Operations Research
Vol. 58, No. 4, Part 1 of 2, July–August 2010, pp. 849–864
ISSN 0030-364X | EISSN 1526-5463 | 10 | 5804 | 0849

informs®

DOI 10.1287/opre.1090.0772

A Stochastic Programming Based Inventory Policy for Assemble-to-Order Systems with Application to the W Model

Mustafa K. Doğru
Alcatel-Lucent Bell Labs, Dublin 15, Ireland, dogru@alcatel-lucent.com

Martin I. Reiman, Qiong Wang
Alcatel-Lucent Bell Labs, Murray Hill, New Jersey 07974 {marty@alcatel-lucent.com, chiwang@alcatel-lucent.com}

We consider assemble-to-order inventory systems with identical component lead times. We use a stochastic program (SP) to develop an inventory strategy that allows preferential component allocation for minimizing total inventory cost. We prove that the solution of a relaxation of this SP provides a lower bound on total inventory cost for all feasible policies. We demonstrate and test our approach on the W system, which involves three components used to produce two products. (There are two unique parts and a common part. Each product uses the common part and its own unique part.) For the W system, we develop efficient solution procedures for the SP as well as the relaxed SP. We define a simple priority allocation policy that mimics the second-stage SP recourse solution and set base-stock levels according to the first-stage SP solution. We show that our policy achieves the lower bound and is, thus, optimal in two situations: when a certain symmetry condition in the cost parameters holds and when the SP solution satisfies a "balanced capacity" condition. For other cases, numerical results demonstrate that our policy works well and outperforms alternative approaches in many circumstances.

Subject classifications: inventory/production: approximations/heuristics, multi-item, assemble-to-order, component commonality; programming: stochastic.
Area of review: Manufacturing, Service, and Supply Chain Operations.
History: Received November 2007; revisions received August 2008, February 2009, June 2009; accepted August 2009. Published online in *Articles in Advance* February 4, 2010.

1. Introduction

In this paper we consider a two-stage multiproduct, multiperiod assemble-to-order (ATO) inventory system that serves stochastic demand. Component lead times are deterministic and identical. Assembly lead times are assumed to be negligible so that assembly operations can be postponed until after observing realized demands. Any unfulfilled demand is backlogged and each backlogged product incurs a product-dependent constant backlog cost per unit of time. Inventories are kept only at the component level, and each item in stock adds a component-specific inventory holding cost per unit of time. Our objective is to find replenishment and allocation policies that minimize the long-run average expected inventory cost.

The system under consideration has many industrial applications and has been the subject of much academic study. Song and Zipkin (2003) provide a survey of work in this area. Previous research suggests that optimizing a multiproduct ATO system is difficult. The difficulty is due to the coupling between replenishment and allocation decisions, neither of which can be made optimally without considering the choice of the other. Further, these decisions, in the optimal policy, may depend on the state of the *pipeline* (ordered, but not yet delivered, components). One way to tackle this is to fix attention on a particular allocation scheme, restrict attention to base-stock replenishment policies, and seek the optimal base-stock levels for all components.

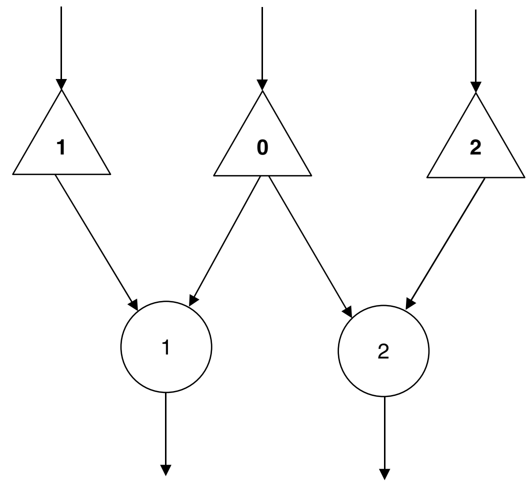
In the literature, first-in-first-out (FIFO) is the main assumption when component allocation is considered. To our knowledge, all studies that analyze continuous review ATO systems follow FIFO allocation, e.g., Song (1998), Song et al. (1999), Song (2002), Song and Yao (2002), Lu et al. (2003), Lu and Song (2005), Zhao and Simchi-Levi (2006), Song and Zhao (2008), and Lu et al. (2008). Further, except for the last two papers, all make the component commitment assumption. (With component commitment, the allocation of each component is done on a FIFO basis, and components are held by a product demand until all required components are acquired.) Assuming FIFO with component commitment, and further assuming a Poisson demand process, Lu and Song (2005) derived the exact optimal base-stock levels, an impressive achievement for multiproduct systems. In periodic-review models, various allocation mechanisms have been considered within a period: a fixed priority rule by Zhang (1997), the FIFO rule by Hausman et al. (1998), a fair-share rule by Agrawal and Cohen (2001), and a product-based allocation rule by Akcay and Xu (2004). However, all these papers assume that backlogs from one period to the next are cleared according to FIFO. In other words, when the length of a period approaches zero (i.e., the model converges to a continuous review model), all these different allocation policies mentioned boil down to FIFO allocation. Although FIFO has the desirable feature of being an allocation method that one can implement in practice, the policy is typically not optimal (especially with component commitment). As observed by Deshpande et al. (2003), for the simpler distribution system, even a mixed use of FIFO and priority leads to inferior backlog clearing decisions when there are both high and low priority backlogs. While for tractability such a policy is assumed in their analysis, they suggest that for implementation a pure priority scheme should be used to clear backlog in conjunction with their replenishment and reservation policies.

In this paper, we depart from the restriction of FIFO and consider allocation schemes that allow preferential treatment of some products to minimize the total inventory cost. We also want to integrate the new allocation approach into replenishment optimization. The exactly optimal form of such an allocation policy is difficult to derive and its performance hard to evaluate. Therefore, we relax a constraint on allocation by assuming an optimal allocation outcome that requires "hindsight" to achieve, in the sense that we are allowed to undo previous allocation decisions. The assumption naturally reduces the problem of optimizing the inventory policy for a multiperiod ATO system to a two-stage stochastic program (SP) with complete recourse that is a variant of the SP previously used for the one-period model, presented in Song and Zipkin (2003). It is also similar to the SP considered in Harrison and Van Mieghem (1999), and is an example of a "newsvendor network" introduced by Van Mieghem and Rudi (2003). As we show in the paper, this SP is not truly a relaxation of the original inventory problem but can be made to be so by imposing appropriate initial backlog levels. We prove that the solution of the latter version of the SP provides a lower bound on the achievable (expected long-run average) cost for the multiperiod inventory system.

Our lower bound is a powerful benchmark: it applies to *any* feasible policy, including state-dependent policies like non base-stock policies, while at the same time, it is attainable in special cases presented in the paper. The second-stage SP recourse solution, which represents an idealized allocation outcome, provides important guidance for the development of an allocation policy. Our lower bound result suggests that for systems with identical lead times, a combined use of an allocation policy that conforms with the recourse solution and a replenishment policy that uses the first-stage SP solution to set base-stock levels should drive the total inventory cost close to its minimum. We test this inventory strategy on the so-called "W" model.

The W system is a basic form of multiproduct ATO systems with commonality that has been discussed in Baker et al. (1986), Gerchak and Henig (1986), Gerchak et al. (1988), de Kok and Visschers (1999), Bernstein et al. (2007a, b), and Song and Zhao (2008). (To the best of our knowledge, Baker et al. 1986 is the first paper to introduce an SP for the one-period W model.) As Figure 1 shows, the W system has two products, each assembled from a common and a unique component. We use the solution of the SP to obtain both replenishment and allocation policies. By exploiting the problem structure, we obtain efficient numerical procedures to solve the SP for both continuous and discrete demand distributions. For replenishment we use a base-stock policy and adopt the first-stage SP solution as the base-stock level. For allocation we implement a simple priority policy that always clears the backlog of the product with a higher unit inventory cost ahead of the one with the lower unit cost. The policy mimics the second-stage recourse SP solution in which the former product always takes precedence in using the common part (part 0) whenever the capacity constraint associated with that part becomes a bottleneck. (We also consider a policy that includes reservation and provide a heuristic to determine a good reservation level.)

Figure 1. An illustration of the W model.



We show that, when a certain symmetry condition in the cost parameters is satisfied or a so-called *capacity balance* holds for the SP solution (a condition defined in Harrison and Van Mieghem 1999 and Van Mieghem 2003: the base-stock level of the common component equals the sum of the base-stock levels of the unique parts, so effectively no sharing takes place), the cost of our policy matches that of the lower-bound SP. As mentioned earlier, the bound applies to any feasible policy, so our policy achieves the true optimum as opposed to being optimal just over a particular set of policies. In the case of balanced capacity, which our numerical results indicate is not rare, there is no value to component commonality and the W system should be run as two separate "V" systems. (A V system is a single-product ATO system with one product assembled from two unique parts.) Note that, although the V system is solvable by standard techniques (e.g., the newsvendor model), without the lower bound to compare against, there is no way to know when using separate V systems is optimal.

For general cases of W systems, we have carried out extensive simulations to evaluate the performance of our SP-based policy, using the percentage deviation from the SP lower bound as the performance measure. We compare our policy with the policy of Lu and Song (2005), which uses FIFO with commitment as an allocation policy along with the associated optimal base-stock levels. To separate contributions from allocation and replenishment decisions of an inventory strategy, we not only compared the original form of the two approaches but also some variants. Varying Lu and Song's approach in simulation, we combine the base-stock levels from their analysis with allocation policies in addition to the one they proposed, including FIFO allocation without commitment and allocation policies that we use for our scheme. Varying our approach, we consider no reservation, the reservation level that is set by our heuristic, and a reservation level set through simulation-based local search.

In all of our numerical results we use independent Poisson processes for the product demand. (This makes possible a direct comparison with Lu and Song 2005.) For cases where our policy achieves the exact optimum, none of the variants involving the base-stock levels of Lu and Song (2005) are optimal, suggesting that finding the optimal policy in these cases is not immediate. For other cases, our policy performs well within a large range of parameter values, as is demonstrated by a representative test bed in the paper. We also identify an extreme parameter region in which the performance of our approach may degrade and discuss how the inefficiency can be compensated by using reservation and more conservative base-stock levels. Finally, we observe that as the demand arrival rates all increase (with all other parameters held fixed), the percentage difference between the inventory cost under our policy and the lower bound provided by the SP converges to zero. The numerical outcome, corroborated by the analysis in Reiman (2008) for a simple distribution system (the W model without any unique components), leads us to believe that our policy is asymptotically optimal; however, proving this conjecture will require significant effort.

As mentioned above, Song and Zhao (2008) and Lu et al. (2008) are the only other papers that consider a policy different than FIFO with commitment in the continuous review multicomponent, multiproduct setting. Thus it is appropriate to provide a comparison between these two papers and the current one. Song and Zhao (2008) discuss no hold-back allocation rules (the same as our myopic policies, to be discussed below in §3.3.1) with an emphasis on FIFO without commitment, which they call modified FIFO (MFIFO). Their focus is on fill rates, while our focus in this paper is on minimizing inventory plus backlog cost, so their results are not directly comparable to ours. Lu et al. (2008) consider both fill rates and inventory/backlog costs. They show (in their Theorem 1) that for the W model operating under a base-stock policy, under the symmetric cost condition, any no hold-back rule is an optimal allocation rule for the inventory/backlog cost minimization problem. This result is related to our Theorem 3.4, but neither subsumes the other. Our results hold for a W model with identical lead times, while theirs hold for nonidentical lead times as well. On the other hand, they do not provide a way to evaluate the total inventory/backlog cost of these policies, while we show that this cost is provided by the solution of the second stage of our stochastic program. The latter quantification provides a stepping stone that allows us to prove (in Theorem 3.5) that when the base-stock levels are set by the first-stage solution of our stochastic program, the resulting replenishment and allocation policy is actually optimal among all feasible policies for this problem (Lu et al. 2008 do not consider replenishment optimization). Moreover, for asymmetric cost cases, we prove that our priority allocation policy dominates all other myopic policies.

All of the results in this paper are for ATO systems in which all component lead times are identical. This strong assumption applies to several situations in practice. For instance, if all parts can be manufactured quickly in an outsourced facility in a foreign country, then a common transportation delay can dominate all lead times. For some electronic products, inventory control is mainly over several expensive circuit boards. These boards normally go through similar production and testing processes and hence are subject to a similar manufacturing delay. Extending our SP approach to ATO systems with nonidentical lead times represents an important topic for future research. Based on the results of Rosling (1989) for single product ATO systems with nonidentical lead times, it is likely that achieving good performance will require the consideration of replenishment policies that are more complicated than base-stock policies. Also, although we introduce a SP for a general ATO structure, in this paper we focus on applying it to the W model. While it seems clear that the SP solution provides base-stock levels for more general ATO systems, the task of translating the SP solution into an allocation policy for more general ATO systems appears to be more difficult. This is also an important topic for future research. A further discussion on both of these important extensions—to nonidentical lead times and more general ATO systems is presented in §5.

The rest of the paper is organized as follows. In §2, we introduce the continuous review version of the general ATO inventory system, formulate our SP, and prove that a relaxation of this SP provides a lower bound. (The associated periodic review system is discussed in the online appendix. An electronic companion to this paper is available as part of the online version that can be found at http://or.journal.informs.org/.) In §3, we focus on the W system, relating the two stochastic programs. We show how our SP can be efficiently solved. We also introduce our allocation policy. Numerical results are contained in §4, and we conclude the paper in §5. All proofs are given in the online appendix.

2. Model and Approach

2.1. The Model

We consider the following assemble-to-order system. There are m products and n components. Define a_{ij} as the amount of component j $(1 \leqslant j \leqslant n)$ needed to assemble one unit of product i $(1 \leqslant i \leqslant m)$. All components have a common, deterministic replenishment lead time L. We consider a continuous-review model. Our results carry over to the periodic-review (discrete time) formulation; these are presented in Appendix II.

Define

$\mathscr{D}_i(t) =$ total demand for product i $(1 \leqslant i \leqslant m)$ in the interval $[0, t]$, $t \geqslant 0$;
$\mathscr{D}(t) \equiv (\mathscr{D}_1(t), \ldots, \mathscr{D}_m(t))$.

We assume that $\{\mathscr{D}(t), t \geqslant 0\}$ is a compound Poisson process and $\mathsf{E}[\mathscr{D}_i(1)] < \infty$ $(1 \leqslant i \leqslant m)$. We allow dependence between demands for different products: it is possible to have $\mathsf{E}[\mathscr{D}_i(1)\mathscr{D}_j(1)] \neq \mathsf{E}[\mathscr{D}_i(1)]\mathsf{E}[\mathscr{D}_j(1)]$ for $1 \leqslant i, j \leqslant m$. Upon arrival, a demand is either served or backlogged. The cost of holding a unit of component j $(1 \leqslant j \leqslant n)$ in the inventory is h_j per unit of time. The backlog cost of a unit of product i $(1 \leqslant i \leqslant m)$ is b_i per unit of time.

We develop an inventory management scheme, which is composed of a replenishment policy γ and an allocation policy p. Define

$\mathscr{R}_j(t) =$ total replenishment orders for component j $(1 \leqslant j \leqslant n)$ placed with the supplier in the interval $[-L, t]$, $t \geqslant -L$;
$\mathscr{R}(t) \equiv (\mathscr{R}_1(t), \ldots, \mathscr{R}_n(t))$;
$\mathscr{Z}_i(t) =$ total product i demand $(1 \leqslant i \leqslant m)$ served in the interval $[0, t]$, $t \geqslant 0$;
$\mathscr{Z}(t) \equiv (\mathscr{Z}_1(t), \ldots, \mathscr{Z}_m(t))$;
$D_i(t) =$ lead-time demand for product i $(1 \leqslant i \leqslant m)$, $D_i(t) = \mathscr{D}_i(t) - \mathscr{D}_i(t-L)$;
$R_j(t) =$ lead-time replenishment orders for component j $(1 \leqslant j \leqslant n)$, $R_j(t) = \mathscr{R}_j(t) - \mathscr{R}_j(t-L)$;
$Z_i(t) =$ total amount of product i demand $(1 \leqslant i \leqslant m)$ served in the time interval $(t-L, t]$, $Z_i(t) = \mathscr{Z}_i(t) - \mathscr{Z}_i(t-L)$.

Then the backlog levels at time $t \geqslant L$ satisfy

$$B_i(t) = B_i(t-L) + D_i(t) - Z_i(t), \quad 1 \leqslant i \leqslant m, \tag{1}$$

and the inventory levels at time $t \geqslant L$ satisfy

$$I_j(t) = I_j(t-L) + R_j(t-L) - \sum_{i=1}^{m} a_{ij} Z_i(t), \quad 1 \leqslant j \leqslant n. \tag{2}$$

Equations (1) and (2) describe the inventory and backlog levels at time t in terms of these quantities a lead time earlier at time $t-L$. It is also useful to describe the evolution of these processes from one moment to the next. Let

$$\Delta\mathscr{D}(t) \equiv \mathscr{D}(t) - \mathscr{D}(t^-), \quad t \geqslant 0, \tag{3}$$

$$\Delta\mathscr{Z}(t) \equiv \mathscr{Z}(t) - \mathscr{Z}(t^-), \quad t \geqslant 0, \tag{4}$$

$$\Delta\mathscr{R}(t) \equiv \mathscr{R}(t) - \mathscr{R}(t^-), \quad t \geqslant -L, \tag{5}$$

where we define $\mathscr{D}(0^-) = 0$, $\mathscr{Z}(0^-) = 0$, and $\mathscr{R}(-L^-) = 0$. Then, for $t \geqslant 0$

$$B_i(t) = B_i(t^-) + \Delta\mathscr{D}_i(t) - \Delta\mathscr{Z}_i(t), \quad 1 \leqslant i \leqslant m, \tag{6}$$

$$I_j(t) = I_j(t^-) + \Delta\mathscr{R}_j(t-L) - \sum_{i=1}^{m} a_{ij} \Delta\mathscr{Z}_i(t), \quad 1 \leqslant j \leqslant n. \tag{7}$$

To complete the description, we need to provide initial conditions. These are

$$I_j(0^-), \quad 1 \leqslant j \leqslant n, \quad B_i(0^-), \quad 1 \leqslant i \leqslant m, \quad \text{and}$$
$$\{\mathscr{R}_j(t),\ -L \leqslant t \leqslant 0\}, \quad 1 \leqslant j \leqslant n.$$

The goal is to choose γ and p to minimize the long-run average expected total cost $C^{\gamma, p}$, defined as

$$C^{\gamma, p} \equiv \limsup_{T \to \infty} \frac{1}{T} \mathsf{E}\left[\int_0^T \left\{\sum_{i=1}^{m} b_i B_i(t) + \sum_{j=1}^{n} h_j I_j(t)\right\} dt\right]. \tag{8}$$

Both $B_i(t)$ and $I_j(t)$ depend on the inventory policy, though to reduce clutter, we usually do not attach superscripts γ, p to the two variables as we did to cost C.

For a policy to be feasible, it is necessary that the policy does not serve demands that do not exist, consume components that are not available, or use advanced information that an inventory manager is not supposed to have. More specifically, a feasible policy must satisfy the following.

1. For $t \geqslant 0$, $B_i(t) \geqslant 0$, $1 \leqslant i \leqslant m$. From (1), this implies that

$$Z_i(t) \leqslant B_i(t-L) + D_i(t), \quad 1 \leqslant i \leqslant m, \tag{9}$$

i.e., the amount of product i demand served in excess of existing backlog cannot exceed the amount of new demand.

2. For $t \geqslant 0$, $I_j(t) \geqslant 0$, $1 \leqslant j \leqslant n$. From (2), this implies that

$$\sum_{i=1}^{m} a_{ij} Z_i(t) \leqslant I_j(t-L) + R_j(t-L), \quad 1 \leqslant j \leqslant n, \tag{10}$$

i.e., the amount of component j consumed in excess of existing inventory cannot exceed the total number of components received in replenishment.

3. For all $t \geqslant 0$, $\mathscr{R}(t)$ and $\mathscr{Z}(t)$ are chosen using only the available information: $\mathbf{I}(0^-)$, $\mathbf{B}(0^-)$, $\{\mathscr{D}(s), 0 \leqslant s \leqslant t\}$, $\{\mathscr{R}(s), -L < s < t\}$, and $\{\mathscr{Z}(s), 0 \leqslant s < t\}$.

The equations simplify under specific policy assumptions. Under a base-stock policy where the base-stock levels are y_j, we have

$$y_j = I_j(t-L) + R_j(t-L) - \sum_{i=1}^{m} a_{ij} B_i(t-L), \quad 1 \leqslant j \leqslant n, \tag{11}$$

i.e., the inventory position at time $t-L$ (= right-hand side of (11)) equals y_j. Combining equation (11) with (2), we have

$$I_j(t) = y_j + \sum_{i=1}^{m} a_{ij}[B_i(t-L) - Z_i(t)], \quad 1 \leqslant j \leqslant n. \tag{12}$$

Hence the requirement that $I_j(t) \geqslant 0$ implies that

$$\sum_{i=1}^{m} a_{ij} Z_i(t) \leqslant y_j + \sum_{i=1}^{m} a_{ij} B_i(t-L), \quad 1 \leqslant j \leqslant n. \tag{13}$$

2.2. The Stochastic Program and Its Application

We develop an inventory management strategy based on the SP that is essentially the one described by Song and Zipkin (2003) in the context of a one-period ATO model. The major change is that in place of their one-period demand, we consider lead-time demand. The SP is a two-stage stochastic linear program (LP) with complete recourse. In the first stage, order quantities (= base-stock levels) y_j, $1 \leqslant j \leqslant m$, are chosen. The lead-time demand D_i, $1 \leqslant i \leqslant n$, is then observed. In the second stage, allocation quantities z_i, $1 \leqslant i \leqslant n$, are chosen. Note that variables z_i are allocation quantities over the common lead time L. The second-stage objective function is linear, as are the second-stage constraints. Complete recourse means that, for any y and D, there is a feasible solution to the second-stage LP. Specifically, the first stage of the SP is to choose $\mathbf{y} \geqslant 0$ to minimize

$$C_s(\mathbf{y}) \equiv \mathsf{E}[\varphi(\mathbf{y})] + \sum_{j=1}^{n} h_j y_j, \tag{14}$$

where φ is the optimal objective function of the second-stage recourse problem

$$\begin{aligned}\varphi(\mathbf{y}) &= \min_{\mathbf{z}\geqslant 0} \left\{ \sum_{i=1}^{m} b_i(D_i - z_i) - \sum_{j=1}^{n} h_j \sum_{i=1}^{m} a_{ij} z_i \,\middle|\, z_i \leqslant D_i, 1 \leqslant i \leqslant m, \sum_{i=1}^{m} a_{ij} z_i \leqslant y_j, 1 \leqslant j \leqslant n \right\} \\ &= \sum_{i=1}^{m} b_i D_i - \max_{\mathbf{z}\geqslant 0} \left\{ \sum_{i=1}^{m} c_i z_i \,\middle|\, z_i \leqslant D_i, 1 \leqslant i \leqslant m, \sum_{i=1}^{m} a_{ij} z_i \leqslant y_j, 1 \leqslant j \leqslant n \right\},\end{aligned} \tag{15}$$

and $c_i = b_i + \sum_{j=1}^{n} a_{ij} h_j$ is the *unit inventory cost* of product i ($1 \leqslant i \leqslant m$). The optimal value of the SP is

$$\begin{aligned}C_s^* &= \min_{\mathbf{y}\geqslant 0} \left\{ \mathsf{E}\left[\sum_{i=1}^{m} b_i D_i + \sum_{j=1}^{n} h_j y_j - \max_{\mathbf{z}\geqslant 0} \left\{ \sum_{i=1}^{m} c_i z_i \,\middle|\, z_i \leqslant D_i, 1 \leqslant i \leqslant m, \sum_{i=1}^{m} a_{ij} z_i \leqslant y_j, 1 \leqslant j \leqslant n \right\} \right] \right\} \\ &= \sum_{i=1}^{m} b_i \mathsf{E}[D_i] + \min_{\mathbf{y}\geqslant 0} \left\{ \sum_{j=1}^{n} h_j y_j - \mathsf{E}\left[\max_{\mathbf{z}\geqslant 0} \left\{ \sum_{i=1}^{m} c_i z_i \,\middle|\, z_i \leqslant D_i, 1 \leqslant i \leqslant m, \sum_{i=1}^{m} a_{ij} z_i \leqslant y_j, 1 \leqslant j \leqslant n \right\} \right] \right\}.\end{aligned} \tag{16}$$

Let $\mathbf{y}^* = \arg\min_{\mathbf{y}\geqslant 0} C_s(\mathbf{y})$, so that $C_s^* = C_s(\mathbf{y}^*)$.

The above SP involves a relaxation of the natural constraints placed on allocation in the multiperiod inventory problem. The second-stage recourse problem (15) is equivalent to allowing an inventory manager in a multiperiod system to freely retract decisions made over the previous lead time and reallocate parts according to the complete observation of realized demand. Because of this relaxation, our SP based approach may not lead to the exact optimum. Although this SP involves relaxing the constraint on allocation in the multiperiod inventory problem, it is not a true relaxation of the multiperiod inventory problem because it imposes a constraint (zero initial backlog) that may not be present in the inventory problem.

In the case of identical component lead times, Theorem 2.1 below shows that for policies that keep no backlog older than one lead time (e.g., FIFO in Lu and Song 2005 and MFIFO in Lu et al. 2008), the solution of the one-period problem (16) is a lower bound on the cost $C^{\gamma,p}$ of the multiperiod system. On the other hand, for policies that are not subject to this requirement, there is no time limit for backlog to remain in the system and affect the inventory cost. To set a lower bound, the one-period model needs to start from a "favorable" backlog level, and requires the following relaxation of (16) that is again a two-stage stochastic LP with complete recourse.

Let $\alpha_i \geqslant 0$, $1 \leqslant i \leqslant m$, represent (possibly) nonzero initial product backlog and define

$$\varphi(\mathbf{y}, \boldsymbol{\alpha}) = \sum_{i=1}^{m} b_i D_i - \max_{\mathbf{z}\geqslant 0} \left\{ \sum_{i=1}^{m} c_i z_i \,\middle|\, z_i \leqslant D_i + \alpha_i, 1 \leqslant 1 \leqslant m, \sum_{i=1}^{m} a_{ij} z_i \leqslant y_j, 1 \leqslant j \leqslant n \right\} \tag{17}$$

as the optimal objective function of the second-stage recourse LP. The first stage of the relaxed SP is to choose $\mathbf{y} \geqslant 0$, $\boldsymbol{\alpha} \geqslant 0$ to minimize

$$\underline{C}_s(\mathbf{y}, \boldsymbol{\alpha}) \equiv \mathsf{E}[\varphi(\mathbf{y}, \boldsymbol{\alpha})] + \sum_{j=1}^{n} h_j y_j + \sum_{i=1}^{m} b_i \alpha_i. \tag{18}$$

Let

$$\underline{C}_s^* \equiv \inf_{\mathbf{y}\geqslant 0,\, \boldsymbol{\alpha}\geqslant 0} \underline{C}_s(\mathbf{y}, \boldsymbol{\alpha}).$$

Note that $\underline{C}_s^* \leqslant C_s^*$. The proof of the following result is in Appendix I.

Theorem 2.1. *Let* (γ, p) *be a feasible policy and let* $C^{\gamma,p}$ *be the cost as defined in* (8). *Suppose* $\{\mathscr{D}(t), t \geqslant 0\}$ *is a compound Poisson process. Then*

$$\underline{C}_s^* \leqslant C^{\gamma,p}. \tag{19}$$

If for all $i = 1, \ldots, m$ *and* $t \geqslant L$, *where* L *is the common lead time of all components,*

$$B_i^{\gamma,p}(t-L) \leqslant Z_i^{\gamma,p}(t) \tag{20}$$

(*hence the policy clears all backlog of product* i *older than one lead time*), *then*

$$C_s^* \leqslant C^{\gamma,p}. \tag{21}$$

This result also holds for the periodic-review model, as described in the proof. If we define $C^{\gamma,p}$ using lim inf rather than lim sup, (19) would still hold. An analog of Theorem 2.1 holds for more general, stationary demand if the set of replenishment policies considered is restricted to base-stock policies, but we do not pursue this generalization in this paper.

The *original* SP (14)–(15) is simpler than the *relaxed* SP (17)–(18). In addition, the interpretation of the solution of the original SP and its use in the ATO inventory control problem is immediate: just as the solution to the simple newsvendor problem provides the base-stock level for the single-item-single-demand inventory model, the solution $\mathbf{y}^*$ to the original SP provides base-stock levels for the n components in the ATO system. As we will see later in the paper, the solutions of the original SP (14)–(15) and the relaxed SP (17)–(18) coincide in many situations. However, as the following example shows, there are cases where the two differ, and the average cost of an ATO system is less than C_s^*, requiring $\underline{C}_s^*$ to serve as the lower bound.

Consider an ATO system with two products, each of which uses one unit of a common part. The system essentially has a distribution structure. The part has a unit replenishment lead time and a holding cost h. The backlog costs of the two products satisfy $b_1 \geqslant b_2$, so product 1 should have priority. Let $x^+ \equiv \max\{x, 0\}$ for $x \in \mathbb{R}$. Specializing the two SPs to this instance (see Appendix I for details),

$$C_s^* = \min_{y\geqslant 0} \big\{ b_1 \mathsf{E}[(D_1 - y)^+] + b_2 \mathsf{E}[(D_2 - (y - D_1)^+)^+] + h\mathsf{E}[(y - D_1 - D_2)^+] \big\} \tag{22}$$

$$\underline{C}_s^* = \min_{y\geqslant 0} \big\{ b_2 \mathsf{E}[(D_1 + D_2 - y)^+] + h\mathsf{E}[(y - D_1 - D_2)^+] \big\}, \tag{23}$$

where (23) is a result of letting $\alpha_1 = 0$ and $\alpha_2 \to \infty$ in (17), corresponding to a situation where the system starts with a huge backlog of product 2. To clear the backlog, the part is ordered in a large quantity, sufficient to serve any realization of high-cost product 1 demand. Hence in no circumstance should there be product 1 backlog, making (17) equivalent to a standard newsvendor model of product 2 (with total demand $D_1 + D_2$). Solving (22) and (23) for extreme cost parameters $h = 10$, $b_1 = 0.5$, $b_2 = 0.35$, unit lead time ($L = 1$) and demand processes forming Poisson processes with arrival rates $\lambda_1 = \lambda_2 = 4$ yields $y^* = 3$,

$$C_s^* = 2.129 \quad \text{and} \quad \underline{C}_s^* = 1.927.$$

The two values differ because the initial backlog level is restricted to 0 for the original SP but optimized for the relaxed SP. We simulated the system that operates under a base-stock replenishment policy (base-stock level $= 3$) and the priority allocation policy. The average cost is 2.054 ± 0.002, which is between $\underline{C}_s^*$ and C_s^*. The result indicates that the original SP solution C_s^* does not serve as a lower bound for this extreme scenario.

3. The W System: Solution and Analysis

The W system is one of the most basic forms of the ATO model. There are two products ($i = 1, 2$) and three components ($j = 0, 1, 2$). (For a better presentation, we let component index j start from 0 instead of 1 as in the general formulation.) Component 0 is the common part used by both products while components 1 and 2 are the unique parts used in products 1 and 2, respectively. Let

$$a_{10} = a_{20} = 1, \quad a_{11} = a_{22} = 1, \tag{24}$$

i.e., each product uses one unit of the common part and one unit of its unique part. Therefore, the unit inventory cost is $c_i = h_0 + h_i + b_i$ ($i = 1, 2$). Without loss of generality, we assume $c_1 \geqslant c_2$. It is straightforward to extend our analysis to cases of using multiple units as long as $a_{10} = a_{20}$.

We use the W system as an example to both demonstrate and test our approach of solving ATO inventory problems using SP. We first set a benchmark in §3.1 by simplifying our lower bound result for the W system. We then introduce our inventory policy, presenting the replenishment solution in §3.2 and the allocation policy in §3.3. Our replenishment scheme is a base-stock policy. As mentioned in §2.2, the solution $\mathbf{y}^*$ of the original SP provides base-stock levels in a natural manner. Using the SP solution to design an allocation policy is more involved. As we mentioned earlier, the formulation of the second-stage recourse LP implies that an inventory manager can freely retract an allocation decision made during the previous lead time, which is not possible in real inventory systems. Therefore, the optimal recourse solution generally cannot be replicated by a feasible allocation policy. Nevertheless, values of z^* do provide important information about which product (or combination of products) should be served first under different component availability conditions. In the W system, the solution indicates that product 1 should always have precedence in using the common part, which leads to the static priority policy considered in this paper. In §3.4 we compare the inventory cost of our policy with its lower bound under special circumstances. We prove that the two values coincide, hence our policy achieves the exact optimum in those cases. A numerical evaluation for the general case is delayed to §4.

3.1. The Lower Bound

For the multiperiod W model, the lower-bound SP (17)–(18) specializes to

$$\inf_{\mathbf{y}\geqslant 0,\ \boldsymbol{\alpha}\geqslant 0}\left\{\sum_{j=0}^{2}h_j y_j+\sum_{i=1}^{2}b_i(\alpha_i+\mathsf{E}[D_i]) -\sum_{i=1}^{2}c_i\mathsf{E}[z_i(\mathbf{y},\boldsymbol{\alpha})]\right\}, \tag{25}$$

where

$$\mathbf{z}(\mathbf{y},\boldsymbol{\alpha})=\arg\max\left\{\sum_{i=1}^{2}c_i z_i \,\middle|\, z_1+z_2\leqslant y_0,\ 0\leqslant z_i\leqslant y_i,\ z_i\leqslant D_i+\alpha_i,\ i=1,2\right\}.$$

The simple structure of the W model enables an explicit solution for $\mathbf{z}(\mathbf{y},\boldsymbol{\alpha})$. In particular, because $c_1\geqslant c_2$, product 1 should always be served as much as possible, which leads to the following explicit optimal allocation solution:

$$z_1(\mathbf{y},\boldsymbol{\alpha})=(D_1+\alpha_1)\wedge y_1\wedge y_0, \tag{26}$$

$$z_2(\mathbf{y},\boldsymbol{\alpha})=(D_2+\alpha_2)\wedge y_2\wedge (y_0-(D_1+\alpha_1)\wedge y_1\wedge y_0), \tag{27}$$

where, for $x, w\in\mathbb{R}$, $x\wedge w\equiv\min(x,w)$. Substitution of these two solutions gives a simpler and more explicit form of (25).

LEMMA 3.1. *Let $\underline{C}_s^*$ be the solution to the relaxed stochastic LP* (25). *Then*

$$\underline{C}_s^*=\min_{\mathbf{y}\geqslant 0}\{\underline{C}_s(\mathbf{y})\}, \tag{28}$$

where

$$\underline{C}_s(\mathbf{y})\equiv\sum_{j=0}^{2}h_j y_j+\sum_{i=1}^{2}b_i\mathsf{E}[D_i]-c_1\mathsf{E}[y_1\wedge D_1] - c_2\mathsf{E}[D_2\wedge y_2\wedge(y_0-D_1\wedge y_1)]. \tag{29}$$

In terms of the underlying one-period model, y_j $(j=0,1,2)$ in (29) is the inventory position of component j, which differs from its interpretation in (25), where it is the total quantity ordered. For product 1, $D_1\wedge y_1$ is the amount of new demand served after clearing all of its backlog at the beginning of the period. If $y_0\geqslant D_1\wedge y_1$, then after satisfying the needs of product 1, there are still some common parts left for product 2 to clear all its initial backlog and to fill its new demand by an amount of $D_2\wedge y_2\wedge(y_0-D_1\wedge y_1)$. If $y_0<D_1\wedge y_1$, then product 1 will consume so many common parts that not only no new demand of product 2 will be served, but also an amount of $D_1\wedge y_1-y_0$ of the latter's initial backlog remains uncleared, resulting in a positive increment

$$-c_2(D_2\wedge y_2\wedge(y_0-D_1\wedge y_1))=c_2(D_1\wedge y_1-y_0)>0$$

to the inventory cost.

Following the above insight, (28) can branch out to two cases. On one branch, we impose the restriction $y_1\leqslant y_0$ to ensure that $y_0\geqslant D_1\wedge y_1$, in which case all backlogs of both products will be cleared. In this case the right-hand side of (28) becomes

$$\min_{\mathbf{y}\geqslant 0}\left\{\sum_{j=0}^{2}h_j y_j+\sum_{i=1}^{2}b_i\mathsf{E}[D_i]-c_1\mathsf{E}[y_1\wedge D_1] -c_2\mathsf{E}[D_2\wedge y_2\wedge(y_0-D_1\wedge y_1)]\,\middle|\, y_1\leqslant y_0\right\}.$$

Consider the SP (14) specialized to the W model. As in the solution (26) and (27) to the second-stage recourse LP of (25), in this case we have $z_1=y_0\wedge y_1\wedge D_1$ and $z_2=y_2\wedge D_2\wedge(y_0-y_0\wedge y_1\wedge D_1)$. Substituting them into (16) leads to

$$C_s^*=\min_{\mathbf{y}\geqslant 0}\left\{\sum_{j=0}^{2}h_j y_j+\sum_{i=1}^{2}b_i\mathsf{E}[D_i]-c_1\mathsf{E}[y_0\wedge y_1\wedge D_1] - c_2\mathsf{E}[y_2\wedge D_2\wedge(y_0-y_0\wedge D_1\wedge y_1)]\right\}.$$

By the same argument as in the proof of Lemma 3.1 we can restrict to $y_1\leqslant y_0$ without loss of optimality, so that

$$C_s^*=\min_{\mathbf{y}\geqslant 0}\left\{\sum_{j=0}^{2}h_j y_j+\sum_{i=1}^{2}b_i\mathsf{E}[D_i]-c_1\mathsf{E}[y_1\wedge D_1] -c_2\mathsf{E}[y_2\wedge D_2\wedge(y_0-D_1\wedge y_1)]\,\middle|\, y_1\leqslant y_0\right\}. \tag{30}$$

Thus the branch $y_1\leqslant y_0$ corresponds to the standard one-period ATO model.

On the other branch, we require $y_1\geqslant y_0$, allowing a maximum amount of y_1-y_0 of product 2 backlog to remain uncleared. In this case, (28) becomes

$$\min_{\mathbf{y}\geqslant 0}\left\{\sum_{j=0}^{2}h_j y_j+\sum_{i=1}^{2}b_i\mathsf{E}[D_i]-c_1\mathsf{E}[y_1\wedge D_1] - c_2\mathsf{E}[y_2\wedge D_2\wedge(y_0-D_1)^+] + c_2\mathsf{E}[(D_1-y_0)^+\wedge(y_1-y_0)]\right\}. \tag{31}$$

The solution of (28), which is a lower bound on the average cost of the W model, takes the smaller outcome from the two branches. Either is possible. For instance, suppose the first branch (30) gives y_0^* as the optimal solution for the common component. Is it beneficial to move to the second branch by setting y_1 above y_0^*? Consider the marginal case with discrete demand. Setting y_1 one unit above y_0^* increases the holding cost by h_1, but if D_1 exceeds y_0^*, there is a benefit for serving one more unit of product 1 (c_1) at the cost of clearing one less backlog of product 2 (c_2). Consequently, if $(c_1 - c_2)\Pr\{D_1 > y_0^*\} > h_1$, then the solution of (28) should be that of (31).

The possibility that the lower bound can rest on the second branch reveals an interesting managerial insight. A persistent existence of backlogs in an inventory system is not necessarily a bad situation. When the backlog cost of the low-value product is small, carrying some of its backlog in the system can be a cheaper alternative to holding excessive inventory of an expensive common part to fulfill the demand of high-value product.

On the other hand, we argue that with long enough lead times the solution to (28) will always be on the branch $y_1 \leqslant y_0$. Consider the situation where the distribution of the compound Poisson process $\mathcal{D}$ is held fixed and L grows large. Let $\Delta_i \equiv \mathsf{E}[\mathcal{D}_i(1)]$. By the strong law of large numbers $L^{-1}D_i \to \Delta_i$ almost surely as $L \to \infty$. We conjecture that, as a consequence, $L^{-1}y_0 \to \Delta_1 + \Delta_2$, $L^{-1}y_1 \to \Delta_1$, and $L^{-1}y_2 \to \Delta_2$ as $L \to \infty$. Thus, $L^{-1}(y_0 - y_1) \to \Delta_2$, so that $y_0 > y_1$ for L large enough. In the next subsection we use the SP associated with the branch $y_1 \leqslant y_0$ to obtain base-stock levels. In our numerical study, we verify that in all cases we consider, the solution to (28) is on this branch.

3.2. Solution to the Stochastic Programs

We adopt a base-stock policy for replenishment and solve the standard one-period model (30) to set the base-stock levels. This is a continuous optimization problem if (y_0, y_1, y_2) are optimized as continuous variables and the distribution of lead-time demands has a density (so, in particular, it has no point mass anywhere including at $D_i = 0$, $i = 1, 2$). Harrison and Van Mieghem (1999) show that the optimal solution is reached at the point where holding costs h_j $(j = 0, 1, 2)$ equal expected duals of the capacity constraint associated with y_j in the second-stage recourse problem. They also formulate the second-stage solution $\mathbf{z}$ as a function of $\mathbf{y}$, and hence reduce the two-stage SP to a one-stage optimization problem. As Song and Zipkin (2003) point out, this is a convex minimization problem.

For lead-time demands that follow Poisson or compound Poisson distributions and are independent between the two products, we discuss solving (30) as a discrete optimization problem to obtain the exact solution. Our discussion extends to the relaxed model (28) that is used to set the lower bound.

From (30), it is easy to see that at the optimum, $y_0 \leqslant y_1 + y_2$. Define $\mathbf{y}$ as a regular solution if $\mathbf{y}$ is a local minimum and satisfies $y_1 > 0$, $y_2 > 0$, and $y_0 < y_1 + y_2$, and as a boundary solution if

$$\mathbf{y} = \arg\min\{C_s(\mathbf{y}) \mid y_1 = 0\}, \quad \mathbf{y} = \arg\min\{C_s(\mathbf{y}) \mid y_2 = 0\}, \text{ or}$$

$$\mathbf{y} = \arg\min\{C_s(\mathbf{y}) \mid y_1 + y_2 = y_0\}.$$

The optimal solution is the regular or boundary solution that gives the lowest value of $C_s(\mathbf{y})$. The same applies to the optimization of $\underline{C}_s(\mathbf{y})$. In both cases, finding boundary solutions is equivalent to solving standard newsvendor models (see Appendix I for details). Thus we present here only the procedure for finding the regular solution.

Let $F_i(k) = \Pr\{D_i \leqslant k\}$ and $\bar{F}_i(k) = 1 - F_i(k)$ $(i = 1, 2$ and $k \geqslant 0)$. For notational convenience, for $k < 0$, define $F_i(k) = 0$ and $\bar{F}_i(k) = 1$ $(i = 1, 2)$. Let

$$Y_0^{\min} = \min\left\{k > 0\colon \bar{F}_2(k-1) \leqslant \frac{c_1 - h_1/\bar{F}_1(0)}{c_2}\right\} \vee \min\left\{k > 0\colon F_1(k-1) \geqslant \frac{h_2}{c_2\bar{F}_2(0)}\right\},$$

$$Y_0^{\max} = \min\left\{k > 0 : \bar{F}_1(k-1) < \frac{h_1}{c_1}\right\} + \min\left\{k > 0\colon \bar{F}_2(k-1) < \frac{h_2}{c_2}\right\}, \tag{32}$$

where $x \vee w \equiv \max\{x, w\}$ for $x, w \in \mathbb{R}$. Observe that $0 \leqslant Y_0^{\min}$, $Y_0^{\max} < \infty$ if $c_i\bar{F}_i(0) > h_i$ $(i = 1, 2)$. As we show in the theorem below, the latter condition is necessary for the regular solution to exist. Assuming this is the case, define

$$s_1(y_1 \mid y_0) = h_1 - \bar{F}_1(y_1)[c_1 - c_2\bar{F}_2(y_0 - y_1)],$$

$$s_2(y_2 \mid y_0) = h_2 - c_2F_1(y_0 - y_2)\bar{F}_2(y_2 - 1). \tag{33}$$

It is easy to verify that, if $Y_0^{\min} \geqslant Y_0^{\max}$, then for each $y_0 \in [Y_0^{\min}, Y_0^{\max}]$, $s_i(y_i \mid y_0)$ $(i = 1, 2)$ weakly increases in y_i. Also,

$$s_1(0 \mid y_0) \leqslant 0 \leqslant s_1(Y_0^{\max} \mid y_0) \quad \text{and}$$

$$s_2(0 \mid y_0) \leqslant 0 < s_2(y_0 + 1 \mid y_0).$$

Therefore, for $y_0 \in [Y_0^{\min}, Y_0^{\max}]$, there exist

$$Y_i(y_0) \equiv \max\{y_i\colon s_i(y_i \mid y_0) \leqslant 0\} \quad (i = 1, 2), \tag{34}$$

where $0 \leqslant Y_1(y_0)$, $0 \leqslant Y_2(y_0) \leqslant y_0$, and values of $Y_i(y_0)$ $(i = 1, 2)$ come directly from a one-dimensional bisection search.

Theorem 3.2. *In both* (28) *and* (30), *a regular solution exists only if* $c_i\bar{F}_i(0) > h_i$ $(i = 1, 2)$. *If* (y_0^*, y_1^*, y_2^*) *is a regular solution to* (28) *or* (30), *then* $y_0^* \in [Y_0^{\min}, Y_0^{\max}]$. *Furthermore, if* (y_0^*, y_1^*, y_2^*) *is a regular solution to* (28), *then*

$$\underline{C}_s(y_0^*, Y_1(y_0^*), Y_2(y_0^*)) = \underline{C}_s(y_0^*, y_1^*, y_2^*),$$

and if (y_0^*, y_1^*, y_2^*) *is a regular solution to* (30), *then*

$$C_s(y_0^*, Y_1(y_0^*) \wedge y_0^*, Y_2(y_0^*)) = C_s(y_0^*, y_1^*, y_2^*).$$

Following the theorem, to find a regular solution, we just need to enumerate over $y_0 \in [Y_0^{\min}, Y_0^{\max}]$, determine $Y_1(y_0)$ and $Y_2(y_0)$, and for cases where $y_0 < Y_1(y_0) + Y_2(y_0)$, evaluate and compare $C_s(y_0, Y_1(y_0) \wedge y_0, Y_2(y_0))$ or $\underline{C}_s(y_0, Y_1(y_0), Y_2(y_0))$. The evaluation can be conducted directly with the following formula (see Appendix I for derivation):

$$\underline{C}_s(\mathbf{y}) = \begin{cases} \displaystyle\sum_{i=1}^{2} b_i \mathsf{E}[D_i] + \sum_{j=0}^{2} h_j y_j - c_1 \sum_{k=0}^{y_1-1} \bar{F}_1(k) \\ \displaystyle\quad - c_2 \left[\sum_{k=0}^{y_0-y_1-1} \bar{F}_2(k) + \sum_{k=y_0-y_1}^{y_2-1} F_1(y_0-k-1)\bar{F}_2(k) \right], \\ \qquad\qquad \text{if } y_1 \leqslant y_0, \\ \displaystyle\sum_{i=1}^{2} b_i \mathsf{E}[D_i] + \sum_{j=0}^{2} h_j y_j - c_1 \sum_{k=0}^{y_1-1} \bar{F}_1(k) + c_2 \sum_{k=y_0}^{y_1-1} \bar{F}_1(k) \\ \displaystyle\quad - c_2 \sum_{k=0}^{y_2-1} F_1(y_0-k-1)\bar{F}_2(k), \quad \text{if } y_1 > y_0. \end{cases} \tag{35}$$

In addition, $C_s(\mathbf{y})$ is defined only for $y_1 \leqslant y_0$, in which case $C_s(\mathbf{y}) = \underline{C}_s(\mathbf{y})$.

The same approach applies to solving (28) and (30) as continuous problems, and the implementation is simpler. As we show in Appendix III, because the continuous formulation yields smooth objective functions, both $C_s(\mathbf{y})$ and $\underline{C}_s(\mathbf{y})$ can be reduced to convex functions of y_0. Consequently, we can find y_0^* by a one-dimensional bisection search instead of enumerating y_0 over $[Y_0^{\min}, Y_0^{\max}]$.

3.3. Priority-Based Allocation Policy

Given the base-stock levels $\mathbf{y}$, the inventory and backlog levels are determined by the allocation of the common part. Denote an allocation policy by p and as before, let $D_i(t)$ $(i = 1, 2)$ be demand of product i that arrived during the lead time ending at time t, and let $Z_i^p(t)$ be the amount of product i served during that period under policy p. Let $B_i^p(t)$ $(i = 1, 2)$, and $I_j^p(t)$ $(j = 0, 1, 2)$ be the corresponding backlog and inventory levels at time t. Then, the balance equations (1) and (12) yield

$$I_0^p(t) = y_0 + B_1^p(t-L) + B_2^p(t-L) - Z_1^p(t) - Z_2^p(t), \tag{36}$$

$$I_i^p(t) = y_i + B_i^p(t-L) - Z_i^p(t) \quad (i = 1, 2), \tag{37}$$

$$B_i^p(t) = D_i(t) + B_i^p(t-L) - Z_i^p(t) \quad (i = 1, 2), \tag{38}$$

for $t \geqslant L$. We consider policies that prioritize common part allocation: the more valuable product 1 is served whenever all its required parts are available, and the less valuable product 2 is served when all required parts are available *and* the service does not interfere with the need of product 1. Before describing two versions of priority policies we first introduce the notion of a *myopic policy*.

3.3.1. Myopic Policies. A policy is said to be *myopic* if neither product 1 nor product 2 have a backlog when the product's required parts have positive inventories, i.e.,

$$[I_i^p(t) \wedge I_0^p(t)] B_i^p(t) = 0, \quad i = 1, 2. \tag{39}$$

We denote the collection of such policies by $\mathcal{P}_m$. In addition to our priority-based backlog clearing policy introduced below, $\mathcal{P}_m$ also includes other policies, such as FIFO without commitment, which is referred to as modified FIFO by Song and Zhao (2008). However, FIFO with commitment (considered by Lu and Song 2005, Lu et al. 2003, etc.) is not myopic.

Given base-stock levels $\mathbf{y}$, denote the second-stage solution of (30) by $z_i^*(\mathbf{y})$, so $D_i - z_i^*(\mathbf{y})$ is the amount of unserved demand of product i $(i = 1, 2)$. In the multiperiod inventory system, let $D_i(t)$ $(t \geqslant L)$ be the amount of product i demand arriving within the lead time that ends at time t, and let $B_i^p(t)$ be the backlog of product i $(i = 1, 2)$ at that time under allocation policy p.

Lemma 3.3. *For $p \in \mathcal{P}_m$ and $D_i(t) = D_i$ $(i = 1, 2)$, if $y_1 + y_2 < y_0$, then*

$$B_i^p(t) = (D_i(t) - y_i)^+, \quad i = 1, 2. \tag{40}$$

If $y_0 \leqslant y_1 + y_2$, then

$$\begin{aligned} B_1^p(t) + B_2^p(t) &= (D_1(t) + D_2(t) - y_0)^+ \vee (D_1(t) - y_1)^+ \\ &\quad \vee (D_2(t) - y_2)^+ \\ &= D_1 + D_2 - z_1^*(\mathbf{y}) - z_2^*(\mathbf{y}). \end{aligned} \tag{41}$$

Equation (40) indicates that the base-stock level of the common part should never exceed the sum of base-stock levels of unique parts. Any excess increases holding cost but has no effect on backlogs. This property is well known for the one-period W system (Harrison and Van Mieghem 1999, Bernstein et al. 2007a), but may not be true under an allocation policy that is not myopic in multiperiod models. As the numerical examples in §4 demonstrate, under the FIFO allocation policy (with commitment), it can be optimal to have $y_0 > y_1 + y_2$ to alleviate a common part shortage caused by commitment.

In (41), given the same base-stock levels, the total amount of backlog under any myopic policy at any given time coincides with the total unserved demand in the solution to the SP, and hence is invariant over the set of myopic policies that use base-stock levels $\mathbf{y}$. As a consequence, the total amount of demand served is also invariant. Note that the mix of the two products in the total amount may still differ from one policy to another. The term to the right of the first equality defines the maximum shortage of all components. Therefore, of all feasible allocations, a myopic policy keeps the total backlog level of all products at the minimum.

3.3.2. Priority-Based Backlog Clearing (PBC). In previous discussions, we have mentioned that we are using priority allocation, and we have given a general idea about how it works. It is a myopic policy that uses priority to clear backlogs. In subsequent discussions, we formally refer to it as Priority-based Backlog Clearing (PBC); the precise definition is given next.

The allocations in PBC are (for $t > 0$)

$$\Delta\mathscr{Z}_1^{PBC}(t) = \min\{B_1^{PBC}(t^-) + \Delta\mathscr{D}_1(t), I_0^{PBC}(t^-) + \Delta\mathscr{R}_0(t-L),\ I_1^{PBC}(t^-) + \Delta\mathscr{R}_1(t)\}, \quad (42)$$

$$\Delta\mathscr{Z}_2^{PBC}(t) = \min\{B_2^{PBC}(t^-) + \Delta\mathscr{D}_2(t), I_0^{PBC}(t^-) + \Delta\mathscr{R}_0(t-L) - \Delta\mathscr{Z}_1^{PBC}(t), I_2^{PBC}(t^-) + \Delta\mathscr{R}_2(t-L)\}. \quad (43)$$

Given base-stock levels, PBC is not necessarily optimal in general. Nevertheless, the following result shows that it is optimal when $c_1 = c_2$, and for $c_1 > c_2$ it performs at least as well as any myopic policy that uses the same base-stock levels. With p denoting an allocation policy, let $C^p(\mathbf{y})$ denote the long-run average expected cost using a base-stock replenishment policy with base-stock levels $\mathbf{y}$ and allocation policy p.

Theorem 3.4. *Let $\underline{C}_s(\mathbf{y})$ be defined as in* (29) *when the base-stock levels are $\mathbf{y} \geqslant 0$. Recall the unit inventory cost $c_i = b_i + h_0 + h_i$ $(i = 1, 2)$. For all $p \in \mathscr{P}_m$,*

$$\text{if } c_1 = c_2, \text{ then } C^p(\mathbf{y}) = \underline{C}_s(\mathbf{y}), \quad (44)$$

$$\text{if } c_1 > c_2, \text{ then } \underline{C}_s(\mathbf{y}) \leqslant C^{PBC}(\mathbf{y}) \leqslant C^p(\mathbf{y}). \quad (45)$$

The result holds for both continuous and periodic-review systems.

This result can be interpreted as follows. With $c_1 = c_2$, each product carries the same inventory cost, so the mix of products in the total amount served does not matter. Since the total amount of demand served is the same for any myopic policy and the SP, their costs should also be the same. In a more recent study, Lu et al. (2008) show that in a W system with general lead times and symmetric unit costs, any myopic policy (which they call "no hold-back" policy) dominates any other allocation scheme under the same base-stock levels. Equation (44) implies their result for the case of identical lead times because $\underline{C}_s(\mathbf{y})$ is a lower bound on the inventory cost for given $\mathbf{y}$. Moreover, the equation also indicates that by solving the second stage of the SP, one can exactly determine the inventory cost ($C^p(\mathbf{y})$) under such policies.

When the unit inventory costs of the two products are not equal, the mix does matter. A myopic policy may not be able to achieve the lower bound of expected long-run average inventory cost given by the solution of the SP because the latter assumes a perfect mix that only exists in hindsight. Still, in such cases, our PBC policy manages to achieve the lowest long-run average inventory cost of all myopic policies because by giving priority to product 1 in backlog clearing, the policy results in the best possible product mix. Following this line of thought, we next explore the possibility of achieving an even more favorable product mix at the expense of reducing the total amount of demand served by any given time, i.e., policies that reserve the common part for product 1.

3.3.3. Priority with Reservation (PR). Our Priority with Reservation (PR) policy differs from the PBC policy by including reservation of the common part in favor of product 1. As with PBC, the common part goes to product 1 when both products have backlog. At other times, the common part will still be withheld from product 2 when its on-hand inventory drops to or below the reservation level K. The policy is determined by the reservation level K and we present a heuristic to set this value for the case when demands form independent Poisson processes (with rates λ_1 and λ_2).

Our heuristic is based on modeling the product 1 backlog as the backlog in an M/M/1 make-to-stock queue with base-stock level K. The demand arrival rate is λ_1 and the production rate is $\lambda_1 + \lambda_2$. Roughly speaking, this corresponds to assuming that the unique part is always available (false), that the product 1 demand arrives in a Poisson process with rate λ_1 (true), and that replenishments arrive in a Poisson process of rate $\lambda_1 + \lambda_2$ (true) which is independent of the demand arrival stream (false).

Let Q denote the queue length in the above M/M/1 queue. With a base-stock (= reservation) level of K, the product 1 backlog is $(Q-K)^+$. Using the standard analysis of the M/M/1 system, the expected backlog cost for product 1 demand is

$$g_1(K) = b\mathsf{E}[Q-K]^+ = b\sum_{n=K+1}^{\infty}(1-\rho)\rho^n(n-K) = b\rho^{K+1}(1-\rho), \quad (46)$$

where $\rho = \lambda_1/(\lambda_1+\lambda_2)$. Clearly $g_1(K)$ decreases with K, so it quantifies the benefit of having reservation.

On the other hand, reservation means additional inventory cost that would not be incurred if the common component is used to serve available product 2 demand and thus leaves the system. The cost is proportional to the reservation level, and we denote it by $g_2(K) = hK$.

We set the reservation level by minimizing the sum of the two costs,

$$g(K) = g_1(K) + g_2(K) = b\rho^{K+1}(1-\rho) + hK. \quad (47)$$

For $K \geqslant 0$, let

$$\Delta g(K) \equiv g(K+1) - g(K) = h - b\rho^{K+1}. \quad (48)$$

Let K^* denote $\arg\min\{g(K)\}$. If $\Delta g(0) > 0$, then $K^* = 0$. If $\Delta g(0) \leqslant 0$, we let $K^* = \min\{K \in \mathbb{Z}^+ \mid \Delta g(K) > 0\}$, which yields

$$K^* = \left\lfloor \frac{\ln(h) - \ln(b)|}{\ln(\rho)} \right\rfloor. \quad (49)$$

This K^* is our proposed reservation level, where we substitute $h = h_0 + h_2 + b_2$ and $b = b_1$. In the special case where $c_1 = c_2$,

$$\Delta g(0) = h - b\rho > h - b = c_2 - b_1 = c_1 - b_1 > 0,$$

so $K^* = 0$. (This is an important case, as indicated in Theorems 3.4 and 3.5.)

Of course, the benefit of reservation is realized only when there is backlog of product 1, while the cost is incurred when there is product 2 backlog that can otherwise be served. The M/M/1 queue analysis is also oblivious to the constraint imposed by the availability of unique parts. Therefore, as we pointed out previously, the approach we described above is not exact. Nevertheless, in many cases, reserving common components does improve the total inventory cost, and, as we demonstrate in §4, our heuristic often chooses a good reservation level. (We also provide some guidance there on when the heuristic should not be used.)

We extend our conjecture on asymptotic optimality in §1 to any fixed reservation level. Our numerical results support this conjecture. Although in some sense this extended conjecture (if true) says that the use of reservation, as well as the choice of reservation level, is of secondary importance to the choice of base-stock levels and the use of priority, the numerical results also indicate that using the reservation level provided by our heuristic can reduce the total cost in "nonasymptotic" cases.

3.4. Sufficient Conditions for Optimality

We have proposed an inventory management strategy that uses an SP to set base-stock levels and implements priority-based component allocation. The following result shows that, under certain conditions, our policy (without reservation) is actually optimal. To the best of our knowledge this is the first result presenting an optimal policy (and proving optimality) for any nontrivial special case of any multiproduct ATO model.

Theorem 3.5. *Let* $\mathbf{y}^*$ *be the optimal solution of* (30). *If either* $c_1 = c_2$ *or* $y_0^* = y_1^* + y_2^*$, *then for all* $p \in \mathcal{P}_m$,

$$C^p(\mathbf{y}^*) = C^{PBC}(\mathbf{y}^*) = \underline{C}_s^*.$$

The condition $c_1 = c_2$ refers to a symmetry relationship in unit inventory costs. Even though the backlog costs for the products and holding costs for the unique parts may be different ($h_1 \neq h_2$, $b_1 \neq b_2$), as long as $h_1 + b_1 = h_2 + b_2$, the condition holds. In this case, as we have learned from Theorem 3.4, given the same base-stock levels, any myopic policy is optimal for allocation. Theorem 3.5 tells us that, by complementing myopic allocation with base-stock levels set by our SP, we can drive the inventory cost to its lower bound and hence achieve *true optimality* over any feasible policy, including those using different base-stock levels and those subscribing to non base-stock replenishment policies.

The condition $y_0^* = y_1^* + y_2^*$ is commonly referred to as *balanced capacity* in the study of the one-period W model (Harrison and Van Mieghem 1999, Van Mieghem 2003, Bernstein et al. 2007a). It is an important concept in commonality analysis: when the condition applies, there is no risk pooling at the optimum, implying that component commonality has no benefit. For the one-period model, Bernstein et al. (2007a) define a region of cost parameters within which the condition holds, indicating that the occurrence is not rare. (Our numerical results also indicate this.) Our theorem significantly amplifies the implication of these previous discussions: if commonality does not help in a one-period W model, then it will not generate any benefit in a multiperiod W system with identical component lead times.

For other parameter values, we resort to numerical studies to evaluate the performance of our policy in the next section.

4. Numerical Study

We evaluate the performance of our approach (using the solution of the SP to set base-stock levels and PBC for allocation) by its application in the W system. Our numerical comparison with other policies highlights the strength and weakness of our approach under different circumstances. We choose the aforementioned policy of Lu and Song (2005), which uses the exact optimal base-stock levels under FIFO with commitment, as our counterpart. We compare not only the original form of the two approaches but also their variations. As shown in Table 1, our approach is denoted by SPR (using Stochastic programming for replenishment and PRiority for allocation) with three variations in allocation policy: SPR(0), which reserves no common part for the high-value product; SPR(K^*), which reserves the common part and the reservation level is set by the heuristic described in §3.3.3; and SPR(K_s), which searches and uses a locally optimal reservation level. We denote the FIFO policy (with commitment) of Lu and Song (2005) by LS and consider two variations, LSNC and LSP, which

Table 1. The inventory policies considered in the numerical study.

Inventory policy	Base-stock levels	Allocation policy	Reservation level
SPR(K^*)	SP model	PBC	K^* (heuristic in §3.3.3)
SPR(K_s)		PBC	K_s (local search)
SPR(0)		PBC	No reservation
LS	Analysis by Lu and Song (2005)	FIFO with commitment	No reservation
LSP		PBC	No reservation
LSNC		FIFO with no-commitment	No reservation

use the same base-stock levels as LS but differ in allocation policy. LSNC allocates components according to FIFO without commitment (Song and Zhao 2008 refer to this allocation policy as MFIFO) and LSP allocates the common part according to our PBC policy.

For a given policy p, the gap between the inventory cost (denoted by C_p), which is estimated by the simulation or, in the case of LS, calculated exactly, and the lower bound $\underline{C}_s^*$, defined as

$$\Delta_p = 100\frac{C_p - \underline{C}_s^*}{\underline{C}_s^*},$$

is used as the performance measure. In all the scenarios presented in this numerical study, $\underline{C}_s^*$ coincides with C_s^* (the cost of the original SP).

We have conducted extensive numerical experiments and our findings are illustrated here through a representative test bed. In all cases, we consider continuous review and assume that the two demand streams form independent Poisson processes (as in Lu and Song 2005). Without loss of generality, we normalize the holding cost of the common part to $h_0 = 1$, set component lead time $L = 1$, and vary all other parameters to generate many different scenarios. In the test bed, we fix the arrival rates at $\lambda_1 = \lambda_2 = 25$ and generate 27 scenarios by exhausting all combinations of

$$h_1 = 0.2, 1, 5; \quad h_2 = 0.2, 1, 5; \quad b_1 = 2(1+h_1), 5(1+h_1);$$

$$\text{and} \quad b_2 = (1+h_2), 2(1+h_2),$$

and excluding cases where $c_1 < c_2$. After reporting results from this test bed in §4.1, we conduct a *stress test* in §4.2 to identify cases when our policy becomes less effective and discuss partial remedies. We also conduct numerical experiments to assess the asymptotic behavior of our approach in §4.3. A summary of our findings is given in §4.4.

Table 2. Δ values for a test bed of 27 scenarios with $\lambda_1 = \lambda_2 = 25$ and sorted by c_1/c_2.

#	h_1	h_2	b_1	b_2	c_1/c_2	$\Delta_{SPR(0)}$	Δ_{LS}	Δ_{LSNC}	Δ_{LSP}
1	1.0	1.0	4.0	4.0	1.00	0.0	7.3	1.9	1.9
2	0.2	0.2	2.4	2.4	1.00	0.0	3.1	0.8	0.8
3*	1.0	5.0	10.0	6.0	1.00	0.0	9.3	2.6	2.6
4*	5.0	5.0	12.0	12.0	1.00	0.0	8.0	4.6	4.6
5	0.2	1.0	6.0	4.0	1.20	0.6	6.4	2.2	1.2
6	0.2	0.2	2.4	1.2	1.50	3.5	14.0	11.1	4.9
7	1.0	1.0	4.0	2.0	1.50	0.5	12.7	3.2	2.0
8*	5.0	5.0	12.0	6.0	1.50	0.0	10.8	6.4	6.4
9	1.0	0.2	4.0	2.4	1.67	1.6	9.0	5.4	2.4
10	0.2	1.0	6.0	2.0	1.80	3.6	17.5	10.6	6.7
11	1.0	1.0	10.0	4.0	2.00	1.4	9.7	4.4	2.9
12*	5.0	5.0	30.0	12.0	2.00	0.0	6.0	3.3	3.3
13	0.2	0.2	6.0	2.4	2.00	6.0	15.0	12.9	7.4
14	1.0	0.2	4.0	1.2	2.50	4.1	17.6	12.6	7.2
15	0.2	0.2	6.0	1.2	3.00	13.5	34.5	31.1	19.1
16	1.0	1.0	10.0	2.0	3.00	4.6	18.3	10.4	7.5
17	5.0	1.0	12.0	4.0	3.00	0.4	8.3	2.2	2.2
18*	5.0	5.0	30.0	6.0	3.00	0.0	9.1	5.2	5.2
19	1.0	0.2	10.0	2.4	3.33	6.6	15.5	12.9	8.4
20	5.0	1.0	12.0	2.0	4.50	0.7	12.1	4.3	4.3
21	1.0	0.2	10.0	1.2	5.00	15.2	27.0	23.2	15.4
22	5.0	0.2	12.0	2.4	5.00	2.9	9.2	6.4	4.5
23	5.0	1.0	30.0	4.0	6.00	3.6	8.8	3.6	3.6
24	5.0	0.2	12.0	1.2	7.50	5.2	13.2	9.9	7.0
25	5.0	1.0	30.0	2.0	9.00	5.9	12.6	7.5	7.5
26	5.0	0.2	30.0	2.4	10.00	8.1	10.6	8.3	6.4
27	5.0	0.2	30.0	1.2	15.00	16.3	14.7	11.5	10.0

Notes. The scenarios for which (i) SP yields balanced capacity are marked with asterisks, (ii) LS yields base-stock levels satisfying $y_0 > y_1 + y_2$ are marked in bold.

4.1. Comparing the Percentage Gap for the Test Bed Cases

Table 2 shows a comparison of the gap figures for SPR(0) and FIFO-based approaches (LS, LSNC, LSP) for the 27 cases in the test bed. The results suggest the following points:

1. As predicted by Theorem 3.5, SPR(0) has zero optimality gap in all scenarios where either both products have the same unit inventory cost (scenarios 1–4) or the balanced capacity condition applies (scenarios 3, 4, 8, 12, and 18). Observe that in each of these scenarios, all other policies have a positive optimality gap that ranges from a minimum of 0.8% to a maximum of 10.8%, suggesting that obtaining the exact optimum in these special cases (as our policy does) is not a simple task.

2. In all scenarios (and also in all other cases we have examined),

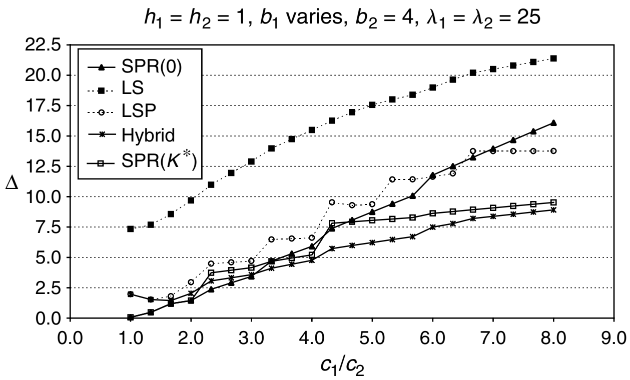
$$\Delta_{LS} > \Delta_{LSNC} \geqslant \Delta_{LSP}.$$

The first inequality is consistent with the message in Song and Zhao (2008) that for the W system, removing the commitment requirement in FIFO improves the performance. The second inequality is expected from (45) of our Theorem 3.4: priority-based backlog clearing dominates all myopic allocation policies in the W system.

3. Table 2 illustrates the fact that the results are influenced by individual parameter values. Consider the scenarios with $c_1/c_2 = 3$. Even though h_0, L, arrival rates and c_1/c_2 ratios are all the same among the four scenarios, the individual values of the holding costs for the product-specific components and the backlog costs lead to a great swing in the performance of the inventory management schemes considered. As an example, observe that $\Delta_{SPR(0)}$ varies between 0 and 13.5%.

4. The results also demonstrate the advantage of using SP-based base-stock levels. SPR(0) and LSP use the same allocation policy and differ only in how the base-stock levels are set. The former outperforms the latter in all but three cases, with a tie in one of these three cases. The advantage of our approach can be explained by the difference in replenishment optimization: SPR(0) assumes product 1 always gets the common part first, which reflects priority allocation better than the FIFO assumption used in LSP. LSP outperforms SPR(0) in scenarios 26 and 27, where the two products have highly asymmetric unit inventory costs. These exceptions motivate us to conduct an additional *stress test* to be discussed next.

Figure 2. Δ vs. c_1/c_2 for a set of scenarios with $h_1 = h_2 = 1$, $\lambda_1 = \lambda_2 = 25$, $b_2 = 4$ and b_1 varying between 4 and 46.



4.2. A Stress Test in Regions of High Cost Disparity

Cases 26 and 27 indicate that our policy does not uniformly dominate alternative approaches. To examine these exceptions in detail, we conduct a stress test. For the purpose of identifying the weakness of our approach, we focus on the particular parameter region that is unfriendly to our policy, the region of high cost disparity.

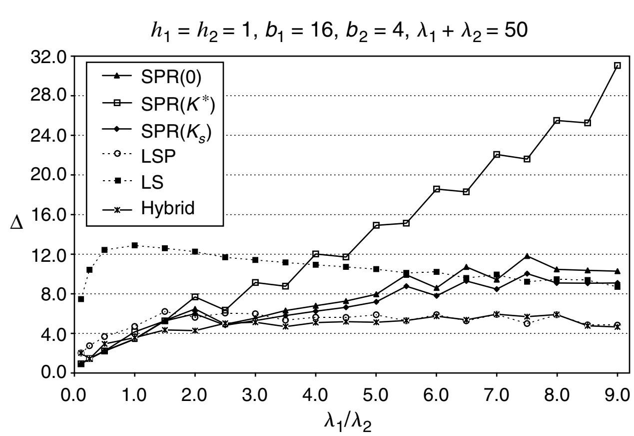
In scenario 1 of Table 2, we vary b_1 from 4 to 46 while keeping the other parameters intact to increase the ratio of the unit inventory cost (c_1/c_2) to an extreme degree. Figure 2 shows that SPR(0) initially outperforms LSP and the trend continues until $c_1/c_2 = 6$, beyond which point LSP performs better. The results suggest that priority allocation is essential to achieve good performance, but SP-based base-stock levels become inadequate for cases of large cost disparity. This can be explained by noting that the second-stage SP resource solution idealizes the allocation outcome, and thus may underestimate the backlog costs and lead to a (too) low base-stock level of the common part. In most cases, such as the ones in the test bed and many cases in Figure 2, it is better to have a simplistic characterization of priority allocation than to not reflect such benefit at all. Therefore, SPR(0) outperforms approaches that optimize base-stock levels by FIFO. However, when b_1 becomes extremely large, running out of the common part for product 1 becomes extremely costly, and SPR(0) pays a price for being too optimistic.

Reservation offers a partial remedy. By withholding a common part from the low-value product, reservation is consistent with the SP's assumption that the common part is always used to serve the high-value product first. Figure 2 shows the performance of SPR(K^*), our reservation heuristic described in §3.3.3. The scheme outperforms LSP in every case and demonstrates a dominant advantage over SPR(0) when the high-value product has a very large backlog cost (b_1), reducing the performance gap from above 15% to less than 10%. We find the heuristic also works well in many other cases. For example, in scenarios 26 and 27 of Table 2, $\Delta_{SPR(K^*)}$ is 6.8% and 7.5%, respectively. Nevertheless, additional stress tests reveal that our reservation heuristic becomes ineffective when there is (i) a high level of cost asymmetry (i.e., large c_1/c_2), *and* (ii) a high composition of product 1 arrival (i.e., large λ_1/λ_2). For instance, in scenario 1 of Table 2, we change b_1 to 16 to make c_1/c_2 large, and vary λ_1/λ_2 from 1/9 to 9 while keeping $\lambda_1 + \lambda_2 = 50$. Figure 3 shows that when $\lambda_1 \gg \lambda_2$, the performance of SPR(K^*) is rather poor because in these cases, $\ln(\rho) = \ln(\lambda_1/(\lambda_1 + \lambda_2))$ in (49) gets close to 0, making the calculation of K^* unstable and sensitive to the inaccuracy of M/M/1 assumption.

Alternatively, we can be more conservative in setting base-stock levels. LSP gives a pessimistic estimate of backlog cost because it assumes a common part can be committed to a low-value product and held back from high-value ones. To alleviate this "self-inflicted" shortage cost, the algorithm tends to set a (too) high base-stock level of the common part, in some cases to the extent that y_0 exceeds $y_1 + y_2$, which is observed in scenarios 4, 8, 12, 18, and 25 of Table 2. As an interesting experiment, we consider a new policy that we call *hybrid*, which sets base-stock levels at the middle point between the levels of SPR(0) and LSP (rounded up to the nearest integer) and allocates components according to the priority-based allocation rule PBC. The simulation results demonstrate that the hybrid policy performs remarkably well in both scenarios shown in Figures 2 and 3.

To distinguish the effectiveness of reservation from that of a particular heuristic, we consider another reservation scheme SPR(K_s), where the reservation level K_s is obtained from a local search (a computationally expensive procedure guaranteeing that K_s is at least locally optimal). Figure 3 shows SPR(K_s) performs better than SPR(0), but the improvement is small in comparison with hybrid or LSP. Hence, we conclude that in cases where cost disparity is high *and* the arrival of the high-value product dominates the total traffic, setting base-stock levels conservatively helps

Figure 3. Δ vs. λ_1/λ_2 for a set of scenarios with $h_1 = h_2 = 1$, $b_1 = 16$, $b_2 = 4$, varying λ_1, and $\lambda_2 = 50 - \lambda_1$.



more than developing a reservation scheme. On the other hand, we caution that the observed benefit of using the *hybrid* approach does not hold in all cases. The scheme does not achieve exact optimality as SPR(0) does in the aforementioned special cases and, as we show next, its performance as well as the performance of LSP falls behind that of either SPR(0) or SPR(K^*) as the lead time grows or the arrival rates increase.

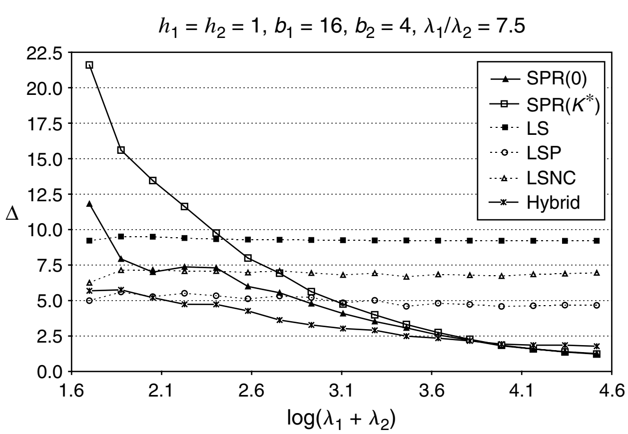
4.3. Asymptotic Behavior

Our last experiment shows that the aforementioned weakness of our SP-based approach dissipates in the presence of a sufficiently long lead time or high demand arrivals. (Scaling up the arrival rate of a Poisson process with the lead time fixed is equivalent to scaling up the lead time with the arrival rate fixed.) We consider a scenario in Figure 3 with $\lambda_1/\lambda_2 = 7.5$ where LSP outperforms both SPR(0) and SPR(K^*) initially. We scale up the arrival rate by a factor of 1.5 in each step and compare our inventory policy (with and without reservation heuristics) against FIFO-based approaches (LS, LSNC, LSP) in Figure 4. The optimality gap of both SPR(0) and SPR(K^*) keep diminishing while those under FIFO-based approaches stabilize at strictly positive values. This implies that the FIFO-based base-stock levels are not asymptotically optimal. Thus, although the hybrid policy performs very well, as the figure shows, its optimality gap also eventually surpasses that of SPR(0) and SPR(K^*). The findings demonstrate by existence that many good policies like LSP are not asymptotically optimal, and they motivate future efforts to prove that our SP-based policy has this desirable property.

4.4. Summary

Our numerical studies corroborate analytical results that the priority allocation policy dominates all myopic policies in

Figure 4. Δ vs. $\log(\lambda_1 + \lambda_2)$ for a set of scenarios with $h_1 = h_2 = 1$, $b_1 = 16$, $b_2 = 4$, $\lambda_1/\lambda_2 = 7.5$, $L = 1$ and $(\lambda_1 + \lambda_2)$ increases by a factor of 1.5 starting from 50.



the W system. They also show that our entire inventory policy (including both allocation and replenishment) achieves the exact optimality in special cases while FIFO-based policies fail to do so. Our policy performs well within reasonable parameter regions. The performance degrades in cases of extreme cost disparity because the base-stock levels set by the SP model are too low in these cases. This weakness can be mitigated by our reservation heuristic when the arrival of high-value product demands does not dominate the total traffic. However, reservation is of secondary importance in comparison with setting base-stock levels. As we observe, choosing the middle point between those from the SP model and FIFO-based optimization results in better performance in these cases. On the other hand, our experiments illustrate that the advantage of hybrid base-stock levels disappears as the lead time grows or arrival rates increase. The optimality gap of our policy asymptotically approaches zero while those of other policies (considered in the study) stabilize at some positive levels.

5. Concluding Remarks

We conclude our paper with a relevant quotation from Song and Zipkin (2003):

> …As indicated above, little is known about the forms of optimal policies for multi-period models. The research to date mostly assumes particular policy types. It would be valuable to learn more about truly optimal policies. Even partial characterizations would be interesting. Also, better heuristic policy forms would be useful.

Our work is a step in this direction. We have introduced a new approach, based on stochastic programming, for solving ATO systems. We chose the W model, a simple and much-studied ATO model, as our initial example to both demonstrate and test our approach. Perhaps most significantly, we have identified a policy that is "truly optimal" for the W system with identical component lead times in two situations: when the balanced capacity condition holds and when both products have the same unit inventory costs. In those circumstances, the cost of our policy provably reaches the lower bound that applies to *any* feasible inventory policy, including those that do not subscribe to the commonly assumed base-stock replenishment policy. In other words, the dominance of our policy is not confined to any particular policy type. To the best of our knowledge, no such strong optimality result exists in the literature for multi-product inventory systems, even for the restrictive cases that we consider.

For general W systems, our approach provides an approximate solution. Although the advantage of priority allocation is intuitively clear, the difficulty involved in analyzing the performance of an ATO inventory system under a priority allocation policy has presented a severe impediment to finding optimal base-stock levels to use in conjunction with such a policy. While our approach does not, in general, provide optimal base-stock levels, the base-stock levels it provides do reflect the priority allocation policy. To see the importance of this, consider, in particular, the comparison between LSP and SPR(0). Both schemes use the same priority-based allocation, but differ in base-stock levels. Our numerical study shows that in a majority of cases, especially in the asymptotic regime, SPR(0), which sets base-stock levels by our stochastic program, has a clear advantage over LSP, which does so by assuming FIFO with commitment. The result indicates that the benefit of priority-based allocation comes from not only implementing the policy itself, but also integrating the policy and reflecting its outcome in replenishment optimization.

As we mentioned, the lower bound is attained by our policy when $c_1 = c_2$ or $y_0^* = y_1^* + y_2^*$. In both of these cases the use of hindsight in allocation, which allows the common component to be taken from a previously assembled low-value product and used for a high-value product, provides no advantage. Thus, the true optimum coincides with the SP solution in these cases. Typically, however, allocation under a feasible policy differs from the hindsight optimum. Nevertheless, our priority scheme can make this gap small if the flow of common components arriving from the pipeline is sufficiently large to quickly clear the high-value product backlog. This will be the case if demands of low-value products arrive sufficiently frequently so that the flow from the pipeline generated (one lead time later) by these demands is high enough to dominate the demand uncertainty of high-value products. This happens when λ_2 is sufficiently large relative to λ_1. It also happens when the total demand over a lead time is large, so that the extra flow in a fixed fraction of the lead time, which is $O(\lambda_2 L)$, overwhelms the fluctuations of high-value product demand over that time, which is $O(\sqrt{\lambda_1 L})$. The gap should also be small when $c_1 - c_2$ is small. The simulation results support all of these assertions. It also seems intuitively clear that reserving the common component can sometimes help to reduce the gap. Interestingly, while our simulation results support this view, they also indicate that, with large total demand over a lead time, the use of reservation is of secondary importance to a good choice of base-stock levels.

Our analysis in this paper is limited to identical component lead times. Also, with the exception of the lower-bound result, we have focused on the W system. Obviously, the next steps involve extending our approach beyond these restrictions—challenging tasks that require substantial effort. Before discussing these extensions we mention that, under the assumption of equal component lead times, the W system is actually general enough to cover a larger set of two-product ATO systems. In particular, due to the equal component lead-time assumption, we can lump all components that are unique to product i, $i = 1, 2$, into a single component. (Take the holding cost of this "super-component" to be the sum of the holding costs of all of the original components.) If the common components are used in the same quantity by both products, then we can similarly lump them into a single common super-component. This yields a W model. If some common components are used in different quantities by the two products, then the lumping to yield a single common super-component cannot be carried out. Many other papers that cover the W model, such as Gerchak et al. (1988), Bernstein et al. (2007a, b), Lu and Song (2005), Song and Zhao (2008), and Lu et al. (2008), also exclude this case. We now outline several important issues to be addressed in extending our results:

- We were able to obtain a reasonably explicit, exactly computable solution to the SP for the W model because the second-stage recourse LP has an explicit solution, and the problem is of low dimension. Although explicit solutions to the recourse LP may be available for other simple ATO structures, it seems unlikely that problems with a general bill of material can be solved in the same manner in which we solved the W system. Thus, some other solution technique is needed. Sample Average Approximation (SAA), which is a sampling-based approach, is a conventional method for solving SPs. This approach has been used by Akcay and Xu (2004) to solve a two-stage SP associated with an ATO model where the objective is to maximize the reward of serving demands within a response time window, subject to a budget constraint on component base-stock levels. It thus seems clear that SAA is a promising approach for solving our SP as well. Although in principle the application of SAA is straightforward, it should be pointed out that, because it is a sampling-based method, convergence is slow. This was acknowledged by Akcay and Xu (2004) and has also been noted in the SP literature (see, e.g., Shapiro and Nemirovsky 2005). An additional point is also worth making: SAA generates an estimate of the solution of the SP (along with a confidence interval) rather than an exact solution. Without an exact solution to the SP for the W model, we could not verify, with certainty, for example, when the capacity balance condition $y_0^* = y_1^* + y_2^*$ holds. This motivates further effort on exact solutions for our SP.
- For the W system, we have developed the priority allocation policy to imitate the optimal solution of the second-stage SP recourse problem. Since the SP solution is a lower bound on the inventory cost, for other ATO systems with common lead times, the recourse solution should also be used as guidance for the development of allocation policies. However, for systems with more than two products and/or a more general structure, a static priority policy may not capture important implications of the optimal recourse solution. One may conceive of situations in which the solution recommends serving a combination of products instead of a single product that has a higher unit inventory cost (c) than individual products in the bundle. The recourse problem may also determine the quantity of a product to serve based on availability of parts that the product does not use or demands of other products that share no part with the said product. Moreover, when a high-value product consumes a large set of components, it may not be possible to follow the recourse solution effectively in a multiperiod system if parts are not withheld for that product before its demand

has arrived. We found that the M system discussed in Lu and Song (2005) provides a perfect example that entails these issues. The system contains three products and two components (products 1 and 2 consume one unit of part 1 and 2, respectively, while the other product consumes one unit of both parts), and we are using it to explore other allocation approaches (Doğru et al. 2010). Nevertheless, how to develop a unified allocation policy to closely match the recourse solution in a general ATO system remains an open and challenging problem.

• The two-stage SP is a result of assuming identical component lead times, and the first-stage solution is used quite naturally as base-stock levels in the corresponding multiperiod ATO system. Although it seems that we can accommodate nonidentical lead times by including more stages in the SP model, mapping this solution into a replenishment policy that is implementable in practice is an open problem. Complexities arise because in the SP solution, the optimal quantities of parts with short lead times should vary according to the observation of partially realized demand over the longer lead times, suggesting that a non base-stock policy should be followed in the corresponding multiperiod ATO system. This will surely give rise to a very different paradigm from the one in a majority of the literature where the base-stock policy is the cornerstone.

In summary, based on our study of the W system, we believe that the SP-based approach to ATO inventory systems is a promising one. An extension to more general cases requires us to address some fundamental problems of inventory management, hence leads to a very exciting future research agenda.

6. Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at http://or.journal.informs.org/.

Acknowledgments

The authors thank the anonymous associate editor and two referees for their helpful comments. The work of Mustafa K. Doğru is partly supported by the Industrial Development Agency, Ireland.

References

Agrawal, M., M. A. Cohen. 2001. Optimal material control and performance evaluation in an assembly environment with component commonality. *Naval Res. Logist.* **48** 409–429.

Akcay, Y., S. H. Xu. 2004. Joint inventory replenishment and component allocation optimization in an assemble-to-order system. *Management Sci.* **50** 99–116.

Baker, K. R., M. J. Magazine, H. L. W. Nuttle. 1986. The effect of commonality on safety stock in a simple inventory model. *Management Sci.* **32**(8) 982–988.

Bernstein, F., G. DeCroix, Y. Wang. 2007a. Incentives and commonality in a decentralized multiproduct assembly system. *Oper. Res.* **55**(4) 630–646.

Bernstein, F., G. DeCroix, Y. Wang. 2007b. Allocation policies based on demand aggregation in an assemble-to-order system. Working paper, the Fuqua School of Business, Duke University, Durham, NC.

Birge, J., F. Louveaux. 1997. *Introduction to Stochastic Programming*. Springer, New York.

de Kok, A. G., J. W. C. H. Visschers. 1999. Analysis of assembly systems with service level constraints. *Internat. J. Production Econom.* **59** 313–326.

Deshpande, V., M. A. Cohen, K. Donohue. 2003. A threshold inventory rationing policy for service-differentiated demand classes. *Management Sci.* **49**(6) 683–703.

Doğru, M., M. I. Reiman, Q. Wang. 2010. Applications of stochastic programming to setting allocation rules for some assemble-to-order systems. In preparation.

Gerchak, Y., M. Henig. 1986. An inventory model with component commonality. *Oper. Res. Lett.* **5**(3) 157–160.

Gerchak, Y., M. J. Magazine, A. B. Gamble. 1988. Component commonality with service level requirements. *Management Sci.* **34**(6) 753–760.

Harrison, J. M., J. A. Van Mieghem. 1999. Multi-resource investment strategies: Operational hedging under demand uncertainty. *Eur. J. Oper. Res.* **113** 17–29.

Hausman, W. H., H. L. Lee, A. X. Zhang. 1998. Order response time reliability in a multi-item inventory system. *Eur. J. Oper. Res.* **109** 646–659.

Lu, Y., J.-S. Song. 2005. Order-based cost optimization in assemble-to-order systems. *Oper. Res.* **53**(1) 151–169.

Lu, Y., J.-S. Song, D. D. Yao. 2003. Order fill rate, leadtime variability, and advance demand information in an assemble-to-order system. *Oper. Res.* **51**(2) 292–308.

Lu, Y., J.-S. Song, D. D. Yao. 2005. Backorder minimization in multiproduct assemble-to-order systems. *IIE Trans.* **37** 763–774.

Lu, Y., J.-S. Song, Y. Zhao. 2008. No-holdback allocation rules for continuous-time assemble-to-order systems. Working paper, the Fuqua School of Business, Duke University, Durham, NC.

Reiman, M. I. 2010. An inventory model with two demand classes. In preparation.

Rosling, K. 1989. Optimal inventory policies for assembly systems under random demands. *Oper. Res.* **37** 565–579.

Shapiro, A., A. Nemirovsky. 2005. On complexity of stochastic programming problems. V. Jeyakumar, A. M. Rubinov, eds. *Continuous Optimization: Current Trends and Applications.* Springer, New York, 111–144.

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., D. D. Yao. 2002. Performance analysis and optimization of assemble-to-order systems with random lead times. *Oper. Res.* **50**(5) 889–903.

Song, J.-S., Y. Zhao. 2008. The value of component commonality in a dynamic inventory system with lead times. Working paper, the Fuqua School of Business, Duke University, Durham, NC.

Song, J.-S., P. Zipkin. 2003. Supply chain operations: Assemble-to-order systems. A. G. de Kok, S. C. Graves, eds. *Supply Chain Management. Handbooks in Operations Research and Management Science*, Vol. 11. Elsevier, B.V., Amsterdam, 561–596.

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

Van Mieghem, J. A. 2003. Capacity management, investment, and hedging: Review and recent developments. *Manufacturing Service Oper. Management* **5**(4) 269–302.

Van Mieghem, J. A., N. Rudi. 2003. Newsvendor networks: Inventory management and capacity investment with discretionary activities. *Manufacturing Service Oper. Management* **4**(4) 313–335.

Zhang, A. X. 1997. Demand fulfillment rates in an assemble-to-order system with multiple products and dependent demands. *Production Oper. Management* **6** 309–324.

Zhao, Y., D. Simchi-Levi. 2006. Performance analysis and evaluation of assemble-to-order systems with stochastic sequential lead times. *Oper. Res.* **54**(4) 706–724.