Optimal Admission Control and Sequencing in a Make-To-Stock/Make-To-Order Production System

Scott Carr
Izak Duenyas
Department of Industrial & Operations Engineering
University of Michigan
Ann Arbor, MI 48109

Production control — Mathematical models.
Production scheduling — Mathematical models.
Queuing theory.

Optimal Admission Control and Sequencing in a Make-To-Stock/Make-To-Order Production System

Scott Carr and Izak Duenyas

Department of Industrial and Operations Engineering

University of Michigan, Ann Arbor, MI 48109

Optimal Admission Control and Sequencing in a Make-To-Stock/Make-To-Order Production System

Scott Carr and Izak Duenyas

Department of Industrial and Operations Engineering

University of Michigan, Ann Arbor, MI, 48109

Abstract

In this paper, we address the problem of admission control and sequencing in a production system which produces two classes of products. The first class of products is made-to-stock, and the firm is contractually obliged to meet demand for this class of products. The second class of products is made-to-order and the firm has the option to accept (admit) or reject a particular order. The problem is motivated by suppliers in many industries who sign contracts with large manufacturers to supply them with a given product and also can take on additional orders from other sources on a make-to-order basis.

We model the joint admission control/sequencing decision in the context of a simple two class M/M/1 queue to gain insight into the following problems:

1. How should a firm decide a) when to accept or reject an additional order; and b) which type of product to produce next?

2. How should a firm decide what annual quantity of orders to commit to when signing a contract to produce the make-to-stock products?

We fully characterize the structure of the optimal admission control and sequencing decisions and also show how changes in problem parameters (such as holding costs, order arrival rates, production rates) affect these decisions. Finally, we compare the performance of simple policies to the performance of the optimal policy, and explore the effects of decreases in demand or production process variability on firm profits and the optimal quantity of orders the firm would be willing to commit to.

1 Introduction

Rarely does a manufacturing firm produce a single product for sale to a single customer. Rather, a typical firm is likely to use its manufacturing facilities to produce a variety of different products for a variety of different customers and market segments. The firm's customers in different market segments may have different requirements and the relationship between producers and customers may vary across different segments. For example, the firm may sign a large contract to supply products to a major customer on a make-to-stock basis and accept additional orders from other customers occasionally when it has extra capacity to be able to fill those orders.

This work is motivated by a major glass manufacturer that fabricates laminated and tempered windows for installation into automobiles, trucks, and farm implements. This firm's products are sold to automotive assembly plants (the original equipment manufacturer (OEM) market segment) for installation into new vehicles and to the automotive glass replacement (aftermarket) segment for installation into older vehicles. The two segments consume nearly identical products but are otherwise of very different character.

Sales to OEMs are based on long-term contracts that include large contractually specified penalties if products are not available when required by the OEM's just-in-time production; the manufacturer is required, through these penalties, to dedicate a portion of its manufacturing capacity to the OEM segment. The long-term contracts are appealing to automotive suppliers because they result in high guaranteed capacity utilization. The tradeoff is that the unit profits generated by these contracts are typically low, and the potential for large penalties dictate that the manufacturer maintain a suitable finished goods inventory to buffer against production delays and to allow production runs of aftermarket items. As noted in Duenyas et al., (1996), although the JIT literature widely espouses the virtues of providing supplier plants with visibility to production schedules, in reality, many suppliers have very poor information about future demand. Duenyas et al. discuss this issue in detail and give examples from several industries. One result of this poor visibility to demand is that suppliers often have to carry significant amounts of inventory to meet the demands of their customers "just-in-time". In this paper, we consider an additional reason for the supplier to choose to carry inventory, namely, to allow acceptance and production of more

profitable orders.

In contrast to the OEM demand, the aftermarket is characterized by higher profit margins but more occasional and limited demand. A manufacturing facility may produce items for a single OEM contract but many different aftermarket items. The OEM contract is served in a make-to-stock fashion while the uncertain arrival pattern and the very large variety of aftermarket items necessitates make-to-order service. Prices are fairly standard in the aftermarket; quick service is the primary dimension of competition. Notably the manufacturer is not contractually required to accept an aftermarket order; an order can be either refused or sourced from another firm.

Several interesting topics emerge from this situation. A long-term strategic question is the optimal proportion of production capacity that the firm should dedicate to the OEM segment. Once the firm commits to a certain volume of OEM production, how should it decide whether to take on additional work? Finally, how should a firm decide to sequence work? In this paper, we provide simple models to gain insight into the nature of these problems.

We first address these problems in the context of a simple multi-class $M/M/1$ queue. The first class of jobs (representing OEM jobs), are made-to-stock. Demands arrive randomly and there is a penalty if demand is not satisfied immediately. The second class of jobs represents possible aftermarket orders. These can be accepted or rejected by the firm depending on the state of the system. The firm also has to decide whether to idle, or to process a job of the first or second class at a given point in time. Thus, we are addressing the joint problem of admission control/sequencing in a make-to-order/make-to-stock $M/M/1$ queue. (Similar analysis can be performed at the cost of significant notational complexity, without additional insight, for multi-class $M/G/1$ queues; as in other recent papers focusing on structural insight including Veatch and Wein (1996) and Ha (1993, 1994, 1995), we prefer the $M/M/1$ case. However, in Section 5, we numerically explore the effects of lower variability in processing or demand arrival times by allowing Erlang distributions).

There is a rich literature on admission control and sequencing although most papers consider these problems in isolation and not jointly. Stidham (1985) reviews the literature on customer admission to single item make-to-order queues. In a recent paper, Stidham and

Weber (1993) review the literature on control of arrivals, routing control and scheduling for a network of queues. In the routing-control models, arrivals are either immediately routed to several parallel servers or join a common buffer and are routed to one of the queues only when the job reaches the head of the queue (see Stidham and Weber for a review of this literature). There has been a significant amount of recent interest in control and performance evaluation of single server systems that produce multiple classes of products. While past work has focused on the optimality of simple index rules for make-to-order systems where set-ups are not required to switch from one class to another (e.g., Varaiya et al. (1985), Walrand (1988) and the references therein), more recent work has focused on systems where a set-up is required to switch from one product to another and the dynamic stochastic economic lot sizing problem (SELSP) arising from the production of multiple make-to-stock products. Duenyas and Van Oyen (1996) and Reiman and Wein (1994) focus on control of make-to-order systems with set-ups. Gallego (1990) and Bourland and Yano (1995) develop heuristic policies for SELSP. Anupindi and Tayur (1994) develop a simulation-based approach to obtain effective base-stock policies, and Federgruen and Katalan (1996) analyze the performance of cyclic base-stock policies for this problem. Other recent papers on the SELSP include Markowitz et al. (1995), Sox and Muckstadt (1995), and Qiu and Loulou (1995), and Ha (1993). However, in all of these papers, the only problem addressed is the dynamic production/sequencing problem; admission control is not considered.

Several recent papers have considered joint admission control and production decisions in single product settings. Li (1992) considers a firm producing a single product and explores the economic environmental factors that lead the firm to hold inventory. Ha (1994) considers stock rationing in a single-item, make-to-stock production system with several demand classes and lost sales, while Ha (1995) considers the case with backordering. Our work is differentiated from these papers in that we consider customer classes that require different products, whereas in these papers all customers require the same product, and therefore we have to address the resulting sequencing problem jointly with the admission control problem.

Duenyas (1995) considers the joint problem of quoting due dates and sequencing jobs

4

in a make-to-order queue with multiple customer classes, where a customer's probability of placing orders depends on the due date quoted. Thus, admission control can be indirectly achieved by quoting a very high due date when more jobs are not desired. However, Duenyas (1995) mainly focuses on developing heuristic solutions and does not characterize optimal policy structure. Also, our paper considers a make-to-stock/make-to-order system.

Finally, recent work has focused on performance evaluation of make-to-stock/make-to-order systems as well as on the problem of how a firm decides whether to operate in a make-to-stock or make-to-order mode. Nguyen (1995) develops fluid and diffusion approximations for the performance of a make-to-order/make-to-stock system with FIFO service discipline. Federgruen and Katalan (1994) use the framework of a cyclic polling system to address the decision of how many products to make to order and how many to make to stock. Their analysis can be used by a firm to decide which products to offer, whether to make them to stock or to order and in what annual quantities. Our paper differs in that we assume the decision of which product to make to stock or to order has already been made (in the environment described above, this decision is severely constrained by contractual considerations for the OEM market, and the difficulty of making to stock the huge variety of products that is demanded by the aftermarket demand segment) and we allow for dynamically deciding which particular orders to accept/reject and which class of jobs to produce next.

The rest of this paper is organized as follows. In Section 2, we introduce the notation and formulation of the problem. In Section 3, we completely characterize the structure of the optimal admission control and sequencing policies. These policies are characterized by monotonic switching curves. We then consider how changes in problem parameters are reflected in the switching curves and overall profitability. Since optimal policies are rather complex, in Section 4, we describe a simple, and easily implementable admission control and sequencing policy and show that, for a wide variety of examples, this policy performs very well. We then demonstrate the importance of considering all three decisions that the firm faces simultaneously, and show how making these decisions without considering their interdependence can result in significant loss of profitability. In Section 5, we explore the effects of lower variability in the demand arrival or production processes on profits as well

as the optimal level of OEM demand the firm should commit to. Section 6 concludes the paper.

2 Problem Formulation

We consider the optimal control of a single server queue which produces two types of products. Demand for product of type i, $(i = 1, 2)$, arrives according to a Poisson distribution with rate λ_i. (We allow Erlang distributions in Section 5). Type 1 products are made to stock, and demand for them is satisfied from inventory. It is not possible to backorder demand for type 1 products; if there is no type 1 product in stock when a demand occurs, the firm satisfies the demand by purchasing the product from another supplier at an additional cost (penalty) of π for each such occurrence.

Type 2 products are produced to order, and the firm has the option of to accept or reject each type 2 order. (We will later generalize to the case where there are multiple classes of products made to order which are only differentiated by price; for simplicity of presentation, we first present the case where there is a single product type which is made-to-order). Each accepted order generates a revenue (alternatively, margin) of R_2. Inventory costs are incurred at the rate of c_1 per unit per unit time for each unit of type 1 product in stock, and at the rate of c_2 per unit time for each unit of type 2 product in process (c_2 may also serve as a proxy for the cost to the firm of delaying production of type-2 orders and thus becoming uncompetitive in that market). We assume that production of a unit of either type 1 or type 2 product takes an exponentially distributed amount of time with mean $\frac{1}{\mu}$. (In Section 5, we allow Erlang processing time distributions and permit type-1 and type-2 production to have different processing time distributions). We further assume that preemptions are permitted, and that no setup is required to switch from producing one type of product to another.

The set of decision epochs correspond to the set of all arrival epochs, service completion epochs, and instances of idling. A policy specifies at each decision epoch that the server idle or produce a unit of type 1 or type 2. Furthermore, in decision epochs corresponding to the arrival of a customer for a type 2 product, a decision must be made whether to accept the customer's order. Accepting the order generates revenue but also increases the rate at which

6

waiting costs accrue. The objective is to find a policy that maximizes the average profit per unit time, where profit is the revenue from type 2 orders minus the costs associated with the inventory costs for type 1 and type 2 products and the penalties incurred when type 1 demands can not be met from stock. (Note that all demand for type 1 products is satisfied. Therefore, adding the revenue associated with type 1 products, R_1, to the objective function has no effect on the optimal control policy).

We can formulate the optimal admission and production control problem as a Markov Decision Process and characterize the structure of optimal order acceptance and sequencing policies. We define the state (n_1, n_2) as the number of units of type-1 inventory in stock and the number of units of type 2 products in process. We let $v(n_1, n_2)$ be the relative value function of being in state (n_1, n_2) and $g(\lambda_1)$ denote the average reward per transition (where a transition is an arrival or service completion). Then, using uniformization, we can write

$$
\begin{aligned}
g(\lambda_1) + v(n_1, n_2) = \frac{1}{\Lambda}\Big\{ &- c_1 n_1 - c_2 n_2 + \lambda_1 [v(n_1 - 1, n_2) \cdot I_{(n_1 > 0)} + (v(0, n_2) - \pi) \cdot I_{(n_1 = 0)}] \\
&+ \lambda_2 \max[v(n_1, n_2 + 1) + R_2, v(n_1, n_2)] \\
&+ \mu \max[v(n_1 + 1, n_2), v(n_1, n_2 - 1) \cdot I_{(n_2 > 0)} + v(n_1, n_2) \cdot I_{(n_2 = 0)}] \Big\}
\end{aligned}
$$
$$(2.1)$$

where $I_{(.)}$ denotes the indicator function, and $\Lambda = \lambda_1 + \lambda_2 + \mu$.

In eqn. (2.1), the cost terms $(\frac{1}{\Lambda}\{-c_1 n_1 - c_2 n_2\})$ represent the expected holding and waiting costs per decision epoch and the terms multiplied by λ_1 represent transitions and penalties associated with the arrival of a type 1 customer, the terms multiplied by λ_2 represent transitions and revenues generated by the arrival of a type 2 order, and the terms multiplied by μ represent transitions generated by a service completion opportunity. We note that since preemptions are allowed, idling when $n_2 > 0$ can not be optimal, and $g(\lambda_1)$ is the profit per transition when the manufacturer faces a long-run class-1 demand rate of λ_1. Since transitions occur with rate Λ, $g(\lambda_1)\Lambda$ represents the profit per unit time.

The second problem that the firm faces is that of choosing the optimal level of λ_1. In the situation we described in the previous section, the firm makes this decision only once (for example, when signing a contract to provide products for the OEM market). The problem

of choosing the optimal level of λ_1 is formulated as

$$\max_{\lambda_1} g(\lambda_1)(\lambda_1 + \lambda_2 + \mu) + R_1\lambda_1 \qquad (2.2)$$

Equation (2.2) shows the interdependence between a firm's decision of the optimal level of λ_1 that the firm will set and the levels of λ_2, the production rate as well as the admission control and sequencing policies that the firm will use. In the situation we described in the introduction, the level of less profitable OEM demand that the firm is willing to commit to providing depends on the level of more profitable aftermarket demand and how much capacity the firm has to satisfy either type of demand. Therefore, the choice of λ_1 is interrelated to the solution of (2.1). It is therefore important to understand the structure of the optimal solution to (2.1) and how this structure is affected by problem variables. We first focus on the solution to this problem.

3 The Structure and Monotonicity of the Optimal Admission Control and Sequencing Policies

In this section, we focus on the problem described by (2.1) and characterize the structure of the optimal admission control and sequencing policy that the firm would use given that the optimal level of λ_1 has already been fixed. The main result of this section is the following theorem which completely describes the structure of the optimal policy.

Theorem 1 *The optimal production policy is defined by a switching curve $h(n_2)$ such that for $n_1 > h(n_2)$, the optimal policy is to produce type-2 items if $n_2 > 0$ and to idle if $n_2 = 0$; for $n_1 \leq h(n_2)$ the optimal policy is to produce type-1 items. We refer to this switching curve as the* **production threshold curve**. *Furthermore, $h(n_2)$ is nonincreasing in n_2.*

The optimal type-2 order acceptance policy is also defined by a switching curve, $f(n_2)$ such that if $n_1 > f(n_2)$, the optimal policy accepts type-2 orders but otherwise rejects them. We refer to this switching curve as the **acceptance threshold curve**. *Furthermore, $f(n_2)$ is nondecreasing in n_2.*

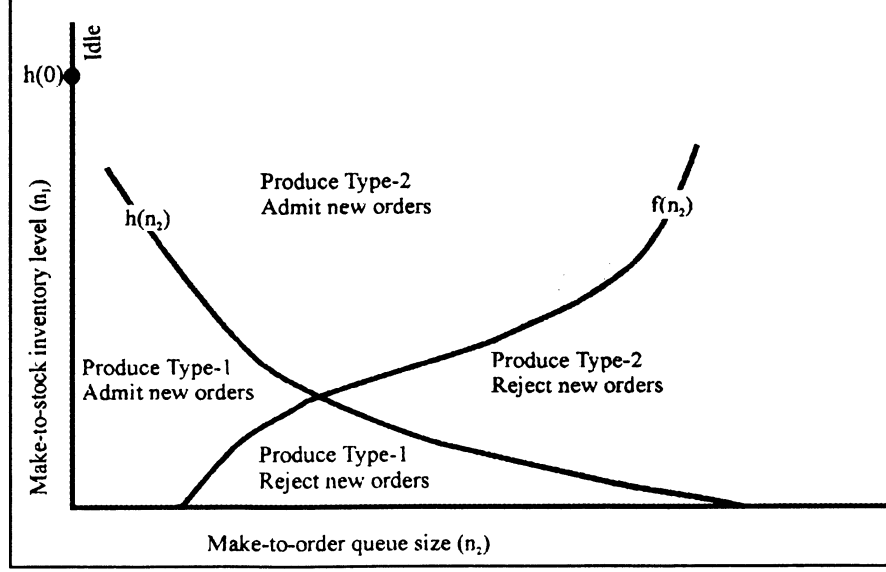**Proof:** The proof is provided in the Appendix.

Figure 1: The structure of the optimal admission and sequencing policies

The structure of the optimal policy in a typical problem is illustrated in Figure 1. The optimal policy is characterized by four regions corresponding to the combinations of producing either type 1 or type 2 items and either accepting or rejecting arriving type 2 customers. We conjecture that Theorem 1 holds even when the production rate for products 1 and 2 are different. However, although we have not been able to find any numerical counterexamples to this conjecture in the very large number of examples we have solved, at this time we are not able to show that Theorem 1 holds without the assumption that the production rate for products 1 and 2 is the same.

Having characterized the structure of the optimal policy for (2.1), we next proceed to analyze how the optimal profit and admission and sequencing decisions change as a function of the problem parameters. Our first result, characterizing the monotonicity of the expected profits as a function of the problem parameters is intuitive.

Theorem 2 *The optimal average profit per unit time* $g(\lambda_1)\Lambda$ *of (2.1) is increasing in* μ, λ_2, R_2, *and decreasing in* π, c_1, c_2, λ_1.

Proof: See appendix.

We note that the formulation in (2.1) does not consider revenues for type-1 products but considers overall profits from acceptance of type-2 products and sequencing decisions once the choice of λ_1 has been made. Therefore, the expected profit in (2.1) is decreasing

in λ_1 since the higher the value of λ_1, the fewer type-2 orders the firm can accept. Having characterized the monotonicity of the average profit per unit time in (2.1), we next proceed to characterize the monotonicity of the switching curves with respect to the input parameters. Consider two instances of the admission control and sequencing problem described by (2.1). To differentiate the second instance from the first one, we use the prime ($'$) symbol for costs and switching curves in the second case.

Theorem 3   *1. Suppose that $\lambda_1 = \lambda_1'$, $\lambda_2 = \lambda_2'$, $c_1 = c_1'$, $c_2 = c_2'$, $\pi = \pi'$, $\mu = \mu'$, and*

$R_2' > R_2$. *Then, $f'(n_2) \leq f(n_2)$ for all n_2, and $h'(n_2) \leq h(n_2)$ for $n_2 \geq 1$, and*
$h'(0) \geq h(0)$.

2. Suppose that $\lambda_1 = \lambda_1'$, $\lambda_2 = \lambda_2'$, $c_1 = c_1'$, $c_2 = c_2'$, $\pi = \pi'$, $R_2' = R_2$, and $\pi' > \pi$. Then,
$f'(n_2) \geq f(n_2)$ *and $h'(n_2) \geq h(n_2)$.*

Proof: see appendix

In words, the second part of Theorem 3 states as π gets larger, it may become optimal to switch from accepting type-2 orders to rejecting orders in any given state and to switch from producing type-2 orders to type-1 orders in states with $n_2 > 0$, and finally to switch from idling to producing type-1 orders in states with $n_1 = 0$. This result can be explained intuitively. As stocking out of type-1 units now costs more, the policy switches towards giving higher priority to type-1 orders, keeping more type-1 units in stock and accepting fewer type-2 orders.

The first part of Theorem 3 states that as R_2 gets larger, it may become optimal to switch from rejecting type-2 orders to accepting them in any given state, to switch from producing type-1 orders to producing type-2 orders in states with $n_2 > 0$, and to switch from idling to producing type-1 orders in states with $n_1 = 0$. As type-2 orders become more profitable when R_2 is increased, the policy tends to switch to accepting more of them and to give higher priority to their production rather than the production of type-1 orders. However, when there are no type-2 orders to work on, the policy works towards building a larger finished goods inventory of type-1 units so that it can have the opportunity to accept and work on type-2 units when orders for them arrive. Figure 2 displays how the switching curves shift as a function of a change in R_2.
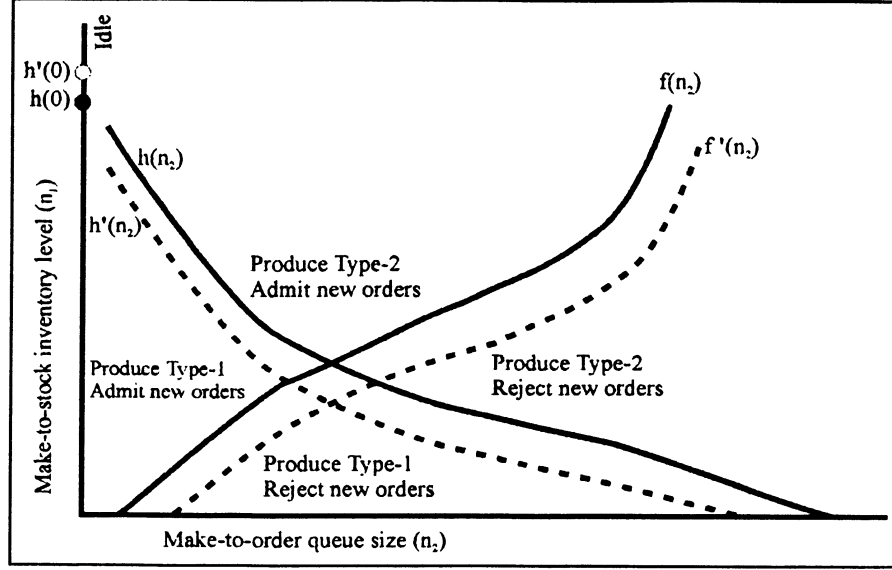
Figure 2: Change in policy as a result of an increase in R_2

Although the switching curves, f and h have nice monotonicity properties with respect to π and R_2, they are not necessarily monotonic with respect to the other parameters. We show this by way of counterexample. Consider the following problem which we will refer to as "base case". The "base case" has the following values: $R_2 = 20$, $\pi = 35$, $c_1 = 0.5$, $c_2 = 2$, $\lambda_1 = 1.5$, $\lambda_2 = 1$ and $\mu = 2$. We will analyze the optimal decisions in states $S_1 = (2, 2)$ and $S_2 = (1, 6)$ under the base case and other cases. Under the base case, the optimal production decision is to produce type-1 items in S_1 and type-2 items in S_2. Changing only λ_1 to 0.75 from 1.5, but keeping all the other parameters the same, will in the optimal policy result in producing type-2 items in S_1 and type-1 items in S_2. Similarly, changing only μ to 5 from 2 will result in once again the optimal policy producing type-2 items in S_1 and type-1 items in S_2. These examples clearly demonstrate that the switching curves are not necessarily monotonic with respect to all parameters.

We conclude this section by pointing out that a simple extension of our model also has a nice structure. In (2.1), we considered a problem with only two classes, where one is the make-to-stock class and the second is the make-to-order class. An obvious extension is to consider a single make-to-stock but an arbitrary number of make-to-order classes. Assume type-1 orders are the make-to-stock orders and type-2 through N orders are make-to-order. If we further assume that all products have the same mean production rate and that all
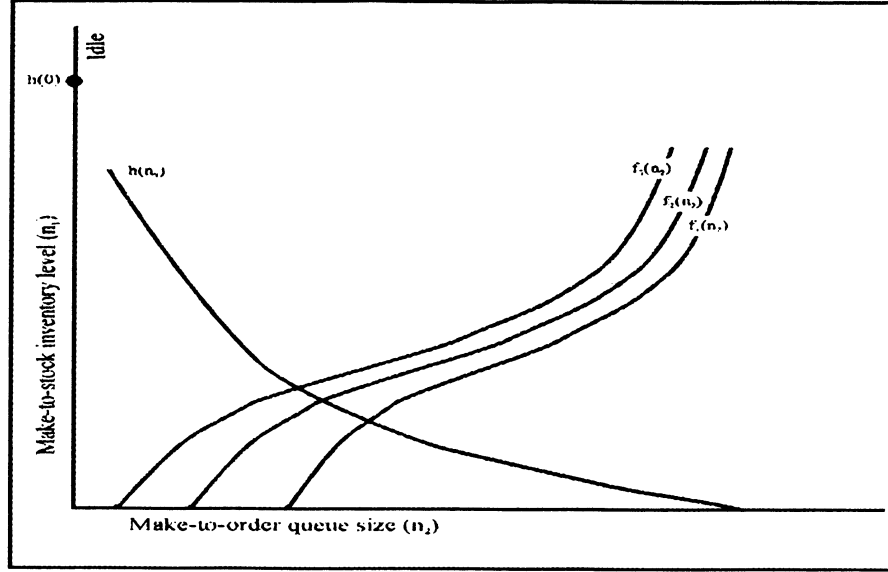
Figure 3: The structure of the optimal admission and sequencing policies- multiple make-to-order types

the make-to-order products have the same waiting cost c_2, and are numbered such that $R_2 > R_3 > \ldots > \ldots R_n$, we are again able to characterize the structure of the policy. First of all, we note that since waiting costs for all make-to-order products are the same, once we accept a make-to-order product, we do not need to keep differentiating it from other make-to-order products. Thus, we can represent the state once again as (n_1, n_2) where n_1 is the number of units of type-1 in stock, and n_2 denotes the sum of type-2 through N orders accepted but not yet processed. Then using the same approach used for proving Theorem 1, it is easy to show that there exists a production threshold curve $h(n_2)$ such that if $n_1 < h(n_2)$ then the optimal policy is to produce a type-1 unit and otherwise to serve any of the make-to-order units waiting. Similarly, for each make-to-order class there exits an acceptance switching curve $f_i(n_2)$ such that if $n_1 > f_i(n_2)$ then a type-2 order will be accepted, and it will be rejected otherwise. Furthermore, $h(n_2)$ and $f_i(n_2)$ are monotonic with respect to n_2 as in Theorem 1. Finally, it is easy to show that $f_i(n_2) < f_j(n_2)$ for $j > i$. Figure 3 demonstrates the structure of the optimal policy in this case. We note that all of the production rates and waiting costs being equal is a crucial assumption in this case since if those strong assumptions did not hold, one would need to differentiate between types thereby increasing the dimensionality of the state space.

12

We close this section by noting that the firm can find the optimal level of λ_1 by solving (2.1) by discretizing the interval between $[0, \mu)$ and solving (2.1) for each different level of λ_1. The value obtained for $g(\lambda_1)$ can then be input into (2.2) and the value of λ_1 that maximizes the value of (2.2) can be picked. We note that this procedure is very different from procedures used by most firms, including the one that motivated this study. The above-described way of setting λ_1 levels takes into account the opportunities for satisfying aftermarket (type-2) demand as well as optimal production and acceptance policies. For example, a typical firm will have separate and relatively uncoordinated OEM sales, aftermarket sales and operations. In our numerical study in the next section, we focus on the importance of coordinating all three functions by showing how even a firm which is uncoordinated with respect to one of these functions (say, OEM sales) operates significantly suboptimally compared to a firm which coordinates all three functions.

4   Numerical Study

The optimal policies described in the previous section are rather complex. In particular, the production and acceptance policies are defined by switching curves which make their actual implementation problematic. However, they also point to the importance of considering these decisions simultaneously, i.e., "coordinating" these decisions. In this section, we first explore the performance of a simpler, more implementable policy for production and acceptance decisions. However, this policy still considers these decisions simultaneously. We then give an example of a relatively sophisticated, but somewhat uncoordinated, policy for controlling the three decisions that ultimately fails.

We first focus on the sequencing and acceptance decisions after λ_1 is set. Consider the following simple policy: type-2 orders are accepted if $n_2 < N_2$, and production of type-1 inventory is allowed so long as $n_1 < N_1$. Given this basic structure, two questions remain: 1) What are the best values of N_1 and N_2, and 2) how does one decide whether to produce type-1 or type-2 units at any point in time? We first consider a policy which gives total priority to the production of type-1 items as long as $n_1 < N_1$; for any given N_1 and N_2, this gives rise to a Markov chain with an associated profit that is easily calculated. We perform a two-dimensional search for N_1 and N_2. We then consider a policy which gives

total priority to the production of type-2 production as long as $n_2 > 0$. We once again find the best N_1, and N_2 values and the profit associated with these values under this sequencing rule. We then pick the sequencing rule and corresponding N_1, N_2 values resulting in the highest profit. Although this is computationally inefficient, it is certainly a very simple policy to describe and implement in practice. In actual operation, this policy appears to be uncoordinated at first sight since in deciding whether to accept or reject a type-2 order, the sales department does not need to know the level of current inventory. However, we note that the operating parameters of N_1 and N_2 were in fact set in a coordinated fashion.

Table 1 compares the profits obtained by the optimal policy and by the simple heuristic policy described for 22 examples. Example 1 represents the base case, and in Examples 2 through 16, we systematically increased or decreased one of the problem parameters to test the performance of the described policy under a variety of conditions including high versus low type-1 or type-2 demand, and high or low holding and penalty costs. We also added six other examples to cover other cases. The results in Table 1 clearly show that this simple, implementable policy performs very well. The average difference between the performance of the optimal and heuristic policies was only 1.8 %. Although the computation of N_1 and N_2 is somewhat time consuming, in practice this would need to be done only once; and the policy is very easy to describe to both manufacturing and sales departments. The last column of Table 1 indicates which class the heuristic gives priority to. It is interesting to note that in the majority of the cases, the heuristic gives priority to the production of type-2 orders.

However, we note that the above policy still considered sequencing and acceptance decisions simultaneously in setting the parameters for the heuristic. As we noted in the previous section, in many companies these decisions as well as the decision of how much OEM demand to contract for are uncoordinated. In fact, many companies, including the one with which we are most familiar, have separate divisions for OEM and aftermarket sales, but common manufacturing facilities where both types of products are produced. Furthermore, the divisions have an incentive to maximize division profits. In many other firms, OEM contracting is performed first and the aftermarket sales are subsequently considered. We show through a simple example how such uncoordination may be very costly. Consider

Example	R_1	R_2	π	c_1	c_2	λ_1	λ_2	μ	Optimal	Heuristic	% dif.	Priority
1	10	10	25	1	2	1	1	2	11.57	11.35	1.9%	2
2	10	10	50	1	2	1	1	2	10.50	10.23	2.6	2
3	10	50	25	1	2	1	1	2	49.74	49.04	1.4	2
4	10	10	25	1	5	1	1	2	10.00	9.86	1.4	2
5	5	10	25	1	2	1	1	2	6.58	6.35	3.4	2
6	50	10	25	1	2	1	1	2	51.58	51.35	0.4	2
7	10	5	25	1	2	1	1	2	7.97	7.81	2.0	2
8	10	10	4	1	2	1	1	2	15.03	15.02	0.0	2
9	10	10	100	1	2	1	1	2	9.48	9.16	3.4	2
10	10	10	25	.5	2	1	1	2	13.51	12.82	5.1	1
11	10	10	25	2	2	1	1	2	9.04	8.62	4.7	2
12	10	10	25	1	.5	1	1	2	13.51	13.40	0.8%	2
13	10	10	25	1	2	.5	1	2	9.66	9.32	3.5	2
14	10	10	25	1	2	1.75	1	2	12.54	12.48	0.5	2
15	10	10	25	1	2	1	0.5	2	9.44	9.16	3.0	2
16	10	10	25	1	2	1	5	2	14.73	14.37	2.4	2
17	5	25	500	2	2	1	3	10	72.64	72.65	0.0	1
18	5	25	500	5	2	1	3	10	64.92	64.92	0.0	1
19	5	25	500	.5	2	1	3	10	77.08	76.86	0.3	1
20	5	25	32	.5	4	4	3	6	57.37	57.25	0.2	2
21	50	25	10	10	5	5	4	10	320.3	318.8	0.5	2
22	10	10	25	1	2	1.5	1	2	12.61	12.52	0.8	2

Table 1: Comparison of Optimal and Heuristic Admission and Sequencing Policies

a firm's OEM sales department which decides how much OEM demand to accept (i.e., the optimal level of λ_1), without taking into consideration the existence of aftermarket demand. Assuming this division was doing everything else optimally, it would solve (2.1) and (2.2) by setting $\lambda_2 = 0$, and find the optimal level of λ_1. Furthermore, assume that the aftermarket sales and scheduling departments are then notified of the choice of λ_1 and make their acceptance and sequencing decisions optimally and in coordination by solving (2.1) (with the value of λ_1 given to them) and implementing the optimal policy for sequencing and aftermarket order acceptance. In this example, all decisions except the initial decision of the choice of λ_1 by the OEM sales department are coordinated and optimal.

We numerically tested the performance degradation of this behavior on several of the examples from Table 1. In Example 20, the OEM sales department would actually choose a λ_1 value of 4.60 if it behaved in the above described manner, whereas the optimal value of λ_1 is 2. Even though all decisions are made optimally thereafter, the uncoordinated behavior of this one department would result in a 38 % decrease in firm profits! We note that in Example 20, aftermarket sales are nearly five times as profitable as a OEM sales. However, even in Example 1, where aftermarket and OEM margins are equal, the decrease in profits would be nearly 6 %. Not unexpectedly, we have observed that as aftermarket margins become significantly higher than OEM margins, (exactly the scenario observed in industry), the profit discrepancy increases. Finally, we note that in these cases, overestimating the optimal value of λ_1 usually results in a more significant profit degradation than underestimating it.

5   Extension to Erlang Arrival and Service Distributions

In the previous sections, we assumed that all distributions were exponential. However, in some cases, signing a large contract with a firm might also result in a less variable demand pattern. It is easy to extend our formulation to handle Erlang distributions for interarrival and service times. This allows us to handle service and arrival distributions which have significantly less variability than the exponential. For example, if we assume that the arrival process of type-1 demand is Erlang-z, with all other distributions still exponential, we can define $v(n_1, n_2, k)$ to be the relative value function of having n_1 type-1 units in stock, n_2 type-2 orders, and k phases until the next type-1 arrival, where each phase is of duration
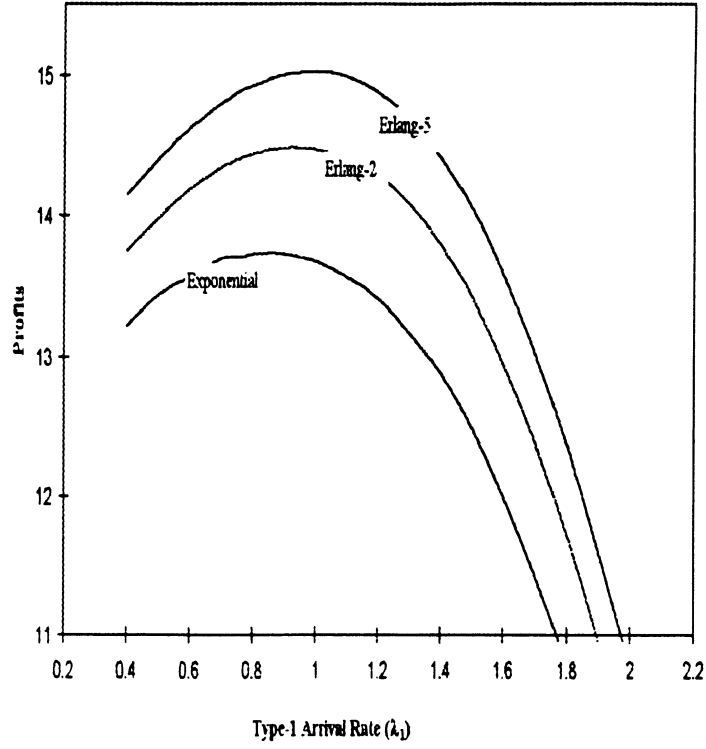
16

Figure 4: Optimal profits as a function of λ_1 with changes in arrival variability

$1/(k\lambda)$. We can then write the following recursive optimality equation,

$$
\begin{aligned}
g(\lambda_1) + v(n_1, n_2, k) = \frac{1}{\Lambda}\Big\{ &- c_1 n_1 - c_2 n_2 + z\lambda_1[v(n_1, n_2, k - 1) \cdot I_{(k>1)} \\
&+ v(n_1 - 1, n_2, z) \cdot I_{(k=1, n_1>0)} + (v(0, n_2, z) - \pi) \cdot I_{(k=1, n_1=0)}] \\
&+ \lambda_2 \max[v(n_1, n_2 + 1, k) + R_2, v(n_1, n_2, k)] \\
&+ \mu \max[v(n_1 + 1, n_2, k), v(n_1, n_2 - 1, k) \cdot I_{(n_2>0)} + v(n_1, n_2, k) \cdot I_{(n_2=0)}]\Big\}
\end{aligned}
$$

$$(5.3)$$

where $g(\lambda_1)$ denotes the average reward per transition as before, and $\Lambda = z\lambda_1 + \lambda_2 + \mu$. In a similar manner, we can easily extend the formulation to allow Erlang processing times for type-1 or type-2 production. This also allows us to formulate situations where type-1 and type-2 items do not have the same processing time distribution.

We are interested in how improvements in arrival or production process variability affect the firm's profitability and the size of contracts the firm would be willing to sign with an OEM. Intuition would suggest that as the arrival process of type-1 demand becomes less variable, perhaps through better customer-supplier communication or contractual agree-

ment, the firm's profits would increase and the firm would also be willing to dedicate more of their capacity to that customer. Similarly, the large cumulative production volumes resulting from a long-term relationship with the OEM customer might lead to a more efficient, less varible production process for the product demanded by the OEM. It is also common practice for some firms, such as Toyota, to assist their suppliers in variability reduction. Intuition would again suggest that such improvements in type-1 production process would have the same results as the reduction in the variability of the type-1 demand arrival process. Numerical examples have verified our intuition. We give two examples of this behavior.

As a base case, consider the situation where all interarrival and processing times have exponential distributions and $R_1 = 8, R_2 = 15, \pi = 25, c_1 = 1, c_2 = 2, \lambda_2 = 1$, and $\mu = 2$. Figure 4 shows the impact of decreases in the variability of type-1 demand arrivals. In this figure, we plotted the profit of the firm as a function of its choice of λ_1 when the type-1 demand arrival process was expontial, Erlang-2, and Erlang-5. As the arrival process becomes less variable, the profit level at any λ_1 increases, and as we expected the optimal level of λ_1 also increases. In particular, the optimal level of λ_1 is 0.85 when the arrival process is exponential and it increases to 0.93 for Erlang-2 and to 0.99 for Erlang-5 distributions. Similarly, Figure 5 demonstrates the effect of reducing type-1 processing time variability. Once again, we plotted the optimal profits as a function of λ_1 for exponential, Erlang-2 and Erlang-5 type-1 processing time distributions (the processing time for type-2 was fixed to be exponential with the same mean for all examples. Once again, we note that as the production process becomes less variable, both profits and the optimal value of λ_1 increase. In particular, the optimal λ_1 changes from 0.85 to 1.00 for Erlang-2 and 1.02 for Erlang-5 distributions.

These results clearly show the beneficial effects to the supplier of decreasing arrival or processing time variability. In fact, as lower variability leads to higher profits at all levels of λ_1, a supplier might be willing to offer lower prices to the OEM in exchange for guarantees in demand variability or improvements in the design of products which lead to demonstrated decreases in production variability.
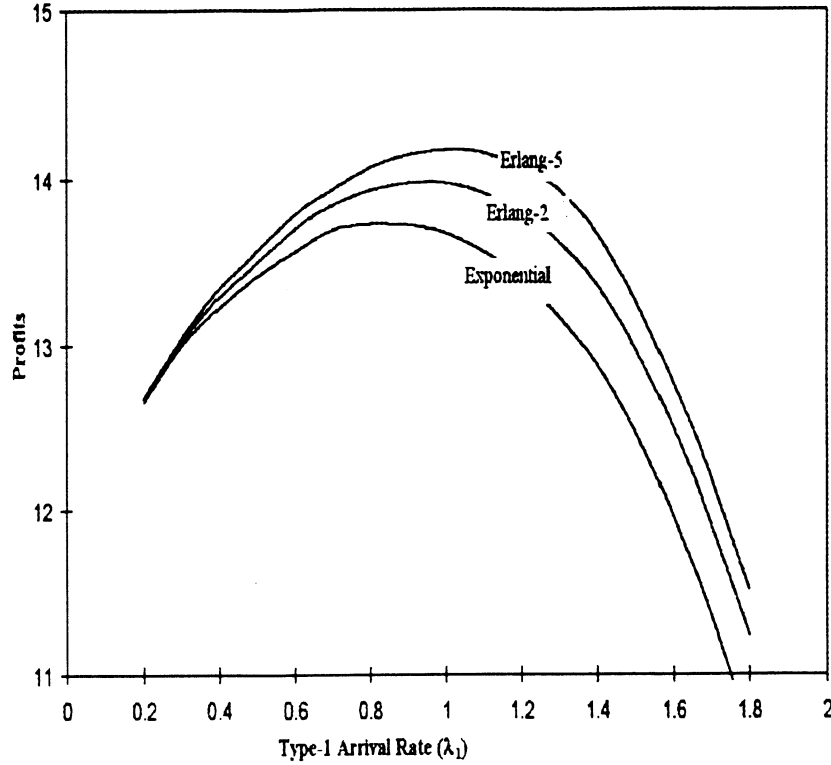
Figure 5: Optimal profits as a function of λ_1 with changes in process variability

6 Conclusions and Further Research

In this paper, we addressed the joint admission control and sequencing decisions faced by a firm producing products for multiple segments. We were able to characterize the structure and sensitivity of optimal policies in the context of a multi-class M/M/1 queue. Since the structure of the optimal policy is rather complex, we explored the performance of a simpler policy and also presented examples showing the importance of simultaneously considering these decisions.

Many open problems remain to be explored. First, our formulation easily extends to general processing times with different rates for different products and we conjecture that our structural results continue to hold. Second, in many cases in practice, switching from one type of product to another requires a set-up; exploring how optimal admission control and sequencing decisions are affected by the presence of setups is a very interesting question. For example, is a firm more likely to accept orders for a product for which its machines are already set-up if set-up are very long? Third, we have so far explored these questions only

in the context of a single server queue. It would be interesting to consider more complicated multi-server or network settings.

Appendix - Proofs of Theorems

Proof of Theorem 1:

Consider a value iteration algorithm to solve for the optimal policy in which $v_0(n_1, n_2) = 0$ for every state (n_1, n_2) and:

$$
\begin{aligned}
v_{k+1}(n_1, n_2) = \frac{1}{\Lambda}\Big\{ &- c_1 n_1 - c_2 n_2 + \lambda_1[v_k(n_1 - 1, n_2) \cdot I_{(n_1 > 0)} + v_k(0, n_2) \cdot I_{(n_1 = 0)} - \pi \cdot I_{(n_1 = 0)}] \\
&+ \lambda_2 \max[v_k(n_1, n_2 + 1) + R_2, v_k(n_1, n_2)] \\
&+ \mu \max[v_k(n_1 + 1, n_2), v_k(n_1, n_2 - 1) \cdot I_{(n_2 > 0)} + v_k(n_1, n_2) \cdot I_{(n_2 = 0)}] \Big\}
\end{aligned}
$$
(6.4)

Here, $v_k(n_1, n_2)$ can also be viewed as the optimal value function when the problem is terminated after k transitions. In order to prove Theroem 1, we first show that (2.1) has a well-defined solution, and the $v_k(n_1, n_2)$ values in (6.4) converge to $v(n_1, n_2)$ in (2.1).

Lemma 1 *There exist bounded functions, $v(n_1, n_2)$, $n_1, n_2 \geq 0$, and a bounded constant $g(\lambda_1)$ satisfying (2.1). Furthermore, $v_k(n_1, n_2) \to v(n_1, n_2)$ as $k \to \infty$.*

Proof: By Theorem 2.2 of Hernandez-Lerma (1989), the following are sufficient conditions: 1) compact action spaces, 2) bounded one-period rewards, 3) expected future reward functions continuous in actions, 4) a "uniformly accessible state" which can be reached from any state with nonzero probability. Finiteness of the action spaces guarantee conditions (1) and (3). Without loss of optimality, we can add the constraint that we can never accept a customer when $R_2 < c_2 n_2/\Lambda$, and we can not produce another unit of type 1 when $\pi < c_1 n_1/\Lambda$. These are both straightforward. (For example, if $c_1 n_1/\Lambda > \pi$, this indicates that the expected amount of holding cost incurred by holding the n_1 units of type-1 products in stock until the next transition is greater than the penalty cost that would be incurred if no type-1 products were available and the next event was the arrival of a demand for a type-1 product). Therefore, we can convert the problem (2.1) to one with finite state spaces, and the finiteness of c_1, c_2, R_2 guarantees condition (2). Finally, we note that the

state (0, 0) is reachable from any state. This immediately follows from the fact that idling is not allowed as long as $n_2 > 0$ $\square$.

To prove Theorem 1, we will analyze (6.4). We first note that the following conditions on the value function of (6.4) are sufficient for the structure of the policy to hold at every iteration of the value iteration algorithm.

C1a) $v_k(n_1, n_2 + 1) - v_k(n_1, n_2)$ is non-decreasing with n_1 and non-increasing with n_2,

C1b) $v_k(n_1 + 1, n_2) - v_k(n_1, n_2 - 1)$ is non-increasing with both n_1 and n_2, and

C1c) $v_k(n_1 + 1, n_2) - v_k(n_1, n_2)$ is non-increasing with n_1 and non-decreasing with n_2.

To see this, we note that if it is optimal to accept a customer in state (n_1, n_2) (i.e. $v_k(n_1, n_2 + 1) + R_2 \geq v_k(n_1, n_2)$), then condition (C1a) implies that $v_k(n_1 + 1, n_2 + 1) + R_2 \geq v_k(n_1 + 1, n_2)$, and it is optimal to accept customers in state $(n_1 + 1, n_2)$. Similarly, if it is optimal to reject a customer in state (n_1, n_2) then is also optimal to reject a customer in state $(n_1, n_2 + 1)$ by the fact that $v_k(n_1, n_2 + 1) - v_k(n_1, n_2)$ is non-increasing with n_2. The existence and monotonicity of the production threshold curve is similarly implied by conditions (C1b), and (C1c).

Lemma 2 *For every $k = 0, 1, 2, \ldots$, conditions (C1a, C1b, C1c) hold.*

Proof: $v_0(n_1, n_2) = 0$ for every (n_1, n_2), so the lemma holds trivially in the first iteration. We make the induction assumption that the entire lemma holds at arbitrary iteration k and show that the lemma holds in the $(k + 1)^{st}$ iteration. That is, we show that the structure described by the lemma survives each iteration of the algorithm. For brevity, we present the proof that $v_{k+1}(n_1 + 1, n_2) - v_{k+1}(n_1, n_2 - 1)$ is non-increasing with n_1 for the case of both n_1 and n_2 greater than zero (i.e., the first part of condition C1b). Proofs of the other cases and parts of the lemma are entirely similar. Suppose that $n2 \geq 2$. Relating $v_{k+1}(n_1 + 1, n_2) - v_{k+1}(n_1, n_2 - 1)$ to terms from the kth iteration by applying equation

(6.4), it is sufficient that:

$$\frac{\lambda_1}{\Lambda}[v_k(n_1, n_2) - v_k(n_1 - 1, n_2 - 1)]$$

$$+\frac{\lambda_2}{\Lambda}\{max[v_k(n_1 + 1, n_2 + 1) + R_2, v_k(n_1 + 1, n_2)] - max[v_k(n_1, n_2) + R_2, v_k(n_1, n_2 - 1)])\}$$

$$+\frac{\mu}{\Lambda}\{max[v_k(n_1 + 2, n_2), v_k(n_1 + 1, n_2 - 1)] - max[v_k(n_1 + 1, n_2 - 1), v_k(n_1, n_2 - 2)]\}$$

$$(6.5)$$

is non-increasing with n_1. A sufficient condition for (6.5) is for all three of the following conditions to be satisfied:

1. $v_k(n_1, n_2) - v_k(n_1 - 1, n_2 - 1)$ is non-increasing in n_1 ,

2. $max[v_k(n_1 + 2, n_2 + 1) + R_2, v_k(n_1 + 2, n_2)] - max[v_k(n_1 + 1, n_2) + R_2, v_k(n_1 + 1, n_2 - 1)] - (max[v_k(n_1 + 1, n_2 + 1) + R_2, v_k(n_1 + 1, n_2)] - max[v_k(n_1, n_2) + R_2, v_k(n_1, n_2 - 1)]) \le 0$, and

3. $max[v_k(n_1 + 2, n_2), v_k(n_1 + 1, n_2 - 1)] - max[v_k(n_1 + 1, n_2 - 1), v_k(n_1, n_2 - 2)]$ is non-increasing in n_1.

Conditions 1 and 3 follow directly from the induction assumption. Condition 1 is true as (C1b) is assumed to hold true for v_k. The terms in condition 3 have possible values of: (1) $v_k(n_1 + 2, n_2) - v_k(n_1 + 1, n_2 - 1)$, (2) $v_k(n_1 + 2, n_2) - v_k(n_1, n_2 - 2)$, (3) zero, or (4) $v_k(n_1 + 1, n_2 - 1) - v_k(n_1, n_2 - 2)$; the non-zero values are non-increasing with n_1 by the induction assumption (C1b).

Showing that condition 2 holds is a bit more involved. We use the acronym LHS to denote the left-hand side of this condition. Since each of the max functions can take one of two values, LHS has 16 possible values. For example, LHS would take the value:

$$v_k(n_1 + 2, n_2) - [v_k(n_1 + 1, n_2) + R_2] - v_k(n_1 + 1, n_2 + 1) + v_k(n_1, n_2)$$

in the case of:

$$\mathbf{v_k(n_1 + 2, n_2 + 1) + R_2} < \mathbf{v_k(n_1 + 2, n_2)}, \; v_k(n_1 + 1, n_2) + R_2 \ge v_k(n_1 + 1, n_2 - 1)$$

$$\mathbf{v_k(n_1 + 1, n_2 + 1) + R_2} \ge \mathbf{v_k(n_1 + 1, n_2)}, \text{ and } v_k(n_1, n_2) + R_2 \ge v_k(n_1, n_2 - 1)$$

Note that the first and third requirements for this case (expressed in boldface) contradict condition (C1a) which is assumed to hold by induction; this case is therefore infeasible.

Of the 16 possible cases, 10 can be eliminated from consideration in this manner. The remaining six are considered individually.

i) $LHS = v_k(n_1 + 2, n_2 + 1) - v_k(n_1 + 1, n_2) - v_k(n_1 + 1, n_2 + 1) + v_k(n_1, n_2) \leq 0$ by condition (C1b).

ii) $LHS = v_k(n_1 + 2, n_2) - v_k(n_1 + 1, n_2 - 1) - v_k(n_1 + 1, n_2) + v_k(n_1, n_2 - 1) \leq 0$ again by condition (C1b).

iii) $LHS = v_k(n_1 + 2, n_2 + 1) - 2v_k(n_1 + 1, n_2) + v_k(n_1, n_2 - 1) \leq 0$ again by condition (C1b).

iv) $LHS = v_k(n_1 + 2, n_2) - 2v_k(n_1 + 1, n_2) + v_k(n_1, n_2) \leq 0$ by condition (C1c).

v) $LHS = v_k(n_1 + 2, n_2) - [v_k(n_1 + 1, n_2) + R_2] - v_k(n_1 + 1, n_2) + v_k(n_1, n_2 - 1)$ One of the requirements for this case is that $v_k(n_1 + 1, n_2) + R_2 \geq v_k(n_1 + 1, n_2 - 1)$. Thus, LHS is less than or equal to $v_k(n_1 + 2, n_2) - v_k(n_1 + 1, n_2 - 1) - v_k(n_1 + 1, n_2) + v_k(n_1, n_2 - 1)$ which is itself less than or equal to zero by condition (C1b).

vi) $LHS = v_k(n_1 + 2, n_2 + 1) - 2v_k(n_1 + 1, n_2) + v_k(n_1, n_2) + R_2$ A requirement for this case is that $v_k(n_1 + 1, n_2 + 1) + R_2 \leq v_k(n_1 + 1, n_2)$. Thus, $LHS \leq v_k(n_1 + 2, n_2 + 1) - v_k(n_1 + 1, n_2) - v_k(n_1 + 1, n_2 + 1) + v_k(n_1, n_2) \leq 0$ by condition (C1b).

Now suppose that $n_2 = 1$. The first two conditions remain unchanged, but the third takes a slightly different form because production of a type-2 item is disallowed in state $(n1, 0)$ while idleness is now possible. The third condition is now that $max[v_k(n_1 + 2, 1), v_k(n_1 + 1, 0)] - max[v_k(n_1 + 1, 0), v_k(n_1, 0)]$ is non-increasing in $n1$. This condition can be equivalently stated as: $\{[v_k(n_1 + 2, 1) - v_k(n_1 + 1, 0)]^+ + v_k(n_1 + 1, 0)\} - \{[v_k(n_1, 0) - v_k(n_1 + 1, 0)]^+ + v_k(n_1 + 1, 0)\}$ where $[\cdot]^+$ denotes $max[\cdot, 0]$. After canceling terms, this becomes $[v_k(n_1 + 2, 1) - v_k(n_1 + 1, 0)]^+ - [v_k(n_1, 0) - v_k(n_1 + 1, 0)]^+$. This is non-increasing in $n1$ by conditions (C1b) and (C1c).

We have therefore shown that the first part of condition (C1b) holds. The proofs for the other conditions are similar and omitted $\square$.

23

To complete the proof of the theorem, we note that the conditions sufficient for the Theorem hold in every iteration of the value iteration algorithm by Lemma 2. By Lemma 1, $v_k(n_1, n_2) \to v(n_1, n_2)$ which proves the result. $\square$.

Proof of Theorem 2 The results for R_2, c_1, c_2 and π directly follow from the fact that using the policy which is optimal before the change in these parameters after the change as well will result in higher profits. We give the proof for λ_2; the proofs for μ and λ_1 are similar.

Consider two systems, labeled System 1 and System 2, that are identical except for differing type-2 order arrival rates, $\lambda_2^{[1]}$ and $\lambda_2^{[2]}$ for systems 1 and 2 where $\lambda_2^{[1]} < \lambda_2^{[2]}$. Suppose that System 2 is operated in the following manner:

1. Whenever a type-2 order arrives, it is immediately rejected with probability $\frac{\lambda_2^{[2]} - \lambda_2^{[1]}}{\lambda_2^{[2]}}$.

2. Acceptance/rejection of the the orders not rejected in step 1; as well as the sequencing of orders is done using the optimal policy for System 1

Under this policy, System 2 now exactly replicates the optimal operation of System 1 in all respects and will thus realize the same profitability as System 1. Since the profitability for System 1 is feasible to System 2, the optimal profitability for System 2 must be greater than or equal to that of System 1. $\square$.

Proof of Theorem 3

We give the proof for the policy sensitivity with respect to R_2. The proof for the sensitivity with respect to π is similar. We again consider two systems, denoted System A and System B, that are identical except for differing type-2 profits that are respectively R_2 and R_2'; $R_2 < R_2'$. We apply equation (6.4) to both systems simultaneously; the algorithms are initialized with $v_0(\cdot) = v_0'(\cdot) = 0$. We first prove the following lemma:

Lemma 3 *For every $k = 0, 1, \cdots$:*

C3a) $[v_k'(n_1, n_2 + 1) + R_2' - v_k'(n_1, n_2)] - [v_k(n_1, n_2 + 1) + R_2 - v_k(n_1, n_2)] \geq 0$

C3b) $[v_k'(n_1, n_2) - v_k'(n_1 + 1, n_2 + 1)] - [v_k(n_1, n_2) - v_k(n_1 + 1, n_2 + 1)] \geq 0$

C3c) $[v_k'(n_1 + 1, n_2) - v_k'(n_1, n_2)] - [v_k(n_1 + 1, n_2) - v_k(n_1, n_2)] \geq 0$

Proof of condition C3a: The lemma holds trivially for $k = 0$. We assume that the lemma applies for all states at arbitrary iteration k and now show that it holds at iteration $(k+1)$. In the proof, we will also use the fact that previously proved conditions (C1a-C1c) also hold for both v_k and v'_k at each iteration of the value iteration. For brevity, we only present the proof of (C3a) in the case $n_1, n_2 > 0$; the proof is similar for the other cases and its presentation here is not additionally illustrative. Relating $v_{k+1}(\cdot)$ and $v'_{k+1}(\cdot)$ to $v_k(\cdot)$ and $v'_k(\cdot)$ through equation (6.4) yields the condition that showing (C3a) for the $k+1^{st}$ iteration is equivalent to showing that

$$
\begin{aligned}
0 \leq \quad & \tfrac{\lambda_1}{\Lambda} \quad \{[v'_k(n_1 - 1, n_2 + 1) - v'_k(n_1 - 1, n_2)] - [v_k(n_1 - 1, n_2 + 1) - v_k(n_1 - 1, n_2)]\} \\
& + \tfrac{\lambda_2}{\Lambda} \quad \{max[v'_k(n_1, n_2 + 2) + R'_2, v'_k(n_1, n_2 + 1)] - max[v'_k(n_1, n_2 + 1) + R'_2, v'_k(n_1, n_2)] \\
& \qquad - max[v_k(n_1, n_2 + 2) + R_2, v_k(n_1, n_2 + 1)] + max[v_k(n_1, n_2 + 1) + R_2, v_k(n_1, n_2)]\} \\
& + \tfrac{\mu}{\Lambda} \quad \{max[v'_k(n_1 + 1, n_2 + 1), v'_k(n_1, n_2)] - max[v'_k(n_1 + 1, n_2), v'_k(n_1, n_2 - 1)] \\
& \qquad - max[v_k(n_1 + 1, n_2 + 1), v_k(n_1, n_2)] + max[v_k(n_1 + 1, n_2), v_k(n_1, n_2 - 1)]\} + R'_2 - R_2
\end{aligned}
$$

$$(6.6)$$

We again extract three conditions that, when simultaneously satisfied, are sufficient for (6.6) to hold.

1. $v'_k(n_1 - 1, n_2 + 1) - v'_k(n_1 - 1, n_2)] - [v_k(n_1 - 1, n_2 + 1) - v_k(n_1 - 1, n_2)] + R'_2 - R_2 \geq 0,$

2. $max[v'_k(n_1, n_2+2)+R'_2, v'_k(n_1, n_2+1)] - max[v'_k(n_1, n_2+1)+R'_2, v'_k(n_1, n_2)] - max[v_k(n_1, n_2+2) + R_2, v_k(n_1, n_2 + 1)] + max[v_k(n_1, n_2 + 1) + R_2, v_k(n_1, n_2)] + R'_2 - R_2 \geq 0,$

3. $max[v'_k(n_1 + 1, n_2 + 1), v'_k(n_1, n_2)] - max[v'_k(n_1 + 1, n_2), v'_k(n_1, n_2 - 1)] - max[v_k(n_1 + 1, n_2 + 1), v_k(n_1, n_2)] + max[v_k(n_1 + 1, n_2), v_k(n_1, n_2 - 1)] + R'_2 - R_2 \geq 0.$

The first of these conditions holds by the induction assumption that C3a is assumed to hold for the kth iteration. We let $LHS^{[2]}$ denote the left-hand side of condition 2 and let $LHS^{[3]}$ denote the left-hand side of condition 3. $LHS^{[2]}$ and $LHS^{[3]}$ can each take 16 values of which 10 can be eliminated in the same manner as in the proof of Lemma 2. We give the six feasible values of $LHS^{[3]}$ showing which terms come out of the four maximums in $LHS^{[3]}$ in each case.

The six feasible values of $LHS^{[3]}$ are:

25

i $LHS^{[3]} = v'_k(n_1+1, n_2+1) - v'_k(n_1+1, n_2) - v_k(n_1+1, n_2+1) + v_k(n_1+1, n_2) + R'_2 - R_2 \geq$
0 by the induction assumption of C3a.

ii $LHS^{[3]} = v'_k(n_1, n_2) - v'_k(n_1, n_2-1) - v_k(n_1, n_2) + v_k(n_1+1, n_2) + R'_2 - R_2 \geq 0$ by
C3a.

iii $LHS^{[3]} = v'_k(n_1, n_2) - v'_k(n_1+1, n_2) - v_k(n_1+1, n_2+1) + v_k(n_1+1, n_2) + R'_2 - R_2$;
Since $v'_k(n_1, n_2)$ came out of the first maximum in $LHS^{[3]}$ in this case, a requirement
for this case is that $v'_k(n_1+1, n_2+1) \leq v'_k(n_1, n_2)$. Thus,

$$LHS^{[3]} \geq v'_k(n_1+1, n_2+1) - v'_k(n_1+1, n_2) - v_k(n_1+1, n_2+1) + v_k(n_1+1, n_2) + R'_2 - R_2$$

which is greater than or equal to zero by C3a.

iv $LHS^{[3]} = v'_k(n_1, n_2) - v'_k(n_1+1, n_2) - v_k(n_1, n_2) + v_k(n_1+1, n_2) + R'_2 - R_2$, which
we can rewrite as

$$LHS^{[3]} = \{v'_k(n_1, n_2) - v'_k(n_1+1, n_2+1) - v_k(n_1, n_2) + v_k(n_1+1, n_2+1)\}$$

$$+\{v'_k(n_1+1, n_2+1) - v'_k(n_1+1, n_2) - v_k(n_1+1, n_2+1) + v_k(n_1+1, n_2)\} + R'_2 - R_2$$

which is greater than or equal to zero by C3a and C3b.

v $LHS^{[3]} = v'_k(n_1, n_2) - v'_k(n_1, n_2-1) - v_k(n_1+1, n_2+1) + v_k(n_1+1, n_2) + R'_2 - R_2$
or equivalently

$$LHS^{[3]} = \{v'_k(n_1, n_2) - v'_k(n_1+1, n_2+1) - v'_k(n_1, n_2-1) + v'_k(n_1+1, n_2)\}$$

$$+\{v'_k(n_1+1, n_2+1) - v'_k(n_1+1, n_2) - v_k(n_1+1, n_2+1) + v_k(n_1+1, n_2) + R'_2 - R_2\}$$

which is greater than or equal to zero by C1b and C3a.

vi $LHS^{[3]} = v'_k(n_1, n_2) - v'_k(n_1, n_2-1) - v_k(n_1, n_2) + v_k(n_1+1, n_2) + R'_2 - R_2$ Since
$v_k(n_1+1, n_2)$ came out of the fourth maximum of $LHS^{[3]}$ in this case, a requirement
for this case is that $v_k(n_1+1, n_2) \geq v_k(n_1, n_2-1)$. Thus,

$$LHS^{[3]} \geq v'_k(n_1, n_2) - v'_k(n_1, n_2-1) - v_k(n_1, n_2) + v_k(n_1, n_2-1) + R'_2 - R_2 \geq 0$$ by
C3a.

The other 10 cases for $LHS^{[3]}$ are not feasible due to the fact that they contradict the
policy structure at the kth iteration. The proof for the fact $LHS^{[2]}$ is greater than or equal

to zero is similar involving 6 feasible cases out of 16 and is omitted for brevity. Since all three conditions are shown to hold, the proof of the lemma is complete □.

To complete the proof of the theorem, since $v'_k(n_1, n_2) - v_k(n_1, n_2)$ converges to $v'(n_1, n_2) - v(n_1, n_2)$, conditions C3a-C3c will hold for the value functions of the average cost problem for both systems. Now note that condition C3a implies that if acceptance of type-2 orders at state (n_1, n_2) in System A is optimal, it must also be optimal in System B. Similarly, condition C3b implies that if production of a type-2 product is optimal in state (n_1, n_2) with $n_2 > 0$ in system A, then it must also be optimal in the same state in system B. Finally, note that the special case of $n_2 = 0$ of Condition C3c implies that if production of a type-1 unit is preferred over idling in state $(n_1, 0)$ for system A, then it will also be preferred in the same state for system B. This completes the proof □.

Acknowledgements:

Bibliography

[1] Anupindi, R., and S. Tayur, "Managing Stochastic Multi-Product Systems: Model, Measures and Analysis. Technical Report, Graduate School of Industrial Administration, Carnegie Mellon University, Pittsburgh, PA (1994).

[2] Bourland, K.E., and C.A. Yano, "The Strategic Use of Capacity Slack in the Economic Lot Scheduling Problem with Random Demand," *Management Science*, 40, (1995), 1690-1704

[3] Duenyas, I., "Single Facility Due Date Setting with Multiple Customer Classes, " *Management Science*, 41, No:4, (1995), 608-619.

[4] Duenyas, I., W.J. Hopp, and Y. Bassok, "Production Quotas as Bounds on Interplant JIT Contracts," (1996), to appear in *Management Science*.

[5] Duenyas, I., and M. VanOyen, "Heuristic Scheduling of Parallel Heterogeneous Queues with Set-Ups," (1996), *Management Science*, 42, (1996), 814-829.

[6] Federgruen, A and Z. Katalan, "The Stochastic Economic Lot Scheduling Problem: Cyclical Base Stock Policies with Idle Times," *Management Science*, 42, (1996), 783-796.

[7] Federgruen, A and Z. Katalan, "Make-to-Stock or Make-to-Order: That is the Question; Novel Answers to an Ancient Debate," Working paper – Columbia University, 1994.

[8] Gallego, G., "Scheduling the Production of Several Items with Random Demands in a Single Facility," *Management Science*, 36, 12 (1990), 1579-1593.

[9] Ha, A. Y., "Optimal Dynamic Scheduling Policy for a Make-to-Stock Production System," forthcoming in *Operations Research*, 1993.

[10] Ha, A. Y., "Inventory Rationing in a Make-to-Stock Production System with Several Demand Classes and Lost Sales," working paper – Yale School of Management, 1994.

[11] Ha, A.Y., "Stock Rationing Policy for a Make-To-Stock Production System with Two Priority Classes and Backordering," working paper-Yale School of Management, 1995.

[12] Haessler, R. W., "An Improved Extended Basic Period Procedure for Solving the Economic Lot Scheduling Problem," *AIIE Transactions*, 11, 4 (1979), 336-340.

[13] Hernandez-Lerma, O., *Adaptive Markov Control Processes*, Springer-Verlag, New York, (1989).

[14] Li, L., "The Role of Inventory in Delivery-time Competition," *Management Science*, 38, (1992), 182-197.

[15] Markowitz, D.M., M.I. Reiman, L.M. Wein, " The Stochastic Economic Lot Scheduling Problem: Heavy Traffic Analysis of Dynamic Cyclic Policies," Technical Report, Sloan School of Management, M.I.T., Cambridge, MA, (1995).

[16] Nguyen, V., "Fluid and diffusion Approximations of a Two-Station Mixed Queueing Network," *Mathematics of Operations Research*, 20, 2 (1995), 321-354.

[17] Qui, J., and R. Loulou, "Multiproduct Production/Inventory Control under Random Demands," *IEEE Transactions on Automatic Control*, 40, (1995), 350-356.

[18] Reiman, M.I., and L.M. Wein, "Dynamic Scheduling of a Two-Class Queue with Set-Ups," Technical Report, Sloan School of Management, M.I.T., Cambridge, MA (1994).

[19] Sox, C.R., and J.A. Muckstadt, "Optimization-Based Planning for the Stochastic Lot Scheduling Problem," Operations Research and Industrial Engineering Department, Cornell University, Ithaca, NY, (1995).

[20] Stidham, S., "Optimal Control of Admission to a Queueing System," *IEEE Transactions of Automatic Control*, AC-30, 8 (1985), 705-713.

[21] Stidham, S. and R. Weber, "A survey of Markov Decision Models for Control of Networks of Queues," *Queueing Systems: Theory and Applications*, 13, (1993), 291-314.

[22] Varaiya, P., J. Walrand, and C. Buyukkoc, "Extensions of the multiarmed bandit problem: the discounted case," *IEEE Transactions on Automatic Control* AC-30, (1985), 426-439.

[23] Veatch, M.H., and L.M. Wein, "Scheduling a Make-To-Stock Queue: Index Policies and Hedging Points," *Operations Research*, 44, (1996), 634-647.

[24] Walrand, J., *An Introduction to Queueing Networks,*, Prentice-Hall, Englewood-Cliffs, NJ, (1988).