Manufacturing & Service Operations Management

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Managing an Assemble-to-Order System with Returns

Gregory A. DeCroix, Jing-Sheng Song, Paul H. Zipkin,

M&SOM MANUFACTURING & SERVICE OPERATIONS MANAGEMENT
Vol. 11, No. 1, Winter 2009, pp. 144–159
ISSN 1523-4614 | EISSN 1526-5498 | 09 | 1101 | 0144

informs®
DOI 10.1287/msom.1070.0209

Managing an Assemble-to-Order System with Returns

Gregory A. DeCroix
Wisconsin School of Business, University of Wisconsin-Madison, Madison, Wisconsin 53706,
gdecroix@bus.wisc.edu

Jing-Sheng Song
Fuqua School of Business, Duke University, Durham, North Carolina 27708, and
Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai, China,
jssong@duke.edu

Paul H. Zipkin
Fuqua School of Business, Duke University, Durham, North Carolina 27708,
zipkin@duke.edu

We consider a multiproduct assemble-to-order (ATO) system, in which inventory is kept only at the component level and the finished products are assembled in response to customer demands. In addition to stochastic demand for finished products, the system experiences stochastic returns of subsets of components, which can then be used to satisfy subsequent demands. The system is managed over an infinite horizon using a component-level base-stock policy. We identify several ways in which returns complicate the behavior of the system, and we demonstrate how to handle these additional complexities when calculating or approximating key order-based performance metrics, including the immediate fill rate, the fill rate within a time window, and average backorders. We also present a method for computing a near-optimal base-stock policy. We use these results to address managerial questions on both operational and product-design levels. For example, we find that tracking product-based (as opposed to component-based) return information appears to provide much less value than tracking product-based demand information, and we explore the impact of the number of products, component lead times, and different patterns of component returns (joint versus independent returns, returns of common versus dedicated components) on the value of component commonality.

Key words: supply chain management; product returns; assemble to order; reverse logistics; environment
History: Received: October 28, 2005; accepted: October 8, 2007. Published online in *Articles in Advance* May 7, 2008.

1. Introduction

In recent years, a variety of factors have made it more important for many companies to consider reverse material flows when managing their supply chains. For example, the increase in online sales has brought with it an increase in product returns, because customers unable to physically observe items before purchase are more likely to return them. (See, for example, Tedeschi 2001.) In addition, many companies have started taking back products after customers use them. In several countries, particularly in Europe, environmental concerns have led to such take-backs being legally required for products such as automobiles, electronic goods, and packaging. (See, for example, Frankel 1996, Diem 1999, Schenkman 2002, and Thorn and Rogerson 2002.) In other cases, companies voluntarily collect used products to recover residual value by harvesting components, recycling materials, etc. Examples of products that are recovered for this reason include single-use cameras (Kodak, Fuji), photocopiers (Xerox), and communication network equipment (Lucent).

Regardless of why they occur, product returns complicate the management of an inventory system by introducing an uncertain reverse flow of materials. This is particularly true when only a subset of the components of a product can be recovered for reuse. In such cases, managing component inventories to preserve balance among components that are recovered and those that are not can be challenging. In systems with multiple products that share components (some of which are recovered), the task becomes even more complex.

As an example, consider Kodak's single-use camera production and remanufacturing operations. Kodak produces various single-use camera models, and those models share some components. After use, a consumer takes the camera to a photofinisher where the film is removed and the camera is sent to a disassembly facility. There the camera is cleaned and inspected, and reusable components (e.g., the circuit board) are sent to a Kodak production facility. That facility must then manage inventories of new components (e.g., batteries, external packaging) and reusable components (including ordering some new units of such components) to facilitate assembly of the various camera models. (For a more detailed description, see Guide et al. 2003.) As another example, consider an online or catalogue retailer where some items are returned by customers and get placed back into inventory. The retailer must manage inventories of the items to meet the mix of customer demands. In this setting, any combination of items ordered by a customer corresponds to a "product" in the former setting, and an individual item corresponds to a "component." To reflect this, we use the terms "product" and "order" interchangeably to refer to a customer's demand, and we use the terms "component" and "item" interchangeably to refer to units in a product/order.

Our goal is to develop methods for choosing and evaluating inventory management policies for multicomponent, multiproduct systems facing the additional complexity of reverse flows, and to provide insights into how the behavior of such systems differs from traditional nonnegative-demand systems.

Specifically, we study an infinite-horizon, multiproduct, assemble-to-order (ATO) system facing both demands for products and returns of components. Demands and returns arrive according to independent Poisson processes. Demands are satisfied on a first-come, first-served basis, with unsatisfied demands being backordered, and returned components are immediately available to satisfy demands. The system is managed using a component-level base-stock policy. We identify several ways in which returns complicate the behavior of the system, and we demonstrate how to handle these additional complexities when calculating or approximating key order-based performance metrics, including the immediate fill rate, the fill rate within a time window, and average backorders. We also present a method for computing a near-optimal base-stock policy.

The presence of returns introduces several questions that managers need to consider. For example, without returns it is known (see Song 2002) that tracking demand rates at the product level (rather than just the component level) is important. Is this extra effort still worthwhile for returns, or is component-level information sufficient? Another example involves the use of common components across products. How is the value of component commonality affected by factors such as which components are recoverable? We explore such questions using our performance metric results and our proposed heuristic policy.

This paper builds on and contributes to two distinct streams of research. The first consists of research on the management of multi-item ATO systems without returns. Of particular relevance for this paper is the work related to order-based performance metrics and optimization. For example, Song (1998) derives expressions for order-based fill rates in an ATO system without returns, while Song (2002) derives expressions for order-based backorders and Lu and Song (2005) present methods for minimizing the sum of component-based holding costs and order-based backorder costs in such a system. We refer the reader to Agrawal and Cohen (2001) and Song and Zipkin (2003) for reviews of the ATO literature.

The second stream consists of research on inventory management with return flows. For a single-location, finite-horizon system, Heyman and Sobel (1984) note that Scarf's (1960) proof of the optimality of an (s, S) policy also holds for systems with uncertain returns. Fleischmann et al. (2002) extend this result to the infinite-horizon case. When there is no fixed order cost, a base-stock policy is optimal. Cohen et al. (1980) establish conditions under which a base-stock policy is optimal when a fixed fraction of demands in each period is returned after a fixed number of periods. Kelle and Silver (1989) develop a heuristic for managing a similar system that also has fixed order costs and stochastic return times.

Inderfurth (1997) studies a system where returned items must be remanufactured before reuse, and analyzes when the optimal policy structure can be characterized. Mahadevan et al. (2003) and Kiesmüller and Minner (2003) propose heuristic policies for such systems.

Research on multistage or multi-item systems facing product returns has been more limited. Muckstadt and Isaac (1981) study a two-echelon distribution system with demands and returns at the retailer level, and present an algorithm for finding a policy that minimizes an approximate cost function. For a series system with returns, DeCroix et al. (2005) show that an echelon base-stock policy is optimal and present exact and approximate methods for evaluating any such policy. The authors also propose a near-optimal heuristic approach. DeCroix (2006) characterizes the optimal policy structure in a series system that also includes remanufacturing operations. DeCroix and Zipkin (2005) establish conditions under which a single-product, multicomponent assembly system where some used components are recovered is equivalent to a series system with returns, and they present heuristics that can be used when those conditions are not satisfied. For reviews of other research on reverse logistics, see Fleischmann et al. (1997) and Dekker et al. (2004).

The rest of the paper is organized as follows. Section 2 introduces the basics of the model for a single-item system. Section 3 extends the model to a multiproduct, multicomponent ATO system. Sections 4 and 5 present methods for computing or approximating order-based fill rates and order-based backorders, respectively, while §6 discusses a method for computing near-optimal base-stock policies. Section 7 explores the value of component commonality. Section 8 provides some concluding remarks. Appendix A describes a method for computing a particular distribution that is central to the analysis, Appendix B presents exact expressions for certain performance metrics, and Appendix C contains a detailed specification of parameters used in one of the numerical studies (appendixes available online).

2. Single-Item System

To establish concepts and notation, let us review the case of a single-item system. Time is continuous, indexed by $t \geq 0$, and stockouts are backlogged. Denote

$L =$ order lead time, a positive number
$D(t) =$ cumulative demand by time t,
$R(t) =$ cumulative returns by time t,
$N(t) =$ cumulative net demand by time $t = D(t) - R(t)$,
$N[t, u) = N(t+u) - N(t)$
$IN(t) =$ net inventory (inventory minus backlog) at time t
$IP(t) =$ inventory position after demand at time t $= IN(t)$ plus stock in transit at time t
$I(t) =$ on-hand inventory at time t
$B(t) =$ backorders at time t.

The demand process $\{D(t), t \geq 0\}$ is a Poisson process with rate μ. The returns process $\{R(t), t \geq 0\}$ is also a Poisson process, with rate λ. Returned items are immediately placed in inventory, and items are never discarded. The demand and return processes are independent, and we assume $\mu > \lambda$. The standard flow-conservation equation holds:

$$IN(t+L) = IP(t) - N[t, t+L).$$

We study a system with stationary data, operating over an infinite horizon under a stationary base-stock policy with base-stock level s. Let IP, IN, I, B and $N = N[0, L) = D - R$ represent equilibrium versions of the previously defined quantities, where D is a Poisson random variable with mean μL and R is a Poisson random variable with mean λL.

Some of the model assumptions warrant additional discussion. Because items being returned must have been purchased in the past, it may seem more natural to model returns as a function of cumulative sales to date (rather than assume the two are independent). Actually, a similar claim can be made about demand itself. Because no product has an unlimited market, one could argue that current demand should *always* depend on cumulative demand. In practice, this is rarely modeled—the effect is small in most cases, so the added modeling complexity yields little benefit. For similar reasons, the independence assumption is common in the literature on systems with returns. This approximation is most accurate when a significant amount of time elapses between a demand and a subsequent return. Such a time lag would be typical when components are recovered from used products. However, it could also arise in commercial (i.e., new item) returns when retailers have generous returns policies. (See Fleischmann 2000 for a more detailed justification of this assumption in systems with returns, and Zipkin 2000 for a discussion of these issues for nonnegative demand systems. Also, see Kiesmüller and van der Laan 2001 for a single-stage model where returns depend on past demands.) The assumption that returned components can be immediately replaced in inventory is reasonable for commercial returns, but recovering components from used items may require some processing cost and time. Recovery costs can easily be incorporated here, because in the long run they are merely constants independent of the inventory policy. If recovery activities require significant processing time, then the assumption of immediate reuse is equivalent to assuming that inventories of usable components are managed without using information about the state of the component recovery process (e.g., as might be the case if the recovery occurs at a separate facility), so that arrivals from that process are simply viewed as arrivals from customers. Certainly some improvements might be possible if such information is used, but the analysis of that possibility is beyond the scope of this paper. Finally, item disposal may be attractive when return rates are very close to demand rates (so that excess components could accumulate for significant periods of time), or when demand for a product is declining, e.g., near the end of a product's life cycle. The former is generally not an issue under the kinds of moderate return rates typically seen in practice, and the latter effect does not occur here, given the assumption of stationary data.

A base-stock policy in this setting can be described in the usual way. At any demand or return epoch t, if $IP(t) < s$, order the difference $s - IP(t)$ to bring $IP(t)$ up to s; otherwise, order nothing. The presence of returns complicates the behavior of the state variables. Without returns, once $IP(t)$ falls below s, it will stay at or below s in the future, so that $IP(t) = s$ from that point on. In that case, a base-stock policy is a demand replacement policy. Here, returns may cause $IP(t)$ to exceed s, so some additional work is required to describe the behavior of $IP(t)$. Also, when $IP(t)$ exceeds s, a demand does not trigger an order, so the base-stock policy is clearly *not* a demand-replacement policy.

Define $Z(t) = IP(t) - s$, i.e., the amount the inventory position exceeds s due to negative net demands. If a demand occurs at time t, then $Z(t) = [Z(t^-) - 1]^+$, and if a return occurs at time t, then $Z(t) = Z(t^-) + 1$, where $Z(t^-)$ indicates the limit of $Z(s)$ as s approaches t from below. Due to exponential interarrival times, $\{Z(t), t \geq 0\}$ is a birth-and-death process with birth rate λ and death rate μ—in fact, $Z(t)$ is equivalent to the queue-length process in an $M/M/1$ queue with arrival rate λ and service rate μ.

The equilibrium random variables IP, IN, I, and B can be expressed as

$$IP = s + Z, \quad IN = IP - N = s + Z - N,$$
$$I = [IN]^+, \quad \text{and} \quad B = [IN]^-.$$

The fill rate of the system, denoted by f, is defined as the long-run proportion of demand that can be filled immediately from inventory. Because Poisson arrivals see time averages (PASTA) (see Wolff 1989), and due to unit demand,

$$f = \Pr\{IN > 0\} = \Pr\{N < s + Z\}.$$

3. ATO System Model

Now consider a multiproduct, multicomponent ATO system. Let $\mathcal{I} = \{1, 2, \ldots, m\}$ denote the set of component indices. Customer orders arrive according to a stationary Poisson process with rate μ. Each customer order may require several components simultaneously. For any subset of components $K \subseteq \mathcal{I}$, we say a demand is of type K if it requests one unit of component $i \in K$, and zero units in $\mathcal{I} \setminus K$. Although we do not keep stock for any finished product, it is convenient and equivalent to say a demand of type K is a demand for "product" K. We assume that each order's type is independent of the other orders' types and of all other events. Also, there is a fixed probability q^K that an order is of type K, $\sum_K q^K = 1$. Thus, the type-K order stream forms a Poisson process with rate $\mu^K = q^K \mu$.

Let $\mathcal{K}$ be the set of all demand types, that is, $\mathcal{K} = \{K \subset \mathcal{I}: q^K > 0\}$. Note that $\mathcal{K}$ is not necessarily the set of all possible subsets of $\mathcal{I}$. For each component i, let $\mathcal{K}_i$ denote all product types that contain i. The demand process for component i is then a Poisson process with rate $\mu_i = \sum_{K\in\mathcal{K}_i} \mu^K$.

Similarly, assume that returns arrive according to a stationary Poisson process with rate λ, and this process is independent of the demand process. A return is of type S if it contains one unit of component $i \in S \subseteq \mathcal{I}$, and zero units of $\mathcal{I} \setminus S$. A return's type is independent of the other returns' types and all other events. There is a fixed probability r^S that a return is of type S, so the type-S return stream forms a Poisson process with rate $\lambda^S = r^S \lambda$.

Define $\mathcal{S} = \{S\colon \lambda^S > 0\}$ to be the set of all return types, and for each component i, let $\mathcal{S}_i$ denote all return types that contain item i. The return process for component i is a Poisson process with rate $\lambda_i = \sum_{S\in\mathcal{S}_i} \lambda^S$. Assume that returned items are immediately available to satisfy existing backorders or future demands. Then, the net demand process for component i is the difference of two independent Poisson processes, one with rate μ_i and the other with rate λ_i. We assume $\lambda_i < \mu_i$ for all i.

Demands are filled on a first-come-first-served (FCFS) basis. Demands that cannot be filled immediately are backlogged. When a demand arrives and some of its required components are in stock but others are not, we put the in-stock components aside as committed inventory. A demand is considered backlogged until it is satisfied completely.

The inventory of each component i is controlled by a base-stock policy with base-stock level s_i. A base-stock policy has the same interpretation here as in the single-item case.

Let $t \ge 0$ be the continuous time variable, and for each t denote

$IN_i(t) =$ net inventory of item i,
$N^K(t) =$ cumulative net demand for type K by time t,
$D_i(t) =$ cumulative demand for i by time t,
$R_i(t) =$ cumulative returns for i by time t,
$N_i(t) =$ cumulative net demand for i by time $t = D_i(t) - R_i(t)$,
$B^K(t) =$ type-K backorders at t $=$ number of type-K orders that are not yet completely satisfied by t,
$B_i(t) =$ number of backorders for item i at t.

Let L_i be the constant lead time for replenishing component i, and let D_i be the steady-state limit of $D_i(t - L_i, t) = D_i(t) - D_i(t - L_i)$, the lead-time demand of item i. Define R_i similarly. Then, D_i and R_i have Poisson distributions with mean $\mu_i L_i$ and $\lambda_i L_i$, respectively. Let IN_i be the steady-state limit of $IN_i(t)$, and define B^K and B_i similarly. Also, define

$W^K =$ steady-state waiting time of a type-K backorder.

The performance measures of interest are, for any demand type K,

$f^{K,w} =$ type-K order fill rate within a time window w
$=$ probability of satisfying a type-K order within a time window w
$= \Pr\{W^K \le w\}$
$f^K =$ immediate fill rate of type-K demand $= f^{K,0}$
$\bar{B}^K =$ average number of type-K backorders.

When we wish to explicitly represent the dependence of these performance measures on the vector $\mathbf{s} = (s_i)_i$ of base-stock levels and the vector $\mathbf{L} = (L_i)_i$ of component lead times, we will use the notation $f^K(\mathbf{s} \mid \mathbf{L})$ (or simply $f^K(\mathbf{s} \mid L)$ when all lead times equal L) for f^K, and similarly for the other measures. With these order-based performance measures, one can easily obtain the following system performance measures:

$f =$ average (over all demand types) immediate fill rate $= \sum_K q^K f^K$.
$\bar{B} =$ total average order-based backorders $= \sum_K \bar{B}^K$.

At times it will also be useful to refer to the component-based performance measures:

$f_i =$ immediate fill rate of component i,
$\bar{B}_i =$ average number of backorders of component i.

4. Order-Based Fill Rates

From the results for the single-item case,

$$IN_i = s_i + Z_i - N_i,$$

where Z_i is the steady-state queue length in an M/M/1 queue with arrival rate λ_i and service rate μ_i. Also, Z_i is independent of N_i. The order-based performance evaluation involves the joint distribution of the net inventories $(IN_1, \ldots, IN_m)$, which, in turn, depends on the joint distribution of the lead-time net demands $(N_1, \ldots, N_m)$ and the joint distribution of $(Z_1, \ldots, Z_m)$. For example,

$$f^K = \Pr\{IN_i > 0,\ i \in K\} = \Pr\{N_i < s_i + Z_i,\ i \in K\}.$$

To keep exposition simple, for the rest of this section we focus on a two-component system unless otherwise noted. Here, there are three types of demand: A type 1 customer requires one unit of component 1; type 2 requires one unit of component 2; and type 12 (short for type $\{1, 2\}$) asks for one unit of each component. Similarly, there are three types of returns, corresponding to one unit of component 1, one unit of component 2, or one unit of each component. Note that there may not be a perfect match between demand types and return types. For example, due to imperfect recovery yields, a used type 12 unit might provide only a unit of component i for $i = 1$ or 2, and thus it would count as a type i return.

The type i fill rate is exactly component i's fill rate

$$f_i = \Pr\{IN_i > 0\} = \Pr\{N_i < s_i + Z_i\}, \quad i = 1, 2.$$

The type 12 fill rate is

$$\begin{aligned} f^{12} &= \Pr\{IN_1 > 0, IN_2 > 0\} \\ &= \Pr\{N_1 < s_1 + Z_1, N_2 < s_2 + Z_2\}. \end{aligned}$$

4.1. Identical Lead Times

First consider the case where $L_1 = L_2 = L$. Then,

$$N_i = N_i(L) = N^i(L) + N^{12}(L), \quad i = 1, 2.$$

Here, $N^K(L)$ is the difference of two independent Poisson random variables, one with mean $\mu^K L$, another with $\lambda^K L$. Moreover, the $N^K(L)$ are independent across K. By conditioning on $N^{12}(L)$ and then deconditioning, we obtain

$$\begin{aligned} f^{12}(\mathbf{s} \mid L) = \sum_{k=-\infty}^{\infty} &\Pr\{N^{12}(L) = k\} \\ &\cdot \Pr\{N^1(L) < s_1 + Z_1 - k, N^2(L) < s_2 + Z_2 - k\}. \end{aligned}$$

If we further condition on $(Z_1, Z_2) = (z_1, z_2)$ and then decondition, this becomes

$$\begin{aligned} &f^{12}(\mathbf{s} \mid L) \\ &\quad = \sum_{k=-\infty}^{\infty} \Pr\{N^{12}(L) = k\} \\ &\qquad \cdot \Bigg[\sum_{z_1, z_2 \geq 0} \Pr\{Z_1 = z_1, Z_2 = z_2\} \cdot \Pr\{N^1(L) < s_1 + z_1 - k\} \\ &\qquad\qquad \cdot \Pr\{N^2(L) < s_2 + z_2 - k\} \Bigg] \\ &\quad = \sum_{z_1, z_2 \geq 0} \Pr\{Z_1 = z_1, Z_2 = z_2\} \\ &\qquad \cdot \Bigg[\sum_{k=-\infty}^{\infty} \Pr\{N^{12}(L) = k\} \Pr\{N^1(L) < s_1 + z_1 - k\} \\ &\qquad\qquad \cdot \Pr\{N^2(L) < s_2 + z_2 - k\} \Bigg]. \end{aligned} \tag{1}$$

The quantity in brackets has the same form as $f^{12}(\mathbf{s} \mid L)$ for a system without returns and base-stock levels $(s_1 + z_1, s_2 + z_2)$, but in the system without returns the N^K are simply Poisson random variables. In addition to dealing with differences of Poisson random variables, the system with returns requires one additional level of conditioning. This additional conditioning also requires obtaining the joint distribution of (Z_1, Z_2), which can be computed using a matrix-geometric approach similar to that in Song et al. (1999). Details are provided in Appendix A online.

4.2. Nonidentical Lead Times

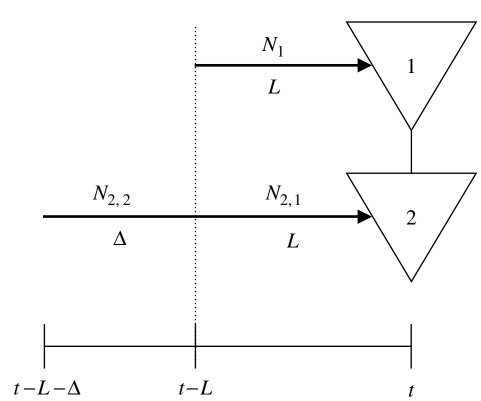
Now consider the case of nonidentical lead times. Without loss of generality, assume that $L_1 < L_2$. Set $L = L_1$ and $\Delta = L_2 - L_1$ so that $L_2 = L + \Delta$. (See Figure 1 for an illustration.) Then,

$$\begin{aligned} IN_2(t) &= s_2 + Z_2(t - L - \Delta) \\ &\qquad - [N_2(t - L - \Delta, t - L) + N_2(t - L, t)]. \end{aligned}$$

By the independent increment property of Poisson processes, $N_2(t - L - \Delta, t - L)$ and $N_2(t - L, t)$ are independent. Letting $t \to \infty$ yields

$$IN_2 = s_2 + Z_2 - [N_{2,2} + N_{2,1}].$$

Figure 1 System with Unequal Lead Times



Here, $N_{2,1}$ stands for the limit of $N_2(t-L, t)$, so it has the same distribution as $N_2(L)$, whereas $N_{2,2}$ stands for the limit of $N_2(t-L-\Delta, t-L)$, which is the difference of two Poisson random variables with means $\mu_2\Delta$ and $\lambda_2\Delta$, respectively. Recall that $IN_1 = s_1 + Z_1 - N_1$. Due to the type 12 demands and returns, N_1 and $N_{2,1}$ are correlated. However, $N_{2,2}$ is independent of both N_1 and $N_{2,1}$.

To compute $f^{12}(\mathbf{s} \mid L, L+\Delta)$, we need to characterize the joint distribution of (IN_1, IN_2) at an arbitrary point in time t when the system is in steady state. Using the method described previously, we can compute the steady-state joint distribution of (Z_1, Z_2), say at time $t-L-\Delta$. Now consider a two-dimensional Markov chain $(X_1(r), X_2(r))$ defined for $0 \le r \le \Delta$ with initial distribution (at $r=0$) equal to the steady-state joint distribution of (Z_1, Z_2). This Markov chain will track relevant information about component inventories from time $t-L-\Delta$ (i.e., $r=0$) to time $t-L$ (i.e., $r=\Delta$). During that time period, orders placed for component 1 can still arrive by time t, so $X_1(r)$ evolves as Z_1—i.e., if a demand or return event occurs at time r, then $X_1(r) = [X_1(r^-)-1]^+$ if that event includes a demand for component 1, $X_1(r) = X_1(r^-)+1$ if it includes a return of component 1, and $X_1(r) = X_1(r^-)$ otherwise. Because orders placed during this time period cannot affect component 2 inventory levels until after time t, these are not included in $X_2(r)$. Instead, that quantity tracks only demands and returns of item 2, so $X_2(r) = X_2(r^-)-1$ if the event at time r includes a demand for item 2, $X_2(r) = X_2(r^-)+1$ if the event includes a return of item 2, and $X_2(r) = X_2(r^-)$ otherwise. The pair $(s_1 + X_1(\Delta), s_2 + X_2(\Delta))$ represents component net inventories plus units on order at time $t-L$ that will be available to satisfy a demand at time t. Given the joint distribution of $(X_1(\Delta), X_2(\Delta))$, it is possible to compute $f^{12}(\mathbf{s} \mid L, L+\Delta)$ similarly to the case of equal lead times in (1), i.e.,

$$f^{12}(\mathbf{s} \mid L, L+\Delta) = \sum_{x_1, x_2 \ge 0} \Pr\{X_1(\Delta)=x_1, X_2(\Delta)=x_2\} \cdot \left[\sum_{k=-\infty}^{\infty} \Pr\{N^{12}(L)=k\} \cdot \Pr\{N^1(L) < s_1+x_1-k,\, N^2(L) < s_2+x_2-k\}\right]. \quad (2)$$

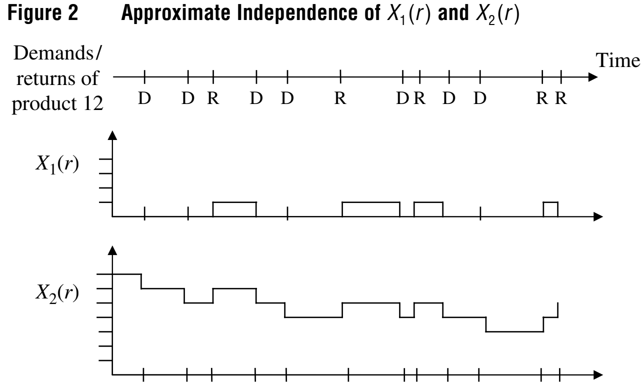
The challenge here is the computation of the joint distribution of $(X_1(\Delta), X_2(\Delta))$, which requires transient analysis of a two-dimensional Markov chain.

Consider a simpler approach. Assume that starting in any state $(X_1(0), X_2(0)) = (x_1, x_2)$, $X_1(r)$ and $X_2(r)$ evolve independently during $(0, \Delta)$—clearly an approximation, because type 12 demands or returns affect both variables. However, whereas $X_2(r)$ tracks all type 12 demands or returns, during $(0,\Delta)$ type 12 demands do not affect $X_1(r)$ when $X_1(r) = 0$. Therefore, each time $X_1(r)$ hits zero, it partially "forgets" past demand and return history (see Figure 2 for an illustration), thus reducing the dependence between the two variables. We would expect these variables to be less correlated when r is large and/or λ_1 is small compared to μ_1 (leading to frequent occurrences of $X_1(r)=0$). Under this assumption,

$$\begin{aligned}
&\Pr\{X_1(\Delta)=x_1, X_2(\Delta)=x_2\} \\
&\quad\approx \sum_{z_1, z_2 \ge 0} \Pr\{Z_1=z_1, Z_2=z_2\} \Pr\{X_1(\Delta)=x_1 \mid Z_1=z_1\} \\
&\qquad \cdot \Pr\{X_2(\Delta)=x_2 \mid Z_2=z_2\} \\
&\quad= \sum_{z_1 \ge 0} \Pr\{Z_1=z_1\} \sum_{z_2 \ge 0} \Pr\{Z_2=z_2 \mid Z_1=z_1\} \\
&\qquad \cdot \Pr\{X_1(\Delta)=x_1 \mid Z_1=z_1\} \Pr\{X_2(\Delta)=x_2 \mid Z_2=z_2\} \\
&\quad= \sum_{z_1 \ge 0} \Pr\{Z_1=z_1\} \Pr\{X_1(\Delta)=x_1 \mid Z_1=z_1\} \\
&\qquad \cdot \sum_{z_2 \ge 0} \Pr\{Z_2=z_2 \mid Z_1=z_1\} \Pr\{X_2(\Delta)=x_2 \mid Z_2=z_2\}.
\end{aligned}$$

Because $X_2(r)$ is a simple birth-and-death process, $X_2(\Delta) = X_2(0) - N_2(\Delta)$ and the distribution of $X_2(\Delta)$ is easy to compute ($X_2(0)$ and $N_2(\Delta)$ are independent). Computing the distribution of $X_1(\Delta)$ given any

Figure 2 **Approximate Independence of $X_1(r)$ and $X_2(r)$**



starting state, however, requires transient analysis of an $M/M/1$ queue. As $\Delta \to \infty$, the distribution of $X_1(\Delta)$ approaches the steady-state queue-length distribution for such a queue regardless of the initial state, so for sufficiently large Δ this should be a good approximation. For tractability, we define a variable X_1 with such a distribution, and then approximate $\Pr\{X_1(\Delta)=x_1 \mid Z_1=z_1\}$ by $\Pr\{X_1=x_1\}$. With this, the above quantity is approximately equal to

$$\begin{aligned}
&\Pr\{X_1=x_1\}\sum_{z_1\ge 0}\Pr\{Z_1=z_1\}\sum_{z_2\ge 0}\Pr\{Z_2=z_2 \mid Z_1=z_1\}\\
&\quad\cdot\Pr\{N_2(t-L-\Delta,t-L)=z_2-x_2\}\\
&\quad=\Pr\{X_1=x_1\}\sum_{z_2\ge 0}\Pr\{N_2(t-L-\Delta,t-L)=z_2-x_2\}\\
&\qquad\cdot\sum_{z_1\ge 0}\Pr\{Z_1=z_1\}\Pr\{Z_2=z_2 \mid Z_1=z_1\}\\
&\quad=\Pr\{Z_1=z_1\}\sum_{z_2\ge 0}\Pr\{N_2(t-L-\Delta,t-L)=z_2-x_2\}\\
&\qquad\cdot\Pr\{Z_2=z_2\}, \qquad (3)
\end{aligned}$$

where the last equality follows from the fact that Z_1 and X_1 have the same distribution. Substituting (3) into (2) yields

$$\begin{aligned}
&f^{12}(\mathbf{s} \mid L, L+\Delta)\\
&\approx\sum_{z_1,z_2\ge 0}\Pr\{Z_1=z_1\}\Pr\{Z_2=z_2\}\sum_{x_2\ge 0}\Pr\{N_{2,2}=z_2-x_2\}\\
&\quad\cdot\left[\sum_{k=-\infty}^{\infty}\Pr\{N^{12}(L)=k\}\Pr\{N^1(L)<s_1+z_1-k\}\right.\\
&\qquad\left.\cdot\Pr\{N^2(L)<s_2+x_2-k\}\right]\\
&=\sum_{z_1,z_2\ge 0}\Pr\{Z_1=z_1\}\Pr\{Z_2=z_2\}\sum_{y=-\infty}^{\infty}\Pr\{N_{2,2}=y\}\\
&\quad\cdot\left[\sum_{k=-\infty}^{\infty}\Pr\{N^{12}(L)=k\}\Pr\{N^1(L)<s_1+z_1-k\}\right.\\
&\qquad\left.\cdot\Pr\{N^2(L)<s_2+z_2-y-k\}\right]\\
&=\sum_{y=-\infty}^{\infty}\Pr\{N_{2,2}=y\}\sum_{z_1,z_2\ge 0}\Pr\{Z_1=z_1\}\Pr\{Z_2=z_2\}\\
&\quad\cdot\left[\sum_{k=-\infty}^{\infty}\Pr\{N^{12}(L)=k\}\Pr\{N^1(L)<s_1+z_1-k\}\right.\\
&\qquad\left.\cdot\Pr\{N^2(L)<s_2+z_2-y-k\}\right]. \qquad (4)
\end{aligned}$$

Table 1 **Parameters for Numerical Test of Order Fill Rate Approximation**

(L_1, L_2)	(μ^1, μ^2, μ^{12})	$(\lambda^1, \lambda^2, \lambda^{12})$	α
$L_1=1$	(8, 8, 4)	$\beta\cdot(\mu^1,\mu^2,\mu^{12})$ for $\beta=0.1, 0.4, 0.75$	0
$L_2=1.2, 1.4, 1.6, 1.8,$	(5, 5, 10)	(2, 0, 0)	0.67
2, 2.5, 3, 3.5, 4	(2, 2, 16)	(6, 0, 0)	1.64
	(7, 3, 10)		

To test the accuracy of this approximation, the results from (4) were compared to simulation results for the set of 540 test problems using all combinations of parameters given in Table 1. Base-stock levels were set as

$$s_i=\lfloor(\mu_i-\lambda_i)L_i+\alpha\sqrt{(\mu_i+\lambda_i)L_i}\rfloor, \qquad (5)$$

where $\lfloor x\rfloor$ represents the largest integer less than x, and the α values in Table 1 reflect 50%, 75%, and 95% component-level fill rates, respectively. (For a justification of this approach to setting s_i, see Song 2002.)

The approximation was quite accurate, with an average relative error of 1.72%. It was most accurate at relatively high fill rates (which is where most systems would likely operate in practice)—the average relative errors were 3.20% for $\alpha=0$, 1.48% for $\alpha=0.67$, and 0.48% for $\alpha=1.64$. (Maximum errors were 9.89%, 4.34%, and 2.31%, respectively.) Perhaps surprisingly, the accuracy is not very sensitive to the difference between the component lead times. This can be seen in Table 2, which reports the average relative error for each combination of α and L_2.

Recall that in a system without returns, we have

$$f^{12}(\mathbf{s} \mid L, L+\Delta)=\sum_{y=0}^{s_2-1}\Pr\{N_{2,2}=y\}f^{12}(\mathbf{s}-y\mathbf{e}_2 \mid L),$$

where $\mathbf{e}_i$ is the unit m-vector whose ith position is 1. Thus, computing the fill rates in the unequal lead-time case is not much harder than in the equal lead-time case. By comparing (4) with (1) we see a similar relationship for systems with returns. Interestingly, under our approximation, conditioning on the Z_i variables in the unequal lead-time case is somewhat simpler than in the equal lead-time case. Here only the marginal distributions of the Z_i variables are required. The difference in the lead times (approximately) eliminates the dependence between Z_1 (at time $t-L$) and Z_2 (at time $t-L-\Delta$).

Table 2 **Average Relative Error (%) for Order Fill Rate Approximation**

	$L_2 = 1.2$	$L_2 = 1.4$	$L_2 = 1.6$	$L_2 = 1.8$	$L_2 = 2$	$L_2 = 2.5$	$L_2 = 3$	$L_2 = 3.5$	$L_2 = 4$
$\alpha = 0$	2.72	3.13	3.20	3.01	3.92	3.29	3.77	2.73	3.05
$\alpha = 0.67$	1.57	1.26	1.44	1.79	1.65	1.58	1.20	1.44	1.35
$\alpha = 1.64$	0.91	0.58	0.38	0.49	0.43	0.38	0.39	0.40	0.38
Avg.	1.73	1.66	1.67	1.76	2.00	1.75	1.79	1.52	1.59

4.3. Fill Rate Within a Time Window

For general m-component systems without returns, Song (1998) shows that for any fixed K and $0 \le w < \max_{i\in K}\{L_i\}$,

$$f^{K,w}(\mathbf{s} \mid \mathbf{L}) = f^{K,0}(\mathbf{s} \mid (L_1 - w)^+, \ldots, (L_m - w)^+), \quad (6)$$

so one only needs to focus on $f^K = f^{K,0}$. This result reflects the fact that a current demand for an order of type K can be satisfied within a time window of length w if and only if for all $i \in K$ there is at least one uncommitted unit of component i due to arrive within w time units. That is equivalent to having at least one unit of each component $i \in K$ in stock in a transformed system with lead times shortened by w.

In a general system with returns, however, (6) no longer holds with equality, but instead provides a lower bound. As before, a current demand can be satisfied if the necessary uncommitted components are within w time units of arriving, and the probability of this is again equal to $f^{K,0}$ in the transformed system. However, even if some necessary components are *not* due to arrive within the time window, it may be possible to satisfy the demand using future returns that arrive during that time window. Using this observation to generalize (6), we can show that, for any K and $0 \le w < \max_{i\in K}\{L_i\}$,

$$f^{K,w}(\mathbf{s} \mid \mathbf{L}) = \sum_{\mathbf{x}} \Pr\{\mathbf{IN}(\mathbf{s} \mid (\mathbf{L} - w\mathbf{e})^+) = \mathbf{x}\} \cdot \Pr\{R_i(w) > -x_i,\ i \in K\}, \quad (7)$$

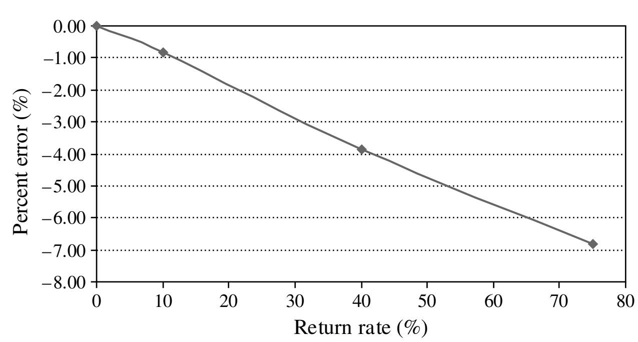
where $\mathbf{IN}(\mathbf{s} \mid (\mathbf{L} - w\mathbf{e})^+) = (IN_i(s_i \mid (L_i - w)^+))_i$. By rewriting (7) as

$$\begin{aligned} f^{K,w}(\mathbf{s} \mid L) &= \sum_{\substack{\mathbf{x} \\ x_i > 0 \text{ for } i\in K}} \Pr\{\mathbf{IN}(\mathbf{s} \mid (\mathbf{L} - w\mathbf{e})^+) = \mathbf{x}\} \\ &\quad \cdot \Pr\{R_i(w) > -x_i, i \in K\} \\ &\quad + \sum_{\substack{\mathbf{x} \\ x_i \le 0 \text{ for some } i\in K}} \Pr\{\mathbf{IN}(\mathbf{s} \mid (\mathbf{L} - w\mathbf{e})^+) = \mathbf{x}\} \\ &\quad \cdot \Pr\{R_i(w) > -x_i, i \in K\} \\ &= f^{K,0}(\mathbf{s} \mid (L_1 - w)^+, \ldots, (L_m - w)^+) \\ &\quad + \sum_{\substack{\mathbf{x} \\ x_i \le 0 \text{ for some } i\in K}} \Pr\{\mathbf{IN}(\mathbf{s} \mid (\mathbf{L} - w\mathbf{e})^+) = \mathbf{x}\} \\ &\quad \cdot \Pr\{R_i(w) > -x_i, i \in K\}, \end{aligned}$$

it is easy to see that (6) provides a lower bound for $f^{K,w}(\mathbf{s} \mid \mathbf{L})$ in the presence of returns.

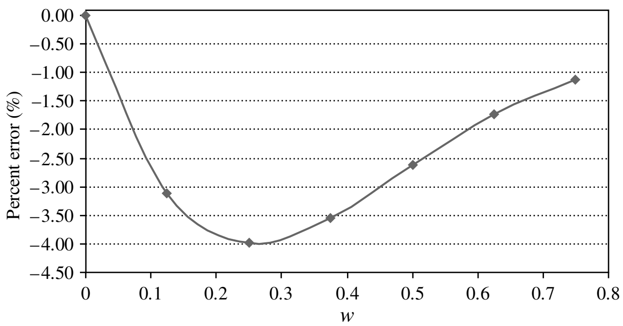
Exact expressions for the fill rate within a time window for products in the two-component system are derived in Appendix B (online). Even for that simple system, evaluating those expressions requires significant computational effort. As a result, it would be useful to know if the lower bound in (6) provides a good estimate for $f^{K,w}(\mathbf{s} \mid \mathbf{L})$. We tested this for a two-component setting using the parameters in Table 1 (with α fixed at 0.67, L_2 fixed at 2, and $(\lambda^1, \lambda^2, \lambda^{12}) = \beta(\mu^1, \mu^2, \mu^{12})$) and time windows $w = 0.125, 0.25, \ldots, 0.75$. The accuracy of the approximation depends on the parameter values. Specifically, as would be expected, the approximation is fairly accurate for small to moderate return rates, but deteriorates as the return rate increases—average relative errors for different return rates are shown in Figure 3. As shown in Figure 4, the approximation is

Figure 3 **Average Relative Performance of $f^{12,w}$ Lower Bound as Function of Return Rate β**



Note. Parameters: $(L_1, L_2) = (1, 2)$, $\alpha = 0.67$, $(\lambda^1, \lambda^2, \lambda^{12}) = \beta(\mu^1, \mu^2, \mu^{12})$ for $\beta = 0.1, 0.4, 0.75$; average taken across (μ^1, μ^2, μ^{12}) values in Table 1 and $w = 0.125, 0.25, \ldots, 0.75$.

Figure 4 Average Relative Performance of $f^{12,w}$ as a Function of the Time Window w



Note. Parameters: $(L_1, L_2) = (1, 2)$, $\alpha = 0.67$, $w = 0.125, 0.25, \ldots, 0.75$; averages taken across $(\lambda^1, \lambda^2, \lambda^{12}) = \beta(\mu^1, \mu^2, \mu^{12})$ for $\beta = 0.1, 0.4, 0.75$ and (μ^1, μ^2, μ^{12}) values in Table 1.

also accurate for short time windows, and the accuracy initially deteriorates as w increases. This makes intuitive sense—as w gets larger, the approximation ignores potentially valuable returns over a longer time interval. However, when w is sufficiently large, the accuracy actually improves, which is somewhat counterintuitive. This is because as w gets larger, more demands are satisfied by uncommitted (ordered) units that will arrive within the time window. As a result, returns play less of a role in achieving the fill rate within the window, making the approximation more accurate.

5. Order-Based Backorders

Now let us consider the evaluation of the average order-based backorders $\bar{B}^K(\mathbf{s} \mid \mathbf{L})$. Again, for simplicity, we discuss the two-component case, and to begin with we assume equal lead times. First, notice that a request for component i is due to a type i order with probability $q^i/(q^i + q^{12})$, so average type i backorders equal

$$\bar{B}^i(s_i \mid L) = \frac{q^i}{q^i + q^{12}} \bar{B}_i(s_i \mid L),$$

where

$$\bar{B}_i(s_i \mid L) = \mathsf{E}[(N_i - s_i - Z_i)^+]$$

represents expected component i backorders.

As in Song (2002), Little's Law implies that average type 12 backorders equal

$$\bar{B}^{12}(\mathbf{s} \mid L) = \mu^{12} \mathsf{E}[W^{12}(\mathbf{s} \mid L)], \tag{8}$$

where $\mathsf{E}[W^{12}(\mathbf{s} \mid L)]$ is the average wait by a type 12 customer. This can be written as

$$\begin{aligned} \mathsf{E}[W^{12}(\mathbf{s} \mid L)] &= \int_0^L \Pr\{W^{12}(\mathbf{s} \mid L) > w\}\, dw \\ &= \int_0^L (1 - f^{12,w}(\mathbf{s} \mid L))\, dw \\ &= L - \int_0^L f^{12,w}(\mathbf{s} \mid L)\, dw, \end{aligned} \tag{9}$$

because $W^{12}(\mathbf{s} \mid L) \le L$ with probability 1. Using a conditioning approach, it is possible to use (9) and the expression for $f^{12,w}(\mathbf{s} \mid L)$ (in Appendix B, online) to derive an expression for $\bar{B}^{12}(\mathbf{s} \mid L)$. However, evaluating that expression is impractical even for the two-component problem. As a result, we are interested in obtaining easy-to-compute bounds and approximations.

To this end, consider again the general m-component system, define B_i^K to be the (steady-state) number of backorders of item i due to demand type K, and denote its average by $\bar{B}_i^K$. Then, just as in a system without returns, we can define bounds

$$\begin{aligned} LB^K &\stackrel{\text{def}}{=} \mu^K \max_{i \in K} \frac{\bar{B}_i}{\mu_i} = \max_{i \in K} \bar{B}_i^K \le \mathsf{E}[\max B_i^K] \\ &= \bar{B}^K \le \mathsf{E}\Big[\sum_{i \in K} B_i^K\Big] = \mu^K \sum_{i \in K} \frac{\bar{B}_i}{\mu_i} \stackrel{\text{def}}{=} UB^K. \end{aligned}$$

Summing these inequalities yields bounds on the total average order-based backorders:

$$LB \stackrel{\text{def}}{=} \sum_K LB^K \le \bar{B} \le \sum_K UB^K \stackrel{\text{def}}{=} UB.$$

A natural approximation for $\bar{B}^K$ is the simple average $AB^K = \frac{1}{2}(LB^K + UB^K)$. In the two-item system, we have

$$AB^{12} = \mu^{12} \max\left\{ \frac{\bar{B}_1}{\mu_1} + \frac{\bar{B}_2}{2\mu_2}, \frac{\bar{B}_2}{\mu_2} + \frac{\bar{B}_1}{2\mu_1} \right\},$$

which yields the estimate of systemwide backorders

$$AB = \bar{B}^1 + \bar{B}^2 + AB^{12} = \bar{B}_1 + \bar{B}_2 - \frac{\mu^{12}}{2} \min\left\{ \frac{\bar{B}_1}{\mu_1}, \frac{\bar{B}_2}{\mu_2} \right\}.$$

We tested the performance of this approximation using the set of two-component sample problems with $L_2 = 2$ described in Table 1. For each of the 60 cases we used simulation to compute an estimate for the true value of $\bar{B}^{12}$, and summed this with the actual $\bar{B}^i$ to obtain a benchmark for comparison. Overall, AB appeared to be a reasonably good estimate for systemwide backorders. The relative error, defined as

$$\frac{|AB - (\bar{B}^1 + \bar{B}^2 + \text{Simulated } \bar{B}^{12})|}{\bar{B}^1 + \bar{B}^2 + \text{Simulated } \bar{B}^{12}} \times 100\%$$

averaged across the 60 cases was 3.6%. Performance was weakest for the 15 problems with the highest demand correlation ($\mu = (2, 2, 16)$)—those cases averaged 5.1% error (with maximum error of 11.6%), whereas the remaining 45 cases averaged 3.1% error (with maximum error of 9.4%). In most cases the average relative error (without the absolute value) became more negative at higher fill rates—i.e., as α increased.

Before concluding this section, we consider an issue related to the value of product-based demand and return information. For simplicity we again restrict attention to the two-component problem. It is known (see Song 2002) that in a system without returns the average type 12 backorders $\bar{B}^{12}$ are sensitive to (and increasing in) the order-based demand rate μ^{12}, so there is value in tracking that information rather than just relying on the component-based demand rates μ_1 and μ_2. The question we ask here is whether there is value in tracking product-based returns, i.e., is $\bar{B}^{12}$ sensitive to the product-based return rate λ^{12}, and if so, does λ^{12} affect $\bar{B}^{12}$ in the same way as μ^{12} does.

To answer that question, assume that the item-based rates μ_i and λ_i are fixed, and inventory is managed using the component base-stock levels in (5). Actual demands and returns occur in a product-based manner, however—i.e., μ^{12} and λ^{12} may be positive. We explored the impact of joint demands and returns for a set of problems with parameters as in Table 1 (with $L_2 = 2$), except now the return rates satisfied $(\lambda_1, \lambda_2) = \beta \cdot (\mu_1, \mu_2)$ for $\beta = 0.1$, 0.2, and 0.4. For each problem we first varied λ^{12} from 0 to $\min\{\lambda_1, \lambda_2\}$ (keeping $\lambda^i = \lambda_i - \lambda^{12}$, $i = 1, 2$), and calculated average backorders (by direct calculation for each $\bar{B}^i$ and by simulation for $\bar{B}^{12}$) for each parameter value. We then repeated this for fixed λ^{12} while varying μ^{12} from $\tilde{\mu}^{12} - \min\{\lambda_1, \lambda_2\}/2$ to $\tilde{\mu}^{12} + \min\{\lambda_1, \lambda_2\}/2$ (keeping $\mu^i = \mu_i - \mu^{12}$, $i = 1, 2$), where $\tilde{\mu}^{12}$ is the original value of μ^{12} in Table 1 (i.e., we consider changes in λ^{12} and μ^{12} of the same absolute size). Because overall demand increases with μ^{12} (for fixed μ_1, μ_2), it is more appropriate to measure the impact of changing μ^{12} using average customer waiting time $\overline{W} = \bar{B}/\mu$, rather than systemwide average backorders—and to facilitate comparison, we also use $\overline{W}$ to measure the impact of changing λ^{12}.

The results of this study indicate that the impact of λ^{12} differs from that of μ^{12} in two ways. First, increasing λ^{12} leads to shorter average customer waits, whereas increasing μ^{12} leads to longer waits. The reason for this is that it is harder to satisfy a demand for both components at the same time than to satisfy demand for one component at a time, because the former case results in a backorder if either component is unavailable. So a higher μ^{12} will result in more backorders and longer waits. A higher value of λ^{12}, however, makes it more likely that both components are returned together. *Receiving* both components at the same time is more valuable than receiving one at a time, because a joint return makes it possible to fill any type of demand. Therefore, a higher λ^{12} leads to fewer backorders and shorter waits. The second difference is in the magnitude of the impact—a change in λ^{12} tends to have less impact on customer waits than the same size change in μ^{12}. Across the 36 cases, the impact of λ^{12} was only 28.4% of the impact of μ^{12} on average. One implication of this is that, in contrast to demand information, there may be little value in obtaining product-based return information. Note that these results are consistent with the fact that the AB approximation, which generally is fairly accurate, is constant as λ^{12} changes (but not as μ^{12} changes).

6. Component-Based Heuristic

The preceding sections focused on performance metrics for given base-stock levels—we now turn to determining base-stock levels that minimize the sum of inventory holding and backorder costs. To that end, let h_i be the unit holding cost rate for component i, and let b^K be the unit backorder cost rate for order type K. The expected cost function to minimize is

$$C(\mathbf{s}) = \sum_{i=1}^{m} h_i \bar{I}_i(\mathbf{s}) + \sum_{K \in \mathcal{K}} b^K \bar{B}^K(\mathbf{s}),$$

where $\bar{I}_i(\mathbf{s})$ is the average on-hand inventory for item i. (In this section we modify notation slightly to emphasize the dependence of the performance metrics on the vector $\mathbf{s}$ of base-stock levels—e.g., $\bar{B}^K(\mathbf{s})$, $IN_i(s_i)$, etc.) On-hand inventory consists of two parts. The first is on-hand inventory that is available to serve new demands, i.e., $[IN_i(s_i)]^+ = [s_i + Z_i - D_i + R_i]^+$. The second is inventory still being held, but that has been set aside for earlier demands (backordered due to shortage of another component), which we denote $J_i(\mathbf{s})$. As with no returns (see Lu and Song 2005), it can be shown that

$$J_i(\mathbf{s}) = \sum_{K\in\mathcal{K}_i} [B^K(\mathbf{s}) - B_i^K(s_i)] = \sum_{K\in\mathcal{K}_i} B^K(\mathbf{s}) - B_i(s_i).$$

Combining this with the fact that $B_i(s_i) = [IN_i(s_i)]^-$, we can write the expected cost as

$$\begin{aligned} C(\mathbf{s}) &= \sum_{i=1}^m h_i \mathsf{E}\bigg[(s_i + Z_i - D_i + R_i + B_i(s_i)) \\ &\qquad + \bigg(\sum_{K\in\mathcal{K}_i} B^K(\mathbf{s}) - B_i(s_i)\bigg)\bigg] + \sum_{K\in\mathcal{K}} b^K \bar{B}^K(\mathbf{s}) \\ &= \sum_{i=1}^m h_i \mathsf{E}[s_i + Z_i - D_i + R_i] + \sum_{K\in\mathcal{K}} \bigg(b^K + \sum_{i\in K} h_i\bigg)\bar{B}^K(\mathbf{s}) \\ &= \sum_{i=1}^m h_i s_i + \sum_{K\in\mathcal{K}} \tilde{b}^K \bar{B}^K(\mathbf{s}) + \sum_i h_i \mathsf{E}[Z_i - D_i + R_i], \end{aligned}$$

where $\tilde{b}^K = b^K + \sum_{i\in K} h_i$. Let $\mathbf{s}^*$ be the vector of base-stock levels that minimizes $C(\mathbf{s})$. (Because the final term is a constant independent of $\mathbf{s}$, it does not affect the optimization.)

Given the difficulty of calculating $\bar{B}^K$, we would like to approximate $C(\mathbf{s})$ with cost functions that are easier to compute, and whose optimal solutions are nearly optimal for $C(\mathbf{s})$. In particular, we are interested in approximate cost functions that can be expressed in terms of component parameters only, rather than orders, i.e., functions of the form

$$\begin{aligned} C^C(\mathbf{s}) &= \sum_{i=1}^m (h_i \bar{I}_i(s_i) + b_i \bar{B}_i(s_i)) \\ &= \sum_{i=1}^m (h_i \mathsf{E}[s_i + Z_i - D_i + R_i + B_i(s_i)] + b_i \bar{B}_i(s_i)) \\ &= \sum_{i=1}^m (h_i s_i + (h_i + b_i)\bar{B}_i(s_i)) + \sum_{i=1}^m h_i \mathsf{E}[Z_i - D_i + R_i], \end{aligned}$$

where again the final term is a constant with respect to $\mathbf{s}$. The advantage to this component-based cost function is that it is separable across i, and for each i it has the same form as the single-item problem with returns discussed earlier, i.e., for each i the minimizer s_i^C is the smallest s_i such that

$$\Pr\{s_i + Z_i - N_i \geq 0\} \geq \frac{b_i}{b_i + h_i}.$$

The question is how to choose the b_i such that $C^C(\mathbf{s})$ and $\mathbf{s}^C = (s_i^C)_i$ are good approximations for $C(\mathbf{s})$ and $\mathbf{s}^*$.

One possibility proposed by Lu and Song (2005) for a system without returns is

$$b_i^u = \sum_{K\in\mathcal{K}_i} \frac{\mu^K}{\mu_i}\tilde{b}^K - h_i = \sum_{K\in\mathcal{K}_i} \frac{\mu^K}{\mu_i}\bigg(b^K + \sum_{j\in K,\, j\neq i} h_j\bigg).$$

They showed that this choice of b_i yields an upper bound on the optimal solution $\mathbf{s}^*$—the following result extends this to systems with returns.

Proposition. *Set $b_i = b_i^u$ in $C^C(s)$, and let the resulting solution be s^u. Then $s^* \leq s^u$.*

The proof is similar to the proof of Proposition 6 of Lu and Song (2005) and is omitted.

Lu and Song (2005) observe that in a system without returns $\mathbf{s}^u$ is often close to $\mathbf{s}^*$, and thus in addition to providing an upper bound, it also serves as a good approximation. Although $\mathbf{s}^u$ is still a reasonably good approximation in the system with returns, we have identified an alternative item-based formulation whose solution yields a better approximation for $\mathbf{s}^*$.

Recall from the previous section that

$$LB^K(\mathbf{s}) \overset{\text{def}}{=} \mu^K \max_{i\in K} \frac{\bar{B}_i(s_i)}{\mu_i} \leq \bar{B}^K(\mathbf{s}) \leq \mu^K \sum_{i\in K} \frac{\bar{B}_i(s_i)}{\mu_i} \overset{\text{def}}{=} UB^K(\mathbf{s}).$$

Define

$$\widehat{LB}^K(\mathbf{s}) \overset{\text{def}}{=} \frac{\mu^K}{|K|} \sum_{i\in K} \frac{\bar{B}_i(s_i)}{\mu_i},$$

where $|K|$ is the cardinality of K. Because $\widehat{LB}^K(\mathbf{s}) \leq LB^K(\mathbf{s})$, this also is a lower bound for $\bar{B}^K(\mathbf{s})$. Now, instead of approximating $\bar{B}^K(\mathbf{s})$ by $AB^K(\mathbf{s}) = \frac{1}{2}(LB^K(\mathbf{s}) + UB^K(\mathbf{s}))$, use

$$\widehat{AB}^K(\mathbf{s}) = \frac{1}{2}\big(\widehat{LB}^K(\mathbf{s}) + UB^K(\mathbf{s})\big) = \nu^K \sum_{i\in K} \frac{\bar{B}_i(s_i)}{\mu_i}, \qquad (10)$$

where $\nu^K = (|K|+1)\mu^K/(2|K|)$. Substituting $\widehat{AB}^K(\mathbf{s})$ into $C(\mathbf{s})$ yields $C^C(\mathbf{s})$ with $b_i = \hat{b}_i$, where

$$\hat{b}_i = \left[\frac{\sum_{K\in\mathcal{K}_i} \nu^K \tilde{b}^K}{\mu_i}\right] - h_i = \left[\frac{\sum_{K\in\mathcal{K}_i} \nu^K (b^K + \sum_{j\in K} h_j)}{\mu_i}\right] - h_i.$$

Let $\hat{\mathbf{s}}$ be the vector of base-stock levels that minimizes $C^C(\mathbf{s})$ when $b_i = \hat{b}_i$.

We tested the performance of $\hat{\mathbf{s}}$ using three different sets of problems. The first set contained the 60 test problems with $L_2 = 2$ in Table 1. However, instead of setting $\mathbf{s}$ based on the three values of α, we considered three holding and backorder cost combinations $(h_1, h_2, b^1, b^2, b^{12}) = (1, 2, 2, 4, 10)$, $(1, 2, 2, 4, 6)$, and $(1, 2, 4, 4, 6)$. For each of the test problems, we computed $\hat{\mathbf{s}}$ directly and then estimated $\bar{B}^{12}(\hat{\mathbf{s}})$ by simulation, which allowed for calculation of $C(\hat{\mathbf{s}})$. Values for $\mathbf{s}^*$ and $C(\mathbf{s}^*)$ were obtained by simulating the system for each $\mathbf{s}$ in the rectangle $\mathbf{0} \le \mathbf{s} \le \mathbf{s}^u$ to estimate $\bar{B}^{12}(\mathbf{s})$, then calculating $C(\mathbf{s})$ for each $\mathbf{s}$ and identifying $\mathbf{s}^*$ as the one yielding the lowest cost.

The policy $\hat{\mathbf{s}}$ performed quite well in the trial. Out of the 60 cases, 17 yielded $\hat{\mathbf{s}} = \mathbf{s}^*$ (versus zero cases yielding $\mathbf{s}^u = \mathbf{s}^*$). The relative amount by which the cost of the policy $\hat{\mathbf{s}}$ (including the constant term) exceeded that of the policy $\mathbf{s}^*$, i.e., $(100\%)\cdot(C(\hat{\mathbf{s}}) - C(\mathbf{s}^*))/C(\mathbf{s}^*)$, averaged 0.78% across the 60 cases (with maximum error of 4.14%), or 0.28% (1.81% maximum) when the constant term was dropped.

To test the robustness of these results, we used two other sets of test problems, both having more components and more complex product structures. We refer to the first of these as the *manufacturing set*, because it is intended to represent a setting in which products are assembled from components. All problems in this set have the same basic product architecture. There are three families of components, with two different components within each family (e.g., a basic version and a premium version), for a total of six components. Each product requires one component from each family, but any such combination is allowed, resulting in $2^3 = 8$ product types. We generated a set of 28 problems by varying the demand rates, return rates, and backorder costs for the products and the holding costs for the components. We included holding (backorder) costs that were the same across components (products), as well as costs that reflected a premium versus basic component and product distinction—e.g., higher holding costs for component 1 (premium) than component 2 (basic), as well as higher backorder costs for products containing component 1 than those containing component 2. Similarly, we considered product demand and return rates that were the same across products, as well as demand (return) rates that were higher (lower) for products containing the premium versions of the components (representing settings where premium components are newer to the market than basic components). Average product return rates ranged from 10% to 40% of average demand rates. A complete specification of the parameter values appears in Appendix C (online).

We refer to the final set of test problems as the *retailer set*, because it is intended to represent a setting where consumers may purchase any combination of items offered by a retailer. Again, all problems in this set have the same basic structure. There are four different items available, and the customer may select any combination of single items for a total of 15 order types. Customers arrive according to a Poisson process with an aggregate rate of 20, and the probability that any given customer arrival includes a demand for item i is given by q_i. Demand for item i is independent of demand for item $j \ne i$. All cases assume that $L_i = i$, $i = 1, \ldots, 4$. The 32 problems consisted of all combinations of parameter values listed in Table 3.

The results for these additional two trial sets were consistent with those for the initial set. Average errors including (omitting) the constant term were 0.64% (0.20%) for the manufacturing set and 0.57% (0.39%) for the retailer set. Maximum errors including (omitting) the constant term were 2.27% (0.54%) and 1.40% (1.14%), respectively.

Table 3 Parameters for Retailer Set—Heuristic Performance Evaluation

q_i	λ^K	Costs
$q_i = 0.2$	$\lambda^K = \beta^K\mu^K$ for $\beta^K = 0.1$, 0.2, 0.4 (all K)	$h_i = 1$ (all i), $b^K =$ 2, 4, 10 (all K)
$q_i = (1 - 0.2i)$	$\lambda^K = \beta^K\mu^K$ for $\beta^1 = \beta^{12} = 0.1$, $\beta^4 = \beta^{34} = \beta^{134} = 0.4$, and $\beta^K = 0.2$ (all other K)	$h = (5, 3, 3, 1)$, $b^K = \sum_{i\in K} h_i$

Although the analysis in this and preceding sections has focused on Poisson demand and return streams, some of the results can in principle be extended to the case of batch demands and returns. In that case, the Z_i would have the same distribution as the number of jobs in a single-server queue with batch arrivals and service. Given that distribution, and the new distribution for net demand during the lead time (the difference of compound Poissons), it would be straightforward to compute the $\bar{B}_i(s_i)$. Song (2002) noted in a system with batch demands and no returns that $\bar{B}_i^K$ is approximately equal to

$$\frac{\mu^K E[Y_i^K]}{\sum_{A\in\mathcal{K}_i}\mu^A E[Y_i^A]}\bar{B}_i(s_i),$$

(where Y_i^K is the random vector of quantities required when a demand of type K occurs), and this continues to hold in the presence of returns. Similar to before, one could then use the average of the approximate $\bar{B}_i^K$ over $i\in K$ as an approximate lower bound for $\bar{B}^K$, and the sum of the approximate $\bar{B}_i^K$ over $i\in K$ as an approximate upper bound. The average of these upper and lower bounds would still be linear in the $\bar{B}_i(s_i)$, and so the expected cost function could still be approximated in component form (with appropriate component-based backorder costs). As a result, the heuristic could be applied as before, with base-stock levels being efficiently computed by optimizing one component at a time.

7. Value of Component Commonality

In addition to providing a method for identifying good inventory policies, the preceding results can be used to explore broader questions related to product design. In this section we explore one such question—the decision whether to use dedicated components for each product or to share common components across multiple products. In multiproduct systems, using common components has the potential to reduce inventory-related costs by taking advantage of the risk pooling of demands. This issue has been studied in settings where lead times are effectively zero (e.g., Baker et al. 1986, Eynan and Rosenblatt 1996, Hillier 1999 and the references therein) and where they are positive (Song 2002, Song and Zhao 2006). All of the previous research has assumed no returns. We now explore the impact of returns on the value of component commonality. Because $\hat{\mathbf{s}}$ is easy to compute and was shown to be near optimal in most cases, the analysis of this section assumes the use of that policy.

We consider two different versions of systems with $|\mathcal{K}| = 2$, 4, and 6 products. The first is a dedicated system, where each product is assembled from two different components used exclusively for that product. As a result, the system with $|\mathcal{K}|$ products is actually $|\mathcal{K}|$ independent subsystems, each with two components and a single product. The second version is a common-component system, where one component from each product is replaced by a common component shared across all products. We compare the performance of the two systems under a variety of parameters. Specifically, in the dedicated systems we assume that every subsystem has the same set of parameters, and those parameters can be any one of the 180 combinations listed in Table 4. Parameters in the common-component system are obtained in the natural way by combining the $|\mathcal{K}|$ dedicated subsystems. (The common component replaces component 2 in each of the subsystems.) For each set of parameters and each $|\mathcal{K}|$ we computed $C(\hat{\mathbf{s}})$ for both the dedicated and common version of the system, and then computed the relative cost improvement from commonality, defined as 100% × [(cost in dedicated system) − (cost in common system)]/(cost in dedicated system).

Although component commonality did not always yield cost improvements, it did in the majority of cases (517 out of 540). By making more detailed comparisons among the cases we can obtain insights

Table 4 Parameters for Component Commonality Study—Two-Component Subsystem of Dedicated System

μ^{12}	h_i	b^{12}	(L_1, L_2)	$(\lambda^1, \lambda^2, \lambda^{12})$
$20/\lvert\mathcal{K}\rvert$	1	2, 10, 20	$L_1 = 1, 2$	$(0, 0, \theta\mu^{12})$ for $\theta = 0.1$, 0.2, 0.3, and 0.4
			$L_2 = 1, 2$	$(0.5\theta\mu^{12}, 0.5\theta\mu^{12}, 0.5\theta\mu^{12})$ for $\theta = 0.1$, 0.2, 0.3, and 0.4
				$(\theta\mu^{12}, \theta\mu^{12}, 0)$ for $\theta = 0.1$, 0.2, 0.3, and 0.4
				$(\theta\mu^{12}, 0, 0.5\theta\mu^{12})$ for $\theta = 0.1$, 0.2, 0.3, and 0.4
				$(0, \theta\mu^{12}, 0.5\theta\mu^{12})$ for $\theta = 0.1$, 0.2, 0.3, and 0.4

regarding when component commonality is most valuable. For example, the average relative cost improvement from commonality increased with the number of products—the gain was 11.7% for $|\mathcal{K}| = 2$, 24.4% for $|\mathcal{K}| = 4$, and 31.9% for $|\mathcal{K}| = 6$. Also, using common components yielded an average gain of 30.6% when the common component had a longer lead time, versus 22.6% when the dedicated components had a longer lead time, and only 14.2% when the lead times were equal. (All 23 cases with higher costs for the common system had equal lead times.) Both of these results can be linked to the risk-pooling benefit of commonality. For the second result, because components with long lead times are most subject to the effects of demand (and return) variability, they present the greatest risk-pooling opportunities. In multi-item systems like this, demand variability has two different effects. First there is a direct effect—variability in demand and/or returns of component 2 in the dedicated subsystems can be mitigated by pooling demands for those components through component commonality. Second, there is an indirect effect related to variability in demand and/or returns of component 1 in the dedicated system. Because both components are required to fill demand for a finished product, managing inventory of component 2 is made more challenging by increased variability of net demand for component 1. A longer lead time for component 1 corresponds to such an increase in variability, and so there is some benefit of pooling demands for component 2 through commonality. The numerical study suggests that the direct effect is more powerful. With respect to the first result, the benefit is greater when variability can be aggregated over a larger number of product demands.

This study also allowed us to explore how the pattern of returns experienced by the system influences the impact of component commonality on system costs. For example, the first three return patterns in Table 4 reflect different degrees of dependence in the return stream. In the first pattern, every return contains one unit of both items (in the dedicated system). By contrast, in the third pattern every return corresponds to a single component, whereas the second stream contains a mix of individual and joint returns. In the numerical trial, component commonality delivered larger average gains when the return pattern reflected more individual returns—i.e., a 20.1% cost reduction for return pattern 1, a 23.0% gain for pattern 2, and a 24.5% gain for pattern 3. One possible explanation for this is related to the fact that commonality achieves risk-pooling benefits by decoupling the one-to-one link between component inventories. Because returns are a form of inventory replenishment, joint returns force some degree of linkage between component inventories, whereas individual returns allow the decision maker to better exploit the potential benefits of commonality.

For the common system, the fourth return pattern reflects significant returns of the dedicated component, whereas the fifth pattern is weighted more toward returns of the common component. In the numerical trial, component commonality provided larger average gains under the fifth pattern (24.8%) than under the fourth pattern (17.2%). Because returns increase the coefficient of variation for component net demands (by increasing variability and decreasing mean net demand), this observation is consistent with the previous observation regarding lead times. Component commonality has more benefit when the demand distribution of the common component (component 2 in the dedicated subsystems) is directly affected than when it is indirectly affected (through higher variability of the dedicated component 1).

8. Conclusions

In this paper we studied an infinite-horizon ATO system facing both demands for products and returns of components. We identified several ways in which returns complicate the behavior of the system, and we demonstrated how to handle these additional complexities when calculating or approximating key order-based performance metrics, including the immediate fill rate, the fill rate within a time window, and average backorders. We also presented a heuristic method for computing a base-stock policy that generally appears to be close to optimal. Because this method uses a component-based formulation, computing the policy is quite easy.

We also obtained a number of insights into the ways that returns affect the behavior of the system. For example, for any given base-stock policy, when joint returns make up a larger fraction of overall returns, average customer waiting times are lower.

This is the opposite of the effect of joint demands. In addition, the impact of joint returns appears to be much smaller than that of joint demands, which suggests that there may be little value in tracking product-based (rather than component-based) information about returns. Finally, we explored the impact of the number of products, component lead times, and different patterns of component returns (joint versus independent returns, returns of common versus dedicated components) on the value of component commonality.

Electronic Companion

An electronic companion to this paper is available on the *Manufacturing & Service Operations Management* website (http://msom.pubs.informs.org/ecompanion.html).

Acknowledgments

The authors would like to thank the anonymous senior editor and referees for their comments and suggestions, which have helped to improve the paper. Song's research was supported in part by Award No. 70328001 from the National Natural Science Foundation of China.

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.

Baker, K., M. Magazine, H. Nuttle. 1986. The effect of commonality of safety stock in a simple inventory model. *Management Sci.* **32** 982–988.

Cohen, M., S. Nahmias, W. Pierskalla. 1980. A dynamic inventory system with recycling. *Naval Res. Logist. Quart.* **27** 289–296.

DeCroix, G. 2006. Optimal policy for a multi-echelon inventory system with remanufacturing. *Oper. Res.* **54** 532–543.

DeCroix, G., P. Zipkin. 2005. Inventory management for an assembly system with product or component returns. *Management Sci.* **51** 1250–1265.

DeCroix, G., J.-S. Song, P. Zipkin. 2005. A series system with returns: Stationary analysis. *Oper. Res.* **53** 350–362.

Dekker, R., M. Fleischmann, K. Inderfurth, L. van Wassenhove, eds. 2004. *Reverse Logistics: Quantitative Models for Closed-Loop Supply Chains*. Springer, New York.

Diem, W. 1999. *Makers Must Dispose of Autos*. Detroit Free Press (September 2).

Eynan, A., M. Rosenblatt. 1996. Component commonality effects on inventory costs. *IIE Trans.* **28** 93–104.

Fleischmann, M. 2000. Quantitative models for reverse logistics. Ph.D. thesis, Erasmus University, Rotterdam, The Netherlands.

Fleischmann, M., R. Kuik, R. Dekker. 2002. Controlling inventories with stochastic item returns: A basic model. *Eur. J. Oper. Res.* **138** 63–75.

Fleischmann, M., J. Bloemhof-Ruwaard, R. Dekker, E. van der Laan, J. van Nunen, L. van Wassenhove. 1997. Quantitative models for reverse logistics: A review. *Eur. J. Oper. Res.* **103** 1–17.

Frankel, C. 1996. The environment. *IEEE Spectrum* **33**(1) 76–81.

Guide, V. D. R., V. Jayaraman, J. Linton. 2003. Building contingency planning for closed-loop supply chains with product recovery. *J. Oper. Management* **21** 259–279.

Heyman, D., M. Sobel. 1984. *Stochastic Models in Operations Research*, Vol. II. McGraw-Hill, New York.

Hillier, M. 1999. Component commonality in a multi-period inventory model with service level constraints. *Internat. J. Production Res.* **37** 2665–2683.

Inderfurth, K. 1997. Simple optimal replenishment and disposal policies for a product recovery system with leadtimes. *OR Spectrum* **19** 111–122.

Kelle, P., E. Silver. 1989. Purchasing policy of new containers considering the random returns of previously issued containers. *IIE Trans.* **21** 349–354.

Kiesmüller, G., S. Minner. 2003. Simple expressions for finding recovery system inventory control parameter values. *J. Oper. Res. Soc.* **54** 83–88.

Kiesmüller, G., E. van der Laan. 2001. An inventory model with dependent product demands and returns. *Internat. J. Production Econom.* **72** 73–87.

Lu, Y., J.-S. Song. 2005. Order-based cost optimization in ATO systems. *Oper. Res.* **53** 151–169.

Mahadevan, B., D. Pyke, M. Fleischmann. 2003. Periodic review, push inventory policies for remanufacturing. *Eur. J. Oper. Res.* **151** 536–551.

Muckstadt, J., M. Isaac. 1981. An analysis of single item inventory systems with returns. *Naval Res. Logist. Quart.* **28** 237–254.

Scarf, H. 1960. The optimality of (s, S) policies in the dynamic inventory problem. K. Arrow, S. Karlin, P. Suppes, eds. *Mathematical Models in the Social Sciences*. Stanford University Press, Stanford, CA.

Schenkman, L. 2002. EU government to enforce e-waste recycling. *Waste Age* **33** 14.

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. 2002. Order-based backorders and their implications in multi-item inventory systems. *Management Sci.* **48** 499–516.

Song, J.-S., Y. Zhao. 2006. Lead times, component costs and allocation rules in the value of component commonality. Working paper, Duke University, Durham, NC.

Song, J.-S., P. Zipkin. 2003. Supply chain operations: ATO systems. T. De Kok, S. Graves, eds. *Supply Chain Management. Handbooks in Operations Research and Management Science*, Vol. 30. North-Holland, Amsterdam.

Song, J.-S., S. Xu, B. Liu. 1999. Order-fulfillment performance measures in an ATO system with stochastic leadtimes. *Oper. Res.* **47** 131–149.

Tedeschi, B. 2001. Online retailers grapple with the age-old problem of handling returned merchandise. *The New York Times* (May 28) C6.

Thorn, B., P. Rogerson. 2002. Take it back. *IIE Solutions* **34** 34–40.

Wolff, R. 1989. *Stochastic Models and the Theory of Queues*. Prentice-Hall, Englewood Cliffs, NJ.

Zipkin, P. 2000. *Foundations of Inventory Management*. McGraw-Hill, New York.