y_1 < R$, by Condition C.2.4 and Part (vii), the optimal policy is to backlog an arriving type 2 demand if $y_1 + y_2 > B$, and to reject otherwise; When $y_1 \geq R$, by Condition C.1.3 and Part (iii), the optimal policy is to backlog an arriving type 2 demand if $y_2 > A(y_1)$, and to reject otherwise. ■

## PROOF OF LEMMA 3

If the inventory is not available, an OEM order cannot be satisfied immediately. The number of OEM orders in the system (including the arriving order) is $y_1^- + 1$, and the expected production time for each unit is $\frac{1}{\mu}$. Therefore, the expected leadtime for the arriving OEM order is $T_1 = (y_1^- + 1) \times \frac{1}{\mu}$. ■

Calculating the expected leadtime for an aftermarket order is more complicated. When $y_1 < R$, in addition to the current orders in the system, we also need to count the difference between $y_1$ and $R$, because the manufacturer will increase the inventory level of product 1 to $R$ before producing any product 2. In addition, since OEM demand has preemptive priority over aftermarket demand, we also need to estimate the expected number of arriving OEM demands before the aftermarket demand is finally served. Lemma 4 shows how to quote the expected due date for an arriving aftermarket order.

## PROOF OF LEMMA 4

When $y_1 < R$, the manufacturer will first produce $R - y_1$ units of OEM product before processing any aftermarket demand. The number of aftermarket orders in the system (including the arriving order) is $y_2^- + 1$. While these $(R - y_1 + y_2^- + 1)$ items are produced, the expected number of new OEM orders is $\lambda_1 \frac{1}{\mu}(R - y_1 + y_2^- + 1)$. Then while these newly arrived OEM orders are produced, it is expected that another $\lambda_1 \frac{1}{\mu}\left[\frac{\lambda_1}{\mu}(R - y_1 + y_2^- + 1)\right]$ OEM orders will arrive, and so on. Overall, the estimated leadtime for an arriving aftermarket demand is, $(1 + \frac{\lambda_1}{\mu} + \frac{\lambda_1^2}{\mu^2} + \cdots)(R - y_1 + y_2^- + 1)\frac{1}{\mu} = \frac{1}{\mu - \lambda_1}(R - y_1 + y_2^- + 1)$. ■

## CALCULATING EXPECTED LEADTIME FOR AFTERMARKET DEMAND WHEN $y_1 \geq R$

To calculate the expected leadtime for a newly accepted demand, we may ignore the aftermarket orders that arrive later, since they will have lower priority than the order under consideration. So we may simplify the system with only two types of events: production completions with rate $\mu$ and OEM arrivals with rate $\lambda_1$. The aftermarket order under consideration is satisfied when $y_2 = 0$ in the system space. Let $T_2(y_1, y_2)$ be the expected leadtime for the state $(y_1, y_2)$, and it can be calculated as,

$$T_2(y_1, y_2) = \frac{1}{\lambda_1 + \mu} + \frac{\mu}{\lambda_1 + \mu} T_2(y_1, y_2 + 1) + \frac{\lambda_1}{\lambda_1 + \mu} T_2(y_1 - 1, y_2), \quad for \;\; y_1 \geq R, y_2 < 0. \tag{12}$$

To calculate $T_2(\bar{y}_1, \bar{y}_2)$, we first use Lemma 4 to calculate the expected leadtime $T_2(R - 1, y_2)$ for $y_2 = 0, \ldots, \bar{y}_2$; Second, we use Equation 12 to calculate the expected leadtime $T_2(R, y_2)$ for $y_2 = 0, \ldots, \bar{y}_2$. We continue this process until $T_2(\bar{y}_1, \bar{y}_2)$ is calculated. ■

# ON-LINE APPENDIX B

## Model with Multiple Aftermarket Classes

Using the same approach as in Carr and Duenyas [3], we can extend the above model to a problem with a single MTS product and multiple MTO products. For this purpose, we use the same assumption as in Carr and Duenyas [3] that all MTO products have the same production rate and the same backlogging penalty. Without loss of generality, the aftermarket products can be indexed from 2 to $n$, such that $p_2 + r_2 \leq \cdots \leq p_n + r_n$.

Since all MTO demands have the same backlogging penalty, it is unnecessary to differentiate them once they are accepted. All MTO demands have the same priority in the waiting queue, and the first-come-first-serve rule will be applied. So the system state can still be described with $(y_1, y_2)$, where $y_1$ is the net inventory level of the OEM product, and $-y_2$ is the total amount of backorders of all MTO products (product 2 to product $n$).

To adjust the model to this problem, we change the function $H_i$, $i = 2, \ldots, n$, as follows:

$$H_i J(y_1, y_2) = \max\Big\{J(y_1, y_2 - 1) + p_i, J(y_1, y_2) - r_i\Big\}.$$

Then we change Properties (iii) and (viii) to,

*(iii)* $\Delta_2 f(y_1, y_2 - 1) \leq p_i + r_i$ for $y_2 > A_i^f(y_1)$, where $A_i^f(y_1) = \max\{z | \Delta_2 f(y_1, z - 1) > p_i + r_i\}$;

*(viii)* $\Delta_2 f(y_1, y_2 - 1) \leq p_i + r_i$ for $y_1 + y_2 > B_i^f$, where $B_i^f = \max\Big\{z | \Delta_2 f(0, z - 1) > p_i + r_i\Big\}$.

**Property 4** *If $f(y_1, y_2) \in \mathcal{C}2 \subset \mathcal{C}1$, then $B_i^f \geq B_j^f$, for $2 \leq i < j$.*

With some modifications in the proof, we can show that Lemmas 1 and 2 hold for the modified function, $H_i$, $i = 2, \ldots, n$, and the updated condition set. The structure of the optimal policy is characterized by the base-stock level, $S$, the rationing level, $R$, and a sequence of admission levels, $A_2(y_1), \ldots, A_n(y_1)$ ($B_2, \ldots, B_n$ when $y_1 < R$), where $A_i(y_1)$ and $B_i$, $i = 2, \ldots, n$, are the admission threshold of product $i$.

The optimal policy for a three-product problem is described in Figure 7. If the total inventory, $y_1 + y_2$, is between $A_2(y_1)$ and $A_3(y_1)$, the manufacturer will reject demands for type 2, and accept demands for type 3; If the total inventory is below $A_3(y_1)$, the manufacturer will reject demands both for type 2 and 3.

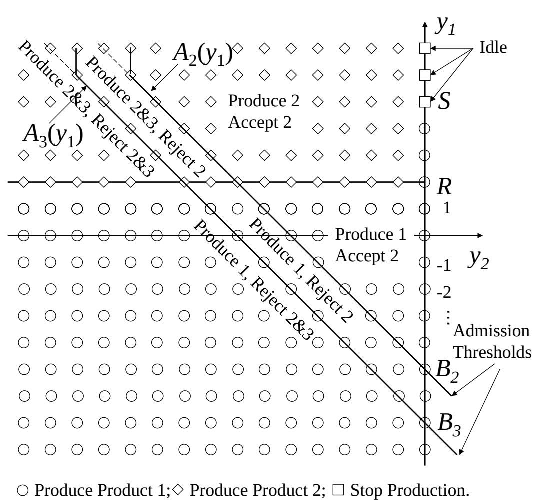


Figure 7: The (S, R, B) Policy with Two Types of Make-to-Order Products

# ON-LINE APPENDIX C

## Analytical Supporting Results For the OEM/Aftermarket Model With Contingent Outsourcing

Before proving Theorem 2, we prove Proposition 3 and Lemma 5 in the following.

### PROOF OF PROPOSITION 3

**Property (ix):** Property (ix) is a direct result of Condition C.1.3. Since $\Delta_2 f(y_1, y_2)$ is non-increasing in $y_2$, for a given value of $y_1$, there exists a threshold level, $K^f(y_1) = \max\{z | \Delta_2 f(y_1, z-1) > p_2 + l_2\}$, such that $\Delta_2 f(y_1, y_2) \leq p_2 + l_2$ for $y_2 > K^f(y_1)$.
**Property (x):** By contradiction, assume $K^f(y_1) > A^f(y_1)$. By Condition C.2.3, $\Delta_2 f(y_1, K^f(y_1)-1) \leq p_2 + r_2 < p_2 + l_2$. This cannot be true, since by Property (ix), we have $\Delta_2 f(y_1, K^f(y_1)-1) > p_2 + l_2$.

The proofs for Properties (xi) and (xii) are similar to those for (iv) and (vi), thus omitted. Property (xiii) is a special case of the result in Property (x) when $y_1 < R^f$. ■

### PROOF OF LEMMA 5

The model with outsourcable aftermarket demand distinguishes original model in operator $H_1$. So we only need to give out the proof that is related with $H_1$. The rest of the proof is the same as for Lemma 2, and thus is omitted.

**Proof for Condition C.1.1:** We prove that $H_1 f$ preserves Condition C.1.1, i.e., $\Delta_i H_1 f(\mathbf{y}) \geq 0$, for $y_i < 0$, $i = 1, 2$. Note that we can write $H_1 f(\mathbf{y}+\mathbf{e_i}) = f(\mathbf{y}+\mathbf{e_i}-\mathbf{e_1}+\mathbf{e_p}) + p_1 + g_1(p)$, and $H_1 f(\mathbf{y}) = f(\mathbf{y}-\mathbf{e_1}+\mathbf{e_q}) + p_1 + g_1(q)$, with $\mathbf{e_p}$ and $\mathbf{e_q}$ equal to $\mathbf{e_0}$ or $\mathbf{e_2}$, and $g_1(0) = 0$ and $g_1(2) = -p_2 - l_2$. We have,

$$
\begin{aligned}
\Delta_i H_1 f(\mathbf{y}) &= H_1 f(\mathbf{y}+\mathbf{e_i}) - H_1 f(\mathbf{y}) && (Definition)\\
&= f(\mathbf{y}+\mathbf{e_i}-\mathbf{e_1}-\mathbf{e_p}) + g_1(p) - f(\mathbf{y}-\mathbf{e_1}-\mathbf{e_q}) - g_1(q) && (Above)\\
&\geq f(\mathbf{y}+\mathbf{e_i}-\mathbf{e_1}-\mathbf{e_q}) + g_1(q) - f(\mathbf{y}-\mathbf{e_1}-\mathbf{e_q}) - g_1(q) && (Max)\\
&= \Delta_i f(\mathbf{y}-\mathbf{e_1}-\mathbf{e_q}) && (Definition)\\
&\geq 0 && (C.1.1)
\end{aligned}
$$

**Proof for Condition C.1.2:** The proof is similar to the proof for Condition C.1.1, and is therefore omitted.

**Proof for Condition C.1.3:** We must prove that $\Delta_1 H_1 f(\mathbf{y})$ is non-increasing in $y_1$ and $y_2$.

**Proof for $\Delta_1 H_1 f(\mathbf{y})$:** For this case, we have,

$$
\Delta_1 H_1 f(\mathbf{y}) = \begin{cases}
f(\mathbf{y}+\mathbf{e_1}-\mathbf{e_1}) + p_1 - f(\mathbf{y}-\mathbf{e_1}) - p_1 = \Delta_1 f(\mathbf{y}-\mathbf{e_1}) & if \quad y_2 > K^f(y_1)\\
f(\mathbf{y}+\mathbf{e_1}-\mathbf{e_1}+\mathbf{e_2}) - f(\mathbf{y}-\mathbf{e_1}+\mathbf{e_2}) = \Delta_1 f(\mathbf{y}-\mathbf{e_1}+\mathbf{e_2}) & if \quad y_2 \leq K^f(y_1+1)\\
f(\mathbf{y}+\mathbf{e_1}-\mathbf{e_1}) - f(\mathbf{y}-\mathbf{e_1}+\mathbf{e_2}) + p_2 + l_2 = \Delta_{12} f(\mathbf{y}-\mathbf{e_1}) + p_2 + l_2 & if \quad K^f(y_1) \geq y_2 > K^f(y_1+1)
\end{cases} \tag{13}
$$

By Condition C.1.3, it is easy to see that $\Delta_1 H_1 f(\mathbf{y})$ is non-increasing in $y_1$ and $y_2$ within each individual case. In the following, we show it is also true across different cases. Let us denote $\mathbf{y^K} = (y_1, y_2)$, such that $K^f(y_1) \geq y_2 > K^f(y_1+1)$. In the following, we will compare $\Delta_1 H_1 f(\mathbf{y^K})$ with $\Delta_1 H_1 f(\mathbf{y^K}+\mathbf{e_1})$, $\Delta_1 H_1 f(\mathbf{y^K}+\mathbf{e_2})$, $\Delta_1 H_1 f(\mathbf{y^K}-\mathbf{e_1})$ and $\Delta_1 H_1 f(\mathbf{y^K}-\mathbf{e_2})$, one by one.

$$
\begin{aligned}
\Delta_1 H_1 f(\mathbf{y^K}) - \Delta_1 H_1 f(\mathbf{y^K}+\mathbf{e_1}) &= \Delta_{12} f(\mathbf{y^K}-\mathbf{e_1}) + p_2 + l_2 - \Delta_1 f(\mathbf{y^K}+\mathbf{e_1}-\mathbf{e_1}) && (Equation\ (13))\\
&\geq \Delta_{12} f(\mathbf{y^K}) + p_2 + l_2 - \Delta_1 f(\mathbf{y^K}) && (C.2.1)\\
&\geq \Delta_{12} f(\mathbf{y^K}) + \Delta_2 f(\mathbf{y^K}) - \Delta_1 f(\mathbf{y^K}) && (Property(ix))\\
&= \Delta_{12} f(\mathbf{y^K}) - \Delta_{12} f(\mathbf{y^K}) && (\Delta_{12} = \Delta_1 - \Delta_2)\\
&= 0
\end{aligned}
$$

$$\begin{aligned}
\Delta_1 H_1 f(\mathbf{y^K}) - \Delta_1 H_1 f(\mathbf{y^K} - \mathbf{e_1}) &= \Delta_{12} f(\mathbf{y^K} - \mathbf{e_1}) + p_2 + l_2 - \Delta_1 f(\mathbf{y^K} - \mathbf{e_1} - \mathbf{e_1} + \mathbf{e_2}) && (Equation\ (13)) \\
&= \Delta_1 f(\mathbf{y^K} - \mathbf{e_1}) - \Delta_2 f(\mathbf{y^K} - \mathbf{e_1}) + p_2 + l_2 - \Delta_1 f(\mathbf{y^K} - 2\mathbf{e_1} + \mathbf{e_2}) && (\Delta_{12} = \Delta_1 - \Delta_2) \\
&\leq \Delta_1 f(\mathbf{y^K} - \mathbf{e_1}) - \Delta_1 f(\mathbf{y^K} - 2\mathbf{e_1} + \mathbf{e_2}) && (Property(ix)) \\
&= f(\mathbf{y^K}) - f(\mathbf{y^K} - \mathbf{e_1}) - f(\mathbf{y^K} - \mathbf{e_1} + \mathbf{e_2}) + f(\mathbf{y^K} - 2\mathbf{e_1} + \mathbf{e_2}) && (Definition) \\
&= \Delta_{12} f(\mathbf{y^K} - \mathbf{e_1}) - \Delta_{12} f(\mathbf{y^K} - 2\mathbf{e_1}) && (Definition) \\
&\leq 0
\end{aligned}$$

Thus, $\Delta_1 H_1 f(\mathbf{y^K})$ is non-increasing in $y_1$ across the adjacent intervals at $\mathbf{y^K}$. We now compare $\Delta_1 H_1 f(\mathbf{y^K})$ with $\Delta_1 H_1 f(\mathbf{y^K} + \mathbf{e_2})$ and $\Delta_1 H_1 f(\mathbf{y^K} - \mathbf{e_2})$ to show that $\Delta_1 H_1 f(\mathbf{y^K})$ is also non-increasing in $y_2$ across the adjacent intervals.

$$\begin{aligned}
\Delta_1 H_1 f(\mathbf{y^K}) - \Delta_1 H_1 f(\mathbf{y^K} + \mathbf{e_2}) &= \Delta_{12} f(\mathbf{y^K} - \mathbf{e_1}) + p_2 + l_2 - \Delta_1 f(\mathbf{y^K} + \mathbf{e_2} - \mathbf{e_1}) && (Equation\ (13)) \\
&= \Delta_{12} f(\mathbf{y^K} - \mathbf{e_1} + \mathbf{e_2}) + p_2 + r_2 - \Delta_1 f(\mathbf{y^K} + \mathbf{e_2} - \mathbf{e_1}) && (C.2.1) \\
&= p_2 + r_2 - \Delta_2 f(\mathbf{y^K} - \mathbf{e_1} + \mathbf{e_2}) && (\Delta_{12} = \Delta_1 - \Delta_2) \\
&= p_2 + r_2 - \Delta_2 f(\mathbf{y^K}) && (C.2.4) \\
&\geq 0 && (Property(ix))
\end{aligned}$$

$$\begin{aligned}
\Delta_1 H_1 f(\mathbf{y^K}) - \Delta_1 H_1 f(\mathbf{y^K} - \mathbf{e_2}) &= \Delta_{12} f(\mathbf{y^K} - \mathbf{e_1}) + p_2 + l_2 - \Delta_1 f(\mathbf{y^K} - \mathbf{e_1} - \mathbf{e_1} - \mathbf{e_2}) && (Equation\ (13)) \\
&\leq \Delta_{12} f(\mathbf{y^K} - 2\mathbf{e_1} - \mathbf{e_2}) + p_2 + l_2 - \Delta_1 f(\mathbf{y^K} - 2\mathbf{e_1} - \mathbf{e_2}) && (C.2.1) \\
&= p_2 + r_2 - \Delta_2 f(\mathbf{y^K} - 2\mathbf{e_1} - \mathbf{e_2}) && (\Delta_{12} = \Delta_1 - \Delta_2) \\
&\leq 0 && (Property(ix))
\end{aligned}$$

Therefore, $\Delta_1 H_2 f(\mathbf{y^K})$ is also non-increasing in $y_2$ across the adjacent intervals at $\mathbf{y^K}$.

**Proof for $\Delta_2 H_1 f(\mathbf{y})$:** We have,

$$\Delta_2 H_1 f(\mathbf{y}) = \begin{cases} f(\mathbf{y} + \mathbf{e_2} - \mathbf{e_1}) + p_1 - f(\mathbf{y} - \mathbf{e_1}) - p_1 = \Delta_2 f(\mathbf{y} - \mathbf{e_1}) & if \quad y_2 > K^f(y_1) \\ f(\mathbf{y} + \mathbf{e_2} - \mathbf{e_1}) - f(\mathbf{y} - \mathbf{e_1} + \mathbf{e_2}) + p_2 + l_2 = p_2 + l_2 & if \quad y_2 = K^f(y_1) \\ f(\mathbf{y} + \mathbf{e_2} - \mathbf{e_1} + \mathbf{e_2}) - f(\mathbf{y} - \mathbf{e_1} + \mathbf{e_2}) = \Delta_2 f(\mathbf{y} - \mathbf{e_1} + \mathbf{e_2}) & if \quad y_2 < K^f(y_1) \end{cases}$$

According to Condition C.1.3, we know that $\Delta_2 f(\mathbf{y})$ is non-increasing in $y_1$ and $y_2$. Therefore, $\Delta_2 H_2(\mathbf{y})$ is non-increasing in $y_1$ and $y_2$ when $y_2 < K^f(y_1)$ or when $y_2 > K^f(y_1)$. Using the same approach as in the proof of $\Delta_1 H_1 f(\mathbf{y})$, it is easy to prove that $\Delta_2 H_1(\mathbf{y})$ is non-increasing in $y_1$ and $y_2$.

**Proof for Condition C.1.4:** We must prove that $\Delta_{12} H_1 f(\mathbf{y})$ is non-increasing in $y_1$ and non-decreasing in $y_2$.

**Proof for $\Delta_{12} H_1 f(\mathbf{y})$:** For this case, we have,

$$\Delta_{12} H_1 f(\mathbf{y}) = \begin{cases} f(\mathbf{y} + \mathbf{e_1} - \mathbf{e_1}) + p_1 - f(\mathbf{y} + \mathbf{e_2} - \mathbf{e_1}) - p_1 = \Delta_{12} f(\mathbf{y} - \mathbf{e_1}) & if \quad y_2 > K^f(y_1 + 1) \\ f(\mathbf{y} + \mathbf{e_1} - \mathbf{e_1} + \mathbf{e_2}) - p_2 - l_2 - f(\mathbf{y} + \mathbf{e_2} - \mathbf{e_1}) = \Delta_1 f(\mathbf{y} - \mathbf{e_1} + \mathbf{e_2}) - p_2 - l_2 & if \quad K^f(y_1) - 1 < y_2 \leq K^f(y_1 + 1) \\ f(\mathbf{y} + \mathbf{e_1} - \mathbf{e_1} + \mathbf{e_2}) - f(\mathbf{y} + \mathbf{e_2} - \mathbf{e_1} + \mathbf{e_2}) = \Delta_{12} f(\mathbf{y} - \mathbf{e_1} + \mathbf{e_2}) & if \quad y_2 \leq K^f(y_1) - 1 \end{cases}$$

Using the same argument as in the proof of $\Delta_1 H_1 f(\mathbf{y})$, it is easy to prove that $\Delta_{12} H_1(\mathbf{y})$ is non-increasing in $y_1$ and non-decreasing in $y_2$.

**Proof for Condition C.2.1:** We prove that $\Delta_{12} H_1 f(\mathbf{y})$ is non-increasing in $y_1$ and is independent of $y_2$. By definition, $\Delta_{12} H_1 f(\mathbf{y}) = H_1 f(\mathbf{y} + \mathbf{e_1}) - H_1 f(\mathbf{y} + \mathbf{e_2})$, so we have,

$$\Delta_{12} H_1 f(\mathbf{y}) = \begin{cases} f(\mathbf{y} + \mathbf{e_1} - \mathbf{e_1}) + p_1 - f(\mathbf{y} + \mathbf{e_2} - \mathbf{e_1}) - p_1 = \Delta_{12} f(\mathbf{y} - \mathbf{e_1}) & if \quad y_1 + y_2 \geq L^f \\ f(\mathbf{y} + \mathbf{e_1} - \mathbf{e_1} + \mathbf{e_2}) - f(\mathbf{y} + \mathbf{e_2} - \mathbf{e_1} + \mathbf{e_2}) = \Delta_{12} f(\mathbf{y} - \mathbf{e_1} + \mathbf{e_2}) & if \quad y_1 + y_2 < L^f \end{cases}$$

Using the same argument as in the proof for Condition C.1.3, it is easy to show that $\Delta_{12} H_1 f(\mathbf{y})$ is independent of $y_2$.

**Proof for Condition C.2.2:** Condition C.2.2 is a direct result of Condition C.2.1.

**Proof for Condition C.2.3:** The proof for Condition C.2.3 is the same as in Lemma 2.

**Proof for Condition C.2.4:** Condition C.2.4 can be derived from Condition C.2.1 as shown in the proof for Lemma 2. ■

**PROOF OF THEOREM 2**

**Proof.** The proof of the production, rationing, and admission control policies are the same as in Theorem 1, so we only discuss the outsourcing control policy.

When an order for type 1 arrives, the outsourcing control for type 2 orders depends on the sign of $p_2+l_2-\Delta_2 J(y_1, y_2)$: outsource an order if the sign is negative, and do not outsource otherwise. When $y_1 < R$, by Conditions C.1.3 and C.4.3, the optimal cancelation policy is to outsource a type 2 order if $y_1 + y_2 \leq L*$; When $y_1 \geq R$, by Conditions C.1.3 and C.2.6, the optimal cancelation policy is to outsource a type 2 order if $y_2 \leq K(y_1)$ ■

# ON-LINE APPENDIX D

## Analytical Supporting Results For the Model with Customized/Preconfigured Products

Before proving Theorem 3, we first define a set $\mathcal{D}$ of functions such that if $f(y_1, y_2) \in \mathcal{D}$, then,

- **Condition D.1:** For $(y_1, y_2) \in \Omega$,

  D.1.1: $\Delta_i f(y_1, y_2) \geq 0$, if $y_i < 0$, $i = 1, 2$.

  D.1.2: $\Delta_1 f(y_1, y_2) \leq p_1 + r_1$, if $y_1 \geq 0$.

  D.1.3: $\Delta_{12} f(y_1, y_2) \leq 0$.

- **Condition D.2:** For $(y_1, y_2) \in \Omega$,

  D.2.1: $\Delta_1 f(y_1, y_2)$ is non-increasing in $y_1$ and $y_2$.

  D.2.2: $\Delta_1 f(y_1, y_2) \geq 0$ for $y_1 < S$, where $S = \min\{z | \Delta_1 f(z, 0) < 0\}$.

- **Condition D.3:** For $(y_1, y_2) \in \Omega$, $y_1 < 0$,

  D.3.1: $\Delta_1 f(y_1, y_2)$ is independent of $y_1$ and $y_2$ as long as $y_1 + y_2$ is fixed.

  D.3.2: $\Delta_1 f(y_1 - 1, y_2) \leq p_1 + r_1$ for $y_1 + y_2 > B$, where $B = \max\{z | \Delta_1 f(z - 1, 0) > p_1 + r_1\}$.

To get some intuition on the above conditions, we apply the condition set $\mathcal{D}$ to the expected profit, $J(y_1, y_2)$. Condition D.1.1 implies that if there are orders in any class, the production line needs to keep producing. Condition D.1.2 implies that if the inventory for type 1 product is available, it is profitable to satisfy demand for type 1 product. Condition D.1.3 implies that an order for type 2 always has a higher priority than for type 1. Condition D.2.1 implies that the marginal benefit of increasing $y_1$ is non-increasing in both $y_1$ and $y_2$. Condition D.2.2 implies that if $y_1 < S$, then it is profitable to produce product 1. Since the sign of $p_1 + r_1 - \Delta_1 J(y_1 - 1, y_2)$ determines whether to reject an order for type 1, Condition D.3.1 suggests that the rejection decisions depend on total inventory level. Condition D.3.2 implies that accepting an arriving demand for type 1 is profitable as long as the level of inventory satisfies $\sum_{i=1}^{n} y_i > B$.

We show in Proposition 5 that under the optimal conditions, the threshold level $B$ is non-positive. Therefore, it is always optimal to satisfy promotional demands when the inventory is available.

**Property 5** *If $f(y_1, y_2) \in \mathcal{D}$, then $B \leq 0$.*

**Proof.** By contradiction, assume $B > 0$, i.e., $B \geq 1$. By Condition D.3.2, $\Delta_1 f(0,0) > p_1 + r_1$, which cannot be true since we know that $\Delta_1 f(0,0) \leq p_1 + r_1$ by Condition D.1.2. ■

The following lemma shows that the structure of the functions in $\mathcal{D}$ is preserved under function $H$.

**Lemma 8** *If $f(y_1, y_2) \in \mathcal{D}$, then $Hf(y_1, y_2) \in \mathcal{D}$.*

**Proof.** We show that function $H$ preserves each condition in $\mathcal{D}$. Proofs for Conditions D.1.1, D.1.2, and D.1.3 are similar to the proof for Condition C.1.1, and are therefore omitted.

**Proof for Condition D.2.1:**

**Proof for $\Delta_1 H_0 f(\mathbf{y})$:** We first prove that $\Delta_1 H_0 f(\mathbf{y})$ is non-increasing in $y_1$ and $y_2$. By definition, $\Delta_1 H_0 f(\mathbf{y}) = H_0 f(\mathbf{y} + \mathbf{e_1}) - H_0 f(\mathbf{y})$, so we have,

$$\Delta_1 H_0 f(\mathbf{y}) = \begin{cases} f(\mathbf{y} + \mathbf{e_1} + \mathbf{e_2}) - f(\mathbf{y} + \mathbf{e_2}) = \Delta_1 f(\mathbf{y} + \mathbf{e_2}) & if \quad y_2 < 0 \\ f(\mathbf{y} + 2\mathbf{e_1}) - f(\mathbf{y} + \mathbf{e_1}) \qquad = \Delta_1 f(\mathbf{y} + \mathbf{e_1}) & if \quad y_2 = 0, \quad y_1 < S-1 \\ f(\mathbf{y} + \mathbf{e_1}) - f(\mathbf{y} + \mathbf{e_1}) \qquad = 0 & if \quad y_2 = 0, \quad y_1 = S-1 \\ f(\mathbf{y} + \mathbf{e_1}) - f(\mathbf{y}) \qquad\qquad = \Delta_1 f(\mathbf{y}) & if \quad y_2 = 0, \quad y_1 > S-1 \end{cases}$$

By D.2.1, $\Delta_1 f(\mathbf{y})$ is non-increasing in $y_1$ and $y_2$, so $\Delta_1 H_0 f(\mathbf{y})$ is non-increasing in $y_1$ and $y_2$ *within* each of the above four sub-condition intervals. In the following, we show that $\Delta_1 H_0 f(\mathbf{y})$ is also non-increasing in $y_1$ and $y_2$ *across* any adjacent intervals. We only present the proof for the case where $y_1 = S-1$ and $y_2 = 0$. The proof for other cases are similar and are therefore omitted.

When $y_1 = S-1$ and $y_2 = 0$, let us denote $\mathbf{y^S} = (S-1, 0)$. Then we have,

$$\begin{aligned} \Delta_1 H_0 f(\mathbf{y^S} + \mathbf{e_1}) &= \Delta_1 f(\mathbf{y^S} + \mathbf{e_1}) \\ \Delta_1 H_0 f(\mathbf{y^S}) &= 0 \\ \Delta_1 H_0 f(\mathbf{y^S} - \mathbf{e_1}) &= \Delta_1 f(\mathbf{y^S}) \end{aligned}$$

By D.2.2, we have $\Delta_1 f(\mathbf{y^S} + \mathbf{e_1}) \leq 0 \leq \Delta_1 f(\mathbf{y^S})$, so $\Delta_1 H_0 f(\mathbf{y})$ is non-increasing in $y_1$ across the adjacent intervals at $\mathbf{y^S}$.

Notice that,

$$\Delta_1 H_0 f(\mathbf{y^S} - \mathbf{e_2}) = \Delta_1 f(\mathbf{y^S})$$

so for the same reason, $\Delta_1 H_0 f(\mathbf{y})$ is non-increasing in $y_2$ across the adjacent intervals at $\mathbf{y^S}$.

**Proof for $\Delta_1 H_1 f(\mathbf{y})$:** Now we prove that $\Delta_1 H_1 f(\mathbf{y})$ is non-increasing in $y_1$ and $y_2$. By definition, $\Delta_1 H_1 f(\mathbf{y}) = H_1 f(\mathbf{y} + \mathbf{e_1}) - H_1 f(\mathbf{y})$, so we have,

$$\Delta_1 H_1 f(\mathbf{y}) = \begin{cases} f(\mathbf{y} + \mathbf{e_1} - \mathbf{e_1}) + p_1 - f(\mathbf{y} - \mathbf{e_1}) - p_1 = \Delta_1 f(\mathbf{y} - \mathbf{e_1}) & if \quad y_1 > 0 \\ f(\mathbf{y} + \mathbf{e_1} - \mathbf{e_1}) + p_1 - f(\mathbf{y} - \mathbf{e_1}) - p_1 = \Delta_1 f(\mathbf{y} - \mathbf{e_1}) & if \quad y_1 + y_2 > B, \quad y_1 \leq 0 \\ f(\mathbf{y} + \mathbf{e_1} - \mathbf{e_1}) + p_1 - f(\mathbf{y}) + r_1 \qquad = p_1 + r_1 & if \quad y_1 + y_2 = B, \quad y_1 \leq 0 \\ f(\mathbf{y} + \mathbf{e_1}) - r_1 - f(\mathbf{y} - \mathbf{e_1}) + r_1 \qquad = \Delta_1 f(\mathbf{y}) & if \quad y_1 + y_2 < B, \quad y_1 \leq 0 \end{cases}$$

By D.2.1, $\Delta_1 f(\mathbf{y})$ is non-increasing in $y_1$ and $y_2$, so $\Delta_1 H_1 f(\mathbf{y})$ is non-increasing in $y_1$ and $y_2$ *within* each of the four sub-condition intervals. In the following, we show that $\Delta_1 f(\mathbf{y})$ is non-increasing in $y_1$ and $y_2$ *across* any two adjacent intervals. We only present the proof for the third interval, since the proofs for other intervals are similar.

Let us denote $\mathbf{y^B} = (y_1, y_2)$, such that $y_1 + y_2 = B$ and $y_1 \leq 0$. With the above results, we have,

$$\begin{aligned} \Delta_1 H_1 f(\mathbf{y^B} + \mathbf{e_1}) &= \Delta_1 f(\mathbf{y^B}) \\ \Delta_1 H_1 f(\mathbf{y^B}) &= p_1 + r_1 \\ \Delta_1 H_1 f(\mathbf{y^B} - \mathbf{e_1}) &= \Delta_1 f(\mathbf{y^B} - \mathbf{e_1}) \\ \Delta_1 H_1 f(\mathbf{y^B} + \mathbf{e_2}) &= \Delta_1 f(\mathbf{y^B} + \mathbf{e_2} - \mathbf{e_1}) \\ \Delta_1 H_1 f(\mathbf{y^B} - \mathbf{e_2}) &= \Delta_1 f(\mathbf{y^B} - \mathbf{e_2}) \end{aligned}$$

Condition D.3.1 and D.3.2 imply that $\Delta_1 f(\mathbf{y^B}+\mathbf{e_2}-\mathbf{e_1}) = \Delta_1 f(\mathbf{y^B}) \le p_1 + r_1 \le \Delta_1 f(\mathbf{y^B}-\mathbf{e_1}) = \Delta_1 f(\mathbf{y^B}-\mathbf{e_2})$, so $\Delta_1 H_1 f(\mathbf{y})$ is non-increasing in $y_1$ and $y_2$ across the adjacent intervals at $\mathbf{y^B}$.

**Proof for $\Delta_1 H_2 f(\mathbf{y})$:** Now we prove that $\Delta_1 H_2 f(\mathbf{y})$ is non-increasing in $y_1$ and $y_2$. We have

$$\Delta_1 H_2 f(\mathbf{y}) = H_2 f(\mathbf{y}+\mathbf{e_1}) - H_2 f(\mathbf{y}) = f(\mathbf{y}+\mathbf{e_1}-\mathbf{e_2}) - f(\mathbf{y}-\mathbf{e_2}) = \Delta_1 f(\mathbf{y}-\mathbf{e_2}).$$

By D.2.1, it is clear that $\Delta_1 H_2 f(\mathbf{y})$ is non-increasing in $y_1$ and $y_2$.

**Proof for $\Delta_1 c(\mathbf{y})$:** Finally for $c_1(\mathbf{y})$ we have

$$\Delta_1 c_1(\mathbf{y}) = c_1(\mathbf{y}+\mathbf{e_1}) - c_1(\mathbf{y}) = \begin{cases} b_1(y_1+1) - b_1 y_1 = b_1 & if \quad y_1 < 0; \\ -h(y_1+1) + h y_1 = -h & if \quad y_1 \ge 0. \end{cases}$$

which is non-increasing in $y_1$ and $y_2$. In Conclusion, $\Delta_1 H f(\mathbf{y})$ is non-increasing in $y_1$ and $y_2$.

**Proof for Condition D.2.2:** Condition D.2.2 is a direct result of Condition D.2.1.

**Proof for Condition D.3.1:** We first prove that $\Delta_{12} f(y_1, y_2) = \Delta_{12} f(y_1 - 1, y_2)$ for $y_1 < 0$. By definition, $\Delta_{12} H_0 f(\mathbf{y}) = H_0 f(\mathbf{y}+\mathbf{e_1}) - H_0 f(\mathbf{y}+\mathbf{e_2})$, so we have,

$$\begin{aligned}
\Delta_{12} H_0 f(\mathbf{y}) &= \begin{cases} f(\mathbf{y}+\mathbf{e_1}+\mathbf{e_2}) - f(\mathbf{y}+2\mathbf{e_2}) = \Delta_{12} f(\mathbf{y}+\mathbf{e_2}) & if \quad y_2 < -1 \\ f(\mathbf{y}+\mathbf{e_1}+\mathbf{e_2}) - f(\mathbf{y}+\mathbf{e_2}+\mathbf{e_1}) = 0 & if \quad y_2 = -1 \end{cases} \\
\Delta_{12} H_1 f(\mathbf{y}) &= \begin{cases} f(\mathbf{y}+\mathbf{e_1}-\mathbf{e_1}) + p_1 - f(\mathbf{y}+\mathbf{e_2}-\mathbf{e_1}) - p_1 = \Delta_{12} f(\mathbf{y}-\mathbf{e_1}) & if \quad y_1+y_2 \ge B \\ f(\mathbf{y}+\mathbf{e_1}) - r_1 - f(\mathbf{y}+\mathbf{e_2}) + r_1 = \Delta_{12} f(\mathbf{y}) & if \quad y_1+y_2 < B \end{cases} \\
\Delta_{12} H_2 f(\mathbf{y}) &= f(\mathbf{y}+\mathbf{e_1}-\mathbf{e_2}) - f(\mathbf{y}+\mathbf{e_2}-\mathbf{e_2}) = \Delta_{12} f(\mathbf{y}-\mathbf{e_2})
\end{aligned}$$

Noting that $\Delta_{12} c_1(\mathbf{y}) = b_1$ is independent of $y_1$ for $y_1 < 0$, and $\Delta_{12} f(y_1, y_2) = \Delta_{12} f(y_1 - 1, y_2)$, so $\Delta_{12} H f(y_1, y_2) = \Delta_{12} H f(y_1 - 1, y_2)$.

Then we have

$$\begin{aligned}
& \Delta_{12} f(\mathbf{y}) &=& \ \Delta_{12} f(\mathbf{y}+\mathbf{e_1}) & \\
\Rightarrow\ & f(\mathbf{y}+\mathbf{e_1}) - f(\mathbf{y}+\mathbf{e_2}) &=& \ f(\mathbf{y}+\mathbf{e_1}+\mathbf{e_1}) - f(\mathbf{y}+\mathbf{e_1}+\mathbf{e_2}) & (Definition) \\
\Rightarrow\ & f(\mathbf{y}+\mathbf{e_1}+\mathbf{e_2}) - f(\mathbf{y}+\mathbf{e_2}) &=& \ f(\mathbf{y}+\mathbf{e_1}+\mathbf{e_1}) - f(\mathbf{y}+\mathbf{e_1}) & (Rearrangement) \\
\Rightarrow\ & \Delta_1 f(\mathbf{y}+\mathbf{e_2}) &=& \ \Delta_1 f(\mathbf{y}+\mathbf{e_1}) & (Definition)
\end{aligned}$$

Therefore, we have $\Delta_1 f(\mathbf{y}) = \Delta_1 f(\mathbf{y}+\mathbf{e_1}-\mathbf{e_2})$, and Condition D.3.1 can be derived by adding 1 to one element and deducing 1 form another element in vector $\mathbf{y}$ in the above result.

**Proof for Condition D.3.2:** Condition D.3.2 can be directly deduced from Conditions D.2.1 and D.3.1. ■

## PROOF OF THEOREM 3

The proof for the existence of the optimal policies is similar to the proof in Theorem 1, and is therefore omitted.

From equation (7), we see that when there are no orders from any classes, the optimal production control only depends on the sign of $\Delta_1 J_1(y_1, y_2)$: produce if it is positive, or do not produce otherwise. So by Condition D.2.2, it is optimal to produce if $y_1 < S$. When there are orders, by Conditions D.1.1 and D.1.2, it is always optimal to produce for the order with the highest priority.

The admission control for an arriving demand for type 1 depends on the sign of $p_1 + r_1 - \Delta_1 J_1(y_1 - 1, y_2)$: satisfy or backlog the demand if the sign is positive; reject otherwise. So, by Condition D.1.2, the optimal admission policy is to satisfy an arriving demand for type 1 from inventory if $y_1 > 0$, and to reject if $y_1 < 0$ and $y_1 + y_2 \le B$ by Condition D.3.2. ■

**PROOF OF LEMMA6**

The proof is similar to that for Lemma 3, and thus is omitted. ■

**PROOF OF LEMMA7**

The proof is similar to that for Lemma 4, and thus is omitted. ■

# ON-LINE APPENDIX E

## Analytical Supporting Results for Customized/Reconfigured Model with Contingent Outsourcing

Before proving Theorem 4, we first add the following condition to $\mathcal{D}$, and denote the new condition set as $\mathcal{D}^{'}$:

- **Condition D.3:** For $(y_1, y_2) \in \Omega$, $y_1 < 0$,

  D.3.3: $\Delta_1 f(y_1 - 1, y_2) \leq p_1 + l_1$ for $y_1 + y_2 > L$, where $L = \max\left\{z|\Delta_1 f(z-1,0) > p_1 + l_1\right\}$.

Condition D.3.3 suggests that outsourcing decision for an order of the type 1 product depends on total inventory level, but not the number of orders of type 1 itself. It implies that outsourcing an order of product 1 is profitable when $y_1 + y_2 \leq L$.

Then we show that under the optimal conditions, threshold level $B$ cannot be lower than $L$.

**Property 6** *If $f(y_1, y_2) \in \mathcal{D}^{'}$, then $L \leq B$.*

**Proof:**By contradiction, assume $L > B$. By Condition D.3.2, $\Delta_1 f(L-1,0) \leq p_1 + r_1 < p_1 + l_1$. This cannot be true, since by Condition D.3.3 we have $\Delta_1 f(L-1,0) > p_1 + l_1$. ■

**PROOF OF THEOREM 4**

The proof of the production and admission control policies are the same as in Theorem 3, so we only discuss the outsourcing control policy.

After accepting a demand for a higher-priority product, the outsourcing control for type 1 orders depends on the sign of $p_1 + l_1 - \Delta_1 J_1(y_1, y_2)$: outsource an order if the value is negative, and do not outsource otherwise. So by Conditions D.2.2 and D.3.2, the optimal cancelation policy is to outsource a type 1 order if $y_1 + y_2 \leq L$. ■

# ON-LINE APPENDIX F
# Heuristic for $(S,R,B)$ Thresholds in OEM/Aftermarket Case

Although the linear $(S,R,B)$ structure greatly simplifies the optimal policy, the value iteration approach itself may require a large amount of computation. In this section, we extend de Véricourt et al.'s algorithm to speed up the searching process for the thresholds.

For a problem with two demand classes and a rationing level $R$, de Véricourt et al.'s algorithm first solves a subproblem of a make-to-stock $M/M/1$ queue with holding cost $h+b_2$ and backorder cost $b_1-b_2$. The average cost minimizer for the subproblem is the optimal rationing level for the main problem. We adopted this idea and treat the OEM product's inventory level as a make-to-stock $M/M/1$ queue with holding cost $h+b_2$ and backorder cost $b_1-b_2$. Given the base-stock level, $R$, and let $\Psi_i$ be the steady state probability when $y_1(t)=i$, the total cost function $g_1(R)$ of the $M/M/1$ subproblem is as follows.

$$
\begin{aligned}
g_1(R) &= (h+b_2)\sum_{i=0}^{R} i\Psi_i + (b_1-b_2)\sum_{i=-\infty}^{-1}(-i)\Psi_i \\
&= (h+b_2)\sum_{i=0}^{R} i\Psi_i + (b_1-b_2)\big[\sum_{i=-\infty}^{R}(-i)\Psi_i - \sum_{i=0}^{R}(-i)\Psi_i\big] \\
&= (h+b_2+b_1-b_2)\sum_{i=0}^{R} i\Psi_i + (b_1-b_2)\sum_{i=-\infty}^{R}(-i)\Psi_i \\
&= (h+b_1)\sum_{i=0}^{R} i\Psi_i + (b_1-b_2)\big[\sum_{i=-\infty}^{R}(R-i)\Psi_i - R\big]
\end{aligned}
$$

Note that $R-y_1(t)$ is the number of customers in an $M/M/1$ queue, and as a result, $\sum_{i=-\infty}^{R}(R-i)\Psi_i$ is the average queue length, which is $\frac{\lambda_1}{\mu_1-\lambda_1}$. Also note that $\sum_{i=0}^{R} i\Psi_i$ is the average queue length for a $M/M/1/R$ queue (where $R$ is the capacity), $L_1=\frac{\lambda[1+R(\lambda/\mu)^{R+1}-(R+1)(\lambda_1/\mu_1)^R]}{(u_1-\lambda_1)(1-(\lambda_1/\mu_1)^{R+1})}$. Denote $R^H=\arg\min\{g_1(R):R\in\mathbf{Z}^+\}$ be the optimal solution and $g_1^*=g_1(R^H)$ as the optimal time-average cost of the subproblem.

We take the optimal base-stock level $R^H$ of the above subproblem as the rationing level of the heuristic algorithm in the main problem. Given base-stock level $S$ $(S\geq R^H)$ and backlog-up-to level $B$ $(B\leq 0)$ , the main problem's instantaneous cost function (not including the rejection penalty for aftermarket demand) is:

$$
c(y_1,y_2)=hy_1^+ + b_1y_1^- - b_2y_2 = (h+b_2)y_1^+ + (b_1-b_2)y_1^- - b_2(y_1+y_2) = c'(y_1)+b_2\big[S-(y_1+y_2)-S\big],
$$

where $c'(y_1)=(h+b_2)y_1^+ + (b_1-b_2)y_1^-$ is the instantaneous cost function for the subproblem. The time-average cost function of the system is

$$
\begin{aligned}
g_2(S,B) &= \mathbf{E}\big[c(y_1(t),y_2(t))\big] + (p_2+r_2)\lambda_2 Prb(y_1(t)+y_2(t)\leq B) \\
&= \mathbf{E}\big[c'(y_1(t))\big] + b_2\mathbf{E}\big[S-(y_1(t)+y_2(t))-S\big] + (p_2+r_2)\lambda_2 Prb(y_1(t)+y_2(t)\leq B) \\
&= \mathbf{E}\big[c'(y_1(t))\big|y_1(t)\leq R^H\big]Prb(y_1(t)\leq R^H) + \mathbf{E}\big[c'(y_1(t))\big|y_1(t)>R^H\big]Prb(y_1(t)>R^H) \\
&\quad +(p_2+r_2)\lambda_2 Prb(y_1(t)+y_2(t)\leq B) + b_2\mathbf{E}\big[S-(y_1(t)+y_2(t))-S\big]
\end{aligned}
$$

To calculate this cost function, we use the following birth-death queue to approximate the original system. Let $\gamma_{ij}$ be the transition rate from state $i$ to state $j$,

$$
\gamma_{ij} = \begin{cases}
\lambda & if \quad j=i+1,\ j\leq S-B \\
\lambda_1 & if \quad j=i+1,\ j>S-B \\
\mu & if \quad j=i-1,\ j<S-B \\
\mu_1 & if \quad j=i-1,\ j\geq S-B \\
0 & if \quad otherwise
\end{cases}
$$

In this birth-death queue, state $i$ corresponds to $y_1 + y_2 = S - i$ in the original system. We use $\lambda = \lambda_1 + \lambda_2$ and $\mu = \frac{(\lambda_1+\lambda_2)\mu_1\mu_2}{\lambda_1\mu_2+\lambda_2\mu_1}$ to approximate the forward transition rate and backward rate when $i \leq S - B$. Then we use the following function to calculate $g_2(S, B)$,

$$g_2(S,B) = g_1^*(1 - \sum_{i=0}^{S-R^H-1} \Theta_i) + (h + b_2) \sum_{i=0}^{S-R^H-1} (i\Theta_i) + (p_2 + r_2)\lambda_2 \sum_{i=S-B}^{\infty} \Theta_i + b_2(L_2 - S)$$

where $L_2$ is the average length of the birth-death queue. Searching for the minimizers of $g_2(S, B)$, we get the heuristic estimate of $S^H$ and $B^H$.

We perform an extensive numerical study with 16,000 cases to evaluate the performance of the heuristic algorithm. The performance measure is defined as

$$PM = \frac{J_{opt} - J_{heu}}{J_{opt}} \times 100\%,$$

where $J_{heu}$ is the system's time-average profit using the heuristic thresholds $(S^H, R^H, B^H)$, and $J_{opt}$ is the system's optimal time-average profit.

Our numerical study shows that, on average, the heuristic algorithm results in 3.6% less profit than the optimal policy. In 10,194 (out of 16,000) cases, the error is within 2%. Larger errors happen when $u_2$ is significantly smaller than $\mu_1$. Intuitively, this is because the heuristic algorithm does not distinguish the difference in processing time for the two types of orders when calculating $g_2(S, B)$.

# ON-LINE APPENDIX G
# Heuristic for $(S, B)$ Thresholds in Customized/Pre-configured Case

For the model with customized and pre-configured products, the optimal policy only controls the production and order admission for the pre-configured product. Therefore, in the heuristic policy, we approximate the system with a birth-death queue only of the pre-configured product. Let $\gamma_{ij}$ be the transition rate from state $i$ to state $j$,

$$\gamma_{ij} = \begin{cases} \lambda_1 & if \quad j = i+1,\ j \le S-B \\ {\mu'}_1 & if \quad j = i-1,\ j < S-B \\ 0 & if \quad otherwise \end{cases}$$

where ${\mu'}_1 = \mu_1 - \lambda_2\mu_1/\mu_2$. In this birth-death queue, state $i$ corresponds to $y_1 = S - i$ in the original system. Let $\rho_1 = \lambda_1/{\mu'}_1$, and notice the following relation between the steady state probability of each state, $\Theta_{i+1} = \rho_1\Theta_i$, for $i = 0, 1, \ldots, S-B-1$, the probability of state 0 is $\Theta_0 = (1-\rho_1)/(1-\rho_1^{S-B+1})$.

The time-average cost function of the birth-death queue is

$$\begin{aligned} g(S,B) &= h\sum_{i=0}^{S}(S-i)\Theta_i + b_1\sum_{i=S+1}^{S-B}(i-S)\Theta_i + (p_1+r_1)\lambda_1\Theta_{S-B} \\ &= \Theta_0\left[h\sum_{i=0}^{S}(S-i)\rho_1^i + b_1\sum_{i=S+1}^{S-B}(i-S)\rho_1^i + (p_1+r_1)\lambda_1\rho_1^{S-B}\right] \end{aligned}$$

Searching for the minimizers of $g(S, B)$, we get the heuristic estimate of $S^H$ and $B^H$.

We perform an extensive numerical study with 16,000 cases to evaluate the performance of the heuristic algorithm. Our numerical study shows that, on average, the heuristic algorithm results in 3.1% less profit than the optimal policy. In 11,812 (out of 16,000) cases, the error is within 2%. Larger errors happen when the ratio $\lambda_2$ is considerately large. Intuitively, this is because the heuristic algorithm does not consider the impact of type 2 orders.

# ON-LINE APPENDIX H
# Impact on Expected Profit When Capacity Is Tight

In this appendix, we represent the impacts of using the $(S, R, B)$ policy and the impact of flexible capacity on the expected profit when production capacity is very tight, i.e., $\rho = 1.2$. When the capacity is insufficient, the expected profit becomes zero or negative, and it is not appropriate to use profit potentials (i.e. functions (9)) or their absolute values to represent profit improvements. For example, when the expected profit under the $(S, R, B)$ policy is 1 and under the simple base-stock policy is $-1$, the absolute value of the profit potential is 200%; In another case where the expected profit under the $(S, R, B)$ policy is still 1 and under the simple base-stock policy is $-2$, the absolute value of the profit potential is 150%. However, apparently the second case has higher profit improvement. Therefore, we use profits rather than profit potentials in the figures.

Figures 8-Left and 8-Right show the impacts of price ratio and demand ratio on the $(S, R, B)$ policy, the simple base-stock policy, and the dedicated two-facility system, when $\rho = 1.2$. We can examine the impact of using the $(S, R, B)$ policy (whose sufficient capacity case is discussed in Section 4.3). We can see the former by comparing the profits under the $(S, R, B)$ policy and under the simple base-stock policy, and the latter by comparing the profits under the $(S, R, B)$ policy and in the dedicated system.

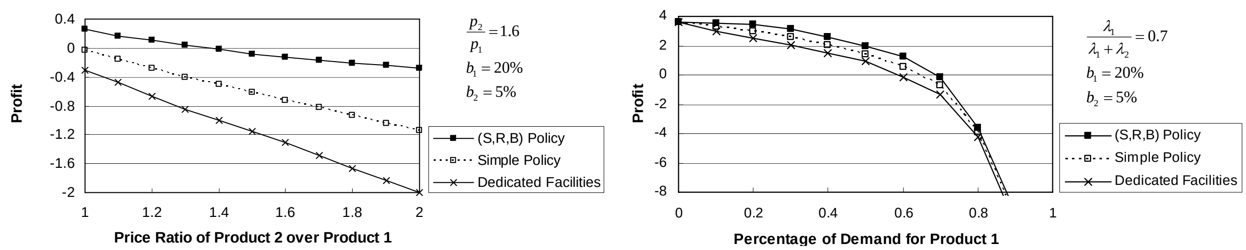


Figure 8: *Left*: Impact of Price Ratio on Profit When Capacity is Tight, $\rho = 1.2$; *Right*: Impact of Demand Ratio on Profit When Capacity is Tight, $\rho = 1.2$.

Figure 8-Left shows that when the production capacity is insufficient, the expected profit decreases as the price for aftermarket product increases. When the production capacity is insufficient, most aftermarket demands are rejected by the manufacturer. In the numerical analysis, we set the rejection penalty proportional to the price. Therefore, the higher price, the higher rejection penalty paid by the manufacturer.

Figure 8-Right shows that when the production capacity is insufficient, the expected profit decreases as the proportion of OEM demand increases. In this situation, most OEM orders will not be satisfied immediately upon their arrivals. So the higher proportion of OEM demand, the higher backlogging cost paid by the manufacturer.

# ON-LINE APPENDIX I
# MDP Formulations for Erlang Model

The system described in Section 5 incurs an inventory holding and backlogging cost at rate,

$$c(x_1, x_2, \cdot, \cdot, \cdot) = -h\lfloor \frac{x_1^+}{m_1} \rfloor - b_1\lfloor \frac{x_1^-}{m_1} \rfloor + b_2\lfloor \frac{x_2}{m_2} \rfloor. \tag{14}$$

Let $\Lambda = \gamma_1 + \gamma_2 + K_1\lambda_1 + K_2\lambda_2$. The optimal profit function $J^*(x_1, x_2, k_1, k_2)$ under the time-discount criterion (defined by replacing $\mathbf{y}(0)$ with $\mathbf{x}(0)$ in Equation (1)) satisfies the following equation:

$$J(x_1, x_2, k_1, k_2) = \frac{1}{\Lambda}\Big\{c(x_1, x_2, \cdot, \cdot) + H_0 J(x_1, x_2, k_1, k_2) + K_1\lambda_1 H_1 J(x_1, x_2, k_1, k_2) + K_2\lambda_2 H_2 J(x_1, x_2, k_1, k_2)\Big\} \tag{15}$$

where $H_0$, $H_1$, and $H_2$ are functions defined by,

$$H_0 J(x_1, x_2, k_1, k_2) = \max \left\{ \begin{array}{l} \gamma_1 J(x_1, x_2, k_1, k_2) + \gamma_2 J(x_1, x_2, k_1, k_2), \\ \gamma_1 J(x_1 + 1, x_2, k_1, k_2) + \gamma_2 J(x_1, x_2, k_1, k_2), \\ \gamma_1 J(x_1, x_2, k_1, k_2) + \gamma_2 J(x_1, x_2 + 1, k_1, k_2 | x_2 < 0). \end{array} \right\}$$

$$H_1 J(x_1, x_2, k_1, k_2) = \left\{ \begin{array}{ll} J(x_1, x_2, k_1 - 1, k_2) & if \quad k_1 > 1 \\ J(x_1 - m_1, x_2, K_1, k_2) + p_1 & if \quad k_1 = 1 \end{array} \right.$$

$$H_2 J(x_1, x_2, k_1, k_2) = \left\{ \begin{array}{ll} J(x_1, x_2, k_1, k_2 - 1) & if \quad k_2 > 1 \\ \max[J(x_1, x_2 - m_2, k_1, K_2) + p_2, J(x_1, x_2, k_1, K_2) - r_2] & if \quad k_2 = 1 \end{array} \right.$$

$H_0$ corresponds to the production decision: the manufacturer can idle, or choose to produce a phase for either type of product. $H_1$ indicates that an OEM demand arrives when the number of phases decreases from $K_1$ to 1, and the work storage level of type 1 is decreased by $m_1$ units as the demand is accepted. $H_2$ indicates that an aftermarket demand arrives when the number of phases decreases from $K_2$ to 1. The manufacturer can either reject or accept an aftermarket demand. If the demand is accepted, the work storage level of type 2 is decreased by $m_2$ units.

The optimal profit function $J^*(x_1, x_2, k_1, k_2)$ under the average-profit criterion (defined by replacing $\mathbf{y}(0)$ with $\mathbf{x}(0)$ in Equation (2)) satisfies the following equation:

$$J(x_1, x_2, k_1, k_2) + g = \frac{1}{\Lambda}\Big\{c(x_1, x_2, \cdot, \cdot) + H_0 J(x_1, x_2, k_1, k_2) + K_1\lambda_1 H_1 J(x_1, x_2, k_1, k_2) + K_2\lambda_2 H_2 J(x_1, x_2, k_1, k_2)\Big\} \tag{16}$$

where $g$ is the optimal average profit.