OPERATIONS RESEARCH
Vol. 54, No. 4, July–August 2006, pp. 706–724
ISSN 0030-364X | EISSN 1526-5463 | 06 | 5404 | 0706

DOI 10.1287/opre.1060.0302

Performance Analysis and Evaluation of Assemble-to-Order Systems with Stochastic Sequential Lead Times

Yao Zhao
Department of Management Science and Information Systems, Rutgers University, 180 University Avenue, Newark, New Jersey 07102, yaozhao@andromeda.rutgers.edu

David Simchi-Levi
Engineering Systems Division and Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307, dslevi@mit.edu

We consider the multiproduct and multicomponent assemble-to-order (ATO) systems where the replenishment lead times of the components are stochastic, sequential, and independent of the system state. The component inventories are either controlled by the continuous-time base-stock policies, namely, a base-stock ATO system, or by the continuous-time batch-ordering policies, namely, a batch-ordering ATO system. This paper develops the following results: First, for a base-stock ATO system with a single end product and renewal demand arrivals, we characterize the probability distribution of the delivery lead time, i.e., the time it takes to satisfy a demand. The exact analysis allows us to provide simple proofs for the important system properties. Second, for a base-stock ATO system with multiple end products and demand following independent Poisson processes, we characterize the dependence among the stockout delays of the components. We show that a multiproduct ATO system can be decomposed into multiple single-product subsystems with each subsystem corresponding to one product. The analysis allows us to develop two numerical methods to evaluate the performance of the base-stock ATO systems of medium to large sizes. A hypothetical example inspired by a real-world problem is presented. Third, for a batch-ordering ATO system, we develop efficient numerical methods for performance evaluation based on Monte Carlo simulation. Given the sample size, the number of products, and the reorder points, the computational complexity of the methods is no more than that of sorting a set of real numbers, where the set size equals to the sum of the batch sizes of all components. Finally, we characterize the impact of the dependence among the components on various ATO systems, and discuss the limits of the approach.

Subject classifications: inventory/production: multi-item, operating characteristics, stochastic.
Area of review: Manufacturing, Service, and Supply Chain Operations.
History: Received July 2003; revisions received December 2003, July 2004, April 2005; accepted June 2005.

1. Introduction

Assemble-to-order (ATO) systems, which keep inventory only for components and assemble the end products after demand is realized, have become prevalent in many industries. An important driver of ATO systems is the short product life cycle, which substantially increases the risk of carrying inventories for the finished goods. ATO systems enable the manufacturers to provide customers with a large variety of products in a timely fashion without carrying expensive finished-goods inventories. Therefore, this approach has the potential of improving customer service while at the same time reducing operating expenses. We refer the reader to Song and Zipkin (2003) for an extensive discussion of examples and motivations for ATO systems.

Because the optimal inventory policies for the general ATO systems are not known, simple but effective heuristic policies, e.g., independent base-stock policies or batch-ordering policies, are often used in practice to manage the component inventories. In this paper, we focus on performance analysis and evaluation of ATO systems with uncapacitated component production, where the inventory of a component is managed either by a continuous-time, independent base-stock policy, referred to as continuous-time base-stock policy hereafter, or by a continuous-time, independent batch-ordering policy, referred to as continuous-time batch-ordering policy hereafter.

Significant advances have been made in recent years in the areas of evaluation and optimization of various ATO systems under these policies; for reviews see Song and Zipkin (2003), Xu (2001), Hausman et al. (1998), and references therein. Most studies of uncapacitated base-stock ATO systems have been based on the assumption of independent and identically distributed (i.i.d.) component replenishment lead times. In this framework, the main challenge is to characterize the joint probability distribution of the outstanding orders and to evaluate the distribution of the maximum of a number of dependent random variables. This is no small task because, as observed by Song and

Zipkin (2003, p. 592), "Many real ATO systems contain hundreds of components and thousands of products" and hence "such a system poses a considerable computational burden on existing models and solution methods."

The performance evaluation of the batch-ordering ATO systems poses an even more significant challenge than that of the base-stock ATO systems. Song (2000) demonstrates that under certain general conditions, the inventory position vector of a batch-ordering ATO system has a uniform equilibrium distribution. Therefore, the performance evaluation of a batch-ordering ATO system can be reduced to those of multiple base-stock systems. Unfortunately, the number of the base-stock systems associated with a batch-ordering system is exponential in the number of components (details are provided in §5). Thus, it is not clear how to efficiently evaluate the batch-ordering ATO systems even if we know how to efficiently evaluate the base-stock ATO systems.

This paper is related to the literature in the following ways: First, instead of assuming i.i.d. lead times, we model the lead times as stochastic sequential random variables that are exogenously determined; for a definition, see Svoronos and Zipkin (1991). In this model, orders for one type of component never cross (that is, orders are filled according to the same sequence they have been issued). Second, instead of the outstanding orders, our analysis focuses on the component delays defined to be the difference between the time at which a component becomes available and the time at which the corresponding demand arrives; see §3. Indeed, this lead-time model allows us to develop a methodology that characterizes the dependent delays among different components. Our objective is to provide analysis of various ATO systems with stochastic sequential lead times, and to develop *efficient numerical methods* that are able to evaluate the base-stock and batch-ordering policies in medium to large-size ATO systems.

For these purposes, we first consider, in §3.1, a base-stock ATO system with a single end product and renewal demand arrivals. We characterize the joint probability distribution of the delays of the components, which leads to an exact expression for the delivery lead-time distribution. The exact analysis allows us to provide simple proofs for some important system properties.

Second, in §4.1 we analyze a base-stock ATO system with multiple end products and independent Poisson demand processes. We characterize the correlations among the delays of the components, and demonstrate that a multiproduct ATO system can be decomposed into multiple single-product subsystems, with each subsystem corresponding to one product. Stochastic inequalities between the multiproduct ATO systems and the analogous systems with independent delays are established.

Based on these analyses, we propose two numerical methods in §§3.2 and 4.2 for the base-stock ATO systems to overcome the difficulty of evaluating the maximum of dependent random variables. The first method is based on a two-moment approximation, while the second method is based on the Monte Carlo simulation. The following performance measures are evaluated: the mean and variance of the delivery lead times, the inventory holding cost of the components, and the order-based fill rates. A numerical study, using a hypothetical example inspired by a real-world problem from the Dell Computer Corporation, reveals that the simulation-based method is capable of evaluating large-size ATO systems, while the method based on the two-moment approximation is applicable to medium-size ATO systems.

Third, in §5 we develop efficient methods based on Monte Carlo simulation to estimate the performance measures of a batch-ordering system by taking advantage of the problem structure. The methods are "efficient" in the sense that the computational times are not exponential in the number of components, rather, the computational times are at most proportional to that of sorting a set of real numbers, where the set size is the sum of the batch sizes of all components.

Finally, in §6 we discuss the limits of our approach by relaxing various assumptions. We also establish some stochastic inequalities between the ATO systems with stochastic sequential lead times and the analogous systems with capacitated component production.

2. The Model

We consider the multiproduct ATO systems with stochastic demand and stochastic sequential lead time. The concept of stochastic sequential lead time, that is, the so-called "transit time," is formally defined by Svoronos and Zipkin (1991). The transit time is an exogenously determined random variable satisfying the following two assumptions of independence: (1) the transit time is independent of the system state (e.g., demand and order placement), and (2) the transit time is independent across components. Under these assumptions, order crossing is not allowed. Thus, unlike the i.i.d. lead times, Palm's theorem (Palm 1938) cannot be applied to the systems with the transit times. As argued by Svoronos and Zipkin (1991), the transit time may be more realistic than the i.i.d. lead time for modeling replenishment lead times from the suppliers in some real-world applications. This is true especially when the manufacturing and the transportation processes are processing orders for many other customers so that the orders placed by the customers in our ATO systems are only negligible portions of the total workload. We refer the reader to Zipkin (2000) for more discussion of this assumption. As we will see later, the second independent condition is not essential for the analysis.

The component inventories are managed either by the base-stock policies or the batch-ordering policies. A continuous-time batch-ordering policy, i.e., an (r, nQ) policy, works as follows: Whenever the inventory position (net inventory plus inventory on order) drops to or below

the reorder point r, an order of size nQ is placed to raise the inventory position up to the smallest integer above r. Clearly, a base-stock policy is a special case of the batch-ordering policies with a batch size $Q = 1$. Because we consider unit demand, an (r, nQ) policy reduces to an (r, Q) policy. If a batch-ordering policy is used for a component, we assume that the supplier always delivers the orders in full, i.e., no split orders. This assumption is reasonable if the supplier keeps inventory in the integer multiples of the batch size.

We assume that the assembly time is negligible compared to the component replenishment lead times. Demand is satisfied on a first-come-first-serve (FCFS) basis whenever all the required components become available. For any demand that cannot be satisfied immediately, we assume that it is fully 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 stock" (we point out that the same assumption is made by Song 2002).

Let $\mathcal{J}$ be the set of components with size $|\mathcal{J}|$, and $\mathcal{I}$ be the set of products with size $|\mathcal{I}|$. Following the convention, we define the following notations:

- $\bar{L} = (L_j,\ j \in \mathcal{J})$: stochastic replenishment lead times with known probability distributions.
- $\bar{r} = (r_j,\ j \in \mathcal{J})$: reorder points of the components.
- $\bar{Q} = (Q_j,\ j \in \mathcal{J})$: batch sizes of the components.
- $\bar{s} = (s_j,\ j \in \mathcal{J})$: base-stock levels of the components. If the base-stock policy is used for a component $j \in \mathcal{J}$, then $s_j = r_j + 1$.
- $\bar{\lambda} = (\lambda^i,\ i \in \mathcal{I})$: demand arrival rates for the products.
- $\bar{h} = (h_j,\ j \in \mathcal{J})$: inventory holding costs per unit of time per unit of the components.
- $\bar{\pi} = (\pi^i,\ i \in \mathcal{I})$: penalty costs per unit of time per unit of the products.
- $A = [a_j^i]$: a matrix characterizing the bill of materials (BOM) between the products and the components, i.e., assembling one unit of product i requires a_j^i units of component j. We assume that a_j^i either equals one when component j is required by product i, or zero otherwise. Extension to the BOM matrix with elements choosing any nonnegative integers is discussed in §6. Finally, we assume that a product requires at least one component, and a component is required by at least one product.
- $\mathcal{I}_j$: set of products that require component j, $\mathcal{I}_j = \{i \in \mathcal{I} \mid a_j^i = 1\}$.
- $\mathcal{J}^i$: set of components required by product i, $\mathcal{J}^i = \{j \in \mathcal{J} \mid a_j^i = 1\}$.

We further define the following performance measures:

- $\bar{X} = (X^i,\ i \in \mathcal{I})$: delivery lead times of the products, i.e., the time that the demand of a product waits to be satisfied.
- $\bar{W} = (W_j,\ j \in \mathcal{J})$: waiting times of the components in inventory for the corresponding demand to arrive and other required components of the demand to be replenished.

Finally, we define the cost function to be the sum of the long-run average penalty costs of all products and the long-run average inventory holding costs of all components, that is,

$$\sum_{i \in \mathcal{I}} \pi^i E(X^i) + \sum_{j \in \mathcal{J}} h_j E(W_j).$$

3. Single-Product Base-Stock Systems

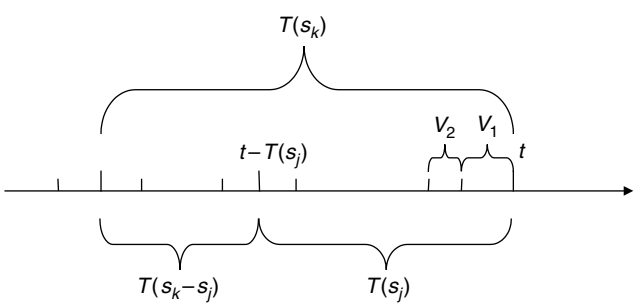
In this section, we consider ATO systems with one final product, that is, $|\mathcal{I}| = 1$. We drop the superscript i accordingly. Following convention, we call these systems single-product assembly systems. The inventory of each component is managed by a continuous-time base-stock policy. The demand arrivals follow a renewal process, where λ is the demand arrival rate. We first provide a probabilistic analysis of the system performances in §3.1 and then present numerical methods in §3.2.

3.1. Performance Analysis

The analysis is based on the following key observation: Suppose that a demand arrives at time t; then the corresponding order of a component $j \in \mathcal{J}$ that satisfies this demand is placed at $t - T(s_j)$, where $T(s_j)$ is determined by starting at time t, counting backwards until the number of demand arrivals reaches s_j. That is, $T(s_j) = \sum_{k=1}^{s_j} V_k$, where V_k is the demand interarrival time, with k being a backward index of the demand arrivals (see Figure1). This is true because an order for the jth component is placed whenever a demand arrives. As observed by Axsater (1993), the order will be used to satisfy the s_jth demand in the future. Thus, the difference between the time at which an order of the component j is placed and the time at which the target demand arrives is exactly the sum of s_j interarrival times. Due to the replenishment lead time, the order placed at time $t - T(s_j)$ will arrive at time $t - T(s_j) + L_j$. Therefore, the delay of component j is $L_j - T(s_j)$.

This observation differs from that of Axsater (1993), because at the time of a demand arrival, we look *backward* to identify the time at which the corresponding order was placed that satisfies this demand (namely, the *backward method*), while Axsater (1993) looks *forward* to identify the

Figure 1. $T(s_j)$ vs. $T(s_k)$ in a single-product assembly system with two components.



time at which the corresponding demand will arrive that is satisfied by the order triggered by the current demand. The difference is important in analyzing assembly systems. For instance, consider an ATO system with two components, j and k. Without loss of generality, assume that $s_j \leqslant s_k$. Suppose that a demand arrives at time t; then the corresponding orders of the components j and k that satisfy this demand are placed at time $t - T(s_j)$ and $t - T(s_k)$, respectively. It is easily seen that $T(s_j)$ overlaps with $T(s_k)$ over the time period $[t - T(s_j), t]$, and therefore $T(s_k) = T(s_j) + T(s_k - s_j)$ (see Figure 1 for a visual aid). The dependence among the arrival times $t - T(s_j) + L_j$, $j \in \mathcal{J}$, is quite intuitive: If the interarrival times are short for recent demands, and as a result $T(s_j)$ is small for all $j \in \mathcal{J}$, then all components are likely to be out of stock.

It is straightforward to extend the observation to systems with an arbitrary number of components. We index the components in the nondecreasing order of their base-stock levels. Given any sequence of $t_1 \leqslant t_2 \leqslant \cdots \leqslant t_{|\mathcal{J}|}$, the joint probability density function of $T(s_j)$, $j \in \mathcal{J}$, is given by

$$\begin{aligned} &P\{T(s_1) = t_1, T(s_2) = t_2, \ldots, T(s_{|\mathcal{J}|}) = t_{|\mathcal{J}|}\} \\ &\quad = P\{T(s_1) = t_1\} P\{T(s_2 - s_1) = t_2 - t_1\} \cdots \\ &\qquad \cdot P\{T(s_{|\mathcal{J}|} - s_{|\mathcal{J}|-1}) = t_{|\mathcal{J}|} - t_{|\mathcal{J}|-1}\}. \end{aligned} \tag{1}$$

For any other sequences of $t_1, t_2, \ldots, t_{|\mathcal{J}|}$, the joint probability density function of $T(s_j)$, $j \in \mathcal{J}$, equals zero. For simplicity, we denote the joint probability density function in Equation (1) as $P(t_1, t_2, \ldots, t_{|\mathcal{J}|})$. A direct application of Equation (1) is the joint probability density function of the delays of all components,

$$\begin{aligned} &P\{L_1 - T(s_1) = \tau_1, L_2 - T(s_2) = \tau_2, \ldots, L_{|\mathcal{J}|} - T(s_{|\mathcal{J}|}) = \tau_{|\mathcal{J}|}\} \\ &\quad = P(L_1 - \tau_1, L_2 - \tau_2, \ldots, L_{|\mathcal{J}|} - \tau_{|\mathcal{J}|}), \end{aligned} \tag{2}$$

where L_j, $j \in \mathcal{J}$, are independent random variables.

By Equation (1), we can characterize the delivery lead-time distribution as follows. For the demand arriving at time t, the latest replenishment time of the components can be expressed as $\max_{j\in\mathcal{J}}\{t - T(s_j) + L_j\}$. Hence, the delivery lead time for this demand equals

$$X = \left[\max_{j\in\mathcal{J}}\{L_j - T(s_j)\}\right]^+. \tag{3}$$

The waiting time of the component k, $k \in \mathcal{J}$, equals the maximum of $t + \max_j\{L_j - T(s_j)\}$ and t subtracting $t - T(s_k) + L_k$, which is

$$\begin{aligned} W_k &= t + \left[\max_j\{L_j - T(s_j)\}\right]^+ - (t - T(s_k) + L_k) \\ &= \left[\max_j\{L_j - T(s_j)\}\right]^+ - L_k + T(s_k) \\ &= X - L_k + T(s_k). \end{aligned} \tag{4}$$

Note that X is correlated with $L_k - T(s_k)$.

We now derive the probability distribution for the delivery lead time, X:

$$\begin{aligned} &P\{X = 0\} = P\left\{\max_{j\in\mathcal{J}}\{L_j - T(s_j)\} \leqslant 0\right\}, \\ &P\{X \leqslant \tau\} = P\left\{\max_{j\in\mathcal{J}}\{L_j - T(s_j)\} \leqslant \tau\right\}, \end{aligned} \tag{5}$$

where $\tau > 0$ is the target service time.

$$\begin{aligned} P\left\{\max_{j\in\mathcal{J}}\{L_j - T(s_j)\} \leqslant \tau\right\} &= P\{L_j - T(s_j) \leqslant \tau,\ j \in \mathcal{J}\} \\ &= P\{T(s_j) \geqslant L_j - \tau,\ j \in \mathcal{J}\} \\ &= P\{T(s_j) \geqslant (L_j - \tau)^+,\ j \in \mathcal{J}\}. \end{aligned} \tag{6}$$

The last equality is due to the fact that $T(s_j)$, $j \in \mathcal{J}$, is always greater than or equal to zero. Conditioning on $\bar{L} = \bar{l} = (l_1, l_2, \ldots, l_{|\mathcal{J}|})$ and utilizing Equation (1),

$$\begin{aligned} &P\{T(s_j) \geqslant (l_j - \tau)^+,\ j \in \mathcal{J}\} \\ &\quad = \int_{(l_1-\tau)^+}^{\infty}\int_{(l_2-\tau)^+}^{\infty}\cdots\int_{(l_{|\mathcal{J}|}-\tau)^+}^{\infty} P(t_1, t_2, \ldots, t_{|\mathcal{J}|})\, dt_{|\mathcal{J}|}\cdots dt_2\, dt_1. \end{aligned} \tag{7}$$

It is easy to see that the dependence among components has an impact on the fill rates because

$$\begin{aligned} &P\{T(s_j) \geqslant (l_j - \tau)^+,\ j \in \mathcal{J}\} \\ &\quad = P\{T(s_1) \geqslant (l_1 - \tau)^+, T(s_1) + T(s_2 - s_1) \\ &\qquad \geqslant (l_2 - \tau)^+, \ldots, T(s_1) + T(s_2 - s_1) \\ &\qquad + \cdots + T(s_{|\mathcal{J}|} - s_{|\mathcal{J}|-1}) \geqslant (l_{|\mathcal{J}|} - \tau)^+\} \end{aligned} \tag{8}$$

is generally not equal to $\prod_{j\in\mathcal{J}} P\{T(s_j) \geqslant (l_j - \tau)^+\}$ (the fill rate obtained by assuming that the delays of all components are independent). Indeed, these two fill rates satisfy the following inequality.

PROPOSITION 3.1. *For any $\bar{l}$ and $\tau \geqslant 0$,*

$$P\{T(s_j) \geqslant (l_j - \tau)^+, j \in \mathcal{J}\} \geqslant \prod_{j\in\mathcal{J}} P\{T(s_j) \geqslant (l_j - \tau)^+\}. \tag{9}$$

That is, the delivery lead time in the single-product assembly system with dependent component delays is stochastically smaller than the delivery lead time in an analogous system with independent component delays.

Note that we use "smaller than" here in its weak form, which means "smaller than or equal to." The intuition behind Proposition 3.1 is that the random component delays, $L_j - T(s_j)$, $j \in \mathcal{J}$, tend to "hang on" together due to the common interarrival times shared by $T(s_j)$, $j \in \mathcal{J}$. We omit the proof of this proposition because a more general result is proven in §4.1 for the multiproduct ATO systems.

The probability defined in Equation (8) is related to a *generalized finite horizon nonruin probability* (see, e.g., De Kok 2003 for the definition). The computation of this probability is discussed in detail in §3.2. It is easily seen from Equation (8) that the second independent condition of the "transit time," i.e., the replenishment lead times are independent across components, is not necessary for the analysis. Equation (8) holds in the cases of dependent replenishment lead times as long as the lead times are independent of the system state. For simplicity, we still assume that the second independent condition holds for the rest of the paper unless otherwise mentioned.

The cost function of the single-product assembly systems can be written as

$$\pi E(X)+\sum_{j\in\mathcal{J}} h_j E(W_j)$$
$$=\left(\pi+\sum_j h_j\right)E(X)-\sum_j h_j E(L_j)+\sum_j h_j E(T(s_j)). \quad (10)$$

Because $\sum_j h_j E(L_j)$ is a constant, the cost function can be simplified as

$$\left(\pi+\sum_j h_j\right)E(X)+\sum_j h_j s_j/\lambda. \quad (11)$$

Our analysis is related to that of Glasserman and Wang (1998), which studies ATO systems with capacitated suppliers. In the proof of their Theorem 1, the delay of a certain component is derived in a way similar to that of this paper. However, the focus of Glasserman and Wang (1998) is on quantifying the limiting trade-off between the inventory levels and the delivery lead time at high fill rates, rather than the dependence structure of the delays among components. For the special case of constant lead times and Poisson demand process, Equation (8) shows that as $\tau\to\infty$, the fill rate tends to 1 for all base-stock levels. Furthermore, the tail distribution of the delivery lead time is Erlang if τ is large enough so that $(l_j-\tau)^+>0$ for only one $j\in\mathcal{J}$.

We now provide simple proofs for the following system properties.

Proposition 3.2. *In the single-product assembly systems,*

(1) *If we keep system parameters unchanged and increase $|\mathcal{J}|$, the cost function is nondecreasing.*

(2) *For two ATO systems, if $L_j\leqslant_{st} L'_j$ for a certain $j\in\mathcal{J}$ while everything else is identical, then $P\{X\leqslant\tau\}\geqslant P\{X'\leqslant\tau\}$, where $\leqslant_{st}$ denotes the stochastic ordering and X (X') is the corresponding delivery lead time of L_j (L'_j, respectively).*

(3) *For two ATO systems, if $L_j\leqslant_{cx} L'_j$ for a certain $j\in\mathcal{J}$ while everything else is identical, then $E(X)\leqslant E(X')$, where $\leqslant_{cx}$ denotes the convex ordering and $E(\cdot)$ is the expectation with respect to both $\bar{L}$ and $T(s_k)$, $k\in\mathcal{J}$.*

Proof. (1) First, note that the expected waiting time of any additional component is always nonnegative. Thus, by Equation (10), we only need to show that $E(X)$ is nondecreasing as $|\mathcal{J}|$ increases, which is straightforward because as $|\mathcal{J}|$ increases, $X=(\max_{j\in\mathcal{J}}\{L_j-T(s_j)\})^+$ is nondecreasing for any realization of $\bar{L}$ and $T(s_j)$, $j\in\mathcal{J}$.

(2) Note that $P\{X\leqslant\tau\mid\bar{l}\}=P\{T(s_j)\geqslant(l_j-\tau)^+\ \forall j\in\mathcal{J}\}$ (by Equations (5)–(6)) is a nonincreasing function of l_j $\forall j\in\mathcal{J}$. Hence, $L_j\leqslant_{st}L'_j$ and the definition of the stochastic ordering (Kulkarni 1995) immediately imply $-P\{X\leqslant\tau\}\leqslant-P\{X'\leqslant\tau\}$.

(3) Note that $(\max_{j\in\mathcal{J}}\{l_j-T(s_j)\})^+$ is a convex function of l_j, $j\in\mathcal{J}$, for any realization of $T(s_j)$, $j\in\mathcal{J}$. Because L_j is independent of $T(s_k)$, $k\in\mathcal{J}$, Equation (3) and the definition of convex ordering immediately imply the requisite result. □

Intuitively, this proposition demonstrates that the larger the number of components, the higher the total cost; the larger the stochastic replenishment lead times, the smaller the fill rates; and the more stochastically variable the replenishment lead times, the longer the expected delivery time. The first property is shown by Kumar (1989) in a model of safety time instead of safety stock. In a model with i.i.d. replenishment lead times, Song and Yao (2002) provides a different proof for the third property under the increasing convex ordering assumption.

3.2. Performance Evaluation

To evaluate the system performances, we rewrite the fill rate in Equation (8) as follows:

$$P\left\{Y_1\geqslant y_1, Y_1+Y_2\geqslant y_2,\ldots,\sum_{j=1}^{|\mathcal{J}|}Y_j\geqslant y_{|\mathcal{J}|}\right\}, \quad (12)$$

where $Y_1=T(s_1)$, $Y_j=T(s_j-s_{j-1})$, $j>1$, and $y_j=(l_j-\tau)^+$, $j\in\mathcal{J}$. Comparing to the finite horizon nonruin probability (see, e.g., Asmussen 2000), the probability in Equation (12) has $\geqslant$ rather than $\leqslant$ for every event.

To calculate the finite horizon nonruin probability, De Vylder and Goovaerts (1988) develop a recursion scheme by conditioning on $Y_1, Y_2,\ldots$ consecutively. De Kok (2003) develops a fast recursive algorithm based on a two-moment approximation for $|\mathcal{J}|\leqslant 20$. For the probability defined in Equation (12), an exact expression can be obtained through a recursive algorithm; see Appendix 1 for details. However, as $|\mathcal{J}|$ increases, the time it takes to derive the exact expression grows very fast even for Poisson demand processes. Further observing that we need to consider all possible values of the lead times to obtain the fill rates, an exact calculation is time consuming for large-scale systems, e.g., systems with hundreds of components. To achieve numerical efficiency, we propose the following methods, which are scalable to evaluate large-size single-product assembly systems with base-stock policies and stochastic lead times.

The first method is based on Clark (1961), which develops an approximate algorithm to calculate the mean and variance of the maximum of dependent random variables. Because the algorithm only uses the mean and covariance of the dependent random variables, it is a two-moment approximation. Based on the assumption of transit time and Equation (1), the covariance of $L_j - T(s_j)$, $j \in \mathcal{J}$, can be calculated as follows:

$$\mathrm{Cov}(L_j - T(s_j), L_k - T(s_k)) = \begin{cases} \mathrm{Var}(T(s_j)) & \text{if } s_j \leqslant s_k, \\ \mathrm{Var}(T(s_k)) & \text{otherwise.} \end{cases} \quad (13)$$

Once the mean and variance of $\max_{j\in\mathcal{J}}\{L_j - T(s_j)\}$ are obtained by Clark's algorithm, we can obtain the approximations of the mean and variance of X by fitting $\max_{j\in\mathcal{J}}\{L_j - T(s_j)\}$ into a normal distribution. The computational complexity of this method is $O(|\mathcal{J}|^2)$ because both the calculation of the covariance and Clark's method require a computing time proportional to $|\mathcal{J}|^2$.

The second method is based on Monte Carlo simulation (Law and Kelton 1991). To obtain a sample of X, we just need to generate a sample of the independent random variables $T(s_1)$, $T(s_j - s_{j-1})$, $1 < j \leqslant |\mathcal{J}|$ and L_j, $j \in \mathcal{J}$, then utilize Equation (3). The computational complexity is $O(|\mathcal{J}|)$ if we fix the sample size. Thus, for relatively small $|\mathcal{J}|$, Clark's method may be more efficient, but for large $|\mathcal{J}|$, the simulation-based method is faster.

To validate our approach, we use the simulation-based method to evaluate the expected backorders for the test example in Song and Yao (2002, p. 897). Because they assume i.i.d. replenishment lead times while we assume transit times, only the cases of deterministic replenishment lead times are comparable. Using a sample size of 40,000, our computational study (not reported here) shows that the results generated by the simulation-based method accurately match those of Song and Yao.

To test the accuracy of the method based on Clark's approximation, we conduct a numerical study to compare the methods based on Clark's approximation and the simulation. We consider the single-product assembly systems with $|\mathcal{J}| = 20, 80$; Poisson demand process, and $\lambda = 1$. We assume that the replenishment lead times follow Erlang distribution because it is a special case of both gamma and continuous phase-type (CPH) distributions, which are often used in practice to approximate lead times (Zipkin 2000). We also point out that Erlang distribution is assumed in the numerical studies of Song and Yao (2002) and Lu et al. (2003).

To generate test instances for a certain $|\mathcal{J}|$, we first fix the lead-time distributions and then vary base-stock levels according to $\bar{s} = \lceil \beta \times \bar{s}' \rceil$, where $\bar{s}'$ is the default base-stock level, and $\beta \in \{0.25, 0.5, 0.75, 1, 1.5, 2, 3, 4, 5, 6, 7, 8\}$. For our randomly generated lead-time distributions and the default base-stock levels, see http://zhao.rutgers.edu/. In the simulation, we set sample size equal to 40,000 for all test examples. The running times of both methods for all test examples are in seconds on a Pentium 1.67 GHZ laptop.

Figure 2. Accuracy of Clark's (1961) method in estimating the expected delivery lead time for single-product assembly systems.

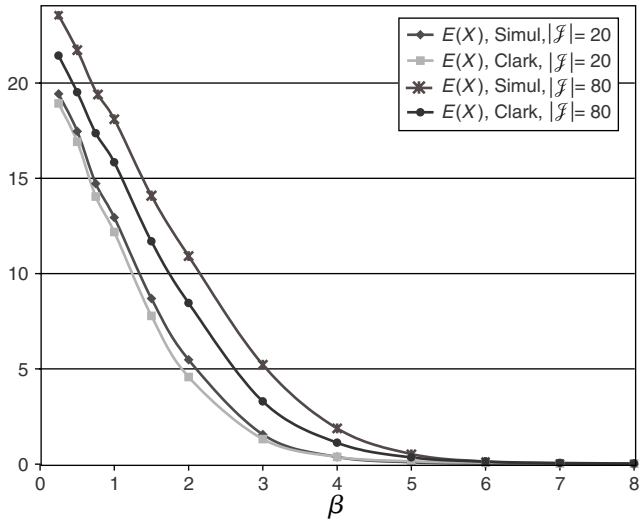


Figure 2 demonstrates that the method based on Clark's approximation is reasonably accurate in the absolute value of the expected delivery lead time when $|\mathcal{J}| = 20$. However, it may be subject to a large percentage error (defined as the absolute difference of the computed $E(X)$s between these two methods divided by the $E(X)$ estimated by the simulation-based method) when β is large, i.e., when $E(X)$ is small. As $|\mathcal{J}|$ increases to 80, both the absolute error and percentage error increase considerably, which implies that the method based on Clark's approximation may only be appropriate for small to medium-size problems. Figure 3 shows that the method based on Clark's approximation has a relatively large error in calculating the standard deviation of the delivery lead times. However, the method is generally quite accurate in terms of the fill rate ($\tau = 0$). Except, in one test example, we observe that the difference of the fill rates between Clark's approximation and the simulation is as large as 8%.

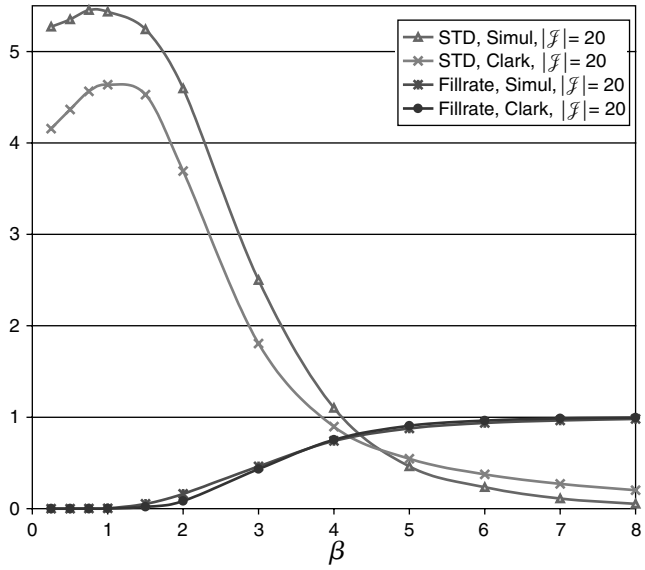
4. Multiple-Product Base-Stock Systems

In this section, we consider the multiproduct ATO systems with the base-stock policies and demands following independent Poisson processes. We first present the decomposition and characterize the correlations of the component delays in §4.1, and then propose numerical methods in §4.2.

4.1. Performance Analysis

The analysis is based on the *backward method* (see §3.1). Assuming that a demand of product type $i \in \mathcal{I}$ arrives at

Figure 3. Accuracy of Clark's (1961) method in estimating the standard deviation of the delivery lead time and the fill rate for single-product assembly systems.



time t, then the corresponding order of a component $j \in \mathcal{F}^i$ that satisfies this demand is placed at time $t - T_j^i(s_j)$, where $T_j^i(s_j)$ is determined by starting at time t, counting backward demand arrivals of all products that require component j until the total number of arrivals reaches s_j. This observation holds as a result of the noncrossing property of the stochastic sequential lead times and the "committed stock" assumption (§2). Due to the replenishment lead time, the order placed at time $t - T_j^i(s_j)$ will arrive at time $t - T_j^i(s_j) + L_j$.

For each product $i \in \mathcal{I}$, given the joint probability density function of $T_j^i(s_j)\ \forall\, j \in \mathcal{F}^i$, we can apply Equations (3) and (4) to determine the delivery lead time, X^i, and the component waiting times W_j^i, $j \in \mathcal{F}^i$ (defined to be the waiting time of the component j if it satisfies a demand of product i), as follows:

$$X^i = \left[\max_{j \in \mathcal{F}^i}\{L_j - T_j^i(s_j)\}\right]^+, \tag{14}$$

$$W_j^i = X^i - L_j + T_j^i(s_j). \tag{15}$$

Clearly, W_j^i may have different probability distributions if the component j satisfies the demand of different products. Conditioning on $\bar{L} = \bar{l} = (l_1, l_2, \ldots, l_{|\mathcal{F}|})$ and utilizing Equation (14) yields

$$P\{X^i \leqslant \tau \mid \bar{l}\} = P\{T_j^i(s_j) \geqslant (l_j - \tau)^+,\ j \in \mathcal{F}^i\}, \tag{16}$$

where $\tau \geqslant 0$ is the target service time. To characterize the impact of the dependence among the component delays, we utilize the concept of "associated" random variables.

Consider random variables $Y_1, Y_2, \ldots, Y_n$, and denote vector $\overline{Y} = (Y_1, Y_2, \ldots, Y_n)$. The following definition is due to Esary et al. (1967); see Tong (1980) and Shaked and Shanthikumar (1994) for reviews.

Definition 4.1. The set of random variables $\{Y_1, Y_2, \ldots, Y_n\}$ is associated, or the random variables $Y_1, Y_2, \ldots, Y_n$ are associated, if

$$\mathrm{Cov}[f(\overline{Y}), g(\overline{Y})] \geqslant 0, \tag{17}$$

or equivalently,

$$E[f(\overline{Y})g(\overline{Y})] \geqslant E[f(\overline{Y})]E[g(\overline{Y})] \tag{18}$$

for all nondecreasing real functions f, g for which $E[f(\overline{Y})]$, $E[g(\overline{Y})]$, and $E[f(\overline{Y})g(\overline{Y})]$ exist.

Lemma 4.2. *Associated random variables have the following properties*:

(1) *Any subset of associated random variables is associated.*

(2) *If two sets of associated random variables are independent of each other, their union is a set of associated random variables.*

(3) *Independent random variables are associated.*

(4) *Nondecreasing (or nonincreasing) functions of associated random variables are associated.*

(5) *Let $Y_1, Y_2, \ldots, Y_n$ be associated random variables; then*

$$P\{Y_1 \leqslant y_1, Y_2 \leqslant y_2, \ldots, Y_n \leqslant y_n\} \geqslant \prod_{k=1}^{n} P\{Y_k \leqslant y_k\}$$

and

$$P\{Y_1 \geqslant y_1, Y_2 \geqslant y_2, \ldots, Y_n \geqslant y_n\} \geqslant \prod_{k=1}^{n} P\{Y_k \geqslant y_k\}$$

for all $(y_1, y_2, \ldots, y_n) \in R^n$.

We refer to Esary et al. (1967) and Tong (1980) for proofs. Intuitively, Part (5) of Lemma 4.2 means that associate random variables are dependent in such a way that they tend to "hang on" together.

The following proposition characterizes the impact of the dependence among the component delays on the multiproduct ATO systems. The proof involves establishing that the set of random variables $\{T_j^i(s_j),\ j \in \mathcal{F}^i\}$ is associated for each product i.

Proposition 4.3. *In a multiproduct ATO system, for any $\bar{l}$ and $\tau \geqslant 0$,*

$$P\{T_j^i(s_j) \geqslant (l_j - \tau)^+,\ j \in \mathcal{F}^i\} \geqslant \prod_{j \in \mathcal{F}^i} P\{T_j^i(s_j) \geqslant (l_j - \tau)^+\}$$
$$\text{for all } i \in \mathcal{I}. \tag{19}$$

That is, the delivery lead time of each product in the multiproduct ATO systems with dependent component delays is stochastically smaller than the delivery lead time of the same product in an analogous system with independent component delays.

PROOF. See Appendix 4 for the details. □

Equations (14) and (15) imply that the multiproduct ATO system can be decomposed into $|\mathcal{J}|$ single-product subsystems with each subsystem corresponding to a product $i \in \mathcal{J}$ and its component set $\mathcal{J}^i$. Each single-product subsystem can be characterized separately. It is straightforward to show that statements (2) and (3) of Proposition 3.2 apply to each single-product subsystem defined here. However, it is important to note that these single-product subsystems are not identical to the single-product assembly systems of §3 because $T_j^i(s_j)$, $j \in \mathcal{J}^i$, are dependent in a different way than $T(s_j)$, $j \in \mathcal{J}$. It is easily seen that $T_j^i(s_j)$ is associated with the superposition of the demand processes of all products that require component j, as well as the arrival time, t, of product i. The random variables, $T_j^i(s_j)$, are dependent across components $j \in \mathcal{J}^i$ because of the common demand process of product i. It is not yet clear how to characterize the joint probability density function of $T_j^i(s_j)$, $j \in \mathcal{J}^i$, because a component $j \in \mathcal{J}^i$ may be required by other products. Thus, we do not attempt to derive the joint probability density function in this paper, but instead, we characterize the covariance matrix of $T_j^i(s_j)$, $j \in \mathcal{J}^i$. For this purpose, we first point out the following facts:

- Poisson arrival processes are reversible in the sense that starting at any time $t \in (-\infty, +\infty)$, the arrival process counting backward is statistically identical to the arrival process counting forward (see, e.g., Kulkarni 1995, Theorem 8.12 and Corollary 8.2).
- The forward and backward counting processes are independent of the starting time t.
- The forward and backward counting processes are still Poisson.

Because we assume that the demand follows independent Poisson processes, $T_j^i(s_j)$ is independent of the starting time t and therefore the product type i. Hence, we can simplify the notation of $T_j^i(s_j)$ by $T_j(s_j)$. Finally, because the covariance of $T_j(s_j)$ and $T_k(s_k)$ is the same for all products $i \in \mathcal{J}_j \cap \mathcal{J}_k$, we focus on the covariance matrix of $T_j(s_j)$, $j \in \mathcal{J}$.

Given two components j and $\jmath$, and their base-stock levels s_j and $s_\jmath$, we can calculate the covariance of $T_j(s_j)$ and $T_\jmath(s_\jmath)$ by considering the following four cases:

Case 1. $\mathcal{J}_j \cap \mathcal{J}_\jmath = \varnothing$.
Case 2. $\mathcal{J}_j = \mathcal{J}_\jmath$.
Case 3. $\mathcal{J}_\jmath \subset \mathcal{J}_j$.
Case 4. $\mathcal{J}_j \cap \mathcal{J}_\jmath \neq \varnothing$, $\mathcal{J}_j \not\subset \mathcal{J}_\jmath$, and $\mathcal{J}_\jmath \not\subset \mathcal{J}_j$.

It is easy to see that $\mathrm{Cov}(T_j(s_j), T_\jmath(s_\jmath)) = 0$ in the first case. The second case reduces to the single-product systems. By Equation (13), $\mathrm{Cov}(T_j(s_j), T_\jmath(s_\jmath)) = \mathrm{Var}(T_\jmath(s_\jmath))$ if $s_j \geqslant s_\jmath$, or $\mathrm{Var}(T_j(s_j))$ otherwise.

To calculate the covariance for the third case, we denote the common demand process for both components to be process 1, and the independent demand process for component j only to be process 2. Define $U_l(m)$ to be the sum of m interarrival times of the process $l = 1, 2$, and γ_l to be the arrival rate of the process $l = 1, 2$. Conditioning on the common demand process 1, let one realization of $U_1(1), U_1(2), \ldots$ be $\bar{u} = (u_1, u_2, \ldots)$, and $\bar{v}$ be $v_1 = u_1$, $v_2 = u_2 - u_1, \ldots,$ where $P\{V = v\} = \gamma_1 e^{-\gamma_1 v}$. For simplicity, denote the joint probability density function of the interarrival times of the process 1, $\bar{V}$, to be $f_1(\bar{v})$; then,

$$\mathrm{Cov}(T_j(s_j), T_j(s_\jmath)) = \int E(T_j(s_j) - E(T_j(s_j)) \mid \bar{v}) \times (u_{s_\jmath} - E(T_\jmath(s_\jmath)))f_1(\bar{v})\, d\bar{v}. \quad (20)$$

To calculate $E(T_j(s_j) - E(T_j(s_j)) \mid \bar{v})$, we only need to condition on $v_1, v_2, \ldots, v_{s_\jmath}$ because $T_j(s_j) \leqslant u_{s_\jmath}$.

For simplicity, we denote $T = T_j(s_j)$; then,

$$\begin{aligned}
&T = u_{s_j} && \text{on the event } \{U_2(1) > u_{s_j}\}\\
&T = U_2(1) && \text{on the event } \{u_{s_j-1} < U_2(1) < u_{s_j}\}\\
&T = u_{s_j-1} && \text{on the event } \{U_2(1) < u_{s_j-1}\} \cap \{U_2(2) > u_{s_j-1}\}\\
&\ldots && \ldots\\
&T = U_2(s_j - 1) && \text{on the event } \{u_1 < U_2(s_j - 1) < u_2\}\\
&T = u_1 && \text{on the event } \{U_2(s_j - 1) < u_1\} \cap \{U_2(s_j) > u_1\}\\
&T = U_2(s_j) && \text{on the event } \{U_2(s_j) < u_1\}.
\end{aligned}$$

Because we assume Poisson processes,

$$\begin{aligned}
&E(T - E(T) \mid \bar{v})\\
&\quad = (u_{s_j} - E(T))e^{-\gamma_2 u_{s_j}} + \int_{u_{s_j-1}}^{u_{s_j}} (t - E(T))\gamma_2 e^{-\gamma_2 t}\, dt\\
&\qquad + (u_{s_j-1} - E(T))e^{-\gamma_2 u_{s_j-1}} \frac{\gamma_2 u_{s_j-1}}{1!}\\
&\qquad + \cdots + \int_{u_{s_j-k-1}}^{u_{s_j-k}} (t - E(T))\gamma_2 e^{-\gamma_2 t} \frac{(\gamma_2 t)^k}{k!}\, dt\\
&\qquad + (u_{s_j-k-1} - E(T))e^{-\gamma_2 u_{s_j-k-1}} \frac{(\gamma_2 u_{s_j-k-1})^{k+1}}{(k+1)!}\\
&\qquad + \cdots + \int_{u_1}^{u_2} (t - E(T))\gamma_2 e^{-\gamma_2 t} \frac{(\gamma_2 t)^{s_j-2}}{(s_j-2)!}\, dt\\
&\qquad + (u_1 - E(T))e^{-\gamma_2 u_1} \frac{(\gamma_2 u_1)^{s_j-1}}{(s_j-1)!}\\
&\qquad + \int_0^{u_1} (t - E(T))\gamma_2 e^{-\gamma_2 t} \frac{(\gamma_2 t)^{s_j-1}}{(s_j-1)!}\, dt.
\end{aligned}$$

After tedious integration and a combination of common terms, we finally arrive at

$$\begin{aligned}
&E(T - E(T) \mid \bar{v})\\
&\quad = \frac{s_j}{\gamma_2} - E(T) - \frac{1}{\gamma_2}(P\{U_2(1) \geqslant u_{s_j}\}\\
&\qquad + P\{U_2(2) \geqslant u_{s_j-1}\} + \cdots + P\{U_2(s_j) \geqslant u_1\}), \quad (21)
\end{aligned}$$

where $E(T) = s_j/(\gamma_1 + \gamma_2)$. Thus, the covariance

$$\begin{aligned}
&\mathrm{Cov}(T_j(s_j), T_J(s_J)) \\
&= -\frac{1}{\gamma_2}\Bigg[\sum_{k=0}^{s_j-1}\int P\{U_2(k+1) \geqslant u_{s_j-k}\}(u_{s_J} - E(T_J(s_J))) \\
&\quad \times P\{U_1(s_j-k) = u_{s_j-k}, U_1(s_J) = u_{s_J}\}\, du_{s_j-k}\, du_{s_J}\Bigg]. \quad (22)
\end{aligned}$$

To calculate the covariance for the fourth case, we define the common demand process for both components j and J to be process 1, the independent process for component j to be process 2, and the independent process for component J to be process 3. Let γ_l be the arrival rate of the process $l = 1, 2, 3$. Conditioning on the interarrival times, $\bar{V} = \bar{v}$, of the common demand process, and following a similar calculation as in the third case, we obtain

$$\begin{aligned}
&\mathrm{Cov}(T_j(s_j), T_J(s_J)) \\
&= \int E(T_j(s_j) - E(T_j(s_j)) \mid \bar{v}) \\
&\quad \times E(T_J(s_J) - E(T_J(s_J)) \mid \bar{v}) f_1(\bar{v})\, d\bar{v} \\
&= \frac{1}{\gamma_2\gamma_3}\int [P\{U_2(1) \geqslant u_{s_J}\} + \cdots + P\{U_2(s_j) \geqslant u_1\}] \\
&\quad \times [P\{U_3(1) \geqslant u_{s_J}\} + \cdots + P\{U_3(s_J) \geqslant u_1\}] \times f_1(\bar{v})\, d\bar{v} \\
&\quad - \frac{s_J}{\gamma_2}\left(\frac{1}{\gamma_3} - \frac{1}{\gamma_1+\gamma_3}\right)\int [P\{U_2(1) \geqslant u_{s_j}\} \\
&\qquad + \cdots + P\{U_2(s_j) \geqslant u_1\}] f_1(\bar{v})\, d\bar{v}. \quad (23)
\end{aligned}$$

For the exact expressions of Equations (22) and (23), see Appendix 2.

Define

$$P_j^i = \frac{a_j^i \lambda^i}{\sum_{k\in\mathcal{F}} a_j^k \lambda^k}$$

to be the probability that the component j satisfies a demand of product i; the cost function of the multiproduct ATO systems can be written as follows:

$$\begin{aligned}
&\sum_{i\in\mathcal{F}} \pi^i E(X^i) + \sum_{j\in\mathcal{J},\, i\in\mathcal{F}} P_j^i h_j E(W_j^i) \\
&\quad = \sum_{i\in\mathcal{F}}\left(\pi^i + \sum_{j\in\mathcal{J}} P_j^i h_j\right) E(X^i) - \sum_{j\in\mathcal{J}} h_j E(L_j) \\
&\qquad + \sum_{j\in\mathcal{J}} h_j E(T_j(s_j)). \quad (24)
\end{aligned}$$

The equality comes from the fact that $\sum_{i\in\mathcal{F}} P_j^i = 1\ \forall\, j \in \mathcal{J}$. The reason that we take the weighted average P_j^i, $i \in \mathcal{F}$, for the holding cost of the component $j \in \mathcal{J}$, is that the expected waiting time of the component j may vary if the component satisfies a demand of different products (Equation 14). Thus, the long-run average inventory holding cost of a component is the weighted average of the counterparts of this component across all products that require the component.

4.2. Performance Evaluation

In this section, we develop numerical methods to evaluate the performances of the multiproduct ATO systems with the base-stock policies. Similarly to the single-product systems, the first method is based on Clark's (1961) approximation, while the second method is based on Monte Carlo simulation.

The method based on Clark's approximation works as follows: Given any two components j and J, we first determine which case applies, and calculate γ_l, $l = 1, 2, 3$. Then, using the analytic expressions derived in §4.1 and Appendix 2, we calculate the covariances among $T_j(s_j)$, $j \in \mathcal{J}$. Lastly, we apply Clark's approximation to evaluate the performances for each single-product subsystem $i \in \mathcal{F}$, as in §3.2. The computational effort is at most proportional to $C^2_{|\mathcal{J}|} \times (|\mathcal{F}| + s^5_{\max}) + |\mathcal{F}| \times |\mathcal{J}|^2$, where $s_{\max} = \max_{j\in\mathcal{J}}\{s_j\}$ and $C^2_{|\mathcal{J}|}$ is defined in Appendix 2. Observe that determining a single covariance of Case 4 requires a computing time proportional to $s^5_{\max}$, the method can only handle problems with small $s_{\max}$.

The second method is based on Monte Carlo simulation. To obtain a sample of all X^i, $i \in \mathcal{F}$, simultaneously, we sample $s_{\max}$ many inter arrival times independently for each product. Then, for each component $j \in \mathcal{J}$, we superimpose the arrival processes of all products that require it, and count the time of s_jth arrival to obtain a sample for $T_j(s_j)$. Finally, we generate a sample of L_j, $j \in \mathcal{J}$, and utilize Equation (14) to obtain a sample of $X^i\ \forall\, i \in \mathcal{F}$. Thus, the simulation-based method has a computational complexity at most proportional to $s_{\max}(|\mathcal{F}| + |\mathcal{J}|(|\mathcal{F}| - 1)) + |\mathcal{F}||\mathcal{J}|$ because sampling interarrival times takes at most $O(|\mathcal{F}|s_{\max})$, and the superposition and counting processes take at most $O(|\mathcal{J}|(|\mathcal{F}| - 1)s_{\max})$.

To summarize, the computational effort of both methods is linear in the number of products. The method based on Clark's approximation may be more efficient when $|\mathcal{J}|$ and $s_{\max}$ are small; but the simulation-based method is faster when either $|\mathcal{J}|$ or $s_{\max}$ is large. To validate our approach, we study the personal computer example of Song (2002). In this example, there are six components and six demand types (products):

- Components are (1) built-in zip drive, (2) standard hard drive, (3) high-profile hard drive, (4) DVD-Rom drive, (5) standard processor, and (6) high-profile processor.
- Demand types result from different combinations of these components, i.e., (1) $\{2, 5\}$, (2) $\{3, 5\}$, (3) $\{1, 2, 5\}$, (4) $\{1, 3, 6\}$, (5) $\{1, 3, 4, 5\}$, and (6) $\{1, 3, 4, 6\}$.

The lead times of all components are deterministic. The BOM matrix A equals

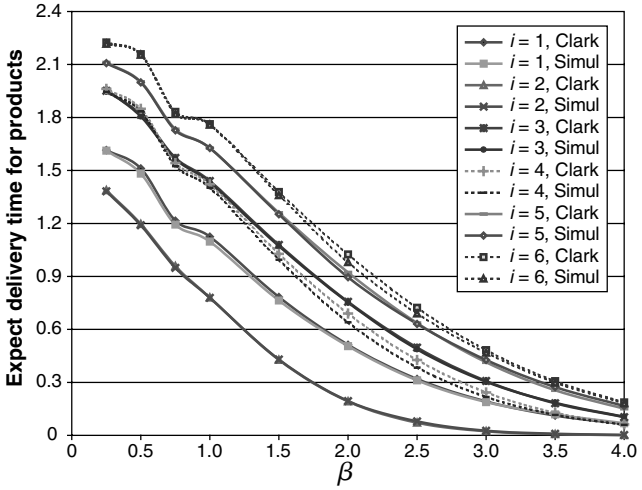
$$\begin{bmatrix} & \text{Component (1)} & (2) & (3) & (4) & (5) & (6) \\ \text{Product (1)} & 0 & 1 & 0 & 0 & 1 & 0 \\ (2) & 0 & 0 & 1 & 0 & 1 & 0 \\ (3) & 1 & 1 & 0 & 0 & 1 & 0 \\ (4) & 1 & 0 & 1 & 0 & 0 & 1 \\ (5) & 1 & 0 & 1 & 1 & 1 & 0 \\ (6) & 1 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}.$$

We refer to Song (2002) for more details on the example. Our numerical study (not reported here) shows that the results of the simulation-based method accurately match those of Song (2002), which are generated by an exact method.

To test the accuracy of the method based on Clark's approximation, we extend the numerical study of this example to include stochastic lead times. We assume the same BOM matrix as in the previous example. Further, we assume that all lead times follow Erlang distribution with randomly generated $\bar{\mu} = (3.96, 3.08, 8.86, 2.6, 4.81, 2.63)$ and $\bar{n} = (7, 5, 11, 5, 7, 5)$ (μ and n are the parameters for an Erlang distribution; see Zipkin 2000 for the definitions). Let the demand arrival rates $\bar{\lambda} = (1, 4, 1.5, 1, 2, 0.5)$. We calculate the mean and the standard deviation of the delivery lead time, as well as the fill rate (at $\tau = 0$) for each base-stock level $\bar{s} = \lceil \beta \times \bar{s}' \rceil$, given $\bar{s}' = (4, 2, 6, 2, 8, 2)$ to be a default base-stock level (randomly chosen) and $\beta \in \{0.25, 0.5, 0.75, 1, 1.5, 2, 2.5, 3, 3.5, 4\}$. In the simulation, the sample size is set to 10,000 for all test examples. The running times of both methods for all test examples are in seconds on a Pentium 1.67 GHZ laptop.

In Figures 4, 5, and 6, we show the mean and standard deviation of the delivery lead time and the fill rate as a function of β for each product. As in the single-product examples, these figures illustrate that the method based on Clark's approximation is quite accurate in terms of the expected delivery lead times. However, for the standard deviation of the delivery lead times, the approximation may be subject to large errors. We also observe that the accuracy varies across different products, e.g., the approximation is more accurate for product 2 than products 5 and 6 in terms of the standard deviation of the delivery lead time and the fill rate. To specify the conditions under which the method based on Clark's approximation works well, we identify the following two sources of errors: first, Clark's two-moment approximation for calculating the mean and standard deviation of $\max_{j \in \mathcal{J}^i}\{L_j - T_j(s_j)\}$; second, the normal approximation of the maximum given its mean and standard deviation. Clark's approximation is originally designed for normal random variables; thus, its accuracy can be poor if it is applied to random variables with other distributions, e.g., exponential or uniform. Consequently, the method based on

Figure 4. Accuracy of Clark's (1961) method in estimating the expected delivery lead time for multiproduct ATO systems.



Clark's approximation is, in general, more accurate if both the lead times L_j and the corresponding order times $T_j(s_j)$ have closer to normal distributions, e.g., the distributions of L_j and $T_j(s_j)$ have larger n in our specific settings.

The numerical results demonstrate that the method based on Clark's two-moment approximation may not be reliably accurate. For other methods of approximating the multivariate probability distributions in ATO systems, we refer the reader to Song and Yao (2002) and Dayanik et al. (2003).

For ATO systems with i.i.d. lead times, Lu et al. (2003) and Lu and Song (2005) provide different methods. Lu et al. (2003) derive the joint probability distribution as well as the first two moments of the outstanding orders, which allows the authors to develop numerical methods based on the factorized normal approximation and pairwise partition

Figure 5. Accuracy of Clark's (1961) method in estimating the standard deviation of the delivery lead time for multiproduct ATO systems.

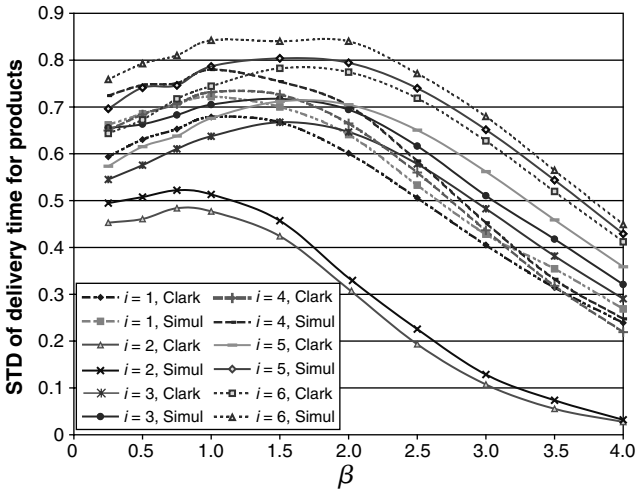
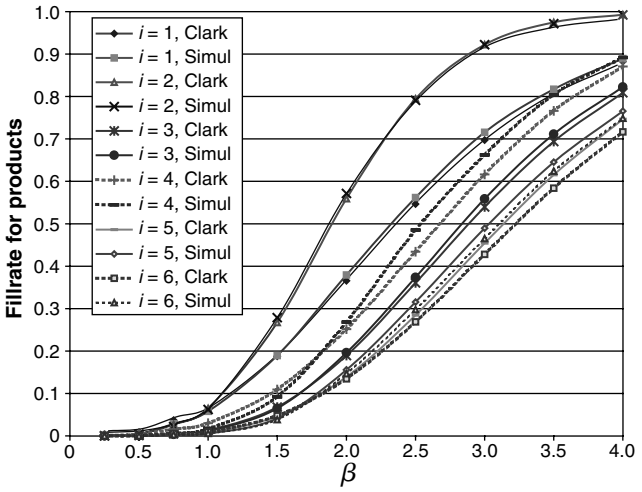


Figure 6. Accuracy of Clark's (1961) method in estimating the fill rate for multiproduct ATO systems.



of the set of components. Lu and Song (2005) develop bounds for the order-based performance measures by the item-based ones. This is done by constructing appropriate independent random variables to replace the dependent outstanding orders and by developing bounds for their maximum. This method is different from our approach because Clark's method directly approximates the first two moments of the maximum of the dependent random variables. It is perhaps worth mentioning that in the special case of constant lead times, the methods developed under the i.i.d. lead-time assumption can also be applied. In particular, Song (1998, 2002) develops exact methods to calculate the order-based fill rates and backorders for ATO systems with constant lead times.

Finally, to test the scalability of the numerical methods, we study a hypothetical example inspired by a real-world problem: the *Dimension 2400 Pentium* model of Dell Computer Corporation. Because the computer model is targeted at individual customers, it is reasonable to assume unit demand. The model has 27 categories of configurable options, with nine of them being software related (for detailed information on this model, see the website printout at zhao.rutgers.edu). In the 18 nonsoftware related categories, each option refers to either a component or not choosing any component in a category. Customer order specifies at most one unit of one component from each category. In these categories, some components may be shipped after the system is shipped, e.g., Dell W1700 LCD TV. For simplicity, we ignore these components. We also ignore six seemingly unpopular categories, e.g., storage device and media, and CD or DVD burner for 2nd bay. In the end, we have 12 categories with a total of 43 options. The categories (the number of options in a category) are, processor (2), memory (4), hard drive (3), floppy (2), CD or DVD (5), keyboard (3), mouse (4), monitor (9), speaker (5), modem (2), network interface (2), and wireless (2).

Because we do not have the real-world demand information, the demand types are created as follows. We first assume that there exists a demand type that chooses all the baseline options. We then assume that for every category, there exists a demand type that chooses an option other than the baseline only in that category; We finally assume that for every two categories, there exists a demand that chooses options other than the baselines only in those two categories. Thus, the total number of demand types (products, equivalently) is 449. These demand types are created based on the assumption that the *Dimension 2400 Pentium* model is targeted at the low-end market, in which it is unlikely for a customer to choose options other than the baselines in all categories. Indeed, we can create more demand types by assuming that for every three categories, there exist demand-choosing options other than the baselines only in those three categories. However, it increases the number of the demand types to 3,684. As a result, the BOM matrix becomes so big that it exceeds the memory limit of our laptop computer. Lastly, we assume that the demands follow independent Poisson processes and the demand arrival rates are randomly generated ($\sim Uniform(2, 12)$). Because we also lack real-world supply information, we created test examples by randomly generating the base-stock levels ($\sim Uniform(0, 50)$) and the parameters of the Erlang distributions for the random lead times. For the option of not choosing any component in a category, we replaced it with an artificial component with zero base-stock level and zero lead time.

Because the base-stock levels can be as high as 50, the problem size is too large for the method based on Clark's approximation. Given a sample size of 10^4, our numerical experiments show that the method based on Monte Carlo simulation takes less than five minutes on a Pentium 1.67 GHZ laptop to evaluate one randomly generated example.

5. Multiple-Product Batch-Ordering Systems

In this section, we consider the multiproduct ATO systems with the batch-ordering policies and demand following independent Poisson processes. We first provide the analysis in §5.1, and then present the numerical methods in §5.2.

5.1. Performance Analysis

The analysis of batch-ordering ATO systems is based on the *backward method* (see §3.1) as well as on Song (2000). In particular, Song (2000) considers batch-ordering ATO systems with constant (not necessarily identical) replenishment lead times and demands following a multivariate compound Poisson process. It is shown that under Assumption 1 of Song (2000) (see below) that the inventory position vector of the components is uniformly distributed in the vector space $\mathcal{S} = \bigotimes_{j\in\mathcal{J}} \mathcal{S}_j$, where $\mathcal{S}_j = \{r_j + 1, r_j + 2, \ldots, r_j + Q_j\}$.

ASSUMPTION 5.1 (ASSUMPTION 1 OF SONG 2000). *The Markov chain of the inventory position vector of the components is irreducible and aperiodic.*

We refer the reader to Song (2000) for the sufficient conditions under which Assumption 5.1 holds.

To understand the impact of Assumption 5.1, we again consider the Dell example of §4.2. An important feature of this example is that a customized order must require at least one type of component from each of the basic categories, e.g., memory, hard drive, and storage drive, etc. Indeed, this example is representative of many real-world ATO systems, in which the components can be classified into different categories, and each customized order specifies at least one component from each category; see Song and Zipkin (2003) for other examples.

In the Dell example, for any component in a given category, it is possible that customer demand requires only one unit of this component together with components from other categories. By Condition (a) of Song (2000), Assumption 5.1 holds for the inventory position vector of the components in each category. However, Assumption 5.1 may not hold for the inventory position vector of the components across all categories. To see this, we consider a simple case in which each category has only one type of component. If, further, each order requires exactly one unit from each category, the model is the single-product assembly system. As Song (2000) points out, the inventory position vector of the components in a single-product assembly system depends on the initial state, and in general does not satisfy Assumption 5.1. Fortunately, these systems can still be analyzed by randomizing the initial state (Song 2000).

The following proposition is based on Theorem 1 of Song (2000).

PROPOSITION 5.2. *Under Assumption* 5.1, *suppose that a demand for a product* $i \in \mathcal{I}$ *arrives at time* t; *then, the corresponding order of a component* j, $j \in \mathcal{J}^i$, *that satisfies this demand is placed at time* $t - T_j^i(S_j)$, *where* $T_j^i(\cdot)$ *is defined in* §4.1, *and the random vector* $(S_j,\ j \in \mathcal{J}^i)$ *is uniformly distributed in* $\mathcal{S}^i = \bigotimes_{j\in\mathcal{J}^i} \mathcal{S}_j$.

PROOF. See Appendix 3 for the details. □

For simplicity, we denote the random vector $(S_j,\ j \in \mathcal{J}^i)$ in Proposition 5.2 by $\bar{S}^i$, and let $\bar{s}^i = (s_j,\ j \in \mathcal{J}^i)$ be a realization of $\bar{S}^i$. Proposition 5.2 implies that for each product $i \in \mathcal{I}$ and each $\bar{s}^i \in \mathcal{S}^i$, we can apply Equations (14)–(15) to determine the delivery lead time, denoted by $X^i(\bar{s}^i)$, and the component waiting times, denoted by $W_j^i(\bar{s}^i)$, $j \in \mathcal{J}^i$, as follows:

$$X^i(\bar{s}^i) = \max_{j\in\mathcal{J}^i}\{(L_j - T_j^i(s_j))^+\}, \tag{25}$$

$$W_j^i(\bar{s}^i) = X^i(\bar{s}^i) - L_j + T_j^i(s_j). \tag{26}$$

Hence, for product $i \in \mathcal{I}$, the batch-ordering ATO system can be reduced to multiple single-product subsystems, each with the product i, component set $\mathcal{J}^i$, and a base-stock level vector $\bar{S}^i$ uniformly distributed in $\mathcal{S}^i$. As in §4.1, we can simplify the notation of $T_j^i(s_j)$ by $T_j(s_j)$ for all $j \in \mathcal{J}^i$ and $i \in \mathcal{I}$. As we point out in §4.1, these single-product subsystems are not identical to the single-product assembly systems of §3 because $T_j(s_j)$, $j \in \mathcal{J}^i$, is associated with the superposition of the demand processes of all products that require component j.

The expected delivery lead time, $E(X^i)$, $i \in \mathcal{I}$, in a batch-ordering system can be characterized by

$$E(X^i) = \frac{1}{\prod_{j\in\mathcal{J}^i} Q_j} \sum_{\bar{s}^i\in\mathcal{S}^i} E(X^i(\bar{s}^i)), \tag{27}$$

the fill rate for a target service time $\tau \geqslant 0$ is given by

$$\begin{aligned} P\{X^i \leqslant \tau\} &= \frac{1}{\prod_{j\in\mathcal{J}^i} Q_j} \sum_{\bar{s}^i\in\mathcal{S}^i} P\{X^i(\bar{s}^i) \leqslant \tau\} \\ &= \frac{1}{\prod_{j\in\mathcal{J}^i} Q_j} \sum_{\bar{s}^i\in\mathcal{S}^i} P\{L_j - T_j(s_j) \leqslant \tau,\ \forall\, j \in \mathcal{J}^i\}, \end{aligned} \tag{28}$$

and the expected inventory holding time for a component $j \in \mathcal{J}$, if it satisfies a demand of product $i \in \mathcal{I}_j$, follows from Equations (26) and (27),

$$\begin{aligned} E(W_j^i) &= \frac{1}{\prod_{j\in\mathcal{J}^i} Q_j} \sum_{\bar{s}^i\in\mathcal{S}^i} E(W_j^i(\bar{s}^i)) \\ &= E(X^i) - E(L_j) + \frac{1}{Q_j} \sum_{s_j\in\mathcal{S}_j} E(T_j(s_j)). \end{aligned} \tag{29}$$

Note that Equations (27) and (28) are consistent with Song (2000). Finally, the cost function of the multiproduct batch-ordering ATO systems follows from Equations (24) and (29),

$$\begin{aligned} &\sum_{i\in\mathcal{I}} \pi^i E(X^i) + \sum_{j\in\mathcal{J},\, i\in\mathcal{I}} P_j^i h_j E(W_j^i) \\ &\quad = \sum_{i\in\mathcal{I}} \left(\pi^i + \sum_{j\in\mathcal{J}} P_j^i h_j\right) E(X^i) - \sum_{j\in\mathcal{J}} h_j E(L_j) \\ &\qquad + \sum_j \frac{h_j}{Q_j} \sum_{s_j\in\mathcal{S}_j} E(T_j(s_j)). \end{aligned} \tag{30}$$

Clearly, Proposition 4.3 and statements (2) and (3) of Proposition 3.2 hold for all products in batch-ordering ATO systems.

5.2. Performance Evaluation

In this section, we develop efficient numerical methods based on Monte Carlo simulation to estimate the key performance measures of the batch-ordering ATO systems. Unfortunately, it is not clear how to apply the method based on the two-moment approximation to these systems.

Observe that the average waiting times, $E(W_j^i)$, are simple functions of the average delivery lead times $E(X^i)$,

$i \in \mathcal{I}$ (Equation (29)). Therefore, we focus on the average delivery lead time (Equation (27)) and the order-based fill-rates (Equation (28)) in the rest of this section.

The performance evaluation of a batch-ordering ATO system poses a substantial challenge because the number of the single-product subsystems associated with each product $i \in \mathcal{I}$ is exponential in the number of the components $j \in \mathcal{J}^i$ (Proposition 5.2). To demonstrate the challenge, consider $E(X^i)$ and $P\{X^i \leqslant \tau\}$ for product $i \in \mathcal{I}$. By Equations (27)–(28), the standard methods based on Monte Carlo simulation (Law and Kelton 1991, §1.8.3) require multiple samples of $X^i(\bar{s}^i)$ for each $\bar{s}^i \in \mathcal{S}^i$. Evidently, there are $\prod_{j\in\mathcal{J}^i} Q_j$ different single-product subsystems (each corresponds to a $\bar{s}^i$) to sample. For instance, if $Q_j = 2\ \forall\, j \in \mathcal{J}^i$, then the number of the single-product subsystems is $2^{|\mathcal{J}^i|}$, where $|\mathcal{J}^i|$ is the cardinality of $\mathcal{J}^i$. Hence, the standard Monte Carlo simulation methods, which need to sample each single-product subsystem multiple times, are inefficient for large-scale batch-ordering ATO systems.

To design methods that do not require computing times exponential in $|\mathcal{J}^i|$, we first rewrite Equations (27) and (28) as follows:

$$E(X^i) = E\left(\frac{1}{\prod_{j\in\mathcal{J}^i} Q_j} \sum_{\bar{s}^i\in\mathcal{S}^i} X^i(\bar{s}^i)\right) = E(\tilde{X}^i), \tag{31}$$

$$P\{X^i \leqslant \tau\} = E\left(\frac{1}{\prod_{j\in\mathcal{J}^i} Q_j} \sum_{\bar{s}^i\in\mathcal{S}^i} 1_{\{X^i(\bar{s}^i)\leqslant\tau\}}\right) = E(\Xi^i(\tau)), \tag{32}$$

where

$$\tilde{X}^i = \frac{1}{\prod_{j\in\mathcal{J}^i} Q_j} \sum_{\bar{s}^i\in\mathcal{S}^i} X^i(\bar{s}^i), \tag{33}$$

$$\Xi^i(\tau) = \frac{1}{\prod_{j\in\mathcal{J}^i} Q_j} \sum_{\bar{s}^i\in\mathcal{S}^i} 1_{\{X^i(\bar{s}^i)\leqslant\tau\}}, \tag{34}$$

and $1_{\{X^i(\bar{s}^i)\leqslant\tau\}}$ is the indicator function of the event $\{X^i(\bar{s}^i) \leqslant \tau\}$. By Equation (25),

$$\Xi^i(\tau) = \frac{1}{\prod_{j\in\mathcal{J}^i} Q_j} \sum_{\bar{s}^i\in\mathcal{S}^i} 1_{\{L_j - T_j(s_j)\leqslant\tau,\ \forall\, j\in\mathcal{J}^i\}}. \tag{35}$$

Let l_j, $t_j(s_j)$, x^i, $x^i(\bar{s}^i)$, $\tilde{x}^i$, and $\xi^i(\tau)$ be a sample (realization) of the random variables, L_j, $T_j(s_j)$, X^i, $X^i(\bar{s}^i)$, $\tilde{X}^i$, and $\Xi^i(\tau)$, respectively. To obtain a sample of $\tilde{X}^i$ and $\Xi^i(\tau)$ for product $i \in \mathcal{I}$, a straightforward method based on Monte Carlo simulation works as follows:

1. Generate the samples l_j and $t_j(s_j)$, $j \in \mathcal{J}^i$, for each single-product subsystem $\bar{s}^i \in \mathcal{S}^i$ independently.

2. Calculate $x^i(\bar{s}^i)$ (according to Equation (25)) and $1_{\{x^i(\bar{s}^i)\leqslant\tau\}}$ for each $\bar{s}^i$.

3. Compute the samples $\tilde{x}^i$ and $\xi^i(\tau)$ by Equations (33) and (34).

Clearly, each of these three steps requires a computing time proportional to $\prod_{j\in\mathcal{J}^i} Q_j$.

In view of Equations (31)–(32), we can generate a common sample of L_j and $T_j(s_j)$ for all single-product subsystems by the following method.

Method of Common Sample

Step 1. We first generate a sample for each L_j, $j \in \mathcal{J}$, independently.

Step 2. For each product $i \in \mathcal{I}$, we sample $\max_{j\in\mathcal{J}^i}\{r_j + Q_j\}$ many interarrival times independently.

Step 3. For each component $j \in \mathcal{J}$, we superimpose the demand arrival processes of products $i \in \mathcal{I}_j$.

Step 4. Finally, we determine the time, $t_j(s_j)$, at which the s_jth demand arrives for each $s_j \in \mathcal{S}_j$ and each $j \in \mathcal{J}$.

The common samples l_j and $t_j(s_j)$ can substantially reduce the computational complexities in Step 1 of the straightforward method (a complexity analysis is provided later). To reduce the computational complexities in Steps 2 and 3 of the method, we make the following observation: Given a product $i \in \mathcal{I}$ and a common sample l_j and $t_j(s_j)$ $\forall\, s_j \in \mathcal{S}_j$, and $\forall\, j \in \mathcal{J}$, different single-product subsystems may have the same delivery lead-time realization. This is true because $x^i(\bar{s}^i)$ is determined by the maximum of $(l_j - t_j(s_j))^+$ over $j \in \mathcal{J}^i$, and the maximum may be shared by other single-product subsystems in a common sample.

However, this observation alone does not guarantee efficient numerical methods. It is the special structure of the problem that allows us to efficiently identify the subsets of $\mathcal{S}^i$ such that $x^i(\bar{s}^i)$ are identical for all $\bar{s}^i$ in each of these subsets. To see this, we sort $(l_j - t_j(s_j))^+$ for all $s_j \in \mathcal{S}_j$ and $j \in \mathcal{J}^i$ into a nonincreasing sequence δ^i_n, $n = 1, 2, \ldots, \sum_{j\in\mathcal{J}^i} Q_j$. That is, $\delta^i_1 = \max_{s_j\in\mathcal{S}_j,\, j\in\mathcal{J}^i}\{(l_j - t_j(s_j))^+\}$ and $\delta^i_{\sum_{j\in\mathcal{J}^i} Q_j} = \min_{s_j\in\mathcal{S}_j,\, j\in\mathcal{J}^i}\{(l_j - t_j(s_j))^+\}$. Instead of enumerating on $\bar{s}^i$ over $\mathcal{S}^i$ as in Step 2 of the straightforward method, we enumerate on δ^i_n and identify the single-product subsystems with $\bar{s}^i \in \mathcal{S}^i$ for which $x^i(\bar{s}^i) = \delta^i_n$.

To highlight the idea, we consider the following simple example with three components. We focus on product 1 and assume, without loss of generality, that $\mathcal{J}^1$ includes all three components. For a common sample of l_js and $t_j(s_j)$s, Table 1 lists the realization, $(l_j - t_j(s_j))^+$, for each component $j \in \mathcal{J}^1$ from $s_j = r_j + 1$ up to $s_j = r_j + Q_j$.

Obviously, for each $j \in \mathcal{J}$, $(l_j - t_j(r_j + 1))^+ \geqslant (l_j - t_j(r_j + 2))^+ \geqslant \cdots \geqslant (l_j - t_j(r_j + Q_j))^+$. Without loss of generality, assume that in a particular sample, $\delta^1_1 = (l_1 - t_1(r_1 + 1))^+$. Because δ^i_n is a nonincreasing sequence, it follows from Equation (25) that for each $\bar{s}^1 = (s_1, s_2, s_3)$ with $s_1 = r_1 + 1$, the corresponding subsystem has a delivery lead-time realization $x^1(\bar{s}^1) = \delta^1_1$. Therefore, in this sample, the number of subsystems with the delivery lead-time realization, δ^1_1, is $Q_2 \times Q_3$.

Table 1. $(l_j - t_j(s_j))^+$ by components.

Component 1	Component 2	Component 3
$(l_1 - t_1(r_1 + 1))^+$	$(l_2 - t_2(r_2 + 1))^+$	$(l_3 - t_3(r_3 + 1))^+$
$(l_1 - t_1(r_1 + 1))^+$	$(l_2 - t_2(r_2 + 1))^+$	$(l_3 - t_3(r_3 + 1))^+$
$(l_1 - t_1(r_1 + 2))^+$	$(l_2 - t_2(r_2 + 2))^+$	$(l_3 - t_3(r_3 + 2))^+$
...	...	...

Now we delete δ_1^1 from the sequence (i.e., delete $(l_1 - t_1(r_1+1))^+$ from Table 1) and consider δ_2^1 for the following three possible cases:

Case 1. $\delta_2^1 = (l_1 - t_1(r_1+2))^+$,

Case 2. $\delta_2^1 = (l_2 - t_2(r_2+1))^+$, and

Case 3. $\delta_2^1 = (l_3 - t_3(r_3+1))^+$.

In Case 1, for all $\bar{s}^1$ with $s_1 = r_1 + 2$, the delivery lead times $X^1(\bar{s}^1)$ have the same realization: δ_2^1. Thus, the number of subsystems with the delivery lead-time realization, δ_2^1, is also $Q_2 \times Q_3$.

In Case 2, for all $\bar{s}^1$ with $s_2 = r_2 + 1$, the delivery lead times $X^1(\bar{s}^1)$ have the same realization: δ_2^1. Note that we delete $(l_1 - t_1(r_1+1))^+$ from Table 1; thus, the number of subsystems with the delivery lead-time realization, δ_2^1, is $(Q_1 - 1) \times Q_3$.

Case 3 is similar to Case 2, i.e., for all $\bar{s}^1$ with $s_3 = r_3 + 1$, the delivery lead times $X^1(\bar{s}^1)$ have the same realization, δ_2^1. Note that we delete $(l_1 - t_1(r_1+1))^+$ from Table 1; thus, the number of subsystems with the delivery lead-time realization, δ_2^1, is $(Q_1 - 1) \times Q_2$.

We continue with this procedure until either $\delta_n^1 = 0$ or we have deleted all $(l_j - t_j(s_j))^+$ for a particular $j \in \mathcal{J}_1$ from Table 1.

Consider a batch-ordering ATO system and a product $i \in \mathcal{I}$. Given a common sample of l_j and $t_j(s_j)$ $\forall s_j \in \mathcal{S}_j$ and $\forall j \in \mathcal{J}^i$, we can generalize the above procedure into the following method to obtain a sample of $\tilde{X}^i$.

Method A

Step 1. Compute $(l_j - t_j(s_j))^+$ for each $s_j \in \mathcal{S}_j$ and $j \in \mathcal{J}^i$.

Step 2. Sort $(l_j - t_j(s_j))^+$ $\forall s_j \in \mathcal{S}_j$ and $\forall j \in \mathcal{J}^i$ into a nonincreasing sequence that is denoted by δ_n^i, $n = 1, 2, \ldots, \sum_{j\in\mathcal{J}^i} Q_j$. If there is a tie across different components, order the numbers in the tie arbitrarily; if there is a tie for the same component across different base-stock levels, order the numbers in the tie in the nondecreasing sequence of the corresponding base-stock levels.

Step 3. Set $n = 1$, $\tilde{x}^i = 0$, and $R_j = Q_j$ $\forall j \in \mathcal{J}^i$.

Step 4. If $\delta_n^i = 0$, output $\tilde{x}^i$ and stop. Otherwise, first identify the component associated with δ_n^i, namely, k_n. Then, add all $x^i(\bar{s}^i)$, which equals δ_n^i by

$$\tilde{x}^i \leftarrow \tilde{x}^i + \delta_n^i \times \prod_{j\neq k_n,\, j\in\mathcal{J}^i} R_j \Big/ \prod_{j\in\mathcal{J}^i} Q_j. \tag{36}$$

Finally, delete δ_n^i (i.e., the corresponding $(l_{k_n} - t_{k_n}(s_{k_n}))$) by

$$R_{k_n} \leftarrow R_{k_n} - 1. \tag{37}$$

If $R_{k_n} = 0$, output $\tilde{x}^i$ and stop. Otherwise, $n \leftarrow n + 1$ and repeat this step.

To obtain multiple samples for $\tilde{X}^i$, one needs to independently generate multiple common samples of l_js and $t_j(s_j)$s and apply Method A to each one. It is easily seen that for different common samples of l_j and $t_j(s_j)$, one may have different subsets of the single-product subsystems with the same delivery lead-time realizations.

Proposition 5.3. *For a product $i \in \mathcal{I}$, given a common sample l_j and $t_j(s_j)$ $\forall s_j \in \mathcal{S}_j$ and $\forall j \in \mathcal{J}^i$ (generated by the Method of Common Sample), let $x^i(\bar{s}^i)$ be the sample of $X^i(\bar{s}^i)$ for $\bar{s}^i \in \mathcal{S}^i$. Method A has the following properties*:

1. *It always stops before the index n reaches $\sum_{j\in\mathcal{J}^i} Q_j$, and*

2. *The output $\tilde{x}^i$ satisfies*

$$\tilde{x}^i = \frac{1}{\prod_{j\in\mathcal{J}^i} Q_j} \sum_{\bar{s}^i\in\mathcal{S}^i} x^i(\bar{s}^i). \tag{38}$$

Proof. See Appendix 3 for the details. □

We now analyze the computational complexity of Method A. For a given sample size, the computing time for sampling the lead times and the interarrival times, and for superimposing the demand processes is at most proportional to $|\mathcal{J}| + |\mathcal{I}| \times \max_{j\in\mathcal{J}}\{r_j + Q_j\} + |\mathcal{I}| \times \sum_{j\in\mathcal{J}}(r_j + Q_j)$. Computing $(l_j - t_j(s_j))^+$ $\forall s_j$ $\forall j \in \mathcal{J}^i$ and sorting these numbers require a computing time proportional to $\sum_{j\in\mathcal{J}^i}(r_j + Q_j)$ and $(\sum_{j\in\mathcal{J}^i} Q_j)\log(\sum_{j\in\mathcal{J}^i} Q_j)$, respectively; Step 4 of Method A takes a computing time at most proportional to $\sum_{j\in\mathcal{J}} Q_j$. Therefore, the overall computing time for generating one sample of $\tilde{X}^i$ for all $i \in \mathcal{I}$ is at most proportional to $|\mathcal{I}| \times \sum_{j\in\mathcal{J}}(r_j + Q_j) + |\mathcal{I}| \times (\sum_{j\in\mathcal{J}} Q_j)\log(\sum_{j\in\mathcal{J}} Q_j)$.

To develop an upper bound for the variance of a random sample of $\tilde{X}^i$ generated by Method A, we first note that the random samples of the delivery lead times of different single-product subsystems are strongly correlated due to the common random sample of L_j and $T_j(s_j)$. However, they are not completely correlated in general due to the different base-stock levels. Hence, it follows from Equation (33) that the variance of a random sample of $\tilde{X}^i$ generated by Method A is bounded from above by the maximum variance of $X^i(\bar{s}^i)$ over $\bar{s}^i \in \mathcal{S}^i$.

We next consider $\Xi^i(\tau)$, $i \in \mathcal{I}$. Given a common sample of l_j and $t_j(s_j)$ $\forall s_j \in \mathcal{S}_j$ and $\forall j \in \mathcal{J}^i$, we design the following method to obtain a sample of $\Xi^i(\tau)$.

Method B

Step 1. For each component $j \in \mathcal{J}^i$, count the number of $s_j \in \mathcal{S}_j$ such that $l_j - t_j(s_j) \leqslant \tau$, and denote it by Q_j',

Step 2. Calculate $\xi^i(\tau) = \prod_{j\in\mathcal{J}^i} Q_j' / \prod_{j\in\mathcal{J}^i} Q_j$.

This method is based on the fact that $1_{\{x^i(\bar{s}^i)\leqslant\tau\}} = 1$ if and only if $l_j - t_j(s_j) \leqslant \tau$ for all $j \in \mathcal{J}^i$ (by Equations (34)–(35)). The following proposition is straightforward; thus, it does not need a proof.

Proposition 5.4. *For a product $i \in \mathcal{I}$, given a common sample l_j and $t_j(s_j)$ $\forall s_j \in \mathcal{S}_j$ and $\forall j \in \mathcal{J}^i$ (generated by the Method of Common Sample), let $x^i(\bar{s}^i)$ be the sample of $X^i(\bar{s}^i)$ for $\bar{s}^i \in \mathcal{S}^i$. The $\xi^i(\tau)$ generated by Method B satisfies*

$$\xi^i(\tau) = \frac{1}{\prod_{j\in\mathcal{J}^i} Q_j} \sum_{\bar{s}^i\in\mathcal{S}^i} 1_{\{x^i(\bar{s}^i)\leqslant\tau\}}. \tag{39}$$

For a given sample size, the computational complexity of Method B for generating a sample of $\Xi^i(\tau)$ for all $i \in \mathcal{I}$ is at most proportional to $|\mathcal{I}| \times \sum_{j\in\mathcal{J}} Q_j + |\mathcal{I}| \times |\mathcal{J}|$. Applying the same logic as for Method A, it follows from Equation (34) that the variance of a random sample of $\Xi^i(\tau)$ generated by Method B is bounded from above by the maximum variance of the indicator functions $1_{\{X^i(\bar{s}^i)\leqslant\tau\}}$ over $\bar{s}^i \in \mathcal{S}^i$.

6. Discussion

Several assumptions of this paper can be relaxed. Given an ATO system, we apply the *backward method* (§3.1) by asking the following key question: Suppose that a demand for a product arrives at time t; when did the system place orders for the corresponding components used to satisfy this demand? Sections 3, 4, and 5 provide answers to this question under the assumptions of Poisson demand processes, the FCFS allocation rule, stochastic sequential lead times, and BOM matrices with zero or one element. Below, we relax some of these assumptions and identify the limit of our approach. Our focus is on the base-stock systems unless otherwise mentioned.

Non-Poisson Demand Processes. The *backward method* can be applied to a base-stock ATO system facing any point unit demand process with either time-varying or temporally correlated arrivals. This is true because answering the key question specified above for each component requires counting backward from time t all demand arrivals of products that require the component until the total number of arrivals reaches the base-stock level of that component.

However, if demand follows arbitrary point processes, the analysis of the base-stock systems becomes more difficult because the backward counting process may depend on the demand arrival. Furthermore, the inventory position vector of the batch-ordering ATO systems may not be uniformly distributed in $\mathcal{S}$, and demand arrivals may not see the system state in time averages.

BOM Matrix with Nonnegative Integer Elements. A similar analysis can be applied to the base-stock ATO systems in which the elements of the BOM matrix, a_j^i, can choose any nonnegative integers. Consider a product $i \in \mathcal{I}$ and its component set $\mathcal{J}^i$, where $a_j^i \geqslant 1$ for $j \in \mathcal{J}^i$. Suppose that a demand of product i arrives at time t; we apply the *backward method* to identify the time at which the corresponding order of a component j is placed that completely satisfies this demand. Let V_k, $k = 1, 2, \ldots$, be the interarrival times of demands faced by the component j if we count backward starting at time t, e.g., V_1 is the most recent interarrival time prior to t, and so on. Denote $T_j(k) = V_1 + V_2 + \cdots + V_k$, and define D_k to be the demand realized for the component j at time $t - T_j(k)$. Clearly, D_k, $k = 1, 2, \ldots$, is a sequence of i.i.d. random variables with probability density function $P\{D_k = a_j^l\} = \lambda^l / \sum_{m\in\mathcal{I}_j} \lambda^m$ for all $l \in \mathcal{I}_j$ (due to the independent Poisson demand processes). It is easily seen that:

- If $a_j^i > s_j$, the corresponding order is placed at time t.
- If $a_j^i \leqslant s_j$, but $a_j^i + D_1 > s_j$, the corresponding order is placed at time $t - T_j(1)$.
- In general, if $a_j^i + D_1 + \cdots + D_{k-1} \leqslant s_j$, but $a_j^i + D_1 + \cdots + D_k > s_j$, the corresponding order is placed at time $t - T_j(k)$, where $k = 1, 2, \ldots, s_j$. In other words, $T_j(k)$ is determined by starting at time t, and counting backward the demand arrivals of all products that require component j until the accumulated demand exceeds $s_j - a_j^i$.

Define $\{N(n), n \geqslant 0\}$ to be a renewal process generated by $\{D_k, k > 0\}$. Then, if $a_j^i \leqslant s_j$, the corresponding order is placed at time $t - T_j(k)$ with probability $P\{N(s_j - a_j^i) = k - 1\}$. Clearly, both the time process, $T_j(k)$, and the volume process, $N(n)$, are dependent across components $j \in \mathcal{J}^i$. Although the dependence structure is much more difficult to characterize than that of §4, the numerical method based on Monte Carlo simulation can be extended to handle this case.

Capacitated Component Production. An ATO system with capacitated component production is different from the analogous ATO system with stochastic sequential lead times because the component delays are dependent in different ways. To highlight the differences and develop insights into their impact, let us reconsider the two components one product example in §3, and assume that the component production processes are capacitated and sequential (as in Song et al. 1999). Without loss of generality, let $s_k \geqslant s_j$. Suppose that a demand arrives at time t; then, the corresponding orders of the components j and k that meet this demand are placed at time $t - T(s_j)$ and $t - T(s_k)$, respectively. By Equation (3), the delivery lead time satisfies

$$\begin{aligned} X &= \max\{(L_j - T(s_j))^+, (L_k - T(s_k))^+\} \\ &= \max\{(L_j - T(s_j))^+, (L_k - T(s_k - s_j) - T(s_j))^+\}, \end{aligned}$$

where L_j (L_k) depends on the demand arrivals prior to $t - T(s_j)$ ($t - T(s_k)$, respectively). See Figure 1.

The dependence structure of an ATO system with capacitated component production differs from that of the analogous system with stochastic sequential lead times in two ways: First, although L_j (or L_k) is independent of $T(s_j)$ (or $T(s_k)$, respectively), the lead times L_j and L_k become positively correlated due to the common demand process. Second, if $s_k \neq s_j$, then L_j becomes negatively correlated with $T(s_k)$. This is true because L_j is negatively correlated with $T(s_k - s_j)$, which is part of $T(s_k)$. We refer the reader to Figure 1 for a visual aid.

The positive dependence among the lead times has an impact on the delivery lead times.

PROPOSITION 6.1. *Consider an ATO system. If* $\{L_j, j \in \mathcal{J}\}$ *is associated, and if* $\{L_j, j \in \mathcal{J}\}$ *is independent of* $\{T_j^i(s_j), j \in \mathcal{J}^i\}$ *for all* $i \in \mathcal{I}$, *then*

$$P\{L_j - T_j^i(s_j) \leqslant \tau_j,\ j \in \mathcal{J}^i\} \geqslant P\{L_j' - T_j^i(s_j) \leqslant \tau_j,\ j \in \mathcal{J}^i\}$$
$$\text{for all } \{\tau_j, j \in \mathcal{J}^i\} \text{ and } i \in \mathcal{I}, \quad (40)$$

where L'_j has the same marginal distribution as L_j for each $j \in \mathcal{J}$, and $\{L'_j, j \in \mathcal{J}\}$ is a set of independent random variables.

In other words, given that the component lead times are sequential and independent of $\{T_j^i(s_j), j \in \mathcal{J}^i\}$ for all $i \in \mathcal{I}$, the delivery lead time of each product in an ATO system with associate lead times (across components) is stochastically smaller than the delivery lead time of the same product in an analogous system with independent lead times (across components).

Proof. See Appendix 4 for the details. □

To give an example of the associated random lead times, let us consider a queueing system with multiple parallel servers facing a common demand stream. Each server corresponds to a capacitated component supplier in the ATO systems, and the waiting times at these servers correspond to the component lead times. The waiting times at different servers are associated if the service times of different servers are associated and the demand interarrival times are independent. This is true because the waiting times are nondecreasing functions of the service times and nonincreasing functions of the interarrival times; see the Lindley equation of a single-server queue (Equations (8.4)–(8.8) in Kleinrock 1975). If the service times are independent of the demand interarrival times, then Theorem 5.2.3 of Tong (1980) immediately implies the association of these random variables. We refer to Xu (1999) for a discussion of the dependence among the waiting times in a multiple-server queueing system where the servers face partially common demand streams. A different example is a single-server queue with associated service times. The waiting times of consecutive items at this server are associated if the demand interarrival times are independent, again because the waiting times are nondecreasing functions of the service times and nonincreasing functions of the interarrival times.

The proposition has an important implication in practice. As one of the referees pointed out, in this age of security concerns, concentration of component production in low-wage regions, and partnership with third-party logistics providers, the production cycle times as well as transportation lead times of various components may be subject to the impact of the same random events, and therefore may be highly positively dependent (e.g., associated). Proposition 6.1 shows that ignoring these dependences may result in overestimating (stochastically) the delivery lead times to end customers.

The cross dependence between the lead times and $T(s_j)$ also has an impact on the delivery lead times to the end customers.

Proposition 6.2. *Consider a single-product assembly system with capacitated component production. Suppose that a demand arrives at time t; let $\overline{V} = (V_1, V_2, \ldots)$ be the vector of the interarrival times that we index backward starting at t. Replace L_j by $L_j(\overline{V})$ and $T(s_j)$ by $T(s_j, \overline{V})$ to show their dependence on $\overline{V}$. Then, the probability distribution of the delivery lead time of this demand satisfies*

$$P\{L_j(\overline{V}) - T(s_j, \overline{V}) \leqslant \tau_j,\ j \in \mathcal{J}\} \geqslant P\{L_j(\overline{V}') - T(s_j, \overline{V}) \leqslant \tau_j,\ j \in \mathcal{J}\} \quad \text{for all } \tau_j,\ j \in \mathcal{J}, \quad (41)$$

where $\overline{V}'$ is an independent copy of $\overline{V}$.

Proof. See Appendix 4 for the details. □

Proposition 6.2 implies that in addition to the positively dependent lead times, the cross dependence between the lead times and the $T(s_j)$s could further reduce the delivery lead time. We conjecture that Proposition 6.2 holds true for more general multiproduct ATO systems. However, due to the complexity of the demand processes, this remains an open problem.

The dependences among the lead times can be incorporated into our analysis without much difficulty (see §3). However, the cross dependence between L_j and $T(s_k)$ significantly complicates the analysis. We refer the reader to Song et al. (1999) for an exact method for small-size capacitated ATO systems, and Dayanik et al. (2003) for approximations and bounds for large-scale capacitated ATO systems.

I.i.d. Lead Times. It is also interesting to compare an ATO system with stochastic sequential lead times to an ATO system with i.i.d. lead times. For simplicity, let us consider the single-product assembly systems. Given a component $j \in \mathcal{J}$ and assuming that a demand arrives at time t, we know from §3 that in the system with stochastic sequential lead times, the corresponding order for this demand is placed at time $t - T(s_j)$. Now, in the system with i.i.d. lead times, the corresponding order for this demand is still placed at one of the demand arrival times, $t - T(K_j)$. However, K_j is an integer-valued random variable which can be either positive, zero, or negative. To see this, let us further assume that at time t, the system has zero on-hand inventory for component j. Then, the corresponding order that satisfies this demand can be placed at any demand arrival before, on, or after time t. In addition, the distribution of K_j depends on s_j, as well as all previous and future demand arrivals if we assume that the replenishment lead times have unbounded domains. It is not clear how to evaluate the performances of ATO systems with i.i.d. lead times using our approach. Indeed, the approaches based on the outstanding orders, e.g., Song and Yao (2002) and Lu et al. (2003), may be more effective.

Priority Rules. Because different products may have different backorder penalty costs, it is desirable to allocate inventory to products according to some priority rules other than the FCFS rule. For periodic review ATO systems, we refer the reader to Zhang (1997) for the study of the *fixed priority* rule, Agrawal and Cohen (2001) for

the *fair share* rule, and Akcay and Xu (2004) for the *OBCA* rule. While the aforementioned papers assume different allocation rules for demand realized in the same time period, they all assume the FCFS rule to process demand realized at different time periods. Exceptions to the FCFS rule are Plambeck and Ward (2003) and Plambeck (2005), which study the dynamic order-sequencing problem in ATO systems with capacitated component production. Plambeck and Ward (2003) proves that if orders can be expedited instantaneously at higher costs, then the control problem of a multicomponent ATO system separates into an independent control problem of each component. Plambeck (2005) further shows that a simple policy with independent control of each component is asymptotically optimal in a high-volume ATO system with fixed transportation cost and nonzero transportation lead times.

To see if the approach proposed in this paper can be extended to handle the priority rules, let us consider a simple single-item inventory system with two-customer classes where the inventory is reviewed continuously in time and is controlled by a (Q, r, K) rationing policy. We assume that demands follow Poisson processes. A (Q, r, K) policy works as follows: When the inventory position (on-hand plus on-order minus backorders) drops to r, then order Q units. Demand from both classes is filled on an FCFS basis, as long as the on-hand inventory level is greater than or equal to K. Once the on-hand inventory level falls below K, class 2 (with lower penalty cost) demand is backordered while class 1 (with higher penalty cost) demand continues to be filled as long as inventory is available. We refer the reader to Deshpande et al. (2003) for a recent review. For simplicity, let $Q = 1$, which corresponds to a base-stock rationing policy. Supposing that a class 1 demand arrives at time t, we count backward starting at t and define $T^1(r+1)$ to be the sum of $r+1$ interarrival times of the demand process of class 1, and $T^{1\oplus 2}(r+1)$ to be the sum of $r+1$ interarrival times of the superimposed demand process of both class 1 and class 2. Clearly, the corresponding order that satisfies this class 1 demand is placed at one of the demand arrival times, T', where $t - T^1(r+1) \leqslant T' \leqslant t - T^{1\oplus 2}(r+1)$. Furthermore, T' is a random variable whose distribution depends on the arrivals and replenishment processes of all demand realized before time t. Similar logic can be applied to class 2 demand, i.e., suppose that a class 2 demand arrives at time t; then the corresponding order for this demand is placed at time T'' with $T'' \geqslant t - T^{1\oplus 2}(r+1)$. However, it is not yet known how to characterize the distribution of T' and T'' exactly using the approach developed in this paper.

Implication for Optimization. Lu and Song (2005) develop an efficient algorithm based on the submodularity property of the objective function to optimize the base-stock levels. Our numerical methods can serve as performance evaluation subroutines for the optimization algorithm. For problems with relatively small size, e.g., $|\mathcal{J}| \leqslant 20$, the method based on Clark's approximation provides reasonably accurate expected delivery lead times, and thus the expected backorders. For large-size problems, the simulation-based method is more appropriate. However, more research needs to be conducted to handle the random errors generated by the simulation-based method.

Appendices 1 and 2

Appendices 1 and 2 can be found at the INFORMS home page in the *Operations Research* online collection at http://or.pubs.informs.org/Pages/collect.html.

Appendix 3

Proof of Proposition 5.2. Our starting point is Theorem 1 of Song (2000), which states that the inventory position vector of the components has a uniform equilibrium distribution in $\mathcal{S}$. The theorem holds for ATO systems with stochastic sequential lead times provided that Assumption 1 of Song (2000) is satisfied. This is true because the inventory position is a function of only the demand process and the inventory control policy, but not the lead time.

Because we assume that demand follows independent Poisson processes, any demand arrival sees the inventory position vector in the uniform equilibrium distribution (PASTA, see, e.g., Heyman and Sobel 1982). Thus, at the time of a demand arrival for product $i \in \mathcal{I}$, the inventory position vector of the required components $j \in \mathcal{J}^i$ is uniformly distributed in $\mathcal{S}^i$.

Let us now focus on a particular product $i \in \mathcal{I}$, and consider a component $j \in \mathcal{J}^i$. Suppose that an order of size Q_j is triggered by a demand at time t_0; we make the following two observations: (1) This order will be used to satisfy the $r_j + 1$st, $r_j + 2$nd, up to $r_j + Q_j$th demand that arrives after t_0. Consequently, relative to the arrival time of the $r_j + q_j$th demand, which occurs at $t_0 + T_j(r_j + q_j)$ ($T_j(\cdot)$ is defined in §4.1), the corresponding order of the component j is placed $T_j(r_j + q_j)$ time earlier. (2) The inventory position right before the arrival of the $r_j + q_j$th demand (after t_0) is exactly $r_j + Q_j - [(r_j + q_j - 1) \bmod Q_j]$, which has a one-to-one relationship with $r_j + q_j$ for $q_j \in \{1, 2, \ldots, Q_j\}$.

It follows from the first observation that if a demand of product i arrives at time t, the corresponding order of component j that satisfies this demand must be placed at one of these times: $t - T_j(r_j + q_j)$ where $r_j + q_j \in \mathcal{S}_j$. By the second observation, the uniform equilibrium distribution of the inventory position vector of components $j \in \mathcal{J}^i$ in $\mathcal{S}^i$ implies a uniform equilibrium distribution of the random vector $(S_j,\ j \in \mathcal{J}^i)$ in $\mathcal{S}^i$. The proof is now complete. □

Proof of Proposition 5.3. 1. The proof is by contradiction. Suppose that $n = \sum_{j\in\mathcal{J}^i} Q_j$. Then, R_j must be zero for all components $j \in \mathcal{J}^i$ except one, because Method A subtracts one of the R_j, $j \in \mathcal{J}^i$, by one in each iteration. Therefore, R_j for some $j \in \mathcal{J}^i$ must reach zero before n reaches $\sum_{j\in\mathcal{J}^i} Q_j$. This creates a contradiction.

2. We only need to show that given a common sample, l_j and $t_j(s_j)$ $\forall s_j \in s_j$ $\forall j \in \mathcal{J}$, Method A adds $x^i(\bar{s}^i) = \max_{j\in\mathcal{J}^i}\{(l_j - t_j(s_j))^+\}$ exactly once for each $\bar{s}^i \in \mathcal{S}^i$. Let us first assume that $\delta^i_n > 0$ for all $n = 1, \ldots, \sum_{j\in\mathcal{J}^i} Q_j$.

For each $\bar{s}^i \in \mathcal{S}^i$, we enumerate $(l_k - t_k(s_k))^+$ over $k \in \mathcal{J}^i$ to identify the one, denoted by $(l_j - t_j(s_j))^+$, with the smallest index n in the sequence of δ^i_n, and that satisfies

$$(l_j - t_j(s_j))^+ = \max_{k\in\mathcal{J}^i}\{(l_k - t_k(s_k))^+\}. \tag{42}$$

We denote this index by $n_{\bar{s}^i}$.

First, $x^i(\bar{s}^i)$ is never added before Method A processes $\delta^i_{n_{\bar{s}^i}}$. This is true because $n_{\bar{s}^i}$ is the smallest index of $(l_k - t_k(s_k))^+$ over $k \in \mathcal{J}^i$; thus, for each $n < n_{\bar{s}^i}$, the corresponding single-product subsystems added must have different base-stock level vectors than $\bar{s}^i$.

Second, the $\delta^i_{n_{\bar{s}^i}}$ will be processed before Method A stops. To see this, we note that when $n = n_{\bar{s}^i}$, $R_k \geqslant 1$ for all components $k \in \mathcal{J}^i$. This is true because Equation (42) and the definition of $n_{\bar{s}^i}$ imply that for each component $k \in \mathcal{J}^i$ but $k \neq j$, $(L_k - T_k(s_k))^+$ has not yet been deleted. When $\delta^i_{n_{\bar{s}^i}}$ is processed by Equation (36), $x^i(\bar{s}^i)$ is added exactly once.

Finally, $x^i(\bar{s}^i)$ is never added after Method A processes $\delta^i_{n_{\bar{s}^i}}$ because $(l_j - t_j(s_j))^+$ is deleted by Equation (37).

If $\delta^i_{n'} = 0$ for $n' < \sum_{j\in\mathcal{J}^i} Q_j$, then $\delta^i_n = 0$ for all $n' \leqslant n$ because the sequence of δ^i_n is nonincreasing. Method A stops because all the single-product subsystems with a delivery lead time equal to δ^i_n, $n' \leqslant n$, do not contribute to $\tilde{x}^i$. □

Appendix 4

Proof of Proposition 4.3. For a given product $i \in \mathcal{I}$, suppose that a demand of this product arrives at time t. For each product k, $k \in \mathcal{I}$, we count backward, starting at t, all the previous interarrival times of demand for this product. Define $\mathcal{V}^k$ to be the set of the most recent $\max\{s_j,\ j \in \mathcal{J}\}$ interarrival times of product k, $k \in \mathcal{I}$. We further define $\mathcal{V}$ to be the union of $\mathcal{V}^k$ for all $k \in \mathcal{I}$. Note that each $\mathcal{V}^k$ is a set of independent random variables and $\mathcal{V}^k$, $k \in \mathcal{I}$, are mutually independent. It follows from parts (3) and (2) of Lemma 4.2 that $\mathcal{V}$ is a set of associated random variables.

Clearly, $T^i_j(s_j)$ for each $j \in \mathcal{J}^i$ is a nondecreasing function in each element of $\mathcal{V}$. Therefore, part (4) of Lemma 4.2 implies that $\{T^i_j(s_j),\ j \in \mathcal{J}^i\}$ is associated for each $i \in \mathcal{I}$. The desirable result now follows immediately from part (5) of Lemma 4.2. □

Proof of Proposition 6.1. For each $i \in \mathcal{I}$, we condition on $T^i_j(s_j) = t^i_j$, $j \in \mathcal{J}^i$. Because $\{L_j,\ j \in \mathcal{J}\}$ is independent of $\{T^i_j(s_j),\ j \in \mathcal{J}^i\}$, we must have

$$\begin{aligned} P\{L_j - t^i_j \leqslant \tau_j,\ j \in \mathcal{J}^i\} &\geqslant \prod_{j\in\mathcal{J}^i} P\{L_j \leqslant \tau_j + t^i_j\} \\ &= \prod_{j\in\mathcal{J}^i} P\{L'_j \leqslant \tau_j + t^i_j\} \\ &= P\{L'_j - t^i_j \leqslant \tau_j,\ j \in \mathcal{J}^i\} \\ &\quad \forall (t^i_j,\ j \in \mathcal{J}^i) \in R^{|\mathcal{J}^i|}, \end{aligned} \tag{43}$$

where the inequality follows by the fact that $\{L_j,\ j \in \mathcal{J}\}$ are associated, and by parts (1) and (5) of Lemma 4.2, the first equality is due to the same marginal distribution of L'_j and L_j, $j \in \mathcal{J}$, and the second equality is due to the fact that L'_j, $j \in \mathcal{J}$, are independent random variables. Unconditioning on $T^i_j(s_j) = t^i_j$ yields the requisite result. □

Proof of Proposition 6.2. Note that for each $j \in \mathcal{J}$, $L_j(\overline{V})$ and $-T(s_j, \overline{V})$ are nonincreasing functions in each element of $\overline{V}$. Furthermore, $T(s_j, \overline{V})$ only depends on $\{V_1, V_2, \ldots, V_{s_j}\}$, while $L_j(\overline{V})$ only depends on $\{V_{s_j+1}, V_{s_j+2}, \ldots\}$. The requisite result now follows immediately from Theorem 4.1 of Goldstein and Rinott (2004) and its Application 1. □

Acknowledgments

The authors thank the associate editor and two referees for their excellent comments and constructive suggestions, which significantly improved the paper. Indeed, §5 and many of the discussions in §6 are motivated by the questions raised by the referees. The hypothetical example was suggested by the associate editor. This research was supported in part by the Singapore-MIT Alliance, ONR contracts N00014-95-1-0232 and N00014-01-1-0146, and by NSF contracts DMI-9732795, DMI-0085683, and DMI-0245352.

References

Agrawal, N., M. A. Cohen. 2001. Optimal material control in an assembly system 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.

Asmussen, S. 2000. *Ruin Probabilities*. World Scientific Publishing Company, Singapore.

Axsater, S. 1993. Continuous review policies for multi-level inventory systems with stochastic demand. S. Graves, A. Rinnooy Kan, P. Zipkin, eds. *Logistics of Production and Inventory*. Elsevier (North-Holland), Amsterdam, The Netherlands.

Clark, C. E. 1961. The greatest of a finite set of random variables. *Oper. Res.* **9** 145–162.

Dayanik, S., J. S. Song, S. H. Xu. 2003. The effectiveness of several performance bounds for capacitated assemble-to-order systems. *Manufacturing Service Oper. Management* **5** 203–251.

De Kok, Ton G. 2003. Ruin probabilities with compounding assets for discrete time finite horizon problems, independent period claim sizes and general premium structure. *Insurance Math. Econom.* **33** 645–658.

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

De Vylder, F., M. J. Goovaerts. 1988. Recursive calculation of survival probabilities. *Insurance: Math. Econom.* **7** 1–7.

Esary, J. D., F. Proschan, D. W. Walkup. 1967. Association of random variables, with applications. *Ann. Math. Statist.* **38** 1466–1474.

Glasserman, P., Y. Wang. 1998. Leadtime-inventory trade-offs in assemble-to-order systems. *Oper. Res.* **46** 858–871.

Goldstein, L., Y. Rinott. 2004. Functional BRK inequalities, and their duals, with applications. Working paper, Department of Mathematics, University of Southern California, Los Angeles, CA.

Hausman, W. H., H. L. Lee, A. X. Zhang. 1998. Joint demand fulfillment probability in a multi-item inventory system with independent order-up-to policies. *Eur. J. Oper. Res.* **109** 646–659.

Heyman, D. P., M. J. Sobel. 1982. *Stochastic Models in Operations Research*, Vol. 1: *Stochastic Processes and Operating Characteristics*. McGraw-Hill, New York.

Kleinrock, L. 1975. *Queueing Systems*, Vol. 1: *Theory*. John Wiley and Sons, New York.

Kulkarni, V. G. 1995. *Modeling and Analysis of Stochastic Systems*. Chapman and Hall, New York.

Kumar, A. 1989. Component inventory costs in an assembly problem with uncertain supplier lead-times. *IIE Trans.* **21** 112–121.

Law, A. M., W. D. Kelton. 1991. *Simulation Modeling and Analysis*, 2nd ed. McGraw-Hill, New York.

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

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

Palm, C. 1938. Analysis of the Erlang traffic formula of busy-signal arrangements. *Ericsson Technics* **5** 39–58.

Plambeck, E. L. 2005. Asymptotically optimal control for an assemble-to-order system with capacitated component production and fixed transport costs. Working paper, Graduate School of Business, Stanford University, Stanford, CA.

Plambeck, E. L., A. R. Ward. 2003. A separation principle for assemble-to-order systems with expediting. Working paper, Graduate School of Business, Stanford University, Stanford, CA.

Shaked, M., J. G. Shanthikumar. 1994. *Stochastic Orders and Their Applications*. Academic Press, New York.

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. 2000. A note on assemble-to-order systems with batch ordering. *Management Sci.* **46** 739–743.

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 in assemble-to-order systems with random lead-times. *Oper. Res.* **50** 889–903.

Song, J. S., P. Zipkin. 2003. Supply chain operations: Assemble-to-order systems. Chapter 11 in *Handbooks in Operations Research and Management Science*, Vol. 11: *Supply Chain Management*. Elsevier, Amsterdam, The Netherlands, 561–596.

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

Svoronos, A., P. Zipkin. 1991. Evaluation of one-for-one replenishment policies for multiechelon inventory systems. *Management Sci.* **37** 68–83.

Tong, Y. L. 1980. *Probability Inequalities in Multivariate Distributions*. Academic Press, New York.

Xu, S. H. 1999. Structural analysis of q queueing system with multiclasses of correlated arrivals and blocking. *Oper. Res.* **47** 264–276.

Xu, S. H. 2001. Dependence analysis of assemble-to-order systems. J. Song, D. Yao, eds. Chapter 11 in *Supply Chain Structures: Coordination, Information and Optimization*. Kluwer Academic Publishers, Norwell, MA.

Zhang, A. X. 1997. Demand fulfillment rates in an assemble-to-order system with multiple products and dependent demands. *Production Oper. Management* **6** 309–323.

Zipkin, P. 2000. *Foundations of Inventory Management*. McGraw Hill, Boston, MA.