Managing an Assemble-to-Order System with Returns

Gregory DeCroix • Jing-Sheng Song • Paul Zipkin

The Fuqua School of Business, Duke University, Durham, NC 27708-0120

decroix@duke.edu, jssong@duke.edu, zipkin@duke.edu

October 2005

Abstract

We consider an 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 nearoptimal base-stock policy. We use these results to address managerial questions on both operational (e.g., the value of product-based return information, the impact of product recovery on inventory-related costs) and product-design (which components should be designed for recovery, the value of using common components across products) levels.
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, since 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, 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 a reasonable amount of balance among those components that are recovered and those that are not can be quite challenging. When the system produces multiple products that share components (some of which are recovered), the task becomes even more complex.

As an example of such a situation, consider Kodak’s single-use camera production and remanufacturing operations. After a consumer drops a used camera off at a photofinisher, the film is removed and the camera is sent to a disassembly facility. There the camera is cleaned and inspected, and the reusable components (e.g., the circuit board) are sent to one of Kodak’s production facilities. That facility must then manage inventories of entirely new components (e.g., batteries, external packaging) and reusable components (including ordering some new units of such components) to have the right supplies on hand to assemble a variety of camera models. (For a more detailed description of Kodak’s operations, see Guide et al. 2003.)

While we will tend to use terminology (e.g., components, products) based on this type of manufacturing/remanufacturing environment, it should be noted that our model and analysis also apply to online or catalogure retailers when customer returns are placed back
into inventory. In that case, a “component” corresponds to an individual item, while a “product” corresponds to a customer order.

Our goal is to develop methods for choosing and evaluating inventory management policies for multi-component, multi-item 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 Assemble-to-Order 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, in systems without returns it is known (see Song 2002) that it is important to track demand rates at the product level rather than just the component level. Is this extra information-gathering effort still worthwhile for returns, or is component-level information sufficient? Also, if companies are choosing to recover products out of economic self-interest, they need to determine whether such a choice is actually profitable. While using recovered components may reduce procurement costs, there is also an impact on the cost of managing component inventories. Do these changes result in a net benefit to the company? This question has implications for product design decisions as well. If a company has some degree of choice over which components to recover (by designing them to be recoverable, etc.), which items should they recover? Given the additional variability introduced by the return stream, is the use of component commonality across products more or less attractive than in a system without returns? How does the answer to that question depend on which components are recoverable? We explore these questions using our fundamental performance-metric and optimality results.

This paper builds on and contributes to two distinct streams of research. The first of these consists of research on the management of multi-item assemble-to-order systems with-
out returns. Of particular relevance for this paper is the recent 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 in the presence of return flows. For a single-location, finite-horizon inventory system, Heyman and Sobel (1984) point out that Scarf’s (1960) proof of the optimality of an (s,S) policy still works when the system faces uncertain returns in addition to demands. 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 approach for managing a similar system that also includes fixed order costs and stochastic return times.

Research on multi-stage or multi-item systems facing product returns has been rather limited. For an infinite-horizon series system, 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 an approximate optimization algorithm for computing a good policy. DeCroix and Zipkin (2005) establish conditions under which a single-product, multi-component 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 multi-product, multi-component ATO system. Sections 4 and 5 present methods for computing or approximating order-based fill rates and order-based backorders, respectively, while Section 6 discusses a method for computing near-optimal base-stock policies. Section 7 explores the value of product recovery and component commonality. Section 8 provides some concluding remarks.
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

\[
\begin{align*}
L &= \text{order leadtime, a positive number} \\
D(t) &= \text{cumulative demand by time } t, \\
R(t) &= \text{cumulative returns by time } t, \\
N(t) &= \text{cumulative net demand by time } t = D(t) - R(t), \\
N[t, u) &= N(t + u) - N(t) \\
IN(t) &= \text{net inventory (inventory minus backlog) at time } t \\
IP(t) &= \text{inventory position after demand at time } t \\
&= IN(t) \text{ plus stock in transit at time } t \\
I(t) &= \text{on-hand inventory at time } t \\
B(t) &= \text{backorders at time } t.
\end{align*}
\]

The demand process \( \{D(t), t \geq 0\} \) is a Poisson process with rate \( \mu \). The returns process \( \{R(t), t \geq 0\} \) is also Poisson process, with rate \( \lambda \). These two processes are independent and we assume \( \mu > \lambda \). The standard flow-conservation equation holds:

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

The assumption that returns are independent of past demands warrants some additional discussion. Returns, by definition, must have gotten to the customers by some means. It is tempting to specify returns as a function of the cumulative sales to date. This approach, however, adds a layer of complexity to the model. Indeed, a similar issue arises for demand itself. No product has an unlimited market. So, one could argue, current demand should always depend on cumulative demand. In practice, we rarely do that. 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. (See Fleischmann 2000 for a more detailed justification of this assumption in that context, 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.)
Now consider a system with stationary data, operating over an infinite horizon under a stationary base-stock policy with base-stock level $s$. Let the following denote equilibrium quantities:

- $IP$ = inventory position
- $IN$ = net inventory
- $I$ = on-hand inventory
- $B$ = backorders
- $N = N[0,L]$ = leadtime net demand $= D - R$

where $D$ is a Poisson random variable with mean $\mu L$ and $R$ is a Poisson random variable with mean $\lambda L$.

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. When returns are present, the behavior of the state variables is more complicated than usual. 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. Let $t' > t$ be the next demand or return epoch, Then, $Z(t') = [Z(t) - 1]^+$ if a demand occurs at $t'$ and $Z(t') = Z(t) + 1$ if a return occurs. Due to exponential interarrival times, $\{Z(t), t \geq 0\}$ is a birth-death process with birth rate $\lambda$ and death rate $\mu$. $Z(t)$ is equivalent to the queue length process in an $M/M/1$ queue with arrival rate $\lambda$ and service rate $\mu$.

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

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

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

$$f = P(IN > 0) = P(N < s + Z).$$
3. Assemble-to-Order System Model

Now consider a multi-product, multi-component Assemble-to-Order system. Let \( I = \{1, 2, \cdots, m\} \) denote the set of component indices. Customer orders arrive according to a stationary Poisson process with rate \( \mu \). Each customer order may require several components simultaneously. For any subset of components \( K \subseteq I \), we say a demand is of type \( K \) if it requests 1 unit of component \( i \in K \), and 0 units in \( I \setminus K \). 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, \( \mathcal{K} = \{K \subset I : q^K > 0\} \). Note that \( \mathcal{K} \) is not necessarily the set of all possible subsets of \( I \). 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 \( \lambda \), and that this process is independent of the demand process. A return is of type \( S \) if it contains 1 unit of component \( i \in S \subseteq I \), and 0 units in \( I \setminus S \). A return’s type is independent of the other returns’ types and of all other events. There is a fixed probability \( r^S \) that a return is of type \( S \), so that the type-\( S \) return stream forms a Poisson process with rate \( \lambda^S = r^S \lambda \).

Define \( \mathcal{S} = \{S : \lambda^S > 0\} \) 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 \( \mu_i \) and the other with rate \( \lambda_i \). 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 either ship the in-stock components or put them aside as committed inventory. However, a demand is considered backlogged until it is satisfied completely.
The inventory of each component is controlled by a base-stock policy, with

\[ s_i = \text{the base-stock level for component } i. \]

A base-stock policy has the same interpretation here as in the single-item case.

Let \( t \geq 0 \) be the continuous time variable, and for each \( t \) denote

\[
\begin{align*}
IN_i(t) & = \text{net inventory of item } i, \\
N^K(t) & = \text{cumulative net demand for type } K \text{ by time } t, \\
D_i(t) & = \text{cumulative demand for } i \text{ by time } t, \\
R_i(t) & = \text{cumulative returns for } i \text{ by time } t, \\
N_i(t) & = \text{cumulative net demand for } i \text{ by time } t = D_i(t) - R_i(t), \\
B^K(t) & = \text{type-}K\text{ backorders at } t \\
& = \text{number of type-}K\text{ orders that are not yet completely satisfied by } t, \\
B_i(t) & = \text{number of backorders for item } i \text{ at } t.
\end{align*}
\]

Let \( L_i \) be the constant lead time for replenishing component \( i \), and let \( D_i \) stand for 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

\[
W^K = \text{steady-state waiting time of a type-}K\text{ backorder}.
\]

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

\[
\begin{align*}
\phi_{K,w} & = \text{type-}K\text{ order fill rate within a time window } w \\
& = \text{probability of satisfying a type-}K\text{ order within a time window } w \\
& = \Pr(W^K \leq w) \\
f^K & = \text{immediate fill rate of type-}K\text{ demand} = \phi_{K,0} \\
\overline{B^K} & = \text{average number of type-}K\text{ backorders}.
\end{align*}
\]

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 \( f^K(\mathbf{s} | \mathbf{L}) \) (or simply \( f^K(\mathbf{s} | \mathbf{L}) \) 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:

\[ f = \text{average (over all demand types) immediate fill rate} = \sum_K q^K f^K. \]

\[ \overline{B} = \text{total average order-based backorders} = \sum_K \overline{B}^K. \]

At times it will also be convenient to make use of the component-based performance measures:

\[ f_i = \text{immediate fill rate of component } i, \]

\[ \overline{B}_i = \text{average number of backorders of component } i. \]

## 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 \( \lambda_i \) and service rate \( \mu_i \). Also, \( Z_i \) is independent of \( N_i \). The order-based performance evaluation involves the joint distribution of the net inventories \((IN_1, ..., IN_m)\), which, in turn, depends on the joint distribution of the leadtime net demands \((N_1, ..., N_m)\) and the joint distribution of \((Z_1, ..., Z_m)\). 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 2-component system unless otherwise noted. Here, there are three types of demand: A type-1 customer requires one unit of component 1 only; type-2 requires one unit of component 2 only; and type-12 (short notation 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

$$f^{12} = \Pr(IN_1 > 0, IN_2 > 0) = \Pr(N_1 < s_1 + Z_1, N_2 < s_2 + Z_2).$$

### 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 parameter $\mu^K L$, 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

$$f^{12}(s|L) = \sum_{k=-\infty}^{\infty} \Pr(N^{12}(L) = k) \Pr(N^1(L) < s_1 + Z_1 - k, N^2(L) < s_2 + Z_2 - k).$$

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

$$f^{12}(s|L) = \sum_{k=-\infty}^{\infty} \Pr(N^{12}(L) = k) \left[ \sum_{z_1, z_2 \geq 0} \Pr(N^1(L) < s_1 + z_1 - k) \Pr(N^2(L) < s_2 + z_2 - k) \right].$$

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

$$f^{12}(s|L) = \sum_{k=-\infty}^{\infty} \Pr(N^{12}(L) = k) \left[ \sum_{z_1, z_2 \geq 0} \Pr(N^1(L) < s_1 + z_1 - k) \Pr(N^2(L) < s_2 + z_2 - k) \right].$$

(1)

The quantity in brackets has the same form as $f^{12}(s|L)$ 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 is the most difficult part of the above computation.
This joint distribution can be computed using a matrix-geometric approach similar to that in Song, et al. (1999). Note that \( \{(Z_1(t), Z_2(t)), t \geq 0\} \) is a two-dimensional Markov chain on the first quadrant. Suppose the current state is \( Z_i(t) = z_i, i = 1, 2 \). Then, the next state for \( Z_i(t) \) is \( (z_i - 1)^+ \) if the next event includes a demand for component \( i \), \( z_i + 1 \) if the next event includes a return of component \( i \), and \( z_i \) otherwise. The chain is stable given the condition \( \lambda_i < \mu_i \). For computational purposes, we truncate the state space so that \( Z_i \) can never exceed some quantity \( M_i \). Then we can write the state space lexicographically, i.e., \([1_0, 1_1, ..., 1_{M_i}]\), where
\[
1_k = [(k, 0), (k, 1), (k, 2), ..., (k, M_2)].
\]
Letting \( p \) be the vector of stationary probabilities for the Markov chain, we can write \( p = (p_0, p_1, ..., p_{M_1}) \), where \( p_k \) is the vector of stationary probabilities associated with the states in \( 1_k \). Using this lexicographic ordering, the generator matrix for the Markov chain can be written as
\[
\tilde{Q} = \begin{bmatrix}
A & A_0 \\
C & A-C & A_0 \\
& C & A-C & A_0 \\
& & ... & ... \\
& & & ... & A_0 \\
& & & & C & A-C \\
& & & & & C & A_1-C
\end{bmatrix},
\]
where
\[
A = \begin{bmatrix}
-\lambda & \lambda^2 \\
\mu^2 + \mu^{12} & -(\lambda + \mu^2 + \mu^{12}) & \lambda^2 \\
... & ... & ... \\
\mu^2 + \mu^{12} & -(\lambda + \mu^2 + \mu^{12}) & \lambda^2 \\
& \mu^2 + \mu^{12} & -(\lambda + \mu^2 + \mu^{12}) \\
& & \mu^2 + \mu^{12}
\end{bmatrix},
\]
\[
A_0 = \begin{bmatrix}
\lambda^1 & \lambda^{12} \\
\lambda^1 & \lambda^{12} \\
\lambda^1 & \lambda^{12} \\
... & ... \\
\lambda^1 & \lambda^{12} \\
\lambda^1 & \lambda^{12}
\end{bmatrix},
\]
\[
C = \begin{bmatrix}
\mu^1 + \mu^{12} & \mu^{12} & \mu^1 & \mu^{12} \\
\mu^{12} & \mu^1 & \mu^{12} & \mu^1 \\
\mu^{12} & \mu^{12} & \mu^1 & \mu^{12} \\
\mu^{12} & \mu^{12} & \mu^1 & \mu^{12}
\end{bmatrix},
\]
and \( A_1 = A + A_0 \).

The stationary probabilities can then be computed recursively as follows. First, define \( V_{M_1} = -A_0(A_1 - C)^{-1} \) and \( V_k = -A_0[(A - C) + V_{k+1}C]^{-1} \) for \( k = M_1 - 1, \ldots, 1 \). With these definitions the balance equations \( p\hat{Q} = 0 \) can be written as \( p_0(A + V_1C) = 0 \), and the normalization equation \( \sum_{k=0}^{M_1} p_k e = 1 \) (where \( e \) is the column vector of ones) can be written as \( p_0 \sum_{k=0}^{M_1} \hat{V}_k e = 1 \), where \( \hat{V}_k \equiv \prod_{i=0}^{k} V_k, k = 0, 1, \ldots, M_1 \). These equations can be solved to find \( p_0 \), after which we compute \( p_k = p_{k-1}V_k \) for \( k = 1, 2, \ldots, M_1 \).

We tried a simpler approximate approach treating the \( Z_i \) as independent random variables, i.e., approximating \( \Pr(Z_1 = z_1, Z_2 = z_2) \) by the product of the marginals \( \Pr(Z_1 = z_1)\Pr(Z_2 = z_2) \). Numerical trials indicate that this approximation can lead to significant errors. This is particularly true when \( \lambda, \lambda^{12} \) or \( \mu^{12} \) is large, i.e., when the return rate is high enough that the \( Z_i \) variables are positive with significant probability, and when joint demands or returns are sufficiently high to introduce substantial correlation between the \( Z_i \).

### 4.2 Non-identical Lead Times

Now consider the case of non-identical 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,

\[
IN_2(t) = s_2 + Z_2(t - L - \Delta) - [N_2(t - L - \Delta, t - L) + N_2(t - L, t)].
\]

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

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

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

*** Figure 1 here***

To compute $f^{12}(s|L, L + \Delta)$ 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 over the time interval $0 \leq r \leq \Delta$ 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 inventories of items 1 and 2 from time $t - L - \Delta$ (i.e., $r = 0$) to time $t - L$ (i.e., $r = \Delta$). During the time period $(t - L - \Delta, t - L)$ orders for component 1 can be placed and still arrive by time $t$, so $X_1(r)$ evolves in the same manner as $Z_1$. That is, for an arbitrary time $r$, let $r'$ be the time of the next demand or return event (of any type). If that event includes a demand for component 1 then $X_1(r') = [X_1(r) - 1]^+$, if the event includes a return of component 1 then $X_1(r') = X_1(r) + 1$, and $X_1(r') = X_1(r)$ otherwise. Since orders of component 2 placed during this time period will not affect inventory levels for that component until after time $t$, these orders are not included in $X_2(r)$. Instead, that quantity tracks only demands and returns of item 2, so that $X_2(r') = X_2(r) - 1$ 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 the two components’ net inventories plus units on order at time $t - L$ that will be available in time to satisfy a demand at time $t$. Given the joint distribution of $(X_1(\Delta), X_2(\Delta))$, it is possible to compute $f^{12}(s|L, L + \Delta)$ similar to the case of equal lead times in (1), i.e.,

$$f^{12}(s|L, L + \Delta) = \sum_{x_1, x_2 \geq 0} \Pr(X_1(\Delta) = x_1, X_2(\Delta) = x_2).$$

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, from any starting state $(X_1(0), X_2(0)) = (x_1, x_2)$, $X_1(r)$ and $X_2(r)$ evolve independently during $(0, \Delta)$. This clearly is not true in
reality, since demands or returns of type-12 affect both variables. However, while $X_2(r)$ tracks all demands or returns of type-12, during $(0, \Delta)$ demands of type-12 do not affect $X_1(r)$ when $X_1(r) = 0$. As a result, each time $X_1(r)$ hits zero, it partially “forgets” past demand and return history. (See Figure 2 for an illustration.) This property reduces the dependence between the two variables. We would expect these variables to be less correlated as $r$ becomes larger and when $\lambda_1$ is relatively small compared to $\mu_1$ (which would result in frequent occurrences of $X_1(r) = 0$). Under this assumption,

$$\Pr(X_1(\Delta) = x_1, X_2(\Delta) = x_2) \approx \sum_{z_1, z_2 \geq 0} \Pr(Z_1 = z_1, Z_2 = z_2) \Pr(X_1(\Delta) = x_1|Z_1 = z_1) \Pr(X_2(\Delta) = x_2|Z_2 = z_2)$$

$$= \sum_{z_1 \geq 0} \Pr(Z_1 = z_1) \sum_{z_2 \geq 0} \Pr(Z_2 = z_2|Z_1 = z_1) \Pr(X_1(\Delta) = x_1|Z_1 = z_1) \Pr(X_2(\Delta) = x_2|Z_2 = z_2)$$

$$= \sum_{z_1 \geq 0} \Pr(Z_1 = z_1) \Pr(X_1(\Delta) = x_1|Z_1 = z_1) \sum_{z_2 \geq 0} \Pr(Z_2 = z_2|Z_1 = z_1) \Pr(X_2(\Delta) = x_2|Z_2 = z_2).$$

***Figure 2 here***

Since $X_2(r)$ is a simple birth and death process, we have $X_2(\Delta) = X_2(0) - N_2(\Delta)$, and the distribution of $X_2(\Delta)$ is relatively easy to compute since $X_2(0)$ and $N_2(\Delta)$ are independent. Computing the distribution of $X_1(\Delta)$ given any starting state, however, requires transient analysis of an $M/M/1$ queue. As $\Delta \to \infty$, however, 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 $\Delta$ this should be a good approximation. For tractability we define a variable $X_1$ with such a distribution, and then approximate $\Pr(X_1(\Delta) = x_1|Z_1 = z_1)$ by $\Pr(X_1 = x_1)$. With this, the above quantity is approximately equal to

$$\Pr(X_1 = x_1) \sum_{z_1 \geq 0} \Pr(Z_1 = z_1) \sum_{z_2 \geq 0} \Pr(Z_2 = z_2|Z_1 = z_1) \Pr(N_2(t - L - \Delta, t - L) = z_2 - x_2)$$

$$= \Pr(X_1 = x_1) \sum_{z_2 \geq 0} \Pr(N_2(t - L - \Delta, t - L) = z_2 - x_2) \sum_{z_1 \geq 0} \Pr(Z_1 = z_1) \Pr(Z_2 = z_2|Z_1 = z_1)$$

$$= \Pr(X_1 = x_1) \sum_{z_2 \geq 0} \Pr(N_2(t - L - \Delta, t - L) = z_2 - x_2) \Pr(Z_2 = z_2)$$

$$= \Pr(Z_1 = z_1) \sum_{z_2 \geq 0} \Pr(N_2(t - L - \Delta, t - L) = z_2 - x_2) \Pr(Z_2 = z_2), \quad (3)$$

14
where the last equality follows from the fact that $Z_1$ and $X_1$ have the same distribution. Substituting (3) in (2) yields

$$f_{12}(s|L, L + \Delta) \approx \sum_{z_1, z_2 \geq 0} \Pr(Z_1 = z_1) \Pr(Z_2 = z_2) \sum_{x_2} \Pr(N_{2,2} = z_2 - x_2) \cdot \left[ \sum_{k=-\infty}^{\infty} \Pr(N_{12}^2(L) = k) \Pr(N_1^1(L) < s_1 + z_1 - k) \Pr(N_2^2(L) < s_2 + z_2 - k) \right]$$

$$= \sum_{z_1, z_2 \geq 0} \Pr(Z_1 = z_1) \Pr(Z_2 = z_2) \sum_{y=-\infty}^{\infty} \Pr(N_{2,2} = y) \cdot \left[ \sum_{k=-\infty}^{\infty} \Pr(N_{12}^2(L) = k) \Pr(N_1^1(L) < s_1 + z_1 - k) \Pr(N_2^2(L) < s_2 + z_2 - y - k) \right]$$

$$= \sum_{y=-\infty}^{\infty} \Pr(N_{2,2} = y) \sum_{z_1, z_2 \geq 0} \Pr(Z_1 = z_1) \Pr(Z_2 = z_2) \cdot \left[ \sum_{k=-\infty}^{\infty} \Pr(N_{12}^2(L) = k) \Pr(N_1^1(L) < s_1 + z_1 - k) \Pr(N_2^2(L) < s_2 + z_2 - y - k) \right]. \tag{4}$$

To test the accuracy of this approximation, the results from (4) were compared to simulation results for the set of 60 test problems using all combinations of parameters given in Table 1. Base-stock levels were set as

$$s_i = \left[ (\mu_i - \lambda_i)L_i + \alpha \sqrt{(\mu_i + \lambda_i)L_i} \right], \tag{5}$$

where $[x]$ represents the largest integer less than $x$, and the $\alpha$ 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 2%. It was most accurate at relatively high fill rates (which is where most systems would likely operate in practice) – the average relative errors were 3.92% for $\alpha = 0$, 1.65% for $\alpha = 0.67$, and 0.43% for $\alpha = 1.64$.

<table>
<thead>
<tr>
<th>$(L_1, L_2)$</th>
<th>$(\mu^t, \mu^\mu, \mu^\mu^2)$</th>
<th>$(\lambda^t, \lambda^\mu, \lambda^\mu^2)$</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1, 2)</td>
<td>(8, 8, 4)</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>(5, 5, 10)</td>
<td>(2, 0, 0)</td>
<td></td>
<td>0.67</td>
</tr>
<tr>
<td>(2, 2, 16)</td>
<td>(6, 0, 0)</td>
<td></td>
<td>1.64</td>
</tr>
<tr>
<td>(7, 3, 10)</td>
<td>(5, 5, 10)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Parameters for numerical trials
Recall that in a system without returns, we have
\[ f^{12}(s|L, L + \Delta) = \sum_{y=0}^{s_{2} - 1} \Pr(N_{2,2} = y) f^{12}(s - y e_{2}|L), \]
where \( e_{i} \) is the unit \( m \)-vector whose \( i \)-th 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. It is interesting to note, however, that under our approximating assumptions, conditioning on the \( Z_i \) variables in the unequal-lead-time case is actually 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 \( t - L - \Delta \)).

### 4.3 Fill Rate within a Time Window

For general \( m \)-component systems without returns, Song (1998) shows that, for any fixed \( K \) and \( 0 \leq w < \max_{i \in K} \{L_i\} \),
\[ f^{K,w}(s|L) = f^{K,0}(s|(L_1 - w)^+, ..., (L_m - w)^+), \tag{6} \]
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 leadtimes truncated 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 \( 0 \leq w < \max_{i \in K} \{L_i\} \),
\[ f^{K,w}(s|L) = \sum_{x} \Pr\{\text{IN}(s| (L - w e)^+) = x\} \Pr\{R_i(w) > -x, i \in K\}, \tag{7} \]
where $\text{IN}(s | (L - w)^+) = (IN_i(s_i | (L_i - w))^+) | i$. By rewriting (7) as

$$f^{K,w}(s|L) = \sum_{x_i > 0 \text{ for } i \in K} \Pr\{\text{IN}(s | (L - w)^+) = x \} \Pr\{R_i(w) > -x_i, i \in K\}$$

$$+ \sum_{x_i \leq 0 \text{ for some } i \in K} \Pr\{\text{IN}(s | (L - w)^+) = x \} \Pr\{R_i(w) > -x_i, i \in K\}$$

$$= f^{K,0}(s | (L_1 - w)^+, ..., (L_m - w)^+)$$

$$+ \sum_{x_i \leq 0 \text{ for some } i \in K} \Pr\{\text{IN}(s | (L - w)^+) = x \} \Pr\{R_i(w) > -x_i, i \in K\}$$

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

For the two-component system with equal lead times, we can write

$$f^{12,w}(s|L) = f^{12,0}(s|L - w)$$

$$+ \sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} \Pr \{IN_1(s_1 | L - w) = -x_1, IN_2(s_2 | L - w) = -x_2\}$$

$$\times \Pr \{R_1(w) \geq x_1 + 1, R_2(w) \geq x_2 + 1\}$$

$$+ \sum_{x_1=0}^{\infty} \sum_{q_2=1}^{\infty} \Pr \{IN_1(s_1 | L - w) = -x_1, IN_2(s_2 | L - w) = q_2\}$$

$$\times \Pr \{R_1(w) \geq x_1 + 1\}$$

$$+ \sum_{q_1=1}^{\infty} \sum_{x_2=0}^{\infty} \Pr \{IN_1(s_1 | L - w) = q_1, IN_2(s_2 | L - w) = -x_2\}$$

$$\times \Pr \{R_2(w) \geq x_2 + 1\}.$$  \hfill (8)

The second factor in the terms in (8) can be written as

$$\Pr \{R_1(w) \geq x_1 + 1, R_2(w) \geq x_2 + 1\}$$

$$= 1 - \Pr \{R_1(w) \leq x_1\} - \Pr \{R_2(w) \leq x_2\} + \Pr \{R_1(w) \leq x_1, R_2(w) \leq x_2\}$$

$$= 1 - P(x_1, \lambda_1 w) - P(x_2, \lambda_2 w) + \sum_{k=0}^{x_1 \wedge x_2} p(k, \lambda^{12} w) P(x_1 - k, \lambda^1 w) P(x_2 - k, \lambda^2 w),$$  \hfill (9)

and

$$\Pr \{R_i(w) \geq x_i + 1\} = P^c(x_i + 1, \lambda_i w), \quad i = 1, 2,$$  \hfill (10)

where $p(\cdot, \xi)$ is the Poisson probability mass function with mean $\xi$, $P(\cdot, \xi)$ is the Poisson cumulative distribution function, and $P^c(\cdot, \xi) = 1 - P(\cdot, \xi)$. The first factor in the terms in (8) can be written using an approach similar to that used in Song (2002). The idea is
to condition on the number of events that can occur in the lead time \( L - w \) (a Poisson random variable with parameter \((\lambda + \mu)(L - w)\)), then, given the total number of events, express the probability of any particular vector (which has a multinomial distribution). To obtain the joint distribution of \( \mathbf{IN}(s|L - w) \) we then just sum over all vectors of events that yield the same net inventory vector. Due to the presence of returns, the approach here also requires conditioning on (and then summing over) all possible values of the excess inventory position vector \((Z_1, Z_2)\). Specifically, if we define \( \beta^K = \mu^K/(\lambda + \mu), \gamma^K = \lambda^K/(\lambda + \mu), \) and \( \delta_i = s_i + z_i + r^{12} - d^{12} + x_i \), we have

\[
\Pr\{IN_1(s_1|L - w) = -x_1, IN_2(s_2|L - w) = -x_2\} = \sum_{z_1=0}^{\infty} \sum_{z_2=0}^{\infty} \Pr\{Z_1 = z_1, Z_2 = z_2\} \\
\times \sum_{d^{12}=0}^{\infty} \sum_{r^{12}=0}^{\infty} \sum_{d^1 = \max\{0, \delta_1\}}^{\infty} \sum_{d^2 = \max\{0, \delta_2\}}^{\infty} \frac{(d^1 + d^2 + d^{12} + (d^1 - \delta_1) + (d^2 - \delta_2) + r^{12})!}{d^1!d^2!d^{12}!(d^1 - \delta_1)!(d^2 - \delta_2)!r^{12}!} \\
\times (\beta^K)^{d^1} (\beta^{12})^{d^{12}} (\gamma^K)^{(d^1 - \delta_1)} (\gamma^{12})^{(d^2 - \delta_2)} (\gamma^{r^{12}})^{r^{12}} \\
\times \left[\frac{[(\lambda + \mu)(L - w)]^{(d^1 + d^2 + d^{12} + (d^1 - \delta_1) + (d^2 - \delta_2) + r^{12})}}{(d^1 + d^2 + d^{12} + (d^1 - \delta_1) + (d^2 - \delta_2) + r^{12})!}\right] \\
\times e^{-(\lambda + \mu)(L - w)}.
\] (11)

Expressions for \( \Pr\{IN_1(s_1|L - w) = -x_1, IN_2(s_2|L - w) = q_2\} \) and \( \Pr\{IN_1(s_1|L - w) = q_1, IN_2(s_2|L - w) = -x_2\} \) are essentially the same, with either \( q_2 \) substituted for \( -x_2 \) or \( q_1 \) substituted for \( -x_1 \), respectively.

Expressions for \( f^{1,w}(s|L) \) and \( f^{2,w}(s|L) \) are similar but simpler, since only component-based information is required. Specifically,

\[
f^{1,w}(s|L) = f^{1,0}(s|L - w) + \sum_{x_1=0}^{\infty} \Pr\{IN_1(s_1|L - w) = -x_1\} \times \Pr\{R_1(w) \geq x_1 + 1\},
\]

where

\[
\Pr\{IN_1(s_1|L - w) = -x_1\} = \\
\sum_{z_1=0}^{\infty} \Pr\{Z_1 = z_1\} \times \sum_{d^{12}=0}^{\infty} \sum_{r^{12}=0}^{\infty} \sum_{d^1 = \max\{0, \delta_1\}}^{\infty} \frac{(d^1 + d^{12} + (d^1 - \delta_1) + r^{12})!}{d^1!d^{12}!(d^1 - \delta_1)!r^{12}!} \\
\times (\beta^K)^{d^1} (\beta^{12})^{d^{12}} (\gamma^K)^{(d^1 - \delta_1)} (\gamma^{12})^{r^{12}} \\
\times \left[\frac{[(\lambda_1 + \mu_1)(L - w)]^{(d^1 + d^{12} + (d^1 - \delta_1) + r^{12})}}{(d^1 + d^{12} + (d^1 - \delta_1) + r^{12})!}\right] \\
\times e^{-(\lambda_1 + \mu_1)(L - w)}
\]

and the expression for component 2 is defined symmetrically.

Given the computational effort required by (11), it would be useful to know whether the lower bound in (6) provides a good estimate for \( f^{K,w}(s|L) \). We tested this for a two-component setting using the parameters in Table 1 (with \( \alpha \) fixed at 0.67) and time windows
\( w = 0.125, 0.25, ..., 0.75 \). 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 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. The reason for this is that, 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, so the approximation becomes more accurate.

***Figure 3 here***

***Figure 4 here***

5. Order-Based Backorders

Now, let us consider the evaluation of the average order-based backorders \( \overline{B}^K (s|L) \). 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 the average type-\( i \) backorders equals

\[
\overline{B}^i(s_i|L) = \frac{q^i}{q^i + q^{12}} \overline{B}_i(s_i|L),
\]

where

\[
\overline{B}_i(s_i|L) = E[(N_i - s_i - Z_i)^+]
\]

represents expected component-\( i \) backorders.

As in Song (2002), Little’s Law implies that the average type-12 backorders equals

\[
\overline{B}^{12}(s|L) = \mu^{12} E[W^{12}(s|L)],
\]

where \( E[W^{12}(s|L)] \) is the average wait experienced by a type-12 customer. This can be expressed as

\[
E[W^{12}(s|L)] = \int_0^L \Pr\{W^{12}(s|L) > w\}dw = \int_0^L (1 - f^{12,w}(s|L))\,dw = L - \int_0^L f^{12,w}(s|L)dw,
\]
since $W^{12}(s|L) \leq L$ with probability 1. To facilitate a more detailed expression for $B^{12}(s|L)$, it is convenient to write $f^{12,0}(s|t)$ as

$$
\begin{align*}
&f^{12,0}(s|t) = \sum_{z_1=0}^{\infty} \sum_{z_2=0}^{\infty} \Pr \{ Z_1 = z_1, Z_2 = z_2 \} \\
&\times \sum_{d^{12}=0}^{\infty} \sum_{r^{12}=0}^{\infty} \sum_{d^1=0}^{\infty} \sum_{d^2=0}^{\infty} \sum_{r^1=\max\{0,d^1-\phi_1\}}^{\infty} \sum_{r^2=\max\{0,d^2-\phi_2\}}^{\infty} \frac{(d + r)!}{d!d^1!d^2!r^{12}!r^1!r^2!} \\
&\times (\beta^1)^{d^1} (\beta^2)^{d^2} (\beta^{12})^{d^{12}} (\gamma^1)^{r^1} (\gamma^2)^{r^2} (\gamma^{12})^{r^{12}} \left[ \frac{((\lambda + \mu)t)^{d+r}}{(d+r)!} \times e^{-(\lambda+\mu)t} \right],
\end{align*}
$$

(14)

where $\phi_i = s_i + z_i + r^{12} - d^{12} - 1$, $d = d^1 + d^2 + d^{12}$, and $r = r^1 + r^2 + r^{12}$. Substituting (9), (10), (11) and (14) into (8), then substituting the result into (13) and then (12) yields

$$
\begin{align*}
\bar{B}^{12}(s|L) &= \mu^{12}L - \mu^{12} \sum_{z_1=0}^{\infty} \sum_{z_2=0}^{\infty} \Pr \{ Z_1 = z_1, Z_2 = z_2 \} \sum_{d^{12}=0}^{\infty} \sum_{r^{12}=0}^{\infty} \sum_{d^1=0}^{\infty} \sum_{d^2=0}^{\infty} \sum_{r^1=\max\{0,d^1-\phi_1\}}^{\infty} \sum_{r^2=\max\{0,d^2-\phi_2\}}^{\infty} \frac{(d + r)!}{d!d^1!d^2!r^{12}!r^1!r^2!} \\
&\times (\beta^1)^{d^1} (\beta^2)^{d^2} (\beta^{12})^{d^{12}} (\gamma^1)^{r^1} (\gamma^2)^{r^2} (\gamma^{12})^{r^{12}} \\
&\int_0^L \left[ \frac{((\lambda + \mu)(L - w))^{d+r}}{(d+r)!} \times e^{-(\lambda+\mu)(L-w)} \right] dw
\end{align*}
$$

$$
-\mu^{12} \int_0^L \left\{ \sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} 1 - P(x_1, \lambda_1w) - P(x_2, \lambda_2w) \\
+ \sum_{k=0}^{x_1 \wedge x_2} p(k, \lambda^2w)P(x_1-k, \lambda^1w)P(x_2-k, \lambda^2w) \sum_{z_1=0}^{\infty} \sum_{z_2=0}^{\infty} \Pr \{ Z_1 = z_1, Z_2 = z_2 \} \\
\times \frac{(d + (d^1 - \delta_1) + (d^2 - \delta_2) + r^{12})!}{d!d^{12}!d^1!d^2!(d^1-\delta_1)!(d^2-\delta_2)!r^{12}!} \\
\times (\beta^1)^{d^1} (\beta^2)^{d^2} (\beta^{12})^{d^{12}} (\gamma^1)^{(d^1-\delta_1)} (\gamma^2)^{(d^2-\delta_2)} (\gamma^{12})^{r^{12}} \\
\times \left[ \frac{((\lambda + \mu)(L - w))^{d+(d^1-\delta_1)+(d^2-\delta_2)+r^{12}}}{(d + (d^1 - \delta_1) + (d^2 - \delta_2) + r^{12})!} \times e^{-(\lambda+\mu)(L-w)} \right] dw \right\}
\right.
$$

20
\[-\mu^{12} \int_0^L \left\{ \sum_{x_1=0}^{\infty} \sum_{q_1=0}^{\infty} P^c(x_1 + 1, \lambda_1 w) \sum_{z_1=0}^{\infty} \sum_{z_2=0}^{\infty} \Pr \{ Z_1 = z_1, Z_2 = z_2 \} \right. \]
\[\times \sum_{d^1=0}^{\infty} \sum_{r^2=0}^{\infty} \sum_{d^1=\max\{0,\delta_1\}}^{\infty} \sum_{d^2=\max\{0,\psi_2\}}^{\infty} \frac{(d + (d^1 - \delta_1) + (d^2 - \psi_2) + r^{12})!}{d!^2 d^2!^2! (d^1 - \delta_1)! (d^2 - \psi_2)! r^{12}!} \times (\beta^1)^{d^1} (\beta^2)^{d^2} (\beta^{12})^{d^{12}} (\gamma^1)^{(d^1 - \delta_1)} (\gamma^2)^{(d^2 - \psi_2)} (\gamma^{12})^{r^{12}} \times e^{-(\lambda + \mu)(L - w)} \left\{ \sum_{z_1=0}^{\infty} \sum_{z_2=0}^{\infty} \Pr \{ Z_1 = z_1, Z_2 = z_2 \} \right. \]
\[\times \sum_{d^1=0}^{\infty} \sum_{r^2=0}^{\infty} \sum_{d^1=\max\{0,\psi_1\}}^{\infty} \sum_{d^2=\max\{0,\delta_2\}}^{\infty} \frac{(d + (d^1 - \psi_1) + (d^2 - \delta_2) + r^{12})!}{d!^2 d^2!^2! (d^1 - \psi_1)! (d^2 - \delta_2)! r^{12}!} \times (\beta^1)^{d^1} (\beta^2)^{d^2} (\beta^{12})^{d^{12}} (\gamma^1)^{(d^1 - \psi_1)} (\gamma^2)^{(d^2 - \delta_2)} (\gamma^{12})^{r^{12}} \times e^{-(\lambda + \mu)(L - w)} \left\} \right. \]
\[\times e^{-(\lambda + \mu)(L - w)} \left. \right\} dw \]

where $\psi_i = s_i + z_i + r^{12} - d^{12} - q_i$.

The presence of returns makes the expression for $B^{12}(s|L)$ more complicated than that in Song (2002) in several ways. First, since returns can be used to satisfy demands that would otherwise not be satisfied within the time window, the expression involves three additional groups of summations. Second, the addition of returns doubles the number of different types of events that must be considered within each group of summations – for each demand type we must also allow for a return of that type – and each new event type adds a summation. Third, since demands and returns cancel each other, it is not possible to put an upper bound on the number of any single type of event – e.g., 5 demands and 3 returns of the same type yield the same inventory level as 1005 demands and 1003 returns, etc. As a result, each summation contains an infinite number of terms. In practice one would need to truncate these summations at some finite limit. Finally, note that each group of summations (i.e., each term beginning with the coefficient $-\mu^{12}$) involves an integral. In the first of these terms, this integral can be pulled inside the summations, and the identity

\[\int_0^L \frac{[\lambda + \mu](L - w)^n}{n!} e^{-(\lambda + \mu)(L - w)} dw = \frac{1}{(\lambda + \mu)^n} P^c(n, (\lambda + \mu)L) \]

(16)

can be used to obtain a closed form expression for that term. The other terms include additional expressions that depend on $w$, however, so there does not appear to be a simple
way to replace the integral with a closed form expression. However, note that

\[ 1 - P(x_1, \lambda_1 w) - P(x_2, \lambda_2 w) + \sum_{k=0}^{x_1 \wedge x_2} p(k, \lambda_1^2 w) P(x_1 - k, \lambda_1 w) P(x_2 - k, \lambda_2^2 w) \]  

and

\[ P^c(x_i + 1, \lambda_i w), \quad i = 1, 2 \]

are increasing in \( w \). As a result, if we replace \( w \) with 0 (\( L \)) in these terms, we obtain a lower (upper) bound for those terms, and thus an upper (lower) bound for \( \bar{B}^{12}(s|L) \). After either of these substitutions, the identity (16) can be used to obtain a closed form expression for the upper and lower bounds for \( \bar{B}^{12}(s|L) \). (Note that \( w = 0 \) corresponds to the bound in (6).)

As one might expect, (15) is computationally intensive even for small problems – even computing the bounds mentioned above (which avoid numerical integration) is too time-consuming to be of practical use. As a result, we are interested in obtaining easy-to-compute bounds and approximations for the average backorders.

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

\[ LB^K \overset{def}{=} \mu^K \max_{i \in K} \frac{\bar{B}_i}{\mu_i} = \max_{i \in K} B^K_i \leq \mathbb{E}[\max B^K_i] = \bar{B}^K \leq \mathbb{E} \left[ \sum_{i \in K} B^K_i \right] = \mu^K \sum_{i \in K} \frac{\bar{B}_i}{\mu_i} \overset{def}{=} UB^K. \]

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

\[ LB \overset{def}{=} \sum_K LB^K \leq \bar{B} \leq \sum_K UB^K \overset{def}{=} UB. \]

A natural approximation for \( \bar{B}^K \) is the simple average \( AB^K = \frac{1}{2}(LB^K + UB^K) \). In the 2-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 system-wide 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, again using the set of two-component sample problems described in Table 1. For each of the 60 cases we used simulation to
compute an estimate for the true value of $\bar{B}_{12}^2$, 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 system-wide 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 ($\mu = (2, 2, 16)$) – the average error for those cases was 5.1%, while the average error for the remaining 45 cases was 3.1%. In most cases the relative error (prior to taking the absolute value) became more negative at higher fill rates – i.e., as $\alpha$ increased. In other words, if $AB$ overestimated backorders for $\alpha = 0$, then the relative amount of that overestimate shrank, or $AB$ began underestimating backorders as $\alpha$ changed to 0.67 and then 1.64. (Or, if $AB$ underestimated backorders for $\alpha = 0$, then the relative amount of that underestimate grew as $\alpha$ changed to 0.67 and then 1.64.)

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 $\mu_{12}$, so there is value in tracking that information rather than just relying on the component-based demand rates $\mu_1$ and $\mu_2$. 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 $\lambda_{12}$, and if so, does $\lambda_{12}$ affect $\bar{B}_{12}$ in the same way as $\mu_{12}$ does.

To answer that question, assume that the item-based rates $\mu_i$ and $\lambda_i$ are fixed, and that inventory is managed using the component-based base-stock levels in (5). Actual demands and returns occur in a product-based manner, however – i.e., $\mu_{12}$ and $\lambda_{12}$ may be positive. We explored the impact of joint demands and returns for a set of problems with parameters as in Table 1 except for the return rates, which we now assumed 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 $\lambda_{12}$ 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 $\lambda_{12}$ while varying $\mu_{12}$ from $\bar{\mu}_{12}^2 - \text{min}\{\lambda_1, \lambda_2\}/2$ to $\bar{\mu}_{12}^2 + \text{min}\{\lambda_1, \lambda_2\}/2$ (keeping $\mu^i = \mu_i - \mu_{12}, i = 1, 2$), where $\bar{\mu}_{12}$ is the original value of $\mu_{12}$ in Table 1. (I.e., we consider changes in $\lambda_{12}$ and $\mu_{12}$ of the same absolute size). Since the overall demand rate increases with $\mu_{12}$ (for fixed $\mu_1, \mu_2$), it is more appropriate to measure the impact of changing $\mu_{12}$ using the
average customer waiting time $W = \bar{B}/\mu$, rather than system-wide average backorders – and to facilitate comparison, we also use $W$ to measure the impact of changing $\lambda^{12}$.

The results of this numerical study indicate that the impact of $\lambda^{12}$ differs from that of $\mu^{12}$ in two ways. First, increasing $\lambda^{12}$ leads to shorter average customer waiting times, while increasing $\mu^{12}$ leads to longer waits. While this difference in behavior may be a bit surprising at first, there is an intuitive explanation for it. It is more difficult to satisfy a demand for both components at the same time than it is to satisfy demands for one component at a time, since in the former case a backorder will occur if either component is unavailable. As a result, a higher $\mu^{12}$ will result in more backorders and longer waits. A higher value of $\lambda^{12}$, however, corresponds to a higher incidence of both components being returned together. Receiving both components at the same time is more valuable than receiving one at a time, since a joint return now makes it possible to fill any type of demand. As a result, a higher $\lambda^{12}$ leads to fewer backorders and shorter waits. The second difference is in the magnitude of the impact – a change in $\lambda^{12}$ tends to have less of an impact on customer waits than the same size change in $\mu^{12}$. Figure 5 shows a case-by-case comparison of the magnitude (absolute value) of the total change in $W$ as $\mu^{12}$ and $\lambda^{12}$, respectively, move from their minimum to their maximum values. Across the 36 cases considered, the impact of $\lambda^{12}$ was only 28.4% of the impact of $\mu^{12}$ 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 the results of this numerical trial are also consistent with the fact that the $AB$ approximation, which performs reasonably well across a range of parameter values, is constant as $\lambda^{12}$ changes (but not as $\mu^{12}$ changes). This behavior is also suggested by the expression in (15). The only parts of that expression that depend on $\lambda^{12}$ rather than just the $\lambda_i$ are the joint distribution of the $Z_i$ and the expression in (17).

***Figure 5 here***

6. Optimization

While the preceding sections focused on calculating or estimating performance metrics for given base-stock levels, we now turn to the task of determining a base-stock vector that minimizes 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
The expected cost function we wish to minimize is

\[ C(s) = \sum_{i=1}^{m} h_i I_i(s) + \sum_{K \in K} b^K \overline{B^K}(s), \]

where \( I_i(s) \) is the average on-hand inventory for item \( i \). (Note that in this section we use slightly modified notation to emphasize the dependence of the performance metrics on the vector \( s \) of base-stock levels – e.g., \( \overline{B^K}(s) \), \( IN_i(s_i) \), etc.) On-hand inventory is made up of two parts. The first part is inventory that is on-hand and available to allocate to new demands, i.e., \( [IN_i(s_i)]^+ = [s_i + Z_i - D_i + R_i]^+ \). The second part is inventory that is still being held, but that has been set aside for earlier demands (which are backordered due to shortage of another component), which we denote \( J_i(s) \). As in the case with no returns (see Lu and Song 2005), it can be shown that

\[ J_i(s) = \sum_{K \in K_i} [B^K(s) - B^K_i(s_i)] = \sum_{K \in K_i} B^K(s) - B_i(s_i). \]

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

\[
C(s) = \sum_{i=1}^{m} h_i \mathbb{E} \left[ (s_i + Z_i - D_i + R_i + B_i(s_i)) + \left( \sum_{K \in K_i} B^K(s) - B_i(s_i) \right) \right] + \sum_{K \in K} b^K \overline{B^K}(s) \\
= \sum_{i=1}^{m} h_i \mathbb{E} [s_i + Z_i - D_i + R_i] + \sum_{K \in K} \left( b^K + \sum_{i \in K} h_i \right) \overline{B^K}(s) \\
= \sum_{i=1}^{m} h_i s_i + \sum_{K \in K} \tilde{b^K} \overline{B^K}(s) + \sum_{i} h_i \mathbb{E} [Z_i - D_i + R_i],
\]

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

Given the difficulty of calculating \( \overline{B^K} \), we are interested in approximating \( C(s) \) with cost functions that are easier to compute, and whose optimal solutions yield near-optimal solutions for \( C(s) \). 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

\[
C^C(s) = \sum_{i=1}^{m} \left( h_i I_i(s_i) + b_i \overline{B}_i(s_i) \right)
\]

\[
= \sum_{i=1}^{m} \left( h_i E[s_i + Z_i - D_i + R_i + B_i(s_i)] + b_i \overline{B}_i(s_i) \right)
\]

\[
= \sum_{i=1}^{m} \left( h_i s_i + (h_i + b_i) \overline{B}_i(s_i) \right) + \sum_{i=1}^{m} h_i E[Z_i - D_i + R_i],
\]

where again the final term is a constant with respect to \( s \). 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 \( s^C_i \) 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 \( C^C(s) \) and \( s^C_i \) are good approximations for \( C(s) \) and \( s^* \).

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

\[
b^u_i = \sum_{K \in K_i} \frac{\mu^K}{\mu_i} b^K - h_i = \sum_{K \in K_i} \frac{\mu^K}{\mu_i} \left( b^K + \sum_{j \in K, j \neq i} h_j \right).
\]

They showed that this choice of \( b_i \) yields an upper bound on the optimal solution \( s^* \). The following result states that this also holds when returns are present. To facilitate the statement of the result, define \( \Delta_i f(x) = f(x + e_i) - f(x) \) for \( i = 1, \ldots, m \). When \( m = 1 \), we simply write \( \Delta f(x) = \Delta_1 f(x) \).

**Proposition:** Set \( b_i = b^u_i \) in \( C^C(s) \), and denote the resulting solution by \( s^u \). Then \( s^* \leq s^u \).

The proof follows the same logic as the proof of Proposition 6 of Lu and Song (2005), and is omitted.

Lu and Song (2005) observe that in a system without returns \( s^u \) is often close to \( s^* \), and thus in addition to providing an upper bound, it also serves as a good approximation. While \( s^u \) 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 \( s^* \).

Recall from the previous section that

\[
LB^K(s) \overset{\text{def}}{=} \mu^K \max_{i \in K} \frac{\overline{B}_i(s_i)}{\mu_i} \leq \overline{B}^K(s) \leq \mu^K \sum_{i \in K} \frac{\overline{B}_i(s_i)}{\mu_i} \overset{\text{def}}{=} UB^K(s).
\]
Define

$$\overline{LB}^K(s) \overset{\text{def}}{=} \frac{\mu^K}{|K|} \sum_{i \in K} \frac{B_i(s_i)}{\mu_i},$$

where $|K|$ is the number of elements in the set $K$. Since $\overline{LB}^K(s) \leq LB^K(s)$, this also serves as a lower bound for $\overline{B}^K(s)$. Now instead of approximating $\overline{B}^K(s)$ by $AB^K(s) = \frac{1}{2}(LB^K(s) + UB^K(s))$, use

$$\overline{AB}^K(s) = \frac{1}{2}(\overline{LB}^K(s) + UB^K(s)) = \frac{(|K| + 1)\mu^K}{2|K|} \sum_{i \in K} \frac{B_i(s_i)}{\mu_i}. \quad (18)$$

Substituting $\overline{AB}^K(s)$ into $C(s)$ yields $C^C(s)$ with $b_i = \overline{b}_i$, where

$$\overline{b}_i = \left[ \frac{\sum_{K \in \mathcal{K}_i} \left( \frac{(|K| + 1)\mu^K}{2|K|} \right) b^K}{\mu_i} \right] - h_i = \left[ \frac{\sum_{K \in \mathcal{K}_i} \left( \frac{(|K| + 1)\mu^K}{2|K|} \right) (b^K + \sum_{j \in K} h_j)}{\mu_i} \right] - h_i.$$

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

To test the performance of $\mathbf{s}$, we again used the 60 test problems described in Table 1. However, instead of setting $s$ based on the three values of $\alpha$, we considered three holding and backorder cost combinations $(h_1, h_2, b^1, b^2, b^{12}) = (1, 2, 2, 4, 10), (1, 2, 2, 4, 6), (1, 2, 4, 4, 6)$. For each of the test problems, we computed $\mathbf{s}$ directly and then estimated $\overline{B}^{12}(\mathbf{s})$ by simulation, which allowed for calculation of $C(\mathbf{s})$. Values for $s^*$ and $C(s^*)$ were obtained by simulating the system for each $s$ in the rectangle $0 \leq s \leq s^*$ to estimate $\overline{B}^{12}(s)$, then calculating $C(s)$ for each $s$ and identifying $s^*$ as the one yielding the lowest cost. (In practice, the search over $s$ was substantially reduced by making use of a lower bound on $C(s)$. Substitute $\overline{LB}^K(s)$ for $\overline{B}^K(s)$ in $C(s)$ and call the result $C(s)$. It is easy to see that $C(s)$ can be written in the component-based form $C^C(s)$ for a particular choice of $b_i$, so it can be computed directly without simulation, and also that $C(s) \leq C^C(s)$. Eliminating from consideration any $s$ such that $C(s) > C(\mathbf{s})$ greatly reduced the search region in many cases.)

The $\mathbf{s}$ policy performed quite well in the trial. Out of the 60 cases, 19 yielded $\mathbf{s} = s^*$ (vs. 0 cases yielding $s^v = s^*$). The relative amount by which the cost of the $\mathbf{s}$ policy (including the constant term) exceeded that of the $s^*$ policy, i.e., $100\% \cdot (C(\mathbf{s}) - C(s^*)) / C(s^*)$, averaged 0.78% across the 60 cases (or 0.28% when the constant term was dropped). Given the strength of this performance and the ease of computing the $\mathbf{s}$ policy, it appears to be a good choice for use in practice.
7. Value of Product Recovery and Component Commonality

In addition to knowing how to manage inventories in a multi-product, multi-component system with returns, managers must make decisions that are somewhat broader in nature. These include deciding which components (if any) to recover from used products, as well as which components to use in which products – e.g., whether to use dedicated components for each product or to share common components across multiple product lines. In this section we explore aspects of system behavior that help answer these kinds of questions. Since $\delta$ is easy to compute and was shown to be near-optimal in most cases, we use that policy throughout this section.

7.1 Value of Product Recovery

In order to decide whether or not to recover a given component, one needs to compare the benefits of recovery (typically, procurement cost savings) with the costs (including recovery and inspection costs, as well as inventory management costs). Many of these costs and benefits – e.g., procurement cost savings, recovery and inspection costs – are easy to calculate once certain system parameters (such as estimated demand and return rates, unit recovery costs, etc.) are estimated. The impact of returns on inventory management costs is not as clear, however, so we performed a numerical experiment to explore this issue.

Specifically, we considered three different versions of a two-item system, corresponding to demand rates equal to the first three demand patterns listed in Table 1. For each of these scenarios, we considered 8 different returns patterns: $(\lambda^1, \lambda^2, \lambda^{12}) = (\theta \mu_1, 0, 0)$ for $\theta = 0, 0.2, 0.4$ and $0.6$, and $(\lambda^1, \lambda^2, \lambda^{12}) = (0, \theta \mu_2, 0)$ for the same values of $\theta$. All cases had lead times $(L_1, L_2) = (1, 2)$, and costs $(h_1, h_2, b^1, b^2, b^{12}) = (1, 1, 4, 4, 10)$.

Two consistent patterns emerged from the results. First, $C(\delta)$ is increasing in the return rate regardless of which component is recovered. Second, the increase in $C(\delta)$ is greater when the component with the longer lead time is recovered. Results for $(\mu^1, \mu^2, \mu^{12}) = (8, 8, 4)$ are shown in Figure 6. Component returns cause the mean net demand rate for that item to decrease, but cause the variance of net demand to increase, leading to a higher coefficient of variation. As is the case in other inventory settings, this increase in variability leads to higher inventory management costs. (Of course, these higher costs might be more than offset by procurement cost savings, which are also increasing in the return rate.) Similar factors
explain the second observation, since the impact of variability on inventory costs tends to increase as the lead time increases. The practical implications of the second observation may be somewhat counterintuitive, however. Since component returns represent a supplemental source of supply (though one that cannot be directly controlled), one might expect it to be more valuable to have this additional source for items that are harder to obtain – i.e., those with longer lead times. It turns out that the reverse appears to be true.

***Figure 6 here***

7.2 Component Commonality

In multi-product systems, using common components across products 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). All of the previous research has assumed no returns. We now explore the impact of returns on the value of component commonality.

We consider two versions of a two-product system. The first version is a dedicated system, where components 1 and 3 are dedicated to product-13 and components 2 and 4 are dedicated to product-24. All lead times are equal to $L$. The second version is a common-component system, where components 3 and 4 are replaced by a common component 5. We compare the performance of the two systems under a variety of parameters. In the dedicated system for product-13, $\mu^{13} = 10$ and $(h_1, h_3, b^1, b^3, b^{12}) = (1, 1, 4, 4, 10)$ for all scenarios. We consider 20 different return patterns: $(\lambda^1, \lambda^3, \lambda^{13})$ is equal to $(0, 0, \theta \mu^{13})$, $(0.5 \theta \mu^{13}, 0.5 \theta \mu^{13}, 0.5 \theta \mu^{13})$, $(\theta \mu^{13}, \theta \mu^{13}, 0)$, $(\theta \mu^{13}, 0, 0.5 \theta \mu^{13})$ or $(0, \theta \mu^{13}, 0.5 \theta \mu^{13})$ for $\theta = 0.1, 0.2, 0.3$ and 0.4. (Parameters for the product-24 system are defined symmetrically.) Parameters in the common system are obtained in the natural way by combining the two dedicated systems. For each set of parameters we compared $C(\bar{s})$ for the common system to that of the dedicated system.

Based on the results of these comparisons, several observations can be made that provide valuable insights for the product and component design process. First, the common system yielded lower costs in all cases, with the relative cost benefit of commonality ranging from 3.3% to 13.5%. Note that this benefit is present even though lead times are equal for all components. This is in contrast to the result in Song (2002) for a system without returns.
where no commonality benefit was found when lead times were equal (though that result may have been due to the way that base-stock levels were set in that study). Second, the commonality benefit tended to increase as the return rate $\theta$ increased (though there were a few exceptions to this, possibly caused by the discrete nature of the policy optimization). This is consistent with the previous observation that returns increase variability, resulting in greater potential benefit from risk pooling. Third, by comparing the first and second return patterns, and then the second and third return patterns, we found that commonality provided greater benefit when a larger fraction of returns were single components rather than entire products. Finally, by comparing the fourth and fifth return patterns, we found that commonality had greater benefit when the “commonality-related” components (3, 4 and 5) were recovered than when the purely dedicated components (1 and 2) were recovered.

8. Conclusions

In this paper we studied an infinite-horizon Assemble-to-Order 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. Since 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. Also, increasing return rates result in higher holding and backorder costs for the system, and this impact is stronger when components with longer lead times are recovered. Finally, we showed that the presence of return flows increases the value of component commonality, and we explored the impacts of different returns patterns on that value.
References


Figure 1: System with unequal leadtimes

Figure 2: Approximate independence of $X_1(r)$ and $X_2(r)$
Figure 3: Performance of $f_{12}^{w}$ lower bound as function of return rate

Figure 4: Performance of $f_{12}^{w}$ lower bound as function of time window
Figure 5: Impact of joint demands and returns on average customer waits

Figure 6: Impact of return rate and lead time on $C(\bar{s})$