

# Production and Inventory Control of a Single Product Assemble-to-Order System with Multiple Customer Classes

Saif Benjaafar

Industrial and Systems Engineering, Department of Mechanical Engineering, University of Minnesota, Minneapolis, Minnesota 55455, saif@umn.edu

Mohsen ElHafsi

The A. Gary Anderson Graduate School of Management, University of California, Riverside, California 92521-0203, mohsen.elhafsi@ucr.edu

We consider the optimal production and inventory control of an assemble-to-order system with  $m$  components, one end-product, and  $n$  customer classes. A control policy specifies when to produce each component and, whenever an order is placed, whether or not to satisfy it from on-hand inventory. We formulate the problem as a Markov decision process and characterize the structure of an optimal policy. We show that a base-stock production policy is optimal, but the base-stock level for each component is dynamic and depends on the inventory level of all other components (more specifically, it is nondecreasing). We show that the optimal inventory allocation for each component is a rationing policy with different rationing levels for different demand classes. The rationing levels for each component are dynamic and also nondecreasing in the inventory level of all other components. We compare the performance of the optimal policy to heuristic policies, including the commonly used base-stock policy with fixed base-stock levels, and find them to perform surprisingly well, especially for systems with lost sales.

*Key words:* assemble-to-order (ATO) systems; production and inventory control; Markov decision processes; make-to-stock queues

*History:* Accepted by Paul H. Zipkin, operations and supply chain management; received October 22, 2004. This paper was with the authors 7 months for 3 revisions.

## 1. Introduction

Assemble-to-order (ATO) production has become a popular strategy for manufacturing firms that seek to be both responsive and cost efficient. ATO production enables a firm to shorten its response time to its customers by staging inventory of components ahead of demand while postponing the final assembly until demand is realized. This strategy is particularly valuable when component supply leadtimes are long or the supply processes are capacitated. Furthermore, by pooling component inventories, ATO can reduce the costs of offering higher product variety, which can be useful when demand for individual end-products is variable. In addition to manufacturing, features of ATO systems are present in several other settings where demand is either correlated across several items (e.g., demand for computers and printers) or orders for multiple items from the same customer must be fulfilled simultaneously (e.g., order fulfillment at e-retailers and mail order catalogs).

An ATO strategy is increasingly being coupled with a customer differentiation strategy. Firms segment customers into different classes to which they offer

the same product but with different levels of service quality. Customers may be differentiated based on the price they pay, the volume they purchase, or the length of contract to which they commit. The quality of service customers are offered may be specified in terms of penalties levied for delays and lost sales, expediting undertaken in case of shortages, or guarantees of a specified service level (e.g., a fill rate) for customers with long-term contracts. Firms affect the quality of service received by different customers via inventory allocation decisions. For example, a firm may choose not to fulfill demand from a low-priority customer to reserve inventory for future demand from higher priority customers.

Examples in industry of ATO systems with multiple customer classes are many. For instance, IBM has recently restructured the manufacturing of its server computers into two stages (Cheng et al. 2005). In the first stage, components with long manufacturing leadtimes are built in a make-to-stock fashion. In the second stage, these components are assembled to order in response to customer demand. The demand for the servers arises from different corporate customers

who can have different fulfillment time requirements and different penalties for delays. Companies such as Solectron and Flextronics that provide outsourcing manufacturing services in the electronics industry to original equipment manufacturers (OEMs) are adopting similar strategies. These contract manufacturers (CMs) often produce in the same plant the same product with minor variations for different OEMs (Benjaafar et al. 2004). To benefit from inventory pooling, sometimes only components, such as printed circuit boards, are produced and stocked ahead of demand, with final assembly occurring when the OEMs place their orders. Depending on their contractual agreements with the OEMs, the CMs can face penalties for delays and stockouts that can vary from one OEM to another.

In spite of their prevalence in practice, however, ATO systems are notoriously difficult to analyze and manage. The difficulty appears to be due to several factors: (a) demands for the different components being correlated, (b) supply leadtimes for different components being different, and (c) order fulfillment being dependent on the availability of multiple components. These factors make it difficult to manage components independently. For example, continuing to produce one component when there is a shortage of another may do little to improve the ability to fulfill demand. Similarly, deciding to stop production of one component without considering the inventory level of other components may not necessarily reduce overall inventory costs. Also, in systems with multiple customer classes, deciding whether or not to fulfill demand from a particular customer class cannot be carried out without considering the inventory level of all components.

In this paper, we consider a system consisting of multiple components, multiple customer classes, and a single end-product. Components are produced one unit at a time on separate production facilities with finite production rates. Components can be produced ahead of demand and held in stock but incur a holding cost that varies by component. To fulfill a customer order, all  $m$  components must be available. Otherwise, the order is lost or backordered if backorders are allowed. An order that is not immediately fulfilled from stock incurs a shortage cost (a lost sale cost or a backorder cost) that can vary from class to class. The system manager must determine when to produce a particular component and, whenever an order is placed, whether or not to satisfy it from on-hand inventory.

We formulate the problem as a Markov decision process (MDP) and characterize the structure of an optimal policy, under both the total expected discounted cost and the average cost criteria. We show that the optimal production policy for each

component is a state-dependent base-stock policy, where the state of the system is specified by the vector of component inventory levels. We show that the base-stock level for each component is nondecreasing in the inventory level of other components with a unit increase in the inventory of any component leading to at most one unit increase in the base-stock level of other components. We show that the optimal inventory allocation policy consists of a multi-level rationing policy. An order from a customer class is satisfied only if the current inventory level at each component is above a certain rationing level. The rationing levels for each class at each component are state-dependent and nonincreasing in the inventory levels of other components. We show using examples that the optimal base-stock and rationing levels are sensitive to system parameters, including production rates and holding costs. However, characterizing analytically the impact of these parameters is difficult in general, and we suspect that in most cases there may not be simple relationships.

Using numerical results, we compare the performance of the optimal policy with those of two simple heuristics. The first heuristic consists of using a fixed base-stock level for each component. The second heuristic monitors the difference in the inventory level between each pair of components and stops the production of a component if this difference is above a target threshold. For both heuristics, inventory allocation is carried out via fixed rationing levels. The first heuristic is motivated by prior studies of ATO systems that have focused on evaluating base-stock policies with fixed base-stock levels. These policies are easy to implement and are common in practice. However, they are static; production decisions are not adjusted to take into account the state of the system, and there is no attempt to coordinate production of the different components once the base-stock levels have been set. The second heuristic is a dynamic policy that monitors the inventory level of all components and attempts to coordinate production in a way that mimics the optimal policy. This heuristic is clearly superior to the first one, and this is confirmed by the numerical results. Perhaps surprisingly, we find both policies to perform well relative to the optimal policy for systems with lost sales. For systems with backorders, there are cases for which the heuristics, and especially the one with fixed base-stock levels, perform poorly.

The rest of this paper is organized as follows. In §2, we offer a brief review of related literature. In §3, we present our model for systems with lost sales. In §4, we extend our analysis to a system with backorders. In §5, we also present numerical results and comparisons with heuristics. In §6, we offer a summary and few concluding comments.

## 2. Literature Review

The literature on ATO systems can be broadly classified into two categories: one dealing with systems with periodic review and one with systems with continuous review; see Song and Zipkin (2003) for a comprehensive literature review. In both streams of literature, little is known about optimal control policies. For periodic review systems, Schmidt and Nahmias (1985) studied the structure of the optimal policy for a system with two components and showed that it does not have a simple form. Rosling (1989) showed that, under some mild conditions on initial inventories, a modified base-stock policy for component replenishment is optimal. However, most other researchers dealing with periodic review have focused on performance evaluation of heuristic policies; see, for example, Hausman et al. (1998), de Kok and Visschers (1999), Cheng et al. (2002), Zhang (1997), Agrawal and Cohen (2001), Frank et al. (2004), and references therein.

The literature on ATO systems with continuous review can be classified itself into two areas based on the underlying component supply process: (1) systems with exogenous and load-independent leadtimes and (2) systems with endogenous and load-dependent leadtimes (our paper belongs to the second category). In the first case, inventory replenishment leadtimes are assumed to be independent of the number of outstanding orders (pure inventory systems). In the second case, replenishment leadtimes are affected by the number of outstanding orders because of limitations in production capacity (integrated production-inventory systems). For both cases, component inventory is typically assumed to be managed using independent base-stock policies with fixed base-stock levels, and the assembly of the final product is instantaneous.

ATO systems with exogenous supply leadtimes (and base-stock control with independent and fixed base-stock levels) can be viewed as a set of infinite server queues—that is,  $G/G/\infty$  queues—with correlated arrivals. Song (1998) considered a system with Poisson demand and deterministic leadtimes modeled as  $M/D/\infty$  queues. She developed an efficient algorithm for computing order fill rates. Song and Yao (2002) extended this analysis to systems with stochastic component leadtimes modeled as  $M/G/\infty$  queues. They examined the impact of supply leadtime variability and evaluated several performance bounds that they used to examine the tradeoff between inventory and service level. Gallien and Wein (2001) also considered ATO systems in which the supply process of each component is modeled as a  $M/G/\infty$  queue. They evaluated a policy consisting of a fixed and common base-stock level for all components and component-dependent order leadtimes, such that

a replenishment order for each component is delayed by a component-dependent leadtime.

ATO systems with load-dependent supply leadtimes and base-stock policies with fixed base-stock levels can also be viewed as queues with correlated arrivals but with a finite number of servers. Song et al. (1999) considered a system in which each component is produced on a separate production facility with a single server and exponentially distributed production times. They also assumed that there is a limit on the number  $N$  of outstanding orders with each facility, so that each facility can be modeled as an  $M/M/1/N$  queue. They showed that exact performance measures can be obtained using the matrix geometric method. Dayanik et al. (2003) developed bounds on the fill rate for systems that can be modeled as correlated  $M/M/1/N$  queues. Glasserman and Wang (1998) considered systems that can be modeled as  $M^X/G/1$  and  $G^X/G/1$  queues and obtained an asymptotic relationship between base-stock levels and target delivery leadtimes. Chen et al. (1993) considered a special ATO system with two components, each produced on a single server facility, and unlimited demand for the final product. They showed that the optimal policy for each facility is to produce as long as its inventory is below a certain threshold or if there is positive inventory at the other facility. This policy ceases to be optimal when the demand rate is finite. Finally, there is a rich literature from queuing theory on the general analysis of systems with correlated arrivals. We refer the reader to Xu (2001) for a recent review.

The literature dealing with production and inventory control with multiple demand classes is relatively limited. Ha (1997a) considered a make-to-stock queue with a single server and single component (no assembly) with Poisson demand and exponentially distributed production times. For a system with multiple customer classes and lost sales, he showed that the optimal policy is of the threshold type, where orders from the lower priority class are fulfilled as long as inventory is above a certain threshold level. Ha (1997b) extended these results to systems with backordering and two customer classes, and de Véricourt et al. (2002) generalized them to systems with  $N$  classes. Benjaafar et al. (2004) considered inventory rationing in a system with multiple products and multiple production facilities.

Our paper fits in the stream of literature on ATO systems with continuous review and load-dependent supply leadtimes. We consider an ATO system with similar characteristics to the ones treated in Song et al. (1999) and Dayanik et al. (2003). Similar to those papers, we assume that demand forms a Poisson process, assembly is instantaneous, and components are produced on separate single server production facilities with exponentially distributed production times. Our focus on an optimal policy is, of

course, different from those papers that assume a specific heuristic policy. Also, those papers treat multiple products, whereas ours considers a system with a single product. Our treatment of systems with multiple customer classes is similar to the one in Ha (1997a, b) and de Véricourt et al. (2002), who also considered systems with single servers, exponentially distributed production times, and Poisson demand processes for each customer class. The focus of those papers is on systems with a single component produced on a single facility. Our paper appears to be the first to consider the optimal control of a system with multiple components, multiple production facilities, and multiple demand classes. It also appears to be the first to compare in this setting the performance of an optimal policy with the performance of heuristics.

### 3. Model Formulation and Structure of the Optimal Policy

We consider a system consisting of a single end-product assembled from  $m$  components and satisfying demand from  $n$  customer classes. The components are produced one unit at a time on  $m$  separate production facilities, and the finished components are placed in inventory. Items in inventory incur a holding cost per unit per unit time that may vary from component to component. Production times are independent and exponentially distributed with mean  $\mu_k^{-1}$ , for component  $k = 1, \dots, m$ . Hence, each facility can be viewed as a single server with finite service rate  $\mu_k$ . Demand for the end-product from each class  $l$ ,  $l = 1, \dots, n$ , takes place continuously over time according to an independent Poisson process with rate  $\lambda_l$ . Demand for the end-product from any class can be satisfied only if all  $m$  components are available. Otherwise, the demand is considered lost (or must be expedited through other means, including overtime or outsourcing to a third party). A demand from class  $l$  that cannot be immediately fulfilled from stock incurs a lost sale cost  $c_l$  per unit, which can vary from class to class. Without loss of generality, we assume  $c_1 \geq c_2 \geq \dots \geq c_n$ .

Because the lost sale costs can be different for different classes, it may not always be optimal to satisfy demand from a class even if all  $m$  components are available in inventory. In fact, it might be more desirable to reject a demand from a class to reserve the available inventory for future demand from a more important class (i.e., one with a higher shortage cost). Consequently, each time an order is placed, the system manager must decide whether or not to satisfy it from on-hand inventory, if any is available. If a decision is made to satisfy an incoming order from available inventory, the product is immediately assembled and shipped to the customer. Our assumption of instantaneous assembly is consistent with most

of the literature on ATO systems (Song and Zipkin 2003). It is in part motivated by the fact that, in many ATO applications, final assembly time is significantly shorter than component production times.

In addition to determining whether or not to reject an incoming order, the system manager must constantly decide whether or not to produce any of the components. If a component is not currently being produced, this means deciding whether or not to initiate its production. If the component is currently being produced, this means deciding whether or not to interrupt its production. If the production of a component is interrupted, it can be resumed the next time the production of that component is initiated (because of the memoryless property of the exponential distribution, resuming production from where it was interrupted is equivalent to initiating it from scratch). We assume that there are no costs associated with interrupting production. This conforms to earlier treatment of production-inventory systems in the literature; see, for example, Ha (1997a, b). This assumption is not restrictive because, as we show in Theorem 1, it turns out that it is never optimal to interrupt production of a component once it has been initiated.

The state of the system at time  $t$  can be described by the vector  $\mathbf{X}(t) = (X_1(t), \dots, X_m(t))$ , where  $X_k(t)$ ,  $k = 1, \dots, m$ , is a nonnegative integer denoting the on-hand inventory for component  $k$  at time  $t$ . We let  $h(\mathbf{X}(t)) = \sum_{k=1}^m h_k(X_k(t))$ , where  $h_k$  is an increasing convex function, denote the inventory holding cost rate when the state is  $\mathbf{X}(t)$ . Note that because of the possibility of interrupting production, it is not necessary to include in the state description whether a component is currently being produced. Furthermore, because both order interarrival times and production times are exponentially distributed, the system is memoryless, and decision epochs can be restricted to only times when the state changes (i.e., the completion of a component or the fulfillment of an order). The memoryless property allows us to formulate the problem as an MDP and to restrict our attention to the class of Markovian policies for which actions taken at a particular decision epoch depend only on the current state of the system.

In each state, the system manager must decide which components to produce and whether or not to accept an order from a particular customer class, should one arise. A policy  $\pi$  specifies for each state  $\mathbf{x} = (x_1, \dots, x_m)$ , the action  $a^\pi(\mathbf{x}) = (u_1, \dots, u_m, w_1, \dots, w_n)$ , where  $u_k = 1$  means produce component  $k$  ( $k = 1, \dots, m$ ),  $u_k = 0$  means do not produce component  $k$ ,  $w_l = 1$  means satisfy demand from class  $l$  ( $l = 1, \dots, n$ ), and  $w_l = 0$  means reject demand from class  $l$ . For example, in a system with two components and two demand classes, the action  $a^\pi(\mathbf{x}) = (1, 0, 0, 1)$

means that whenever the system is in state  $\mathbf{x}$ , produce component 1, do not produce component 2, reject demand from class 1, and accept demand from class 2. As we can see, there are two types of decisions: production decisions indicated by the parameters  $u_k$  and inventory allocation decisions indicated by the parameters  $w_l$ . Production decisions determine for each state whether or not a component should be produced, and inventory allocation decides in which states it is best not to fulfill demand from a particular class to reserve inventory for future demand from other classes. We refer to this type of allocation as *inventory rationing*.

**3.1. The Case of Discounted Cost**

The expected discounted cost over an infinite planning horizon  $v^\pi(\mathbf{x})$  obtained under a policy  $\pi$  and a starting state  $\mathbf{x} = (x_1, \dots, x_m)$  can be written as

$$v^\pi(\mathbf{x}) = E_x^\pi \left[ \sum_{k=1}^m \int_0^\infty e^{-\alpha t} h_k(X_k(t)) dt + \sum_{l=1}^n \int_0^\infty e^{-\alpha t} c_l dN_l(t) \right], \tag{1}$$

where  $\alpha > 0$  is the discount rate and  $N_l(t)$  denotes the number of orders from class  $l$  that have not been fulfilled from on-hand inventory up to time  $t$ . Following Lippman (1975), we work with a uniformized version of the problem in which the transition rate in each state under any action is  $\beta = \sum_{l=1}^n \lambda_l + \sum_{k=1}^m \mu_k$  so that the transition times  $0 = t_0 \leq t_1 \leq t_2 \leq \dots$  are such that the times between transitions  $\{t_{i+1} - t_i : i \geq 0\}$  are a sequence of i.i.d. exponential random variables, each with mean  $1/\beta$ . This leads to a Markov chain defined by  $\{\mathbf{X}_i : i \geq 0\}$ , where  $\mathbf{X}_i \equiv \mathbf{X}(t_i) = (X_1(t_i), \dots, X_m(t_i))$  is the state resulting from the  $i$ th transition. The introduction of the uniform transition rate allows us to transform the continuous time decision process into a discrete time decision process, simplifying the analysis considerably.

If action  $\mathbf{a}$  is selected in state  $\mathbf{x}$ , the next state is  $\mathbf{y}$  with probability

$$p_{\mathbf{x}, \mathbf{y}}(\mathbf{a}) = \begin{cases} \frac{\lambda_l}{\beta} I_{\{\prod_{k=1}^m x_k \neq 0, \text{ and } w_l=1\}} & \text{if } \mathbf{y} = \mathbf{x} - \mathbf{e}_l, \\ \frac{\mu_k}{\beta} I_{\{u_k=1\}} & \text{if } \mathbf{y} = \mathbf{x} + \mathbf{e}_k, \\ \frac{\beta - \sum_{l=1}^n \lambda_l I_{\{\prod_{k=1}^m x_k \neq 0, \text{ and } w_l=1\}} - \sum_{k=1}^m \mu_k I_{\{u_k=1\}}}{\beta} & \text{if } \mathbf{y} = \mathbf{x}, \text{ and} \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where  $I_{\{z\}}$  is an indicator function with  $I_{\{z\}} = 1$  if  $z$  is true and  $I_{\{z\}} = 0$  otherwise, and  $\mathbf{e}_k$  is the  $k$ th unit

vector of dimension  $m$  and  $\mathbf{e} = \sum_{k=1}^m \mathbf{e}_k = (1, 1, \dots, 1)$  is an  $m$ -dimensional vector of ones. Let  $N_l(t_i)$ ,  $l = 1, \dots, n$ , denote the cumulative number of unfilled orders from class  $l$  after the  $i$ th transition. Then,  $v^\pi(\mathbf{x})$  in (1) can be rewritten in the equivalent form

$$v^\pi(\mathbf{x}) = E_x^\pi \left[ \sum_{i=0}^\infty \left( \frac{\beta}{\alpha + \beta} \right)^i \frac{h(\mathbf{x})}{\alpha + \beta} + \sum_{i=1}^\infty \left( \frac{\beta}{\alpha + \beta} \right)^i \cdot \sum_{l=1}^n c_l (N_l(t_i) - N_l(t_{i-1})) \right]. \tag{3}$$

Our objective is to choose a policy  $\pi^*$  that minimizes the expected discounted cost. The optimal cost function  $v^* \equiv v^{\pi^*}$  satisfies the following optimality equation:

$$v^*(\mathbf{x}) = \min_{\mathbf{a}} \left\{ \frac{h(\mathbf{x})}{\alpha + \beta} + \frac{\beta}{\alpha + \beta} \left( \sum_{l=1}^n \frac{\lambda_l}{\beta} c_l (I_{\{\prod_{k=1}^m x_k = 0\}} + I_{\{\prod_{k=1}^m x_k \neq 0, w_l = 0\}}) + \sum_{\mathbf{y}} p_{\mathbf{x}, \mathbf{y}}(\mathbf{a}) v^*(\mathbf{y}) \right) \right\}. \tag{4}$$

Let the operators  $T^l$  and  $T_k$ , for  $l = 1, \dots, n$  and  $k = 1, \dots, m$ , be defined, for any real valued function  $v$  over the state space, as follows:

$$T^l v(\mathbf{x}) = \begin{cases} v(\mathbf{x}) + c_l & \text{if } \prod_{k=1}^m x_k = 0, \\ \min\{v(\mathbf{x} - \mathbf{e}), v(\mathbf{x}) + c_l\} & \text{otherwise,} \end{cases} \tag{5}$$

and

$$T_k v(\mathbf{x}) = \min\{v(\mathbf{x} + \mathbf{e}_k), v(\mathbf{x})\}. \tag{6}$$

Also, without loss of generality, we rescale time so that  $\alpha + \beta = 1$ . Then, the optimality equation in (4) can be rewritten in the following simpler form:

$$v^*(\mathbf{x}) = h(\mathbf{x}) + \sum_{l=1}^n \lambda_l T^l v^*(\mathbf{x}) + \sum_{k=1}^m \mu_k T_k v^*(\mathbf{x}). \tag{7}$$

The operator  $T^l$  is associated with decisions about whether or not to satisfy an order of type  $l$ , and the operator  $T_k$  is associated with decisions about whether or not to produce component  $k$ . Note that whenever there is on-hand inventory for each component, it is optimal to satisfy demand from class  $l$  if  $v^*(\mathbf{x} - \mathbf{e}) \leq v^*(\mathbf{x}) + c_l$ . Similarly, it is optimal to produce component  $k$  when the system is in state  $\mathbf{x}$  if  $v^*(\mathbf{x} + \mathbf{e}_k) \leq v^*(\mathbf{x})$ .

In the following theorem, we characterize the structure of an optimal policy. The proof of this and all subsequent results are included in the appendix available on the *Management Science* website at <http://mansci.pubs.informs.org/ecompanion.htm>.

**THEOREM 1.** *There exists an optimal stationary policy that can be specified as follows. The optimal production policy for each component  $k$ ,  $k = 1, \dots, m$ , consists of a base-stock policy with a vector of state-dependent base-stock levels  $s_k^*(\mathbf{x}_{-k})$  such that it is optimal to produce component  $k$  if  $x_k < s_k^*(\mathbf{x}_{-k})$  for  $k = 1, \dots, m$  and not to produce it otherwise, where  $x_k$  is the on-hand inventory level of component  $k$  and  $\mathbf{x}_{-k} = (x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_m)$  is an  $(m - 1)$  dimensional vector consisting of the on-hand inventory levels for components  $j \neq k$ . The optimal inventory allocation policy consists of a multilevel rationing policy with a vector of state-dependent rationing levels  $\mathbf{r}_k^*(\mathbf{x}_{-k}) = (r_{k,1}^*(\mathbf{x}_{-k}), \dots, r_{k,n}^*(\mathbf{x}_{-k}))$  such that it is optimal to fulfill an incoming order from class  $l$ ,  $l = 1, \dots, n$ , if  $x_k \geq r_{k,l}^*(\mathbf{x}_{-k})$  for  $k = 1, \dots, m$  and not to fulfill it otherwise. The optimal policy has the following additional properties:*

(1) *The optimal base-stock level  $s_k^*(\mathbf{x}_{-k})$  for component  $k$  is nondecreasing in the on-hand inventory levels of other components. That is,  $s_k^*(\mathbf{x}_{-k})$  is nondecreasing in each of the variables  $x_j$ ,  $j \neq k$ .*

(2) *A unit increase in the inventory level of component  $j$  leads to at most a unit increase in the optimal base-stock level for component  $k$  ( $k \neq j$ ). In other words,  $s_k^*(\mathbf{x}_{-k} + \mathbf{e}_j) \leq s_k^*(\mathbf{x}_{-k}) + 1$ .*

(3) *It is never optimal to interrupt the production of a component once it is initiated.*

(4) *The rationing level  $r_{k,l}^*(\mathbf{x}_{-k})$  for class  $l$  at component  $k$  is nonincreasing in the on-hand inventory levels of other components. That is,  $r_{k,l}^*(\mathbf{x}_{-k})$  is nonincreasing in each of the variables  $x_j$ ,  $j \neq k$ .*

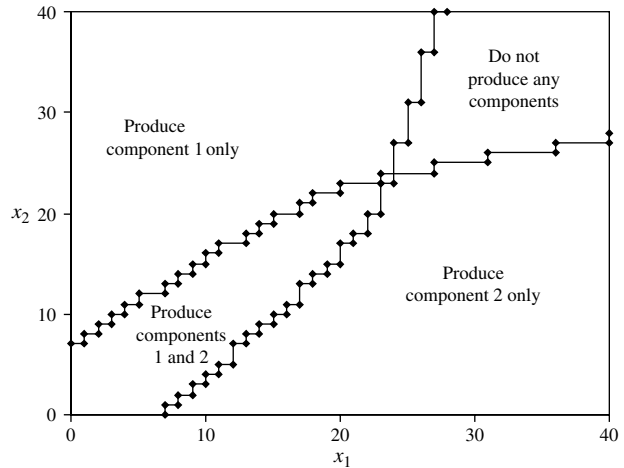
(5) *The rationing levels are ordered as follows:  $r_{k,n}^*(\mathbf{x}_{-k}) \geq \dots \geq r_{k,1}^*(\mathbf{x}_{-k}) = 1$ .*

(6) *It is always optimal to satisfy an order from class 1 if there is on-hand inventory.*

Theorem 1 states that the production of each component is governed by a base-stock policy, with a base-stock level that depends on the inventory level of all the other components such that an increase in the inventory level of one component leads to either an increase in the base-stock level of other components or leaves these base-stock levels unchanged. The theorem also states that inventory allocation is controlled by a multilevel rationing policy that determines whether a demand from a particular class is satisfied. In particular, a demand from a class is satisfied only if the inventory of each component is above its rationing level for that demand class. Note that the rationing levels for a class are not necessarily identical across components. Moreover, an increase in the inventory level of one component could lead to a decrease in the rationing levels at other components.

Theorem 1 applies to two important special cases: (1) systems with multiple components but a single customer class and (2) systems with multiple customer classes but a single component. In the first

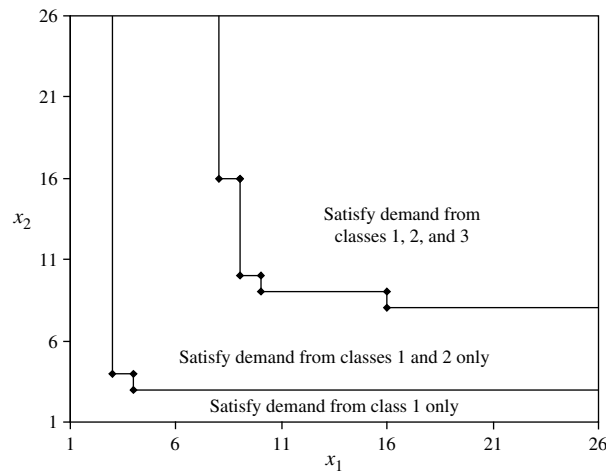
**Figure 1** The Structure of the Optimal Production Policy  
 $(\mu_1 = \mu_2 = 1.0, \lambda_1 = \lambda_2 = \lambda_3 = 0.60, c_1 = 120, c_2 = 60, c_3 = 30, h_1 = h_2 = 1.0, \alpha = 0.0001)$



case, the optimal policy reduces to a production policy only, with state-based base-stock policies for each component, where the base-stock level for each component satisfies properties 1–3 of Theorem 1. In the second case, the optimal production policy is a base-stock policy with a fixed base-stock level, and the optimal inventory allocation policy is a multilevel rationing policy with fixed rationing levels; this is the case treated in Ha (1997a).

The structure of the optimal policy is illustrated in Figures 1 and 2 for a system with three demand classes and two components (details about the computational procedure used to obtain the figures can be found in §5). As we can see from Figure 1, the optimal production policy divides the state space into four regions: region 1 where both components are produced, region 2 where component 1 is produced

**Figure 2** The Structure of the Optimal Allocation Policy  
 $(\mu_1 = \mu_2 = 1.0, \lambda_1 = \lambda_2 = \lambda_3 = 0.60, c_1 = 120, c_2 = 60, c_3 = 30, h_1 = h_2 = 1.0, \alpha = 0.0001)$

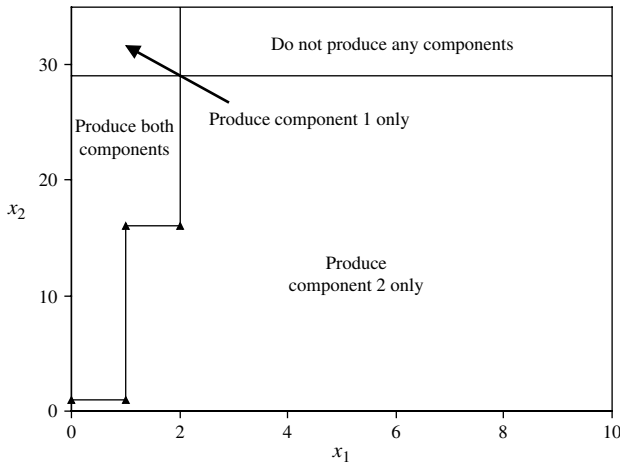


but not 2, region 3 where component 2 is produced but not 1, and region 4 where neither component is produced. It is important to note that other than region 1, the other three regions are transient under the optimal policy. Note also, as suggested by Theorem 1, that the optimal production policy dynamically adjusts the base-stock level of each component based on the inventory level of the other component. In particular, the optimal policy attempts to limit the gap between the inventory levels of the two components. This makes intuitive sense because there is little benefit in continuing to produce one component when the amount of inventory of the other component is significantly smaller.

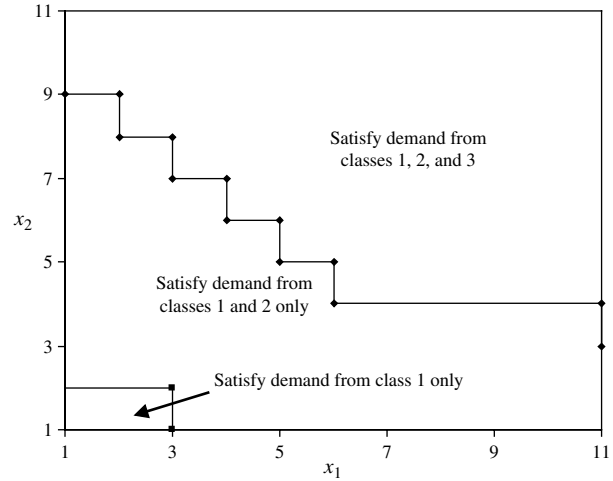
As illustrated in Figure 2, the optimal rationing policy divides the state space into three regions: region 1 where demand from any class is satisfied, region 2 where demand from only classes 1 and 2 is satisfied, and region 3 where only demand from class 1 is satisfied. The boundaries of these regions highlight the fact that the rationing level at one component can be affected in dramatic ways by the inventory level at other components. For example, an increase in the inventory of component 1 from 9 to 10 leads the rationing level for class 2 to drop from 16 to 10. The superposition of Figures 1 and 2 would partition the state space into several subregions, each corresponding to a combination of (a) satisfying one or more demand classes and (b) producing one or both components or not producing at all.

Finally, we must point out that, although the structure induced by the optimal policy is independent of the specific values of system parameters, the curves describing the base-stock and rationing levels are sensitive to these values. For example, in Figures 3 and 4, we show the optimal production and inventory allocation policies for the same system illustrated

**Figure 3** The Structure of the Optimal Production Policy ( $\mu_1 = 5, \mu_2 = 1.0, \lambda_1 = \lambda_2 = \lambda_3 = 0.60, c_1 = 120, c_2 = 60, c_3 = 30, h_1 = 5, h_2 = 1.0, \alpha = 0.0001$ )



**Figure 4** The Structure of the Optimal Allocation Policy ( $\mu_1 = 5, \mu_2 = 1.0, \lambda_1 = \lambda_2 = \lambda_3 = 0.60, c_1 = 120, c_2 = 60, c_3 = 30, h_1 = 5, h_2 = 1.0, \alpha = 0.0001$ )



in Figures 1 and 2 except that the production rate for component 1 is significantly higher than that for component 2 ( $\mu_1 = 5, \mu_2 = 1$ ). As expected, the base-stock level for component 2 is higher than that for component 1. The base-stock level for component 2 in the recurrent region is insensitive to the inventory level for component 1 (component 1 can be replenished relatively quickly and, therefore, does not play a significant role in deciding whether or not component 2 is produced). Not surprisingly, the inventory allocation decisions are also insensitive to the inventory level for component 1 and depend only on that for component 2. Unfortunately, characterizing analytically the impact of component production rates and holding costs on the parameters of the optimal policy appears difficult in general, and we suspect that in most cases there may not be simple relationships.

**3.2. The Case of Average Cost**

In this section, we briefly discuss the case in which the optimization criterion is the average cost per unit time over an infinite planning horizon. The main result is that the structural characteristics of the optimal policy described in the previous section for the discounted cost case carry over to the average cost case. Given a policy  $\pi$ , the average cost is given by

$$J^\pi(\mathbf{x}) = \limsup_{N \rightarrow \infty} \frac{E_x^\pi [\sum_{i=0}^{N-1} (h(\mathbf{X}(t_i)) + \sum_{l=1}^n c_l N_l(t_i))] }{N} \tag{8}$$

The objective is to identify a policy  $\pi^*$  that yields  $J^*(\mathbf{x}) = \inf_\pi J^\pi(\mathbf{x})$  for all states  $\mathbf{x}$ . Such a policy is said to be optimal for the average cost criterion.

**THEOREM 2.** *There exists a stationary policy that is optimal under the average cost criterion. The policy retains*

all the properties of the optimal policy under the discounted cost criterion, namely that (1) the production policy for each component consists of a base-stock policy with a state-dependent base-stock level, (2) the inventory allocation policy consists of a multilevel rationing policy with state-dependent rationing levels, and (3) the base-stock and rationing levels satisfy properties 1–6 in Theorem 1. Furthermore, the optimal average cost is finite and independent of the initial state; that is, there exists a finite constant  $J^*$  such that  $J^*(\mathbf{x}) = J^*$  for all states  $\mathbf{x}$ .

### 3.3. The Case of Systems with Backorders

In this section, we consider systems in which orders that cannot be fulfilled immediately from stock are backordered and not lost. We present results only for the case with a single customer class under the discounted cost criterion. The extension to the average cost case is straightforward, and the structure of the optimal policy remains unchanged from the discounted cost case. Extending the analysis to systems with multiple customer classes appears more difficult, however, and we do not attempt it here (the analysis is more complex because there is a need to include in the state space not only information about net inventory but also backorder level from each class).

When there is a single customer class, the state of the system at time  $t$  can be described by the vector  $\mathbf{Y}(t) = (Y_1(t), \dots, Y_m(t))$ , where  $Y_k(t)$  is an integer that corresponds to the net inventory level of component  $k$  at time  $t$ , such that backorder level  $B(t)$  at time  $t$  is given by  $B(t) = \max_k \{0, -Y_1(t), \dots, -Y_m(t)\}$ , and on-hand inventory of component  $k$  at time  $t$  is given by  $X_k(t) = Y_k(t) + B(t)$ . Let  $b > 0$  denote the backorder cost per unit per unit time. Then, the total cost rate  $z(\mathbf{Y}(t))$  incurred at time  $t$  can be written as

$$z(\mathbf{Y}(t)) = \sum_{k=1}^m h_k(Y_k(t) + B(t)) + bB(t). \quad (9)$$

At any time  $t$ , a control policy  $\pi$  specifies whether or not any of the components should be produced. Let  $\tilde{v}^\pi(\mathbf{y})$  denote the expected discounted cost obtained under a policy  $\pi$  and a starting state  $\mathbf{y} = (y_1, \dots, y_m)$ . Then,  $\tilde{v}^\pi(\mathbf{y})$  is given by

$$\tilde{v}^\pi(\mathbf{y}) = E_{\mathbf{y}}^\pi \left[ \int_0^\infty e^{-\alpha t} z(\mathbf{Y}(t)) dt \right]. \quad (10)$$

Our objective is to choose a policy  $\pi^*$  that minimizes the expected discounted cost. Similar to the case with lost sales, we can restrict our attention to the class of Markovian policies, and we can limit decision epochs to times when the state changes. After rate uniformization and time rescaling, the optimal cost function  $\tilde{v}^* \equiv \tilde{v}^{\pi^*}$  can be shown to satisfy the following optimality equation for any starting state  $\mathbf{y}$ :

$$\tilde{v}^*(\mathbf{y}) = z(\mathbf{y}) + \lambda \tilde{v}^*(\mathbf{y} - \mathbf{e}) + \sum_{k=1}^m \mu_k \tilde{T}_k \tilde{v}^*(\mathbf{y}). \quad (11)$$

The operator  $\tilde{T}_k$  ( $k = 1, \dots, m$ ) is defined for any real valued function  $v$  over the state space as

$$\tilde{T}_k v(\mathbf{y}) = \min\{v(\mathbf{y} + \mathbf{e}_k), v(\mathbf{y})\} \quad (12)$$

and is associated with the decision of whether or not component  $k$  should be produced. This decision is made each time a new order arrives or any component completes production.

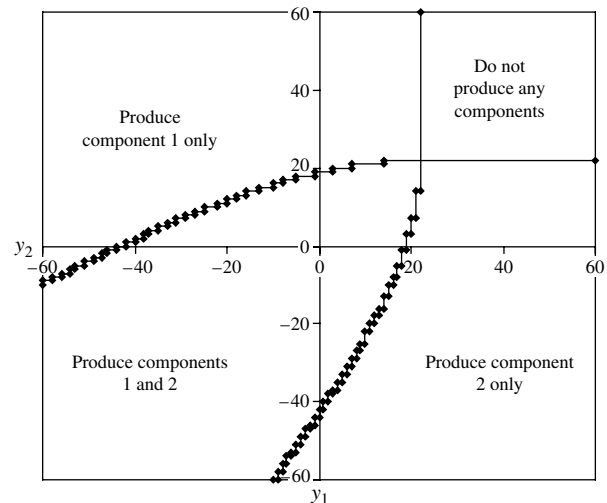
Theorem 3 characterizes the structure of the optimal policy that retains all the properties observed in systems with lost sales.

**THEOREM 3.** *There exists an optimal stationary policy that consists of a state-dependent base-stock policy with a vector of state-dependent base-stock levels  $\mathbf{s}^* = (s_1^*(\mathbf{y}_{-1}), \dots, s_m^*(\mathbf{y}_{-m}))$  such that it is optimal to produce component  $k$  if  $y_k < s_k^*(\mathbf{y}_{-k})$  and not to produce it otherwise, where  $y_k$  is the net inventory of component  $k$  and  $\mathbf{y}_{-k} = (y_1, \dots, y_{k-1}, y_{k+1}, \dots, y_m)$  is an  $(m - 1)$  dimensional vector consisting of net inventory levels for all components  $j \neq k$ . The optimal policy has the following additional properties:*

- (1) *The optimal base-stock level  $s_k^*(\mathbf{y}_{-k})$  for component  $k$  is nondecreasing in the net inventory levels of other components. That is,  $s_k^*(\mathbf{y}_{-k})$  is nondecreasing in each of the variables  $y_j$ ,  $j \neq k$ .*
- (2) *A unit increase in the net inventory of component  $j$  leads to at most a unit increase in the optimal base-stock level for component  $k$  ( $k \neq j$ ). In other words,  $s_k^*(\mathbf{y}_{-k} + \mathbf{e}_j) \leq s_k^*(\mathbf{y}_{-k}) + 1$ .*
- (3) *It is never optimal to interrupt the production of a component once it is initiated.*

The structure of the optimal production policy is illustrated in Figure 5 for a system with two components. Although the structure is similar to the one

**Figure 5** The Structure of the Optimal Production Policy for Systems with Backorders ( $\mu_1 = \mu_2 = 1.0$ ,  $\lambda = 0.9$ ,  $b = 10$ ,  $h_1 = h_2 = 1.0$ ,  $\alpha = 0.0001$ )





with lost sales, notice here that the base-stock levels may not always be positive. It is possible not to produce a component even though its net inventory is negative. This is not surprising because the net inventory of an item can be negative but its on-hand inventory is positive. For example, if  $y_1 = -7$  and  $y_2 = -2$ , then backorder level is 7 and on-hand inventory level of component 1 is zero ( $x_1 = 0$ ), but on-hand inventory level of component 2 is 5 ( $x_2 = 5$ ). Hence, it may not be desirable to produce component 2 until the net inventory level of component 1 catches up sufficiently.

**4. Heuristic Policies**

In this section, we propose two heuristics whose performance we compare with that of the optimal policy in §5. The first heuristic consists of controlling the production of components independently of each other via a vector  $\mathbf{s} = (s_1, \dots, s_m)$  of fixed (i.e., state-independent) base-stock levels and controlling acceptance or rejection of orders via a vector  $\mathbf{r}_l = (r_{1,l}, \dots, r_{m,l})$  of fixed rationing levels for each class  $l$ ,  $l = 1, \dots, n$ . We refer to this policy as the *independent base-stock with rationing* (IBR) policy. The expected discounted cost of the IBR policy can be obtained for a system with multiple classes and lost sales (for a given vector of base-stock levels  $\mathbf{s}$  and a matrix of rationing levels  $\mathbf{r} = [r_{k,l}]$ ) via the following dynamic programming equation:

$$v^{\text{IBR}}(\mathbf{x}) = h(\mathbf{x}) + \sum_{l=1}^n \lambda_l T^{l, \text{IBR}} v^{\text{IBR}}(\mathbf{x}) + \sum_{k=1}^m \mu_k T_k^{\text{IBR}} v^{\text{IBR}}(\mathbf{x}), \tag{13}$$

where

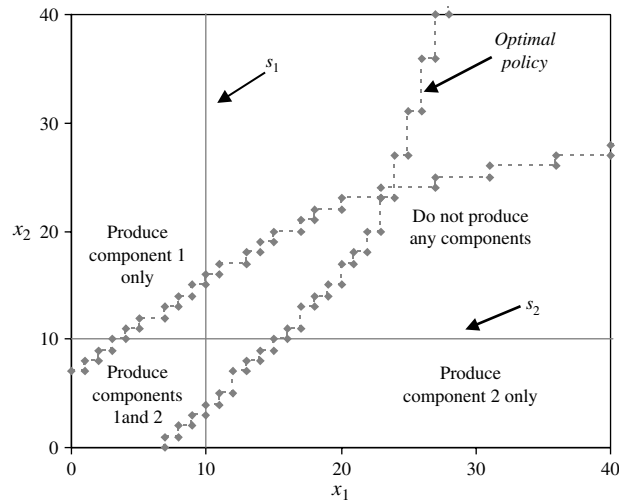
$$T_k^{\text{IBR}} v(\mathbf{x}) = \begin{cases} v(\mathbf{x} + \mathbf{e}_k) & \text{if } x_k < s_k, \\ v(\mathbf{x}) & \text{otherwise,} \end{cases} \quad \text{and} \tag{14}$$

$$T^{l, \text{IBR}} v(\mathbf{x}) = \begin{cases} v(\mathbf{x} - \mathbf{e}) & \text{if } \mathbf{x} \geq \mathbf{r}_l, \\ v(\mathbf{x}) + c_l & \text{otherwise,} \end{cases}$$

with the inequality  $\mathbf{x} \geq \mathbf{r}_l$  taken component-wise, where  $\mathbf{r}_l = (r_{1,l}, \dots, r_{kl})$ .

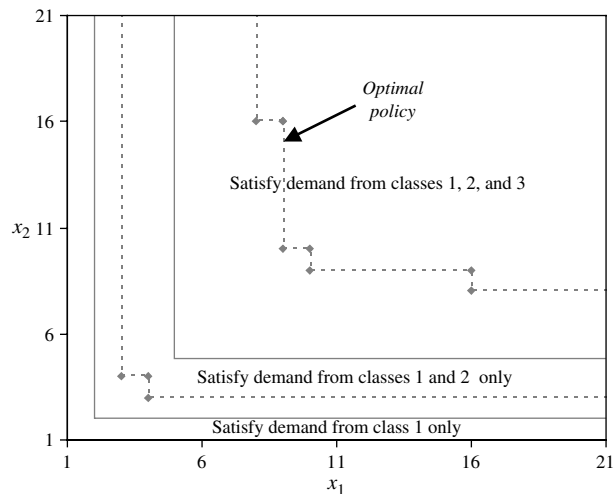
The IBR policy is graphically illustrated in Figures 6 and 7 for a system with two components and three customer classes (for comparison purposes, we also show the optimal policy in the background). Similar to the optimal policy, the base-stock levels define four regions in which we produce both components, one of the components, or none of the components, and the rationing levels define three regions in which we satisfy all customer classes, only customer classes 1 and 2, or only customer class 1. However, in contrast

**Figure 6** The Structure of the IBR Production Policy, with  $s_k, k = 1, 2$ , Indicating the Base-Stock Level Used for Component  $k$  ( $\mu_1 = \mu_2 = 1.0, \lambda_1 = \lambda_2 = \lambda_3 = 0.60, c_1 = 120, c_2 = 60, c_3 = 30, h_1 = h_2 = 1.0, \alpha = 0.0001$ )



to the optimal policy, the base-stock and rationing levels for each component are fixed and not affected by the inventory levels of other components. The advantage of the IBR policy is that it is much simpler to implement than the optimal policy because it does not require the storage of the state-dependent base-stock levels. This is important when the number of components is large and it becomes difficult to compute the optimal policy because of the exponential growth in the size of the state space. However, the IBR policy is clearly suboptimal and lacks the coordination in managing the production of different components carried out under the optimal policy.

**Figure 7** The Structure of the IBR Inventory Allocation Policy ( $\mu_1 = \mu_2 = 1.0, \lambda_1 = \lambda_2 = \lambda_3 = 0.60, c_1 = 120, c_2 = 60, c_3 = 30, h_1 = h_2 = 1.0, \alpha = 0.0001$ )



The second heuristic is similar to the IBR policy, except that we now introduce a coordination parameter  $R$ , such that the production of a component  $k$  is not initiated unless the difference between the inventory level of component  $k$  and the smallest inventory level among the other components is less than  $R$ . We refer to this policy as the *coordinated base-stock with rationing* (CBR) policy. Expected cost for the CBR policy can be obtained via the following dynamic programming equation:

$$v^{\text{CBR}}(\mathbf{x}) = h(\mathbf{x}) + \sum_{l=1}^n \lambda_l T^{l, \text{CBR}} v^{\text{CBR}}(\mathbf{x}) + \sum_{k=1}^m \mu_k T_k^{\text{CBR}} v^{\text{CBR}}(\mathbf{x}), \quad (15)$$

where

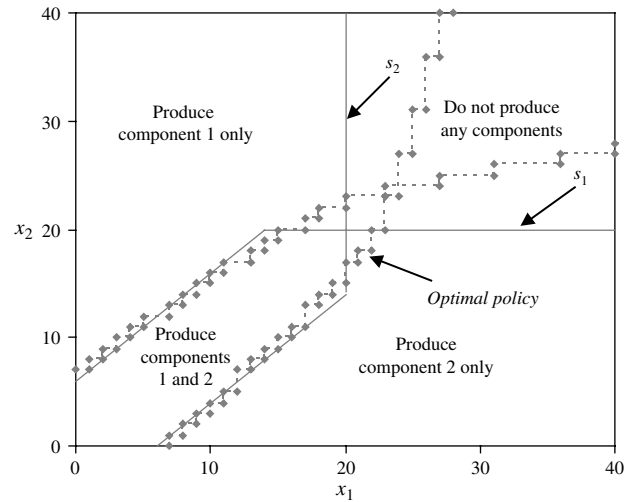
$$T_k^{\text{CBR}} v(\mathbf{x}) = \begin{cases} v(\mathbf{x} + \mathbf{e}_k) & \text{if } x_k < \min_{j \neq k} \{s_j, x_j + R\}, \\ v(\mathbf{x}) & \text{otherwise,} \end{cases} \quad \text{and}$$

$$T^{l, \text{CBR}} v(\mathbf{x}) = \begin{cases} v(\mathbf{x} - \mathbf{e}) & \text{if } \mathbf{x} \geq \mathbf{r}_l, \\ v(\mathbf{x}) + c_l & \text{otherwise.} \end{cases} \quad (16)$$

The CBR policy is motivated by the fact that, under the optimal policy, production of the various components is coordinated. An increase (decrease) in the inventory level of one component tends to increase (decrease) the optimal base-stock level of other components. The CBR policy attempts to mimic the behavior of the optimal policy by monitoring the difference in inventory levels between different components and stopping production of a component when this difference becomes too large; see Figure 8. This is different from the IBR policy, which does not attempt any coordination and manages the production of each component independently. Note that the IBR policy is a special case of the CBR policy and corresponds to the case where  $R$  is greater than or equal to  $\max(s_1, \dots, s_m)$ . Hence, the CBR policy will always outperform the IBR policy, with the cost of the IBR policy providing an upper bound on the CBR. The CBR policy can be further refined by letting the coordination parameter be component-dependent. This leads to a coordination vector  $\mathbf{R} = (R_1, \dots, R_m)$ , where  $R_k$  is the coordination parameter used by component  $k$ ,  $k = 1, \dots, m$ .

When the number of components is small, it is straightforward to evaluate the cost under the IBR and CBR heuristics (using the policy-evaluation version of the value iteration algorithm) for a given combination of the control parameters (i.e., base-stock and rationing levels and coordination parameter). When

**Figure 8** The Structure of the CBR Production Policy ( $\mu_1 = \mu_2 = 1.0$ ,  $\lambda_1 = \lambda_2 = \lambda_3 = 0.60$ ,  $c_1 = 120$ ,  $c_2 = 60$ ,  $c_3 = 30$ ,  $h_1 = h_2 = 1.0$ ,  $\alpha = 0.0001$ )



the number of components is large, exact computation of the cost becomes difficult because of the exponential growth in the size of the state space (these are the same instances in which evaluating the optimal policy is also computationally intractable). However, for a given vector of base-stock levels, a matrix of rationing levels, and a coordination parameter, it is not difficult to evaluate the heuristics using simulation, even when the state space is large. Determining optimal values for the control parameters can be carried out via an exhaustive search over a sufficiently large range of feasible values. Because of the combinatorial nature of the problem, this is possible only when the number of components is small. For large problems, a heuristic search may be needed. This, of course, would not guarantee the optimality of the resulting parameters but may lead to reasonably good approximations.

We close this section by noting that the heuristics proposed for systems with lost sales can be used for systems with backorders. In particular, for a system with a single customer class, the production of each component  $k$  could be managed using an independent fixed base-stock level  $s_k$  so that the component is produced if and only if its net inventory level  $y_k < s_k$ . A coordination parameter  $R$  could be introduced, as in the case of lost sales, so that production of a component  $k$  takes place if and only if  $y_k < \min_{j \neq k} \{s_j, y_j + R\}$ .

## 5. Numerical Results

To test the performance of the optimal policy against the IBR and CBR heuristics, we carried out a series of numerical experiments for a system with two components and for a wide range of parameter values. For each problem instance, numerical results are obtained by solving the corresponding dynamic programming equation using the value iteration method. The value

iteration algorithm we use is a direct adaptation of the algorithm described in Puterman (1994, Chapter 8). We present results for the average cost criterion because they are independent of the starting state and the discount factor. The state space is truncated at  $\{-n_1^{\min}, n_1^{\max}\} \times \{-n_2^{\min}, n_2^{\max}\}$ , where  $n_k^{\min}$  and  $n_k^{\max}$ ,  $k=1,2$ , are positive integers that are gradually increased until the cost is no longer sensitive to the truncation level (for the case of systems with lost sales,  $n_k^{\min}=0$ ). The value iteration algorithm is terminated once five-digit accuracy is obtained. Note that the value iteration algorithm suffers from “the curse of dimensionality” that afflicts dynamic programming algorithms. Because the computational effort grows exponentially in the dimensions of the state space, solving problems with more than two components is generally difficult. See Puterman for further discussion of computational complexity and convergence of the value iteration algorithm.

To identify good values for the control parameters for the IBR and CBR heuristics in the case of lost sales, we use the following heuristic search procedure (a slightly modified version of this procedure is used in the case of backorders): For the base-stock levels, we exhaustively search over the region  $\{0, s_1^{\max}\} \times \{0, s_2^{\max}\}$ , where  $s_k^{\max}$  is the largest base-stock value for component  $k$ ,  $k=1,2$ , that is observed under the optimal policy in the recurrent region. For each combination of base-stock values, we search exhaustively over all feasible values for the rationing levels and coordination parameter. For the rationing levels, this means examining for each component all values that do not exceed the base-stock level for that component. For the coordination parameter, we examine all values that do not exceed the largest of the two component base-stock levels.

### 5.1. Systems with a Single Customer Class

In this section, we present numerical results for systems with a single customer class. Results for systems with multiple classes are presented in §5.2. We first consider the case of lost sales and then discuss the case with backorders. For all problem instances in this section and in §5.2, we use linear holding costs with  $h_1 \geq 0$  and  $h_2 \geq 0$  being the unit holding cost rates for components 1 and 2, respectively. Representative results for systems with lost sales are shown in Tables 1 and 2. The parameter values for the 50 examples shown were generated randomly. The results are consistent with a larger set of examples we examined, including examples in which we systemically varied the value of one parameter while fixing the values of all others (for brevity, results from this more extensive data set are not included but are available from the authors on request). The following two observations can be made. Perhaps surprisingly, the percentage cost difference between the optimal policy and

the CBR heuristic is relatively small. For the instances shown, the average percentage difference is 1.14%, with a minimum of 0% and a maximum of 4.65%. The percentage cost difference between the optimal and IBR heuristic is also small, with an average percentage of 1.41%, a minimum of 0%, and a maximum of 4.75%.

We carried out similar experiments for systems with backorders. In this case, we found the performance of the heuristics to be sensitive to parameter values. Therefore, instead of presenting results with randomly generated data, we present results from three specific sets of examples. In each set, we vary the value of a single parameter while keeping all other parameter values constant. In the first set, we vary the backorder cost  $b$ , in the second set the demand rate  $\lambda$ , and in the third set the ratio  $\mu_1/\mu_2$ . We also include results from a fourth data set that represents rather extreme scenarios where the ratio  $\mu_1/\mu_2$  is high, demand rate is very low, and backorder cost is very small. Representative results from each set are shown in Tables 3 and 4. Based on these results, the following observations can be made.

In the first three data sets, the percentage cost difference between the optimal policy and the heuristics is generally small. However, there are cases for which this difference is significant; nearly 9% for the CBR policy and in excess of 50% for the IBR policy in some examples. In the extreme scenarios of the fourth data set, the heuristics perform poorly with very large percentage cost differences between the optimal policy and the heuristics. However, in these cases, although the percentage cost differences are large, the absolute cost differences are relatively small. The CBR policy significantly outperforms the IBR policy in some cases, with the instances in which the difference between the CBR and IBR policies is large are the same as those in which the difference between the CBR and the optimal policy is also large.

In comparing the numerical results obtained for systems with backorders and systems with lost sales, it appears that the heuristics are less effective in the case of backorders. Also, although there is little difference between the IBR and the CBR policies in the case of lost sales, that difference can be significant in the case of backorders. These differences between the backorder and lost sales case are somewhat surprising. They suggest that the benefit of coordinating production among components, as carried out by either the optimal policy or the CBR heuristic, can be much more valuable for systems with backorders. A possible explanation is that in systems with backorders, the region of the state space over which coordination can take place is larger (the recurrent region is finite for lost sales but not for backorders). This is significant because the difference in inventory

**Table 1** Optimal Policy vs. Heuristics

	$\mu_1$	$\mu_2$	$\lambda$	$h_1$	$h_2$	$c$	Average optimal cost	Percentage difference from optimal cost	
								CBR	IBR
1	3.742	2.707	2.741	7.14	3.73	108.79	79.12	2.309	2.354
2	2.358	7.281	4.405	8.74	8.68	119.12	266.58	1.600	1.600
3	5.469	9.098	8.395	6.80	8.36	132.39	422.11	1.628	1.628
4	4.078	3.608	4.071	5.81	7.54	62.55	99.29	0.344	1.662
5	8.547	6.113	4.334	7.33	5.92	89.53	72.09	0.463	0.495
6	7.251	6.592	8.153	9.61	5.70	176.15	376.90	2.130	2.179
7	2.557	9.818	3.443	3.27	8.88	147.72	154.59	2.370	2.370
8	6.760	3.070	7.130	6.99	2.21	5.48	30.12	4.642	4.745
9	3.560	5.223	1.583	9.90	6.25	85.28	44.85	0.098	0.098
10	5.640	4.006	4.896	3.03	6.22	152.31	182.89	1.867	1.872
11	5.768	6.765	2.882	4.42	8.05	136.49	56.63	0.273	0.397
12	5.150	6.111	8.148	1.53	6.43	11.00	50.67	2.968	3.039
13	9.911	8.100	4.948	5.49	2.93	129.06	54.57	0.119	0.129
14	3.880	9.641	7.540	4.71	7.70	54.32	213.99	1.779	1.779
15	1.470	9.480	2.350	4.46	3.80	34.54	42.47	4.131	4.131
16	4.959	9.400	7.150	2.91	8.55	126.13	300.17	1.654	1.654
17	2.204	2.864	6.465	6.67	4.33	115.45	512.35	0.183	2.257
18	5.063	1.395	1.245	3.81	1.12	77.41	20.75	1.162	1.168
19	7.148	1.836	1.318	6.51	6.48	4.14	5.45	0.000	0.000
20	1.147	2.711	6.282	1.52	4.31	126.66	652.81	0.225	0.225
21	7.459	7.234	1.757	5.09	4.98	71.30	26.24	0.000	0.000
22	2.383	7.081	7.293	7.55	5.31	111.41	557.11	0.555	0.555
23	2.089	5.057	7.443	9.04	3.46	51.70	286.04	0.482	0.482
24	8.790	3.091	8.244	9.18	3.09	48.62	258.93	1.232	1.232
25	1.448	1.706	6.767	2.72	8.60	35.61	202.39	0.729	0.743
26	2.537	9.949	4.958	4.06	3.83	73.65	187.25	1.323	1.323
27	4.539	6.324	2.078	1.34	5.13	174.10	29.19	0.348	0.353
28	9.408	3.380	2.443	8.86	3.14	129.52	51.01	1.935	1.935
29	9.702	6.984	8.833	1.09	2.23	163.93	318.35	0.458	0.458
30	4.872	9.013	7.614	7.19	4.12	34.04	118.28	2.239	2.247
31	2.401	2.720	4.802	8.70	5.41	163.37	437.95	0.217	3.674
32	5.147	5.116	5.056	4.71	9.12	2.11	10.67	0.000	0.000
33	3.677	1.443	7.239	6.85	9.85	110.98	653.74	0.338	0.338
34	4.601	2.789	6.627	7.60	4.38	2.97	19.37	1.452	1.452
35	4.779	7.783	8.145	9.28	8.60	74.18	285.75	2.039	2.043
36	6.587	7.582	2.745	9.14	6.12	126.73	58.71	0.010	0.030
37	3.110	5.939	9.384	4.02	6.90	78.99	505.41	0.660	0.660
38	6.646	7.292	4.575	4.72	6.90	167.68	86.68	0.153	0.311
39	4.345	4.827	6.352	6.09	7.45	102.75	265.91	1.301	1.386
40	7.988	5.404	2.673	7.31	9.84	161.52	70.45	0.097	0.135
41	7.332	5.365	2.032	6.98	4.29	28.87	25.06	1.042	2.130
42	6.101	8.407	7.066	10.00	9.66	12.71	60.43	0.446	2.171
43	4.243	5.937	3.356	6.38	1.44	114.64	57.52	0.170	0.495
44	7.308	9.661	7.755	7.66	4.89	127.22	178.24	1.718	1.738
45	8.227	1.755	9.509	9.24	6.42	51.46	404.26	1.109	1.109
46	8.861	5.621	7.594	4.80	9.65	15.34	66.30	0.417	1.682
47	5.981	3.628	8.722	4.02	7.12	11.64	73.66	0.871	1.618
48	4.210	5.485	4.910	6.06	6.55	23.56	59.74	1.022	1.691
49	9.084	7.791	8.120	8.34	7.03	40.97	111.74	1.139	1.526
50	5.700	9.400	7.420	3.05	5.05	35.27	86.84	3.559	3.561

Note. A sample of examples randomly generated with parameter values drawn from uniform distributions as follows:  $\mu_k \sim U(1, 10)$ ,  $\lambda \sim U(1, 10)$ ,  $h_k \sim U(1, 10)$ ,  $c \sim U(1, 200)$ .

levels between components always stays bounded for lost sales, whereas this difference can grow infinitely large for systems with backorders. For systems with backorders, it is possible for one component to fall behind in production and accumulate a large backlog while the other components have little or no back-

log. This is more likely to occur when the demand rate is high, the production rate for one of the components is low, or the backorder penalty is small, which are indeed the cases for which we observe the largest percentage cost differences. Another important difference between systems with lost sales and those

**Table 2** Control Parameters for the Optimal and Heuristic Policies for the Cases of Table 1

	Optimal [ $s_1^{\max}, s_2^{\max}$ ]	CBR [ $s_1, s_2, R$ ]	IBR [ $s_1, s_2$ ]
1	[5, 10]	[5, 10, 8]	[5, 10]
2	[25, 5]	[18, 2, 18]	[18, 2]
3	[47, 12]	[33, 4, 31]	[33, 4]
4	[7, 7]	[6, 7, 3]	[5, 6]
5	[4, 6]	[4, 6, 5]	[4, 6]
6	[20, 32]	[10, 24, 19]	[10, 23]
7	[36, 4]	[31, 2, 30]	[31, 2]
8	[2, 5]	[1, 4, 3]	[1, 3]
9	[2, 2]	[2, 2, 2]	[2, 2]
10	[12, 23]	[7, 20, 16]	[7, 19]
11	[5, 4]	[5, 4, 3]	[5, 4]
12	[10, 5]	[8, 3, 7]	[8, 3]
13	[5, 7]	[5, 7, 6]	[5, 7]
14	[34, 6]	[22, 2, 21]	[22, 2]
15	[7, 2]	[5, 1, 5]	[5, 1]
16	[59, 8]	[50, 3, 49]	[50, 3]
17	[24, 24]	[15, 4, 12]	[15, 19]
18	[2, 8]	[2, 9, 8]	[2, 9]
19	[0, 0]	[0, 0, 0]	[0, 0]
20	[24, 11]	[24, 1, 24]	[24, 1]
21	[2, 2]	[2, 2, 2]	[2, 2]
22	[24, 8]	[24, 2, 23]	[24, 2]
23	[24, 15]	[12, 2, 11]	[11, 2]
24	[5, 29]	[1, 21, 20]	[1, 20]
25	[21, 16]	[7, 2, 6]	[6, 2]
26	[34, 5]	[32, 2, 31]	[32, 2]
27	[6, 4]	[6, 4, 5]	[6, 4]
28	[3, 7]	[3, 7, 7]	[3, 7]
29	[23, 84]	[8, 84, 84]	[8, 84]
30	[12, 7]	[9, 3, 8]	[9, 3]
31	[29, 29]	[15, 15, 6]	[14, 14]
32	[0, 0]	[0, 0, 0]	[0, 0]
33	[15, 24]	[1, 12, 12]	[1, 12]
34	[1, 1]	[0, 0, 0]	[0, 0]
35	[21, 10]	[13, 3, 11]	[13, 3]
36	[3, 3]	[3, 3, 2]	[2, 2]
37	[34, 14]	[34, 2, 33]	[34, 2]
38	[8, 6]	[8, 6, 5]	[8, 6]
39	[22, 17]	[14, 7, 11]	[13, 7]
40	[3, 4]	[3, 4, 3]	[3, 4]
41	[2, 2]	[2, 2, 1]	[2, 2]
42	[3, 3]	[3, 2, 1]	[2, 2]
43	[7, 7]	[7, 7, 4]	[7, 6]
44	[16, 10]	[15, 8, 12]	[15, 8]
45	[7, 24]	[1, 20, 19]	[1, 20]
46	[4, 4]	[3, 4, 2]	[3, 4]
47	[5, 6]	[3, 4, 2]	[2, 3]
48	[5, 4]	[4, 3, 2]	[4, 3]
49	[6, 8]	[5, 7, 4]	[5, 7]
50	[17, 6]	[15, 4, 14]	[15, 4]

*Note.* For the optimal policy,  $s_k^{\max}$  is the maximum base-stock level observed for components  $k$  in the recurrent region; for the heuristics,  $s_k$  is the selected base-stock level for component  $k$  and  $R$  is the coordination parameter.

with backorders is that in the case of the former, it is always possible to set the base-stock levels at zero and reject demand from all classes. This leads to an average cost of  $\sum_{l=1}^n \lambda_l c_l$ , which provides a lower bound on the cost of the heuristics. No such policy and no

such bound exist in the case of backorders in which the backlog, depending on system parameter values, can be arbitrarily large.

These results suggest four important implications for operations managers in practice: (1) simple heuristics with static parameters can be effective in managing ATO systems; (2) the effectiveness of the heuristics is greater in systems with lost sales than in systems with backorders; (3) for systems with backorders, coordination in the production of different components appears to matter more, and therefore a heuristic such as the CBR heuristic can be more useful; and (4) coordinating production is more important when there is a potential for a significant backlog, such as when the utilization of one or more of the production facilities is high.

**5.2. Systems with Multiple Customer Classes**

For systems with multiple classes and lost sales, we carried out a series of numerical experiments first to evaluate the performance of the heuristics, second to examine the impact of differences in the shortage costs among different classes, and third to assess the benefit of rationing inventory among different classes as indicated by the optimal policy. We consider a system with two components and two demand classes, 1 and 2, with lost sale costs  $c_1$  and  $c_2$ , respectively. We obtain the average cost for the system under (1) the optimal policy, (2) a policy in which orders are fulfilled on a first come–first served (FCFS) basis, regardless of their class, and (3) the IBR and CBR heuristics. Under the FCFS policy, the optimal average cost  $J^{\text{FCFS}}$  is obtained by solving the following dynamic programming equation:

$$f^{\text{FCFS}}(\mathbf{x}) + J^{\text{FCFS}} = h(\mathbf{x}) + \sum_{l=1}^n \lambda_l T_l^{\text{FCFS}} f^{\text{FCFS}}(\mathbf{x}) + \sum_{k=1}^m \mu_k T_k f^{\text{FCFS}}(\mathbf{x}), \tag{17}$$

where

$$T_l^{\text{FCFS}} f(\mathbf{x}) = \begin{cases} f(\mathbf{x}) + c_l & \text{if } \prod_{k=1}^m x_k = 0, \\ f(\mathbf{x} - \mathbf{e}) & \text{otherwise,} \end{cases} \tag{18}$$

and  $T_k$  is as defined in (6).

Representative results from our experiments are shown in Figure 9 and Table 5. In Figure 9(a), we compare the average cost of the optimal policy to that of the FCFS policy as we vary the ratio  $c_1/c_2$  by increasing  $c_1$  and holding  $c_2$  fixed. In Figure 9(b), we do the same except that we vary the ratio  $c_1/c_2$  by increasing  $c_1$  and decreasing  $c_2$  while holding  $c_1 + c_2$  constant. In Table 5, we compare the average cost of

**Table 3** Optimal Policy vs. the Heuristics for Systems with Backorders

	$\mu_1$	$\mu_2$	$\lambda$	$h_1$	$h_2$	$b$	Average optimal cost	Percentage difference from optimal cost	
								CBR	IBR
1	1	1	0.8	1	1	0.1	2.51	8.933	49.851
2	—	—	—	—	—	0.3	3.89	5.604	22.806
3	—	—	—	—	—	0.5	4.98	6.669	14.960
4	—	—	—	—	—	0.7	5.92	4.665	9.799
5	—	—	—	—	—	1	7.12	2.099	7.165
6	—	—	—	—	—	2	10.25	1.096	2.878
7	—	—	—	—	—	4	14.37	0.908	1.328
8	—	—	—	—	—	6	17.18	0.453	0.758
9	—	—	—	—	—	10	21.06	0.287	0.411
10	1	1	0.1	1	1	1	0.27	0.111	1.221
11	—	—	0.2	—	—	—	0.60	0.840	1.882
12	—	—	0.3	—	—	—	1.00	0.110	2.096
13	—	—	0.4	—	—	—	1.52	0.408	2.998
14	—	—	0.5	—	—	—	2.21	1.735	4.594
15	—	—	0.6	—	—	—	3.00	1.577	5.718
16	—	—	0.7	—	—	—	4.41	2.589	6.814
17	—	—	0.8	—	—	—	7.12	2.099	7.165
18	—	—	0.9	—	—	—	15.19	1.977	7.371
19	1.000	1.000	0.6	1	1	5	6.80	0.503	0.832
20	1.024	0.976	—	—	—	—	6.83	0.658	1.044
21	1.048	0.952	—	—	—	—	6.92	1.065	1.607
22	1.130	0.870	—	—	—	—	7.52	1.098	1.749
23	1.149	0.851	—	—	—	—	7.78	1.419	2.109
24	1.216	0.784	—	—	—	—	9.13	2.506	3.792
25	1.245	0.755	—	—	—	—	10.20	3.243	4.509
26	1.333	0.667	—	—	—	—	19.26	5.339	6.804
27	1.355	0.645	—	—	—	—	26.86	6.034	7.463
28	1.0	0.1	0.08	1	1	0.1	0.50	129.5	596.9
29	—	—	—	2	—	—	0.50	250.0	981.8
30	—	—	—	3	—	—	0.50	370.4	1,261.6
31	—	—	—	4	—	—	0.50	490.3	1,479.6
32	—	—	—	5	—	—	0.50	610.7	1,655.9
33	—	—	—	6	—	—	0.50	731.1	1,808.3
34	—	—	—	7	—	—	0.50	851.6	1,940.6
35	—	—	—	8	—	—	0.50	972.0	2,053.8
36	—	—	—	9	—	—	0.50	1,092.4	2,159.7

the optimal policy to the average cost of the heuristics as we vary the ratio  $c_1/c_2$  while holding  $c_1 + c_2$  fixed. Based on these numerical results, the following observations can be made:

- As shown in Table 5, both the CBR and IBR heuristics perform well for a wide range of values of  $c_1/c_2$ . For the cases shown, the maximum percentage difference between any of the two heuristics and the optimal policy is less than 1.5%. This is consistent with the results obtained for systems with a single customer class and supports the argument that simple heuristics can be effective for systems with lost sales.
- As illustrated in Figure 9(a), the difference in average cost between the optimal policy and the FCFS policy increases dramatically as the lost sale cost of one of the demand classes increases. For large values of  $c_1/c_2$ , this difference can be significant. For example, when  $c_1/c_2 = 10$ , the average cost of the FCFS

policy is more than triple that of the optimal policy. Although the average cost increases under both policies, the increase under the optimal policy is relatively modest even when the ratio  $c_1/c_2$  is high, whereas the increase under the FCFS is linear.

- As shown in Figure 9(b), the difference in average cost between the optimal policy and the FCFS policy also increases as one demand class becomes more expensive and the other less expensive, with the differences becoming significant when  $c_1/c_2$  is large. However, in contrast to the results of Figure 9(a), the average cost of the FCFS policy is unaffected by changes in  $c_1/c_2$ , whereas that of the optimal policy actually decreases.

The behavior of the FCFS policy with respect to changes in  $c_1$  and  $c_2$  can be explained as follows. The probability of a rejected order being of class  $l$  is  $\lambda_l / \sum_l \lambda_l$ , and the expected lost sales cost of a

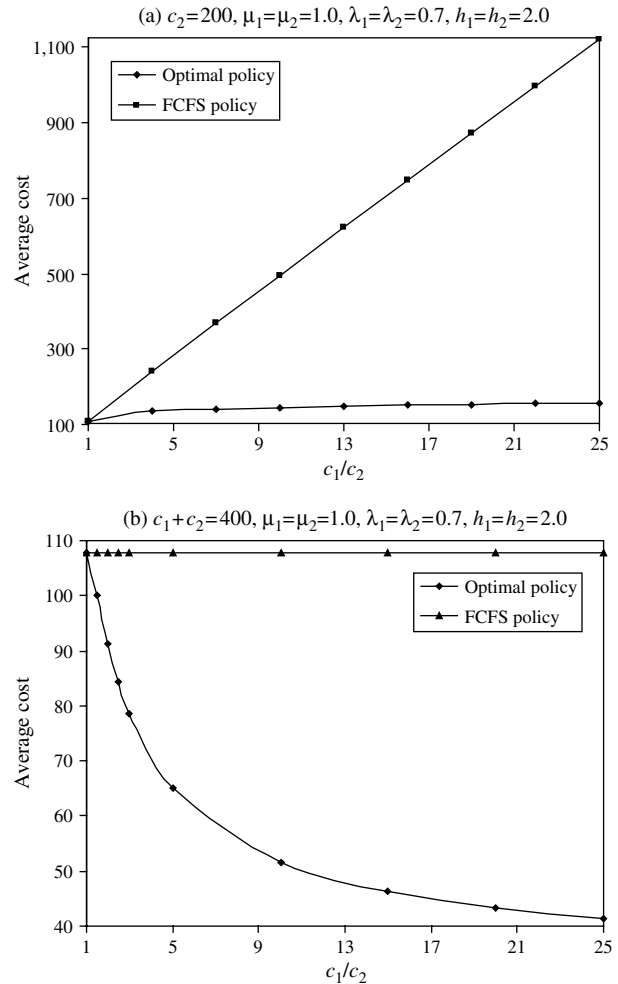
**Table 4** Control Parameters for the Optimal and Heuristic Policies for Systems with Backorders for the Cases in Table 3

	Optimal [ $s_1^{\max}, s_2^{\max}$ ]	CBR [ $s_1, s_2, R$ ]	IBR [ $s_1, s_2$ ]
1	[2,2]	[1, 1, 4]	[0, 0]
2	[2,2]	[1, 1, 4]	[1, 1]
3	[3,3]	[1, 1, 9]	[1, 1]
4	[3,3]	[2, 2, 5]	[2, 2]
5	[4,4]	[3, 3, 7]	[3, 3]
6	[6,6]	[4, 4, 10]	[4, 4]
7	[8,8]	[7, 7, 12]	[6, 6]
8	[9,9]	[8, 8, 15]	[8, 8]
9	[11, 11]	[10, 10, 17]	[10, 10]
10	[1, 1]	[0, 0, 1]	[0, 0]
11	[1, 1]	[0, 0, 1]	[0, 0]
12	[1, 1]	[0, 0, 2]	[0, 0]
13	[1, 1]	[0, 0, 2]	[0, 0]
14	[2, 2]	[0, 0, 3]	[0, 0]
15	[2, 2]	[1, 1, 3]	[1, 1]
16	[3, 3]	[1, 1, 5]	[1, 1]
17	[4, 4]	[3, 3, 7]	[3, 3]
18	[8, 8]	[6, 6, 17]	[6, 6]
19	[4, 4]	[3, 3, 6]	[3, 3]
20	[4, 4]	[3, 3, 6]	[3, 3]
21	[4, 4]	[3, 3, 6]	[3, 3]
22	[3, 5]	[2, 4, 6]	[2, 4]
23	[3, 5]	[2, 4, 6]	[2, 5]
24	[3, 7]	[2, 5, 6]	[2, 6]
25	[3, 8]	[2, 7, 7]	[2, 8]
26	[3, 17]	[2, 16, 16]	[2, 16]
27	[3, 25]	[2, 25, 24]	[2, 25]
28	[1, 1]	[0, 1, 1]	[-1, 1]
29	[1, 1]	[0, 1, 1]	[-3, 1]
30	[1, 1]	[0, 1, 1]	[-4, 1]
31	[1, 1]	[0, 1, 1]	[-5, 1]
32	[1, 1]	[0, 1, 1]	[-6, 1]
33	[1, 1]	[0, 1, 1]	[-7, 1]
34	[1, 1]	[0, 1, 1]	[-7, 1]
35	[1, 1]	[0, 1, 1]	[-8, 1]
36	[1, 1]	[0, 1, 1]	[-8, 1]

rejected order is  $\sum_i c_i (\lambda_i / \sum_i \lambda_i)$ . Consequently, the total expected cost would increase linearly with increases in the lost sale cost of an individual class as long as the costs of the other classes are held fixed. This is the case for the system shown in Figure 9(a). However, the total expected cost remains unchanged if an increase in the lost sales cost of one class is offset by a reduction in the cost of another class so that  $\sum_i c_i (\lambda_i / \sum_i \lambda_i)$  remains constant. This is true for the system shown in Figure 9(b), in which the quantity  $\sum_i c_i (\lambda_i / \sum_i \lambda_i) = c_1 + c_2$  is held fixed.

The behavior of the optimal policy is a little less obvious to explain. It appears that increasing  $c_1$  leads the optimal policy to ration inventory more aggressively in favor of class 1. This reduces the probability that a class 1 demand is rejected, which lessens the impact of the increase in  $c_1$ . In the case of Figure 9(b),

**Figure 9** The Benefit of Inventory Rationing



the increase in  $c_1$  is accompanied by a decrease in  $c_2$ . This allows for even more aggressive rationing in favor of class 1, accompanied with an overall reduction in inventory. The net effect is a reduction in total cost.

For operations managers in practice, the results of this section highlight the importance of customer differentiation and inventory rationing. Regardless of whether the rationing is done optimally or via a heuristic, it is important to differentiate between customers and to make inventory allocation decisions that take into account the cost differences between different customer classes. In fact, inventory allocation among customers seems to be a more critical function (based on the results for systems with lost sales) than production coordination among components.

## 6. Summary and Concluding Comments

In this paper, we considered the optimal production control and inventory allocation of an ATO system consisting of a single product,  $m$  components, and  $n$

**Table 5** Impact of Differences in Lost Sales Costs on Percentage Cost Difference Between Optimal and Heuristic Policies ( $\mu_1 = 1$ ,  $\mu_2 = 1, \lambda_1 = 0.45, \lambda_2 = 0.45, h_1 = 1, h_2 = 1.0$ )

	$c_1/c_2$	FCFS	CBR	IBR
$c_1 + c_2 = 20$	1	0.000	0.000	1.010
	2	0.000	0.000	1.010
	3	1.130	0.281	1.285
	5	4.261	0.000	0.975
	10	6.865	0.000	1.063
	15	13.300	0.000	1.283
	20	15.916	0.000	1.362
	25	18.085	0.000	1.043
$c_1 + c_2 = 100$	30	19.688	0.000	1.033
	1	0.000	0.152	1.048
	2	1.906	0.067	1.196
	3	7.214	0.163	1.319
	5	11.143	0.651	1.369
	10	14.835	0.281	0.857
	15	28.619	0.289	1.356
	20	35.442	0.645	0.974
$c_1 + c_2 = 400$	25	40.345	0.160	0.505
	30	44.300	0.075	0.402
	1	0.000	0.100	0.828
	2	4.047	0.285	0.869
	3	10.254	0.155	0.988
	5	16.317	0.155	0.834
	10	21.340	0.264	1.194
	15	39.514	0.259	0.803
20	51.182	0.476	1.283	
25	58.698	0.655	0.974	
30	64.977	0.458	0.849	

customer classes. We showed that the optimal production policy for each component is a state-dependent base-stock policy, in which the base-stock level for each component is nondecreasing in the inventory level of other components. We showed that the optimal inventory allocation is a state-dependent, multilevel rationing policy in which the component rationing level for each class is nonincreasing in the inventory level of other components. Because obtaining the optimal base-stock levels and the optimal rationing levels can be computing-intensive when the number of components is large, we proposed simple heuristics with fixed parameters. For cases in which it is possible to compare the heuristics to the optimal policy, we found the heuristics to perform well for systems with lost sales. For systems with backorders, there are cases in which the cost differences between the heuristics and the optimal policy are significant.

The results presented in this paper provide a first step toward a better understanding of the structure of the optimal policy for assembly systems and for identifying useful heuristics. There are several avenues for future research. In particular, it would be of interest to consider systems with multiple products, in which each product may require only a subset of the components, systems with variable order sizes in

which the order size distribution varies with product and demand class, and systems with multiple echelons in which the assembly takes place over several stages. We expect the problem to become considerably more difficult with each additional feature, and it is not clear whether the optimal policy would continue to have a recognizable structure. However, it may still be possible to identify effective heuristics. Finally, it would be useful to develop performance evaluation models in which, assuming the structure of the production and inventory allocation policy, closed-form expressions or efficient numerical procedures provide estimates of total cost. For systems with a single component and multiple classes, such procedures are available; see, for example, Ha (1997a) and de Véricourt et al. (2002). However, for systems with multiple components, the task seems difficult even for systems with a single class (Song and Zipkin 2003). Nevertheless, approximate models or numerical procedure might be possible to construct for some special cases.

An online supplement to this paper is available on the *Management Science* website (<http://mansci.pubs.informs.org/ecompanion.html>).

### Acknowledgments

The authors thank Yves Dallery, Jean-Philippe Gayon, Chung-Yee Lee, Francis de Véricourt, and Larry Zhou for useful discussions and Setareh Mardan, Jeannette Song, Yimin Yu, and Dan Zhang for helpful comments on an earlier version of the paper. They are also grateful to two anonymous referees, the associate area editor, and the department editor, Paul Zipkin, for numerous comments and suggestions.

### References

- Agrawal, M., M. Cohen. 2001. Optimal material control and performance evaluation in an assembly environment with component commonality. *Naval Res. Logist.* **48** 409–429.
- Benjaafar, S., M. Elhafsi, F. de Véricourt. 2004. Demand allocation in multi-product, multi-facility make-to-stock systems. *Management Sci.* **50** 1431–1448.
- Chen, H., P. Yang, D. D. Yao, X. Chao. 1993. Optimal control of a simple assembly system. *Oper. Res. Lett.* **14** 199–205.
- Cheng, F., M. Ettl, G. Y. Lin, D. D. Yao. 2002. Inventory-service optimization in configure-to-order systems. *Manufacturing Service Oper. Management* **4** 114–132.
- Cheng, F., M. Ettl, Y. Lu, D. D. Yao. 2005. A two stage push-pull production planning model. Working paper, IBM T. J. Watson Research Center, Yorktown Heights, NY.
- Dayanik, S., J. S. Song, S. H. T. Xu. 2003. The effectiveness of several performance bounds for capacitated production, partial-order-service, assemble-to-order systems. *Manufacturing Service Oper. Management* **5** 230–251.
- de Kok, A., J. Visschers. 1999. Analysis of assembly systems with service level constraints. *Internat. J. Production Econom.* **59** 313–326.
- de Véricourt, F., F. Karaesmen, Y. Dallery. 2002. Stock allocation for a capacitated supply system. *Management Sci.* **48** 1486–1501.



- Frank, K. C., H. Ahn, R. Q. Zhang. 2004. Inventory policies for multi-product systems with different lifecycles. Working paper, Cornell University, Ithaca, NY.
- Gallien, J., L. Wein. 2001. A simple and effective component procurement policy for stochastic assembly systems. *Queueing Systems* **38** 221–248.
- Gerchak, Y., M. Henig. 1989. Component commonality in assemble-to-order systems: Models and properties. *Naval Res. Logist.* **36** 61–68.
- Glasserman, P., Y. Wang. 1998. Leadtime-inventory tradeoffs in assemble-to-order systems. *Oper. Res.* **46** 858–871.
- Ha, A. Y. 1997a. Inventory rationing in a make-to-stock production system with several demand classes and lost sales. *Management Sci.* **43** 1093–1103.
- Ha, A. Y. 1997b. Stock-rationing policy for a make-to-stock production system with two priority classes and backordering. *Naval Res. Logist.* **44** 457–472.
- Hausman, W. H., H. L. Lee, A. X. Zhang. 1998. Joint demand fulfillment probability in a multi-item inventory system with independent order-up-to policies. *Eur. J. Oper. Res.* **109** 646–659.
- Lippman, S. 1975. Applying a new device in the optimization of exponential queueing systems. *Oper. Res.* **23** 687–710.
- Puterman, M. 1994. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley and Sons, New York.
- Rosling, K. 1989. Optimal inventory policies for assembly systems under random demands. *Oper. Res.* **37** 565–579.
- Schmidt, C., S. Nahmias. 1985. Optimal policy for a two-stage assembly system under random demand. *Oper. Res.* **33** 1130–1145.
- Song, J.-S. 1998. On the order fill rate in a multi-item, base-stock inventory system. *Oper. Res.* **46** 831–845.
- Song, J. S., D. D. Yao. 2002. Performance analysis and optimization in assemble-to-order systems with random leadtimes. *Oper. Res.* **50** 889–903.
- Song, J. S., S. H. Xu, B. Liu. 1999. Order fulfillment performance measures in an assembly-to-order system with stochastic leadtimes. *Oper. Res.* **47** 131–149.
- Song, J. S., P. Zipkin. 2003. Assemble-to-order systems. T. de Kok, S. Graves, eds. *Handbooks in Operations Research and Management Science*, Vol. XXX: *Supply Chain Management*, Chapter 13. North Holland, Amsterdam, The Netherlands.
- Weber, R. R., S. Stidham. 1987. Optimal control of service rates in networks of queues. *Adv. Appl. Probab.* **19** 202–218.
- Xu, S. H. 2001. Dependence analysis of assemble-to-order systems. J. Song, D. Yao, eds. *Supply Chain Structures: Coordination, Information and Optimization*. Kluwer Academic, Boston, MA, 359–324.
- Zhang, A. X. 1997. Demand fulfillment rates in an assemble-to-order system with multiple products and dependent demands. *Production Oper. Management* **6** 309–324.