OPERATIONS RESEARCH
Vol. 58, No. 1, January–February 2010, pp. 111–129
ISSN 0030-364X | EISSN 1526-5463 | 10 | 5801 | 0111

DOI 10.1287/opre.1090.0706

Inventory Management with an Exogenous Supply Process

Alp Muharremoglu
Graduate School of Business, Columbia University, New York, New York 10027,
alp2101@columbia.edu

Nan Yang
Johnson School of Management, Cornell University, Ithaca, New York 14853,
ny38@cornell.edu

We study single and multistage inventory systems with stochastic lead times. We study a class of stochastic lead time processes, which we refer to as *exogenous* lead times. This class of lead time processes includes as special cases all lead time models from existing literature (such as Kaplan's lead times with no order crossing or independent and identically distributed lead times with order crossing, among others) but is a substantially broader class. For a system with an exogenous lead time process, we provide a method to determine base-stock levels and to compute the cost of a given base-stock policy. The method relies on relating the cost of a base-stock policy to the cost of a threshold policy in a related single-unit, single-customer problem. This single-unit method is exact for single-stage systems and for multistage systems under certain conditions. If the conditions are not satisfied, the method obtains near-optimal base-stock levels and accurate approximations of cost for multistage systems.

Subject classifications: inventory/production: multiechelon, policies, review/lead times.
Area of review: Manufacturing, Service, and Supply Chain Operations.
History: Received April 2007; revisions received August 2008, November 2008; accepted December 2008. Published online in *Articles in Advance* August 17, 2009.

1. Introduction

One of the core challenges in supply chain management is to manage uncertainty. Frequently, the demand that a firm experiences for its goods is uncertain, necessitating the use of safety stocks. Another main driver of uncertainty is randomness in the supply process. Especially, in today's global supply chains, where companies work with suppliers in different continents or operate facilities that are far away, uncertainty in supply is a fact of life. In the inventory management literature, supply uncertainty is discussed in three main forms: yield uncertainty, lead time uncertainty, and capacity uncertainty. These different ways of thinking about supply uncertainty are related to each other; for example, one can think about capacity shortage at a supplier leading to a longer lead time for the delivery of orders. In this paper, we address the issue of supply uncertainty in the form of uncertainty in lead times. We study *exogenous* lead times, which is a broad class of stochastic lead times, that includes sequential and order-crossing lead time processes. We study both single and serial multistage systems.

There are numerous papers in the inventory management literature that deal with the issue of uncertain lead times. An early reference is the book of Hadley and Whitin (1963, Chapter 5.14), who study an inventory model with independent and identically distributed (i.i.d.) lead times. Further assuming that the orders arrive in the same sequence they were placed, i.e., no order crossing, they characterize the long-run behavior of the system. In particular, their analysis boils down to computing the distribution of total demand during one realization of the random i.i.d. lead time, i.e., the lead time demand. As Hadley and Whitin (1963) point out, the two assumptions of lead times being independent and no order crossing can be satisfied at the same time only under specific conditions. For example, if lead times are smaller than the review period, then the analysis is exact. Another example is the case when the smallest possible lead time and the largest possible lead time differ by at most the length of the review period. Under more general conditions, order crossing can take place, and the analysis becomes substantially more intricate. In fact, i.i.d. lead times can give rise to complicated optimal policies that depend on the state of the pipeline inventory vector. Zalkind (1978) gives an example demonstrating the nonoptimality of base-stock policies when lead times are i.i.d. Still, the predominant approach for dealing with lead time uncertainty, both in the literature and in practice, has been to use the method of Hadley and Whitin (1963), i.e., to assume i.i.d. lead times and to use the lead time demand in the analysis. This approach results in a base-stock policy but does not optimize the system even within the class of base-stock policies. The quantity needed for an exact analysis of a base-stock policy is the distribution of pipeline inventory, instead of the lead time demand. In many cases, lead time demand is a good approximation for pipeline inventory, but in many other cases, using such an approximation leads to substantial suboptimality, as demonstrated in §5.

There are some notable exceptions to Hadley and Whitin's (1963) paradigm that fall into two main categories. The first set of papers assume a certain type of nonorder-crossing stochastic lead time process, initially proposed by Kaplan (1970). Kaplan shows the optimality of base-stock policies in a single-stage, periodic review setting under his proposed lead time model. The underlying assumptions are that orders do not cross and that the arrival probability of a given order at a given period depends only on how long that particular order has been outstanding. Nahmias (1979) and Ehrhardt (1984) use a similar lead time process. Zipkin (1986) extends Kaplan's lead time model to include a more general class of nonorder-crossing stochastic lead time processes, which he denotes as "exogenous, sequential lead times" (Zipkin 2000, Chapter 7.4). Svoronos and Zipkin (1991) evaluate one-for-one replenishment policies in serial systems with such lead times. Muharremoglu and Tsitsiklis (2008) show that echelon base-stock policies are optimal for serial systems with Kaplan's nonorder-crossing stochastic lead time processes. All these papers assume that orders do not cross.

The second set of papers that diverge from Hadley and Whitin's (1963) paradigm are Zalkind (1978), Robinson et al. (2001), Bradley and Robinson (2005), and Robinson and Bradley (2008). These papers study single-stage inventory systems and assume that base-stock policies are utilized. The goal is to find the optimal base-stock levels. The lead times are taken to be independently and identically distributed (i.i.d.). The class of i.i.d. lead times allows for order crossing, so these papers were the first to analyze an inventory system with order crossing. The papers characterize the distribution of pipeline inventory, either through a computational method or through approximations. Zalkind (1978) provides an exact computational method to optimize base-stock levels in single-stage inventory systems with i.i.d. lead times. Robinson et al. (2001) give an approximate method to optimize base-stock levels by matching the first two moments of the inventory shortfall distribution. Bradley and Robinson (2005) and Robinson and Bradley (2008) improve upon this approximation with tighter upper bounds for the variance of the pipeline inventory. Song and Zipkin (1996b) study the effect of lead time variance in an (r, q) system with i.i.d. lead times and discuss the relationship of i.i.d. lead times with exogenous sequential lead times. Note that all these papers assume i.i.d. lead times in a single-stage setting.

We analyze a set of lead time processes that is more general than previously studied lead time models. We refer to this class as *exogenous* lead time processes. In fact, to our knowledge, all lead time models from existing literature are in the class of exogenous processes. For example, both Kaplan's lead times with no order crossing and i.i.d. lead times with order crossing are special cases of exogenous lead time processes, as are the Markov modulated lead time models of Song and Zipkin (1996a) and Chen and Yu (2004), among many others. Lead times that can be modeled as an ARMA process are another example. In terms of ordering strategies, we confine ourselves to the class of base-stock policies. The justification for this restriction is threefold: (i) the truly optimal policy can have a prohibitively complex structure that depends on the state of pipeline inventories, and would be very difficult, if not impossible, to implement, (ii) base-stock policies are very commonly used in practice, and (iii) when applied to multistage systems, (echelon) base-stock policies can be implemented in a decentralized fashion as shown by Axsäter and Rosling (1993), who demonstrated the equivalence of local and echelon base-stock policies. We provide a method to determine base-stock levels and to compute the cost of a given base-stock policy for systems with an *exogenous* lead time process. For single-stage systems, the method finds the optimal base-stock levels and provides the exact cost. For multistage problems, the method is exact under certain conditions. If those conditions are not satisfied, the result provides a very good approximation of the cost and finds near-optimal base-stock levels, as we demonstrate through a numerical study.

We use an idea that has been utilized by Hadley and Whitin (1963) and Zalkind (1978). The idea is to assume that orders are being released at every period, and a stochastic lead time is assigned to every order, but sometimes the order size is zero. Of course, this is just a mathematical tool to analyze the system; no extra costs are incurred due to these zero-sized orders. Let $L_j(t)$ denote the lead time of the order released at time t from stage $j+1$ to stage j. The sequence of values $L_j(t)$, for all t, is the "lead time process of stage j." The joint process of all stages constitutes the "lead time process." A lead time process is exogenous if the lead time process of any stage is independent from the lead time processes of all other stages and independent of the demand process. We also assume that the lead time process is ergodic throughout the paper. (For a formal definition of ergodicity, see §2.) The class of *exogenous* lead time processes is an extension of the *exogenous sequential lead time* class that Zipkin (1986) introduced, in that we relax the sequential restriction. The ergodicity assumption is reasonable if the conditions underlying the lead time process remain stable over a relatively long period, as compared to the frequency of the shipment decisions. The assumption of the lead times being exogenous is reasonable if the orders of our firm are a relatively small portion of the total orders in the supply process. This ensures that the order sizes of our firm are not the primary reason of fluctuations in the supply process.

One can model interesting phenomena under the framework of exogenous and ergodic lead times. For example, if a third-party logistics provider adopts a congestion-dependent dispatch policy, where if more than a certain

Table 1. The relative positioning of this paper compared to existing literature.

	Single-stage	Multistage
Kaplan's lead times	Kaplan (1970)	Muharremoglu and Tsitsiklis (2008)
Exogenous and sequential lead times	Zipkin (1986)	Svoronos and Zipkin (1991)
i.i.d. lead times	Zalkind (1978), Robinson et al. (2001)	This paper[1]
Exogenous lead times	This paper	This paper[1]

[1]The multistage analysis is exact in the four cases outlined in item (c) of the list of contributions given in the introduction, approximate otherwise.

number of trucks are on their way between Chicago and New York, they would start sending new orders by a faster mode of transportation, such as by air. They could also communicate with the truck drivers on the road and tell them to speed up. Note that congestion here refers to the overall congestion in the logistic provider's system, not the congestion due to the order sizes of our firm. Another interesting application of our model is when the supply process corresponds to the production process at a contract manufacturer. In this case, the contract manufacturer is working with many customers, including our firm. The contract manufacturer can use congestion-dependent policies, where congestion refers to the overall workload at the manufacturer, of which our firm's orders are a small part.

The contributions of this paper are:

(a) We provide a method to determine base-stock levels and to compute the cost of a given base-stock policy for systems with exogenous lead times. The method is based on relating the original problem to a single-unit, single-customer problem (Proposition 3.3). In contrast to Muharremoglu and Tsitsiklis (2008), the problem addressed in this paper does not decompose into single-unit problems in the sense of separability of optimal controls of the different units. Still, we are able to obtain a correspondence between the overall cost and the cost of a single-unit, single-customer problem, by interpreting an infinite number of unit-customer pairs on a single sample path as an infinite number of sample paths that one unit-customer pair can experience.

(b) For single-stage systems, the base-stock levels obtained by our single-unit method are optimal within the class of base-stock policies (Proposition 3.10). This extends Zipkin (1986), who studies single-stage systems with exogenous and sequential lead times, by removing the sequential lead times condition. It also extends Zalkind (1978), who studies single-stage systems with i.i.d. lead times, by allowing more general stochastic lead time processes.

(c) For multistage systems, we show that the single-unit method is exact in four cases: (i) when orders do not cross (Proposition 3.7 and Corollary 3.8), or (ii) when order crossing is allowed only at the most upstream stage (Proposition 3.11), or (iii) when the difference between base-stock levels of consecutive stages is sufficiently large (Proposition 3.12), or (iv) in two-stage systems with deterministic upstream lead time, when the difference between base-stock levels is zero (Proposition 3.13). This extends Svoronos and Zipkin (1991), which studies multistage systems with exogenous sequential lead times, by removing the sequential lead times condition.

(d) For multistage systems where the conditions above do not apply, the method is approximate. We demonstrate through a numerical study that the single-unit method produces near-optimal base-stock levels and accurate cost estimates (§5).

We summarize the relative positioning of this paper compared to existing literature in Table 1. Note that all of the lead time models listed in Table 1 assume ergodicity. Kaplan's lead times are a special case of exogenous and sequential lead times, and the first three lead time models are special cases of exogenous lead times.

The rest of this paper is organized as follows. Section 2 formulates the problem. Section 3 contains the main results. Section 4 discusses estimation issues. Section 5 provides a numerical study to test the accuracy of the method in cases where the correspondence between the original problem and the single-unit problem is not exact. We conclude the paper with §6.

2. Problem Formulation

We consider an M-stage serial system with stochastic lead times. Stage M is the manufacturer, who is assumed to have ample supply. We use linear ordering, holding, and backorder costs, with rates c_j, h_j (for all stages $j = 1, \ldots, M-1$), and $b > 0$, respectively. Without loss of generality, we assume that the holding cost rate h_j is decreasing in j. The goal is to determine the optimal policy within the class of base-stock policies. Demands in successive periods are i.i.d random variables and demand occurs only at stage 1. Let $\bar{d} < \infty$ be the expected demand per period. The lead time process is *exogenous*. Let $C(\mathbf{s})$ denote the infinite-horizon average cost of base-stock policy $\mathbf{s}$.

In each time period, there are five successive events. First, deliveries for this period are received. Second, demand arrives. Third, holding and backorder costs are charged. Fourth, orders are placed according to the current echelon inventory positions. Fifth, orders are shipped, and ordering costs are charged.

In the rest of the paper, we use the following convention to refer to stochastic processes and their realizations.

Convention 2.1. *Let* $\{X_j(i)\}$, $i = 1, \ldots, \infty$ *and* $j = 1, \ldots, M-1$ *be a stochastic process. We denote vectors using bold* *letters, i.e.,* $\mathbf{X}(i) = (X_1(i), \ldots, X_{M-1}(i))$ *for all i. We refer to the process as stochastic process X. The realization of this process under sample path* ω *is* $\mathbf{x}(i, \omega)$, $i = 1, \ldots, \infty$. *We use the notation* $\mathbf{X}^{ss}$ *to denote the steady-state random vector of the process X.*

In a system with an exogenous lead time process, an order is released in every period t from each stage $j+1$ (possibly of size zero). Let $L_j(t)$ be the lead time for this order. We now define two other processes that are instrumental in our analysis.

Definition 2.2. (a) *The order-based ordered lead time process* $\hat{L}_j(t)$ is the time between the tth order release from stage $j+1$ and the tth order arrival at stage j. Note that these orders do not have to be the same order because when orders cross, orders do not arrive in the same sequence they were released.

(b) *The unit-based ordered lead time process* $\tilde{L}_j(i)$ is defined as the time between ith unit release from stage $j+1$ and the ith unit arrival at stage j. These units do not have to be the same unit because when orders cross, units do not arrive in the same sequence they were released.

Figure 1 illustrates the order-based ordered lead time process, as well as the unit-based ordered lead time process. Part (a) of the figure depicts the lead times of three successive orders placed at $t = 1$, 2, and 3. Note that order 2 has a size of zero. Part (b) of the figure depicts the order-based order lead time process, as defined in Definition 2.2(a). Note that while the first order release occurs in period 1, the first order arrival occurs in period 9. Therefore, the first order-based ordered lead time goes from period 1 to period 9 in Figure 1(b), and thus equals 8. Figure 1(c) depicts the lead times that the actual units experience. Units 1 and 2 are part of order 1, and therefore have the same lead time as order 1. Unit 3 is part of order 3, and has the same lead time as order 3. Figure 1(d) depicts the unit-based ordered lead time process, as defined in Definition 2.2(b). For example, the third unit release occurs in period 3 as part of order 3, whereas the third unit arrival occurs in period 11 as part of order 1. Therefore, the third unit-based ordered lead time goes from period 3 to period 11 in Figure 1(d), and thus equals 8.

We assume the ergodicity of the orginal lead time process, as well as the ordered lead time processes defined above. In particular, the ergodicity definition and the terminology that we use is the following:

Definition 2.3. The stochastic process X is ergodic if the following two conditions are satisfied (Karlin and Taylor 1975, Theorem 5.6):

$$\text{(a)}\quad \lim_{n\to\infty} \frac{\sum_{i=1}^{n} \mathbb{I}(\mathbf{x}(i,\omega) = \mathbf{x})}{n} = f^X(\mathbf{x})$$

for all $\mathbf{x}$ on all sample paths ω, where $\mathbb{I}(\cdot)$ is the indicator function.

$$\text{(b)}\quad \lim_{i\to\infty} \Pr(\mathbf{X}(i) = \mathbf{x}) = f^X(\mathbf{x})$$

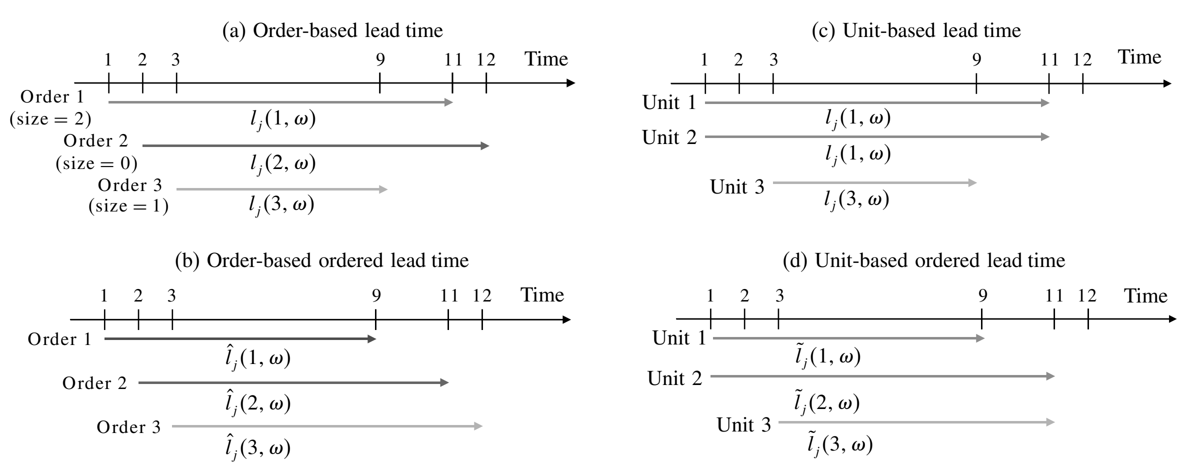
for all $\mathbf{x}$.

The term $f^X(\mathbf{x})$ is the frequency of the value $\mathbf{x}$ in an infinitely long sample path of the ergodic stochastic process X, and we call $f^X(\cdot)$ the steady-state distribution of process X. The steady-state random vector $\mathbf{X}^{ss}$ has the distribution $f^X(\cdot)$.

3. Analysis

In this section, we develop a method to determine base-stock levels and to compute the cost of a given base-stock policy for our problem. We proceed in three stages. We first show that the overall problem can be interpreted as a single-unit problem. Second, we develop the method for the case of sequential lead times. Lead times are sequential if $t + L_j(t)$ is nondecreasing in t for all j, i.e., there is no order crossing. In this case, the single-unit method determines the optimal base-stock levels and computes the exact

Figure 1. A realization of the lead time processes and their ordered versions (under sample path ω).



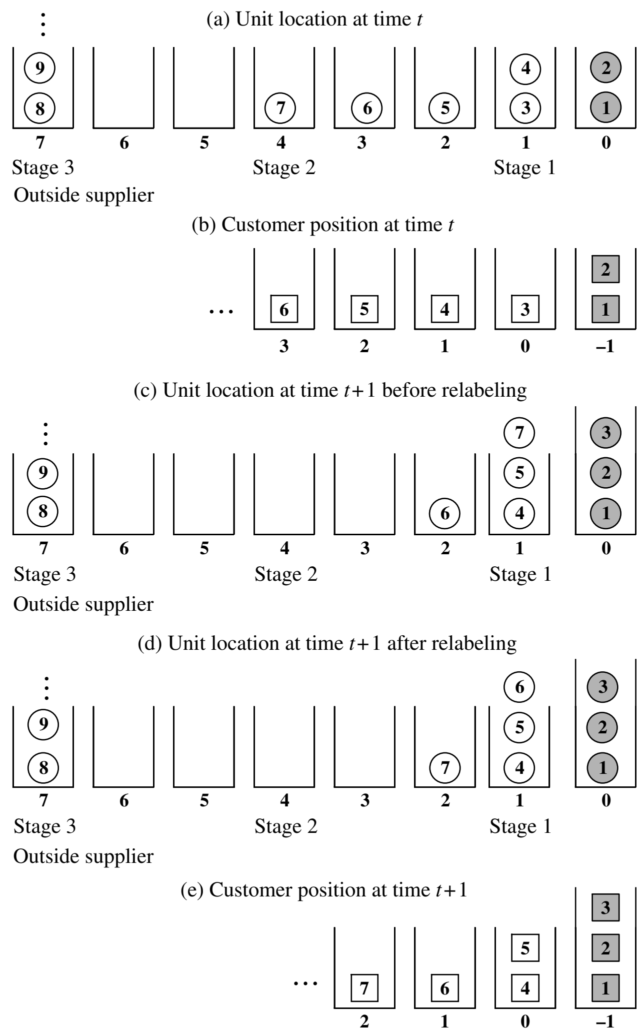
Note. Each arrow starts at the departure time and ends at the arrival time of an order/unit.

cost for both single and multiechelon problems. Finally, we analyze the case of nonsequential lead times. In this case, the single-unit method is exact for single-stage problems and for multistage problems under certain conditions, but is an approximation otherwise. In §5, we study the performance of the algorithm for multistage problems under nonsequential lead times and show that it produces near-optimal base-stock levels and accurate cost estimates.

Our analysis uses a cost accounting scheme that relies on viewing the units and customers as distinct objects and seeing them as forming pairs, as in Muharremoglu and Tsitsiklis (2008). We use the concepts of *the location of a unit* and *the position of a customer* as defined in Muharremoglu and Tsitsiklis (2008). For convenience, we include these definitions in the online appendix. (An electronic companion to this paper is available as part of the online version that can be found at http://or.journal.informs.org/.) Briefly, the location of a unit measures where in the supply chain the unit is. A unit is either in a physical stocking point or in transit between stages or has already been given to a customer. For example, in Figure 2(a), each circle represents a distinct unit. Units 1 and 2 have already been given to a customer and have location 0. Units 3 and 4 are on-hand inventory at stage 1 and have location 1. For units in transit, the location also identifies how long they have been in transit. For example, in Figure 2(a), unit 6 has been in transit for 1 period and unit 5 has been in transit for 2 periods, and the maximum lead time between stage 2 and stage 1 is 3 periods. The position of a customer represents the ranking of the customer in terms of their arrival times of the system. In particular, a customer that has a position of $y > 0$ is the yth next customer to arrive. In Figure 2(b), each rectangle represents a customer. Customers 1 and 2 have already received a unit and have position -1. Customer 3 has arrived, is waiting for a unit, and has a position of 0. Customer 4, which has position 1, is the next customer to arrive.

The cost accounting scheme breaks down the total cost into a sum of costs attributable to pairs of units and customers. There is a particular way in which we pair the units and customers in this paper, which is similar to Muharremoglu and Tsitsiklis (2008), but there is a subtle and significant difference. In short, we update the pairing of the units and customers in every time period taking into account any possible order crossings. In particular, we index the countably infinite pool of units by the nonnegative integers. We assume that the indexing is chosen at the beginning of *every period* in increasing order of their location, breaking ties arbitrarily. This is different from the treatment in Muharremoglu and Tsitsiklis (2008), where the indexing is made at time 0 and is then kept fixed throughout the horizon. In this paper, we relabel the units at the beginning of every period in increasing order of their location. Therefore, when we talk about a unit i in this paper, it may not necessarily refer to the same physical unit in different periods. For example, in Figure 2(a), unit 6 is in location 3 and unit 7 is in location 4 at time t. At this time, unit 7 is released from stage 2 and arrives at stage 1 at time $t+1$. Unit 6, on the other hand, is still in transit between stage 2 and stage 1 and moves to location 2. This means that between period t and $t+1$, unit 7 overtakes unit 6 and this is shown in Figure 2(c). At this point, all units are relabeled in increasing order of their updated locations. After the relabeling, the indices of units 6 and 7 are swapped. Note that this means that when we refer to unit 6 at time t and $t+1$, we refer to different physical units.

Figure 2. Illustration of the unit positions and customer locations and the relabeling of units after order crossing.



Because the labeling of units is updated at the beginning of every period, the unit-to-customer matchings may change throughout the horizon. Nevertheless, at every period t, there is a distinct ith unit and ith customer, i.e., the ith unit-customer pair. From now on, whenever we refer to the ith unit-customer pair, this dynamic pairing of units to customers will be assumed. Let z_t^i be the location of unit i and y_t^i be the position of customer i at time t.

The system incurs linear holding and backorder costs. The total holding and backorder cost at any given period is the sum of holding and backorder costs that can be attributed to distinct unit-customer pairs. A unit-customer pair can incur a holding cost depending on the location of the unit and potentially a backorder cost if the corresponding customer has arrived but has not been served. There are no costs until the unit is released from the outside supplier (we are assuming nonnegative base-stock levels). Then, costs are incurred, until the unit is given to the customer. The cost attributable to a particular unit-customer pair is a function of the movement of the unit and the movement of the customer. The unit is steered along the supply chain, depending on when the customer position crosses the thresholds (base-stock levels) of the different stages. This is how an (echelon) base-stock policy on inventory positions translates into release/don't release decisions for single-unit, single-customer pairs. The timing of these threshold crossings and the lead times that the unit experiences determine the cost attributable to the pair. Next, we define two processes that represent the timing of threshold crossings of the corresponding customer position for different units and the durations between such threshold crossings. All quantities defined below use Convention 2.1 in their notation.

Definition 3.1. Consider a unit-customer pair i. Let $R_j(i) = \min\{t \mid y_t^i < s_j\}$, i.e., this is the moment when the position of customer i crosses the threshold s_j for stage j (the base-stock level for stage j) for the first time, meaning that the corresponding unit can be released from stage $j+1$ to stage j. Let $W_j(i) = R_j(i) - R_{j+1}(i)$, which measures the time between the moment when the position of customer i crosses s_{j+1} for the first time and the moment when it crosses s_j for the first time. Let $\mathbf{W}(i) = (W_1(i), W_2(i), \ldots, W_{M-1}(i))$.

For any unit-customer pair i, let $g_T(\tilde{\mathbf{L}}(i), \mathbf{W}(i))$ be the cost incurred by the pair until period T. Let $g(\tilde{\mathbf{L}}(i), \mathbf{W}(i))$ be the total cost incurred by the pair in infinite horizon (costs will accrue until the unit is delivered to the customer). The fact that we can write costs for unit-customer pairs in this form is a critical observation. This is due to the relabeling of the units in every period and the definition of the unit-based ordered lead time process $\tilde{L}$, which takes into account order crossings among units.

Assumption 3.2. *We assume that the joint process $(\tilde{L}, W)$ is ergodic.*

By the Definition 2.3 of ergodicity, we have

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} \mathbb{1}(\tilde{\mathbf{l}}(i,\omega) = \mathbf{l}, \mathbf{w}(i,\omega) = \mathbf{w})}{n} = f^{\tilde{L},W}(\mathbf{l}, \mathbf{w})$$

for all ω. Note that $f^{\tilde{L},W}(\mathbf{l}, \mathbf{w})$ represents the long-run fraction of unit-customer pairs for which $(\tilde{\mathbf{l}}(i,\omega) = \mathbf{l}, \mathbf{w}(i,\omega) = \mathbf{w})$. The evolution of the process $\mathbf{w}(i)$ depends on the choice of the base-stock vector $\mathbf{s}$; however, to keep the notation simple, we suppress this dependence.

Consider an infinite-horizon sample path ω. We can write the total infinite-horizon average cost for base-stock policy $\mathbf{s}$ as

$$C(\mathbf{s}) = \lim_{T\to\infty} \frac{1}{T} \sum_{i=1}^{\infty} g_T(\tilde{\mathbf{l}}(i,\omega), \mathbf{w}(i,\omega)). \tag{1}$$

The first part of the next proposition states that for any T, we can disregard some initial unit-customer pairs and some later pairs that are unlikely to incur costs within T, and focus on the ones that are likely to incur costs within T. The second part of the proposition states that the summation over the unit-customer pairs can be converted into a summation over values that the process $(\tilde{L}, W)$ can take. Let A be the set of values that this joint process $(\tilde{L}, W)$ can take.

Proposition 3.3.

(a) $$C(\mathbf{s}) = \lim_{T\to\infty} \frac{1}{T} \sum_{i=\lceil \epsilon T+1\rceil}^{\lceil (\bar{d}-\epsilon)T\rceil} g(\tilde{\mathbf{l}}(i,\omega), \mathbf{w}(i,\omega)) + h(\epsilon),$$

where $\lim_{\epsilon\to 0} h(\epsilon) = 0$.

(b) $$C(\mathbf{s}) = \bar{d} \sum_{(\mathbf{l},\mathbf{w})\in A} g(\mathbf{l}, \mathbf{w}) f^{\tilde{L},W}(\mathbf{l}, \mathbf{w}).$$

The intuition for part (b) is that because the cost for pair i is determined by $(\tilde{\mathbf{l}}(i,\omega), \mathbf{w}(i,\omega))$, it is enough to count the number of pairs that experience a certain value $(\mathbf{l}, \mathbf{w})$ and multiply the cost $g(\mathbf{l}, \mathbf{w})$ by this number to account for the cost incurred by all the pairs that experience this value $(\mathbf{l}, \mathbf{w})$. The fraction of pairs that experience a certain $(\mathbf{l}, \mathbf{w})$ value is the same over all sample paths ω by the ergodicity assumption. One then needs to do a summation over all values of $(\mathbf{l}, \mathbf{w}) \in A$ to get to the overall cost.

Let $C = \sum_{(\mathbf{l},\mathbf{w})\in A} g(\mathbf{l}, \mathbf{w}) f^{\tilde{L},W}(\mathbf{l}, \mathbf{w})$, so that the cost for a base-stock policy $\mathbf{s}$ is $\bar{d}C$. This term essentially adds the costs of all the unit-customer pairs in a given sample path by grouping them according to the $(\mathbf{l}, \mathbf{w})$ values they experience. So, it adds the costs of an infinite number of unit-customer pairs along a given sample path. There is an alternative way to interpret this term. If we interpret $f^{\tilde{L},W}(\mathbf{l}, \mathbf{w})$ as a probability distribution, rather than fractions on a sample path, then C can be interpreted as an expected value, rather than the cost on a sample path. In other words, the term C can be interpreted as the expected cost of a single unit-customer pair i, where the probability distribution of $(\tilde{\mathbf{L}}(i), \mathbf{W}(i))$ for this single pair is given by $f^{\tilde{L},W}(\mathbf{l}, \mathbf{w})$ for all $(\mathbf{l}, \mathbf{w}) \in A$. With this interpretation, the cost of the system can be written as the product of the expected demand and the cost of a single unit-customer pair. This is analogous to the result in Muharremoglu and Tsitsiklis (2008), where the system is decomposed into single unit-customer pairs by showing that they can be optimized separately. The result in that paper relies critically on the fact that orders do not cross and that the lead times have a special structure to get the decomposability result. In contrast, in the setting of this paper, the system is certainly not decomposable into unit-customer pairs in the sense of separability of the optimal actions. This is due to order crossing and the relabeling of the units and the interactions that these incidents create. Still, Proposition 3.3 gives us a term analogous to the one in Muharremoglu and Tsitsiklis (2008), this time through algebra, instead of using the decomposability of the problem.

We have $C(\mathbf{s}) = \bar{d}C$, so the cost of the system can be optimized by minimizing C. Using the single-unit interpretation, we can optimize the cost for a single-unit, single-customer problem, and find the optimal base-stock levels. However, note that to even evaluate the cost of a base-stock policy, one needs to determine the fractions $f^{\tilde{L},W}(\mathbf{l},\mathbf{w})$ for all $(\mathbf{l},\mathbf{w}) \in A$. Using the single-unit view, these fractions can also be interpreted as a joint probability distribution. Moreover, because the process W represents the durations between threshold crossings, its evolution depends on the base-stock levels (thresholds). Therefore, in its most general form, it is not easy to "optimize" the single-unit cost C. In the next two sections, we address this challenge in two cases. In §3.1, we study the case where orders do not cross. In §3.2, we study the case where orders are allowed to cross, and give an exact method for single-stage problems. For multistage problems, the method is exact under some conditions. If the conditions are not satisfied, the method is an approximation.

3.1. Sequential Lead Time Processes

In this subsection, we develop an exact method for optimizing the base-stock levels in a multiechelon system where orders do not cross, i.e., systems with sequential lead time processes. Essentially, the method is to solve a related single-unit problem. The method is quite easy to execute, in that it does not require the user to model the entire lead time process, which can be very cumbersome. Rather, it requires a single lead time distribution for a single unit, hence one needs to collect data and fit a distribution for a single random variable for the lead time of each stage of the supply chain and a single random variable for the demand. The lead time distribution needed by the method is the steady-state distribution $\mathbf{L}^{ss}$ of the original lead time process. Systems with ergodic and sequential lead times were studied by Zipkin (1986) and Svoronos and Zipkin (1991) as well. We develop an alternative method for problems with ergodic and sequential lead times that relies on relating the original problem to a single-unit problem. Moreover, this correspondence is used as a building block for problems where orders can cross, which are studied in §3.2.

Proposition 3.4(a) below states that the order-based ordered lead time process and the unit-based ordered lead time process have the same steady-state distribution as the original lead time process if the lead times are sequential. The order-based lead time L and the order-based ordered lead time $\hat{L}$ are of course identical in this case. The sequential nature of the lead times, coupled with the fact that the order sizes for a particular stage are independent of the lead times for that stage, allow us to show that the steady-state distribution $f^{\hat{L}}(\mathbf{l})$ of the order-based ordered lead time and the steady-state distribution $f^{\tilde{L}}(\mathbf{l})$ of the unit-based ordered lead time are identical as well. This leads us to part (b) of Proposition 3.4, which states that the joint steady-state distribution of the process $(\tilde{L}, W)$ can be written in product form, when orders do not cross. Proposition 3.4 is a critical link in establishing the relationship between the cost of a multistage problem and the cost of a single-unit problem.

Proposition 3.4. *If the lead time process is sequential,*
(a) $f^{L}(\mathbf{l}) = f^{\hat{L}}(\mathbf{l}) = f^{\tilde{L}}(\mathbf{l})$ *for all* $\mathbf{l}$,
(b) $f^{\tilde{L},W}(\mathbf{l},\mathbf{w}) = f^{\tilde{L}}(\mathbf{l}) \cdot f^{W}(\mathbf{w})$ *for all* $(\mathbf{l},\mathbf{w})$.

Consider the expression

$$C(\mathbf{s}) = \bar{d} \sum_{(\mathbf{l},\mathbf{w}) \in A} g(\mathbf{l},\mathbf{w}) f^{\tilde{L},W}(\mathbf{l},\mathbf{w})$$

from Proposition 3.3(b) that corresponds to the cost of base-stock policy $\mathbf{s}$. Proposition 3.4(b) states that under no order crossing, the lead time distribution and the W distribution for the single unit are independent. Note that it is relatively easy to determine $f^{\tilde{L}}(\cdot) = f^{L}(\cdot)$, and we discuss this further in §4. On the other hand, the steady-state distribution of W depends on the base-stock levels that are used. Hence, evaluating the expression in Proposition 3.3(b) directly as part of an optimization routine may not be a practical method. Next, we show that the optimization of the base-stock levels can be performed by solving a related single-unit problem whose optimal solution is a set of thresholds that are equal to the optimal base-stock levels for the original overall problem.

Consider the following single-unit problem. The demand process and the cost parameters are the same as the original problem. However, we are interested in only a single customer. The initial position of this customer is y_0, and let y_t be the position of the customer at time t. For example, if $y_0 = 15$, when we refer to "the customer," we are talking about the 15th customer to walk through the door, counting from time 0. Note that we do not know when "the customer" will show up and who she is, we just know that she is the 15th customer to walk through the door. There is a single unit of the good, currently at the outside supplier, that is going to be used to serve the customer (which we refer to as "the unit"). Letting z_t be the location of the unit, we have $z_0 = M$. Let $\mathbf{X} = (X_1, \ldots, X_{M-1})$ be the vector of lead times for the unit, where each X_j is a random variable, and X_j and X_k are independent from each other for all j and k. In other words, once the unit is released from stage j, it takes X_{j-1} periods for it to arrive at stage $j-1$. For each period the unit spends in stage j or between stages j and $j-1$, a cost of h_j is incurred. For each period the customer spends backlogged, a cost of b is incurred. The sequence of events is chosen to be consistent with the sequence of events of the original problem. The goal is to minimize the total cost incurred by choosing the moments to release the unit from its current stage to the next stage. We denote this single-unit problem as $\mathbb{S}1(\mathbf{X})$. The following proposition states that the optimal policy for this single-unit problem is of threshold type for any random lead time vector $\mathbf{X}$. Let the sequence of functions $(\mu_1(y), \mu_2(y), \ldots, \mu_{M-1}(y))$ describe a stationary policy where if the unit is in stage $j+1$, $\mu_j(y_t) \in \{0, 1\}$ corresponds to a hold (0) or release (1) decision for the unit.

Definition 3.5. A threshold policy can be described by a sequence of thresholds $\mathbf{s} = (s_1, \ldots, s_{M-1})$ such that

$$\mu_j(y) = \begin{cases} 1 & \text{if } y \leqslant s_j \\ 0 & \text{if } y > s_j \end{cases} \quad \text{for all } j, \tag{2}$$

where $s_j \in \mathbb{N}_0$ for all j.

Proposition 3.6. *For any random lead time vector* $\mathbf{X}$, *there exists an optimal policy for* $\mathbb{S}1(\mathbf{X})$ *which is a threshold policy with thresholds* $s_1 \leqslant s_2 \leqslant \cdots \leqslant s_{M-1}$.

Consider the version of the single-unit problem where the lead times have the steady-state distribution of the original lead time process, i.e., $\mathbb{S}1(\mathbf{L}^{ss})$. Let $J(y_0, \mathbf{s}, \mathbf{L}^{ss})$ be the total cost of threshold policy $\mathbf{s}$ for $\mathbb{S}1(\mathbf{L}^{ss})$ when the position of the customer at time 0 is y_0. Let $C^S(\mathbf{s}, \mathbf{L}^{ss}) = \lim_{y_0 \to \infty} J(y_0, \mathbf{s}, \mathbf{L}^{ss})$. $C^S(\mathbf{s}, \mathbf{L}^{ss})$ represents the cost of the single-unit problem when the customer is initially far away, i.e., when its initial position goes to infinity. The following proposition relates the infinite-horizon average cost $C(\mathbf{s})$ of a base-stock policy $\mathbf{s}$ in the original problem to the total cost $C^S(\mathbf{s}, \mathbf{L}^{ss})$ of a threshold policy $\mathbf{s}$ in problem $\mathbb{S}1(\mathbf{L}^{ss})$.

Proposition 3.7. *If the lead time process for the original problem is sequential, then*

(a) $C^S(\mathbf{s}, \mathbf{L}^{ss}) = \sum_{(\mathbf{l},\mathbf{w}) \in A} g(\mathbf{l}, \mathbf{w}) f^{\tilde{L}, w}(\mathbf{l}, \mathbf{w})$,

(b) $C(\mathbf{s}) = \bar{d} C^S(\mathbf{s}, \mathbf{L}^{ss})$ *for all* $\mathbf{s}$.

The following corollary is a simple consequence of Proposition 3.7(b).

Corollary 3.8. *Suppose that the lead time process for the original problem is sequential. The base-stock policy* $\mathbf{s}^*$ *is optimal within the class of base-stock policies for the original problem if and only if the threshold policy* $\mathbf{s}^*$ *is optimal for the single-unit problem* $\mathbb{S}1(\mathbf{L}^{ss})$. *Mathematically, the set of minimizers of the two cost functions are equivalent*:

$$\left\{\mathbf{s}' \mid C(\mathbf{s}') = \min_{\mathbf{s}} C(\mathbf{s})\right\} = \left\{\mathbf{s}' \mid C^S(\mathbf{s}', \mathbf{L}^{ss}) = \min_{\mathbf{s}} C^S(\mathbf{s}, \mathbf{L}^{ss})\right\}. \tag{3}$$

Corollary 3.8 enables us to find the optimal base-stock levels for the original problem by solving a related single-unit problem. This is quite striking because we have reduced the task of finding the optimal base-stock levels of the original problem to solving a single-unit single-customer problem if the lead times are sequential. Note that this result was achieved, even though the original problem (even if lead times are sequential) does not decompose into single-unit single-customer problems in the sense of the separability of optimal controls as in Muharremoglu and Tsitsiklis (2008).

The related single-unit problem $\mathbb{S}1(\mathbf{L}^{ss})$ to be solved needs the demand distribution, the cost parameters, and the steady-state distribution $\mathbf{L}^{ss}$. We will discuss issues related to the estimation of the random variable distributions in §4. Suffice it to say that using our single-unit method, we need to fit a distribution to a single random variable for each stage, instead of trying to model and estimate parameters for the whole lead time process, which can be quite complicated. Next on the agenda is to study the case where the lead time process is not sequential.

3.2. Nonsequential Lead Time Processes

In this section, we study systems where orders may cross. The lead time process is assumed to be exogenous and ergodic. Note that Proposition 3.3 still holds even if lead times are not sequential; however, Proposition 3.4, Proposition 3.7, and Corollary 3.8 are no longer valid if orders are allowed to cross. For single-stage systems, we develop analogous results to Proposition 3.7 and Corollary 3.8. In other words, we show that for a single-stage system, the base-stock level can be optimized by solving a related single-unit problem, even if orders can cross. There exists a related single-unit problem for multistage systems as well, but the cost of the original problem and the related single-unit problem can no longer be related in a simple exact form in all cases. The result holds under certain sufficient conditions. If those conditions are not satisfied, the method is not necessarily exact. Still, through a set of numerical experiments, we demonstrate that using the single-unit problem to calculate base-stock levels is an extremely accurate approximation, even in those cases.

Let $V_j(t)$ be the number of outstanding orders between stages $j+1$ and j at time t. Note that we are counting orders (including the ones of size 0), not units. We use Convention 2.1 for this process as well, so $\mathbf{V}(t)$ denotes the vector of outstanding orders for stages 1 through $M-1$ at time t and $\mathbf{V}^{ss}$ is the steady-state random vector of the process V. The results we develop next rely on two important facts:

(a) The process V describing the number of outstanding orders over time has the same steady-state distribution as the order-based ordered lead time process $\hat{L}$.

(b) The infinite-horizon average cost of a single-stage problem depends on the lead time process only through the distribution of $\mathbf{V}^{ss}$, the steady-state random vector of the number of outstanding orders, in other words, only on the distribution of $\hat{\mathbf{L}}^{ss}$, the steady-state random vector of the order-based ordered lead time process.

Propositions 3.9 and 3.10 formally state these facts. (In what follows, $\overset{d}{=}$ means *equal in distribution*.)

Proposition 3.9. (a) $\Pr(\hat{L}_j(t) \leqslant k) = \Pr(V_j(t+k) \leqslant k)$ *for all* j, t, *and* k,

(b) $\mathbf{V}^{ss} \overset{d}{=} \hat{\mathbf{L}}^{ss}$.

Consider the following interpretation of the orders in transit between stages $j+1$ and j as a queueing system. Every period, one order is released, meaning that a "customer" enters the queue. Hence, the arrival rate λ is equal to one. Every time an order arrives, we see that as a service completion. We can think about this queueing system as a first-come-first-serve queue, where the service time for the ith customer is $\hat{L}_j(i)$. The number of customers in the queue at time t then corresponds to the number of outstanding orders $V_j(t)$. Little's law then tells us that in steady state, the expected number of customers waiting is equal to the arrival rate times the expected time each customer spends in the queue. Therefore, we get $E[V_j^{SS}] = 1 \cdot E[\hat{L}_j^{SS}]$. What Proposition 3.9 provides is a much stronger result. Not only are the expected values of these random variables the same, the distributions are completely identical.

The way we relate the original problem to the single-unit problem is to go through an intermediate *surrogate problem*. The surrogate problem is very similar to the original problem, but one in which orders do not cross. The surrogate problem can be described as follows. The demand process and the cost structure are exactly the same as the original problem, the only difference being in the lead time process. In particular, the lead time process for the surrogate problem is defined to be stochastically identical to the order-based ordered lead time process $\hat{L}$ of the original problem. In particular, given a sample path of the lead time process for the original problem, there is a corresponding sample path of the lead time process for the surrogate problem, where the same number of orders arrive in the same periods, but the orders arrive in sequential order. Table 2 illustrates the lead time dynamics of the original problem and the surrogate problem. In period 7, two orders arrive in the original problem. The first order was released at time 5 and the second order at time 6. The size of the first order is 8 units and the size of the second order is 6 units. In the surrogate problem, again two orders arrive in period 7. The first order was released at time 4 and the second order at time 5. The size of the first order is 5 units and the size of the second order is 8 units. In periods 6, 11, and 12, no orders arrive, neither in the original problem, nor in the surrogate problem. The order that was released at time 9 is a zero-sized order, which arrives in period 10 in the original problem and in period 13 in the surrogate problem. In the original problem, the first order is released at time 1 and arrives at time 4, yielding a lead time $L(1) = 3$. The order-based ordered lead time $\hat{L}(1)$ in the original problem is defined as the difference between the time of the first arrival and the time of the first release, which is $3 - 1 = 2$. In the surrogate problem, the first order is released at time 1 and arrives at time 3, yielding a lead time $L^{Surrogate}(1) = 2 = \hat{L}(1)$. In fact, $L^{Surrogate}(i) = \hat{L}(i)$ for all i by the construction of the surrogate problem.

Let $C'(\mathbf{s})$ denote the infinite-horizon average cost of the surrogate problem under base-stock level $\mathbf{s}$. Note that while the original lead time process was not sequential, the lead time process for the surrogate problem is sequential by the definition of the process $\hat{L}$. Therefore, we can apply Proposition 3.7 to the surrogate problem and its related single-unit problem $\mathbb{S}1(\hat{\mathbf{L}}^{ss})$. In the remainder of this section, we show that for any given base-stock level, the cost of this surrogate problem is equal to the cost of the original problem for single-stage systems and for multistage systems under certain conditions.

3.2.1. Single-Stage Systems. Using Proposition 3.9, we obtain the following result.

Proposition 3.10. *For a single-stage problem*:

(a) *The average cost of the system operating under base-stock level* s *depends on the lead time process only through its effect on* $\hat{L}^{ss}$; *i.e., if two systems differ only in their lead time processes and have the same steady-state distribution* $\hat{L}^{ss}$ *of the order-based ordered lead time process, then these two systems have the same average costs under any base-stock level* s.

(b) $C(s) = C'(s) = \bar{d}C^S(s, \hat{L}^{ss})$ *for all* s.

(c) *The set of minimizers for the original problem and the single-unit problem are equivalent*:

$$\left\{s' \mid C(s') = \min_s C(s)\right\} = \left\{s' \mid C^S(s', \hat{L}^{ss}) = \min_s C^S(s, \hat{L}^{ss})\right\}. \quad (4)$$

Part (a) of Proposition 3.10 relies on the fact that the order sizes in a single-stage system under a base-stock

Table 2. An example illustrating the lead time dynamics of the original problem and its surrogate problem.

	Time	3	4	5	6	7	8	9	10	11	12	13	14
Original problem	Release time	2	1	3		5 and 6	4	7	9			8, 11, and 12	10
	Order size	10	3	7		8 and 6	5	1	0			2, 15, and 4	12
Surrogate problem	Release time	1	2	3		4 and 5	6	7	8			9, 10, and 11	12
	Order size	3	10	7		5 and 8	6	1	2			0, 12, and 15	4

Note. Each column represents a time period and shows the release time and the size of the order(s) that arrive during that period, if any, both for the original problem and the surrogate problem.

policy are realizations of the i.i.d. demand distribution, and that the steady-state distribution of the order-based ordered lead time process is equal to the steady-state distribution of the number of outstanding orders. Therefore, given the number of outstanding orders, the total size of the pipeline inventory is stochastically determined. This means that even though in the original problem and the surrogate problem different orders may be outstanding, as long as the total number of outstanding orders are equivalent, the steady-state distributions of pipeline inventories under the two systems are equivalent.

Proposition 3.10(b) is analogous to Proposition 3.7, and it relates the cost of the original problem to the cost of a related single-unit problem by going through an intermediate surrogate problem. Note that the single-unit problem in this case uses $\hat{L}^{ss}$ as its lead time distribution instead of L^{ss}. If orders do not cross in the original problem, these two distributions are identical. However, when orders can cross, they are different, and the correct one to use in the single-unit problem is $\hat{L}^{ss}$ to relate the costs of the two problems. Proposition 3.10(c) is analogous to Corollary 3.8, in that it shows that the original problem and the related single-unit problem are optimized at the same base-stock level. Proposition 3.10 allows us to reduce the task of finding an optimal base-stock level for a single-stage problem to solving a related single-unit problem, even when orders can cross.

3.2.2. Multistage Systems. Note that Proposition 3.9 is valid for multistage systems. Similar to the single-stage case, one can talk about a surrogate problem for a multistage system as well. The surrogate problem has the same steady-state distribution ($\hat{\mathbf{L}}^{ss}$) of the order-based ordered lead time process, and furthermore, the surrogate problem can be optimized by solving its related single-unit problem $\mathbb{S}1(\hat{\mathbf{L}}^{ss})$. In this subsection, we show that the cost of the surrogate problem is equal to the cost of the original problem for multistage systems under certain conditions.

First, consider a case where order crossing can happen only between the outside supplier and the most upstream stage. In this case, we have an analog to Proposition 3.10 for multistage systems.

Proposition 3.11. *Suppose that order crossing happens only in the most upstream stage, i.e., between the outside supplier and stage $M-1$. Then, $C(\mathbf{s}) = C'(\mathbf{s}) = \bar{d}C^S(\mathbf{s}, \hat{\mathbf{L}}^{ss})$, and the set of minimizers of the two cost functions are equivalent*:

$$\left\{\mathbf{s}' \mid C(\mathbf{s}') = \min_{\mathbf{s}} C(\mathbf{s})\right\} = \left\{\mathbf{s}' \mid C^S(\mathbf{s}', \hat{\mathbf{L}}^{ss}) = \min_{\mathbf{s}} C^S(\mathbf{s}, \hat{\mathbf{L}}^{ss})\right\}. \quad (5)$$

The steady-state distribution of the number of outstanding orders $\mathbf{V}^{ss}$ is the same under the original problem and the surrogate problem because the steady-state distribution of the order-based ordered lead time $\hat{\mathbf{L}}^{ss}$ is equal to $\mathbf{V}^{ss}$ (by Proposition 3.9), and the distribution of the order-based ordered lead time is the same in the original problem and the surrogate problem. In both problems, the order sizes between the outside supplier and the most upstream stage $M-1$ are equal to demand realizations under a base-stock policy, which are i.i.d. with the same distribution. It follows that the steady-state distribution of the number of outstanding units between the outside supplier and stage $M-1$ is the same for the original and surrogate problems. Therefore, the expected holding cost incurred at stage $M-1$ is equal for the two problems. Because the distribution of the most upstream pipeline inventory was the same, the steady-state distribution of the internal backlog that stage $M-1$ provides for stage $M-2$ is the same under both problems as well. Because there is no order crossing after this stage, the two systems are identical considering stages $M-2, \ldots, 1$. Therefore, the sum of expected costs incurred at stages $M-2, \ldots, 1$ and the expected backlog costs are equal under the two problems.

The exact relationship between the original problem and the single-unit problem breaks down in the case of a general multistage system. In particular, the costs of the original problem and the surrogate problem are not equal, even though the steady-state distributions of the number of outstanding orders are the same for the two problems. This is due to the fact that in a multistage system, the order sizes are no longer realizations of the i.i.d. demand process, but are modified depending on upstream inventory availability. This causes the order sizes to be dependent over time. This dependency, coupled with the fact that under the two problems the pipeline inventory consists of different orders, means that the steady-state distributions of pipeline inventory in the two systems are no longer equal.

Even though the exact relationship between the original problem and the single-unit problem does not hold in the multistage case, solving the single-unit problem serves as a good approximation to both the optimal base-stock level and the optimal cost. A numerical study is provided in §5. There are two more cases where the single-unit approach does yield exact results. These cases are covered in the following two propositions.

Proposition 3.12. *Consider a problem with base-stock levels $\mathbf{s} = (s_1, \ldots, s_{M-1})$. Suppose that $s_j - s_{j-1} \geqslant L_j^{\max} \cdot D^{\max}$ for all $j = 2, \ldots, M-1$, where $D^{\max}$ is the maximum level of demand in any period and $L_j^{\max}$ is the maximum lead time between stage $j+1$ and j. Then, $C(\mathbf{s}) = C'(\mathbf{s}) = \bar{d}C^S(\mathbf{s}, \hat{\mathbf{L}}^{ss})$.*

The proposition states that in cases where the difference between echelon base-stock levels of adjacent stages is large enough, the costs of the original problem and the surrogate problem are equal, and they are both equal to expected demand times the cost of the single-unit problem. The reason for this result is that under this condition, there is never an internal backlog in the system. Therefore, all order sizes are realizations of demand, which means that the steady-state distribution of pipeline inventories is the same under the original and the surrogate problems. Of course, such base-stock levels are not likely to be optimal. Still, in many systems, the optimal base-stock levels are such that internal backlogs are generally small. This is especially the case if the holding cost differences between adjacent stages are also substantial, i.e., when significant value is added to the product in every stage of the supply chain. Proposition 3.12 provides some theoretical justification for the use of the approximation under conditions when $s_{j+1} - s_j$ is considerable.

Proposition 3.13. *Consider a two-stage problem where the lead time between the outside supplier and stage 2 is deterministic. Then,* $C(\mathbf{s}) = C'(\mathbf{s}) = \bar{d}C^S(\mathbf{s}, \hat{\mathbf{L}}^{ss})$ *if* $s_1 = s_2$.

Under the conditions of Proposition 3.13, one can view the two-stage system as a single-stage system with stochastic lead times, so the result is not surprising and is a simple corollary of Proposition 3.10. However, we explicitly state this result here because when taken in conjunction with Proposition 3.12, it provides some further analytical justification for using the single-unit problem to approximate the overall problem. In particular, consider a two-stage problem with a deterministic lead time between the outside supplier and stage 2. Proposition 3.12 states that for base-stock levels where the difference $s_2 - s_1$ is large, the method is exact. Proposition 3.13 states that for base-stock levels where this difference is zero, the method is exact. This means that for both small and large values of the difference, the method is expected to do well. In §5, we indeed show that the performance of the approximation is quite good over a diverse set of instances.

4. Estimation Issues

In §§3.1 and 3.2, we established a relationship between the original problem and a related single-unit problem. Solving the single-unit problem provides a method to compute base-stock levels. In particular, the single-unit problem to solve is $\mathbb{S}1(\hat{\mathbf{L}}^{ss})$, i.e., a single-unit single-customer problem, where the vector of lead times of the unit are distributed as the random variable $\hat{\mathbf{L}}^{ss}$. The detailed definition of this single-unit problem was given in §3.1. By Proposition 3.6, the solution of the single-unit problem $\mathbb{S}1(\hat{\mathbf{L}}^{ss})$ is a set of thresholds $\mathbf{s_u}$. These thresholds are the base-stock levels that the single-unit method recommends, and they are optimal for single-stage problems and multistage problems under certain conditions. Otherwise, they are not guaranteed to be optimal, but are near optimal, as demonstrated in §5.

To formulate and solve the single-unit problem $\mathbb{S}1(\hat{\mathbf{L}}^{ss})$, we need the following inputs: the echelon holding cost vector $\mathbf{h}$, the backlog cost parameter b, the distribution of one-period demand D, and the steady-state distribution of the order-based ordered lead time process of the original problem, $\hat{\mathbf{L}}^{ss}$. $\mathbf{h}$, b, and the distribution of D are standard quantities that are commonly used in inventory theory. $\hat{\mathbf{L}}^{ss}$ is the new input in our procedure. Conceptually, there are two ways of obtaining an estimation for the distribution of $\hat{\mathbf{L}}^{ss}$. The first approach starts with choosing a model to represent the original lead time process and fitting parameters to that model. One then needs to derive the distribution of $\hat{\mathbf{L}}^{ss}$ from this estimated original lead time process. Such a derivation is already available for the case of i.i.d. lead times. Zalkind (1978) assumes that the original lead time process is i.i.d. with a given distribution, and provides an efficient procedure to compute the steady-state distribution $\mathbf{V}^{ss}$ of the number of outstanding orders. Note that $\mathbf{V}^{ss}$ is equal to $\hat{\mathbf{L}}^{ss}$ in distribution by Proposition 3.9, meaning that the procedure in Zalkind (1978) also provides $\hat{\mathbf{L}}^{ss}$. In cases where the original lead time process is i.i.d. and its distribution is readily available, this approach gives a quick method for computing $\hat{\mathbf{L}}^{ss}$.

The second approach for obtaining an estimate of the steady-state distribution $\hat{\mathbf{L}}^{ss}$ is to directly observe the order-based ordered lead time process and to fit a distribution to the observed realizations. The realizations are obtained by subtracting the time of the ith order release from the time of the ith order arrival between two adjacent stages. The fitted distribution is an estimate of the steady-state distribution $\hat{\mathbf{L}}^{ss}$ of the order-based ordered lead time process, which is all that is needed for the solution procedure. The lead time process L and the order-based ordered lead time process $\hat{L}$ may indeed be very complex processes, involving all kinds of intertemporal dependencies. However, we do not need to estimate either one of these two processes in whole. To optimize the base-stock levels using the outlined procedure, all that we need is a single random variable $\hat{L}_j^{ss}$ for every stage j. If the original lead time process is not necessarily i.i.d. and/or if a model for its dynamics is not available, this second approach may be preferable. Finally, note that the data about arrival and departure times of orders are typically easily accessible in shipping records of most companies.

The discussion above disregards one important factor, which is the existence of zero-sized orders. In the model, these orders also have departure times and arrival times, but of course they are not observed in reality. However, the second method we described above assumes that we can observe all departure and arrival times, even those of zero-sized orders. This necessitates a correction in the estimation procedure. The good news is that there is a simple way to correct for this in cases where our single-unit method is exact. In cases where our single-unit method is not exact, the correction that we propose does not completely resolve the issue caused by the lack of observation of the zero-sized orders. So, this adds a second level of approximation in cases where the method is not exact. The following two processes are observable in reality.

Definition 4.1. *The order-based ordered positive lead time process* $L_j^+(i)$ *is the time between the* i*th positive-sized order release from stage* $j+1$ *and the* i*th positive-sized order arrival at stage* j*. Note that these orders do not have to be the same order.*

Definition 4.2. *The number of positive-sized outstanding orders process* $V_j^+(t)$ is the number of positive-sized outstanding orders between stages $j+1$ and j at time t.

In reality, we can observe the process L^+ and the process V^+, and we can fit a distribution to the sequence of values observed, which gives us estimates of the steady-state distributions $\mathbf{L}^{+ss}$ and $\mathbf{V}^{+\mathbf{ss}}$. To optimize base-stock levels, we need the steady-state distribution $\hat{\mathbf{L}}^{ss}$ of the order-based ordered lead time process $\hat{L}$. There are two cases in which it is possible to compute $\hat{L}_j^{ss}$, given either L_j^{+ss} or V_j^{+ss}, for a given stage j. This is shown in the next proposition. Part (b) of the proposition assumes that there is a maximum lead time between stage $j+1$ and j, which is denoted as $L_j^{\max}$. However, this is not an a priori bound, and can be taken to be the largest lead time value recorded during the observation phase of the estimation procedure. Let

$$\mathbf{f_j} = \big(\Pr(V_j^{+ss}=1), \Pr(V_j^{+ss}=2), \ldots, \Pr(V_j^{+ss}=L_j^{\max})\big)'$$

and

$$\mathbf{g_j} = \big(\Pr(\hat{L}_j^{ss}=1), \Pr(\hat{L}_j^{ss}=2), \ldots, \Pr(\hat{L}_j^{ss}=L_j^{\max})\big)'.$$

Proposition 4.3. (a) *If orders do not cross between stage* $j+1$ *and* j, *then*

$$L_j^{ss} \stackrel{d}{=} \hat{L}_j^{ss} \stackrel{d}{=} \tilde{L}_j^{ss} \stackrel{d}{=} L_j^{+ss}.$$

(b) *Let* $L_j^{\max}$ *be the largest lead time value recorded during the observation phase. If the order sizes between stage* $j+1$ *and* j *are i.i.d. random variables with the distribution of demand, then*

$$\mathbf{g_j} = B_j^{-1}\mathbf{f_j},$$

where B *is the invertible matrix defined as*

$$B_j = \begin{bmatrix} \binom{1}{1}(1-p_0^j) & \binom{2}{1}(1-p_0^j)p_0^j & \binom{3}{1}(1-p_0^j)(p_0^j)^2 & \cdots & \binom{L_j^{\max}}{1}(1-p_0^j)(p_0^j)^{L_j^{\max}-1} \\ 0 & \binom{2}{2}(1-p_0^j)^2 & \binom{3}{2}(1-p_0^j)^2 p_0^j & \cdots & \binom{L_j^{\max}}{2}(1-p_0^j)^2(p_0^j)^{L_j^{\max}-2} \\ 0 & 0 & \binom{3}{3}(1-p_0^j)^3 & \cdots & \binom{L_j^{\max}}{3}(1-p_0^j)^3(p_0^j)^{L_j^{\max}-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \binom{L_j^{\max}}{L_j^{\max}}(1-p_0^j)^{L_j^{\max}} \end{bmatrix}$$

and p_0^j *is the probability of a zero-sized order from stage* $j+1$ *to stage* j *in a given period. This is the same as the probability of a zero-sized demand, i.e.,* $p_0^j = \Pr(D=0)$ *for all* j.

Part (a) of Proposition 4.3 indicates that the steady-state distribution of all the different lead time processes are equal, when lead times are sequential between two stages. When lead times are not sequential, but when the order sizes are distributed as the i.i.d. demand process, Proposition 4.3(b) provides a simple closed-form expression for obtaining the necessary steady-state distribution $\hat{L}_j^{ss}$, given the steady-state distribution V_j^{+ss} of the observable number of positive-sized outstanding orders process.

Using Proposition 4.3, we can resolve the estimation problem that arises due to the lack of observation of zero-sized orders in all cases where our single-unit method is exact. Below, we describe how Proposition 4.3 applies to the different cases.

Case 1 (see Proposition 3.7 and Corollary 3.8). If lead times between all stages are sequential, Proposition 4.3(a) applies to all stages of the system. As a result, it is sufficient to observe the positive-sized order-based ordered lead times and use their steady-state distribution $\mathbf{L}^{+ss}$ in the single-unit problem.

Case 2 (see Proposition 3.10). If the system has a single stage, then all order sizes are equal to demand realizations under a base-stock policy. Therefore, Proposition 4.3(b) is applicable and we can obtain the necessary steady-state distribution $\hat{L}^{ss}$ from the observed steady-state distribution V^{+ss}.

Case 3 (see Proposition 3.11). If order crossing can occur only in the most upstream stage, then order sizes between the outside supplier M and stage $M-1$ are equal to demand realizations. Therefore, Proposition 4.3(b) applies to the most upstream stage. All other stages have sequential lead times, i.e., Proposition 4.3(a) applies to stages $1, \ldots, M-2$.

Case 4 (see Proposition 3.12). If there are no internal backlogs, then order sizes between all stages are equal to demand realizations. Therefore, Proposition 4.3(b) applies to all stages in this case.

Case 5 (see Proposition 3.13). If the upstream lead time in a two-stage problem is deterministic and the difference of the base-stock levels is zero, then all outstanding orders in the system carry a number of units that is a realization of one-period demand. Therefore, Proposition 4.3(b) is applicable.

In multistage problems that do not fall into any one of the cases above, we propose using Proposition 4.3(b). In these cases, p_0^j, the probability of a zero-sized order from stage $j+1$ to stage j, is not equal to the probability of zero-sized demand. However, one can still observe the fraction of periods when an order was not shipped between stages $j+1$ and j, and use that fraction as an estimate for p_0^j. Using the expression $\mathbf{g_j} = B_j^{-1}\mathbf{f_j}$ still implicitly assumes that the order sizes are i.i.d, which is not necessarily true in these systems. Therefore, in cases where using the single-unit problem $\mathbb{S}1(\hat{\mathbf{L}}^{ss})$ to optimize base-stock levels is already an approximation, the estimation procedure for the input $\hat{\mathbf{L}}^{ss}$ also contains an approximation.

To summarize, the new input required by the method developed in this paper is the steady-state distribution $\hat{\mathbf{L}}^{ss}$. If the lead times are i.i.d. and their distribution is readily available, we can analytically compute the distribution of $\hat{\mathbf{L}}^{ss}$ directly using the procedure in Zalkind (1978). Otherwise, we propose estimating this distribution in three steps.

Step 1. Directly observe the positive-sized order-based ordered lead time process L^+ and/or the number of positive-sized outstanding orders process V^+ for a relatively long period (depending on whether and which one of the five cases is applicable).

Step 2. Fit a distribution to the observed values to obtain an estimate of $\mathbf{L}^{+ss}$, the steady-state distribution of the positive-sized order-based ordered lead time process and/or $\mathbf{V}^{+ss}$, the steady-state distribution of the number of positive-sized outstanding orders process.

Step 3a. If one of the above described five cases is applicable, compute the steady-state distribution $\hat{\mathbf{L}}^{ss}$ of the order-based ordered lead time process $\hat{L}$ by using Proposition 4.3. This is the necessary input for the method.

Step 3b. If none of the above described five cases is applicable, observe the fraction of periods with no shipments between stages $j+1$ and j, and use this as the estimate for p_0^j for all j. Then, compute the steady-state distribution $\hat{\mathbf{L}}^{ss}$ of the order-based ordered lead time process $\hat{L}$ by using Proposition 4.3(b) with the observed values of p_0^j.

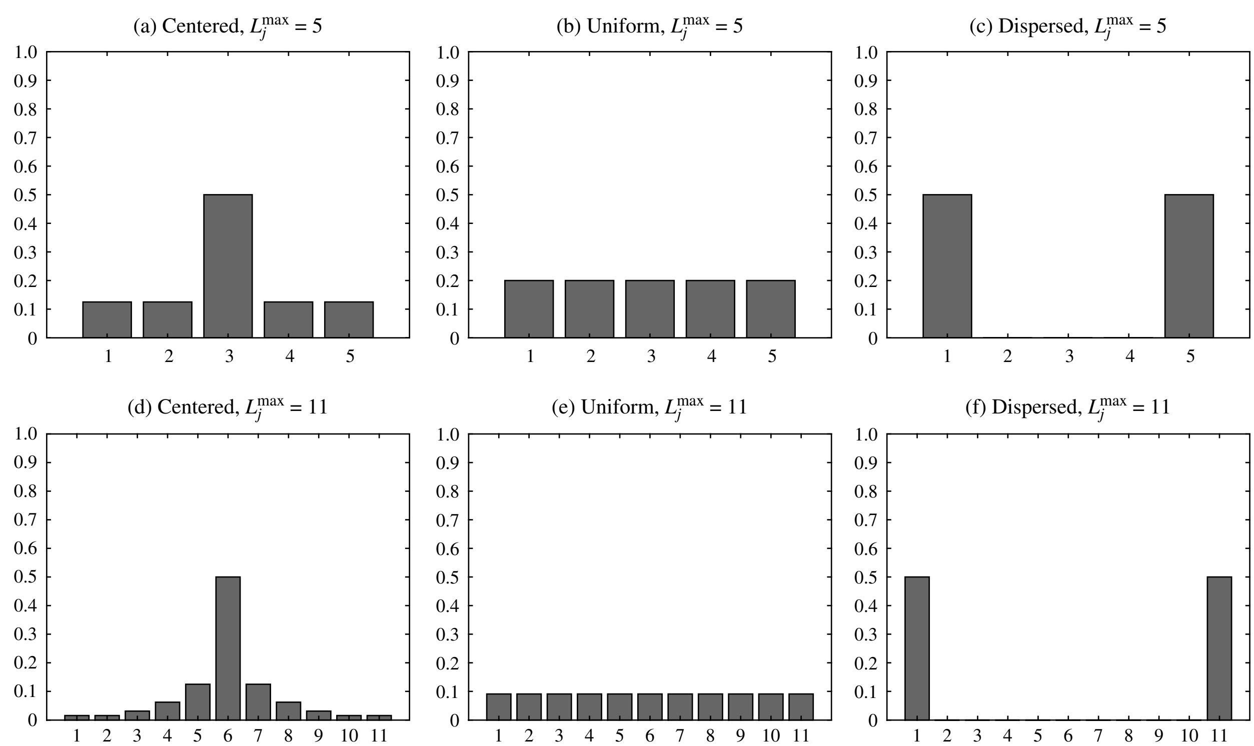
5. Numerical Results

In this section, we study the effectiveness of our single-unit method through a set of numerical experiments. We study multistage problems, where none of the conditions for the optimality of the method are satisfied. Overall, the method performs extremely well, and produces near-optimal base-stock levels.

We ran experiments with two-stage and five-stage systems. For both sets, we changed the cost parameters and the lead time and demand distributions to test a wide variety of problem instances. The lead time process for a given stage is assumed to be an i.i.d process where the lead time for a particular order can vary between 1 and $L_j^{\max}$. $L_j^{\max}$ was chosen to be 5, 11, 101, 201, or 301 for two-stage systems and 5 or 11 for five-stage systems. We studied three types of lead time distributions, which are illustrated in Figure 3. All lead time distributions allow for order crossing and given $L_j^{\max}$ have the same expectation. However, the *centered* type is the least variable and the *dispersed* type is the most variable. In a given instance, the same lead time distribution is used for all stages of the system.

The demand is assumed to have a binomial distribution with mean one. In half of the experiments, we used a binomial distribution with two trials and a success probability of 0.5, and in the other half, we used 10 trials and a success probability of 0.1 to test the effect of the variability of demand. The holding cost rate at the most upstream stage is assumed to be one and the holding cost rate at subsequent stages is assumed to increase linearly with a certain *holding cost increment*. The holding cost increment is taken to be one or four. For example, when the holding cost increment is four, $h_j - h_{j+1} = 4$ for all stages j. Finally, the backlog cost rate is assumed to be proportional

Figure 3. The types of lead time distributions.



to the most downstream holding cost rate, with a ratio of 2 or 10, i.e., $b/h_1 \in \{2, 10\}$. We considered all combinations of the above mentioned choices, which results in 120 two-stage numerical examples and 48 five-stage numerical examples.

In the first set of experiments, we test the effectiveness of the method. For every numerical example, we obtained base-stock levels $\mathbf{s_u}$ by running the single-unit algorithm. We also found the optimal base-stock levels $\mathbf{s}^*$ through simulation. We then simulated the system under both $\mathbf{s_u}$ and $\mathbf{s}^*$ for a sufficiently long period to obtain the cost under both base-stock levels. The loss of optimality is defined as

$$\text{loss of optimality} = \frac{\text{simulated cost of } \mathbf{s_u} - \text{simulated cost of } \mathbf{s}^*}{\text{simulated cost of } \mathbf{s}^*}.$$

Tables 3 and 4 report the results for two-stage and five-stage systems, respectively. The maximum loss numbers were 0.4142% and 0.4394%, respectively. In two-stage systems, the single-unit method found the optimal base-stock levels in 73 out of 120 numerical examples, and in five-stage systems, in 16 out of 48 numerical examples. However, even in cases where the base-stock levels $\mathbf{s_u}$ were not optimal, we observe that they are near optimal. Overall, the average loss of optimality was 0.0348% for two-stage systems and 0.0894% for five-stage systems.

For two-stage systems, the single-unit method found the optimal base-stock levels in 43 out of 48 cases with $L_j^{\max}$ values of 5 or 11, but in 30 out of 72 cases with $L_j^{\max}$ values of 101, 201, or 301. This is due to the difference in the scale of the optimal base-stock levels and the resulting finer grid of available base-stock levels with longer lead times, due to the discreteness of inventory levels. For example, with a maximum lead time of 5, the optimal base-stock levels are on the order of $(5, 10)$, whereas with a maximum lead time of 301, the optimal base-stock levels are on the order of $(150, 300)$. The cost difference between base-stock levels $(5, 10)$ and $(6, 10)$ can be more substantial compared to the difference between base-stock levels $(150, 300)$ and $(151, 300)$, which makes it easier to identify the optimal base-stock levels in cases with smaller lead times. The other implication of the difference in scale is that when the single-unit method fails to find the optimal base-stock level, the loss of optimality is smaller in cases with longer lead times. The average loss of optimality in the five nonoptimal cases with $L_j^{\max}$ values of 5 or 11 is

Table 3. The effectiveness of the single-unit method for two-stage systems.

Demand variability ($n = 60$)	Binomial (10, 0.1)	Binomial (2, 0.5)			
Average loss (%)	0.0380	0.0316			
Maximum loss (%)	0.3410	0.4142			
Number of optimal	34	39			
Max. lead time ($n = 24$)	$L_j^{\max} = 5$	$L_j^{\max} = 11$	$L_j^{\max} = 101$	$L_j^{\max} = 201$	$L_j^{\max} = 301$
Average loss (%)	0.0047	0.0256	0.0543	0.0485	0.0408
Maximum loss (%)	0.1134	0.3410	0.4142	0.2180	0.2224
Number of optimal	23	20	11	10	9
Lead time type ($n = 40$)	Centered	Uniform	Dispersed		
Average loss (%)	0.0018	0.0639	0.0386		
Maximum loss (%)	0.0388	0.4142	0.3410		
Number of optimal	38	18	17		
Holding cost increment ($n = 60$)	$h_1 - h_2 = 1$	$h_1 - h_2 = 4$			
Average loss (%)	0.0150	0.0546			
Maximum loss (%)	0.1258	0.4142			
Number of optimal	37	36			
Backlog/holding cost ($n = 60$)	$b/h_1 = 2$	$b/h_1 = 10$			
Average loss	0.0174	0.0522			
Maximum loss	0.2006	0.4142			
Number of optimal	40	33			
Total ($n = 120$)					
Average loss	0.0348				
Maximum loss	0.4142				
Number of optimal	73				

Notes. For each set of experiments, the reported numbers are the average loss of optimality, the maximum loss of optimality, and the number of numerical examples where the method found the optimal base-stock levels. (n is the number of instances for each block in the corresponding row. For example, there are 24 instances with $L_j^{\max} = 5$.)

Table 4. The effectiveness of the single-unit method for five-stage systems.

Demand variability ($n=24$)	Binomial (10, 0.1)	Binomial (2, 0.5)	
Average loss (%)	0.0712	0.1077	
Maximum loss (%)	0.2976	0.4394	
Total optimal	8	8	
Maximum lead time ($n=24$)	$L_j^{\max}=5$	$L_j^{\max}=11$	
Average loss (%)	0.0768	0.1021	
Maximum loss (%)	0.4394	0.2976	
Total optimal	11	5	
Lead time type ($n=16$)	Centered	Uniform	Dispersed
Average loss (%)	0.0310	0.0694	0.1680
Maximum loss (%)	0.2439	0.2498	0.4394
Total optimal	9	5	2
Holding cost increment ($n=24$)	$h_1-h_2=1$	$h_1-h_2=4$	
Average loss (%)	0.0634	0.1155	
Maximum loss (%)	0.2498	0.4394	
Total optimal	8	8	
Backlog/holding cost ($n=24$)	$b/h_1=2$	$b/h_1=10$	
Average loss (%)	0.0506	0.1283	
Maximum loss (%)	0.2009	0.4394	
Total optimal	11	5	
Total ($n=48$)			
Average loss (%)	0.0894		
Maximum loss (%)	0.4394		
Total optimal	16		

0.1458%, whereas the average loss of optimality in the 42 nonoptimal cases with $L_j^{\max}$ values of 101, 201, or 301 is 0.0821%. Table 10 in the online appendix displays the detailed results of the 47 out of 120 instances where the single-unit method did not produce optimal base-stock levels. In terms of the shape of the lead time distribution, we observe that the effectiveness of the method is best for the *centered* lead time type for both two-stage and five-stage systems. This is intuitive because one expects less order crossing in this case.

In the second set of experiments, we test the accuracy of the cost estimate generated by the single-unit method. We can use the single-unit method to find a set of base-stock levels or to estimate the cost of a given set of base-stock levels. For every numerical example, we consider the optimal base-stock level, as well as some nonoptimal base-stock levels. For every base-stock level considered, we find the cost estimate given by the single-unit method and also a simulated cost. Tables 5 and 6 report the relative absolute error of the cost estimate for two-stage systems and Table 7 does the same for five-stage systems. The error is defined as

$$\text{error} = \frac{|\text{estimated cost} - \text{simulated cost}|}{\text{simulated cost}}. \tag{6}$$

The experiments show that the single-unit method provides good estimates for the cost of the system. For two-stage systems, the average error is 0.76%, and for five-stage systems, the average error is 0.94%. 90% of the time, the error was below 1.47% for two-stage systems, and below 1.87% for five-stage systems. The maximum error over all instances is 2.31% and 5.85%, respectively.

Table 5. Accuracy of cost estimation for two-stage systems, instances with maximum lead time of 5 or 11.

Demand variability ($n=264$)	Binomial (10, 0.1) (%)	Binomial (2, 0.5) (%)	
Average	0.75	0.78	
Median	0.67	0.70	
90th percentile	1.41	1.51	
Maximum	2.04	2.31	
Maximum lead time ($n=264$)	$L_j^{\max}=5$ (%)	$L_j^{\max}=11$ (%)	
Average	0.70	0.82	
Median	0.66	0.74	
90th percentile	1.31	1.60	
Maximum	2.30	2.31	
Lead time type ($n=176$)	Centered (%)	Uniform (%)	Dispersed (%)
Average	0.52	1.04	0.72
Median	0.51	1.08	0.72
90th percentile	0.89	1.87	1.41
Maximum	1.45	2.31	2.18
Holding cost increment ($n=264$)	$h_1-h_2=1$ (%)	$h_1-h_2=4$ (%)	
Average	0.74	0.78	
Median	0.66	0.70	
90th percentile	1.45	1.47	
Maximum	2.30	2.31	
Backlog/holding cost ($n=264$)	$b/h_1=2$ (%)	$b/h_1=10$ (%)	
Average	0.72	0.80	
Median	0.66	0.70	
90th percentile	1.34	1.55	
Maximum	2.31	2.30	
Total ($n=528$)	(%)		
Average	0.76		
Median	0.68		
90th percentile	1.47		
Maximum	2.31		

Note. The percentages reported are the error values as defined in Equation (6).

Table 6. Accuracy of cost estimation for two-stage systems, instances with maximum lead time of 101, 201, or 301.

Demand variability ($n=72$)	Binomial (10, 0.1) (%)	Binomial (2, 0.5) (%)	
Average	0.78	0.74	
Median	0.94	0.84	
90th percentile	1.42	1.43	
Maximum	1.82	2.00	
Maximum lead time ($n=48$)	$L_j^{\max}=101$ (%)	$L_j^{\max}=201$ (%)	$L_j^{\max}=301$ (%)
Average	0.85	0.74	0.68
Median	1.06	0.90	0.81
90th percentile	1.60	1.42	1.24
Maximum	2.00	1.61	1.61
Lead time type ($n=48$)	Centered (%)	Uniform (%)	Dispersed (%)
Average	0.08	1.26	0.92
Median	0.06	1.24	0.97
90th percentile	0.20	1.62	1.30
Maximum	0.29	2.00	1.59
Holding cost increment ($n=72$)	$h_1-h_2=1$ (%)	$h_1-h_2=4$ (%)	
Average	0.63	0.89	
Median	0.65	1.09	
90th percentile	1.20	1.58	
Maximum	1.64	2.00	
Backlog/holding cost ($n=72$)	$b/h_1=2$ (%)	$b/h_1=10$ (%)	
Average	0.68	0.83	
Median	0.73	1.01	
90th percentile	1.35	1.53	
Maximum	1.66	2.00	
Total ($n=144$)	(%)		
Average	0.76		
Median	0.88		
90th Percentile	1.42		
Maximum	2.00		

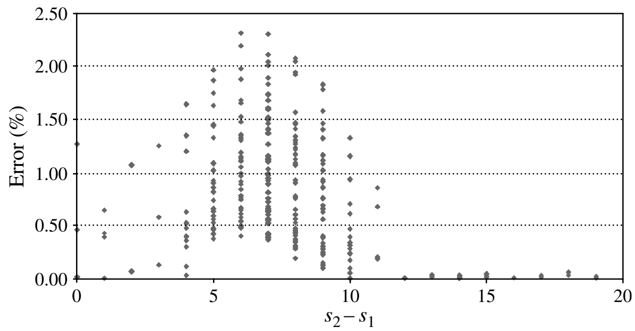
Table 7. Accuracy of cost estimation for five-stage systems.

Demand variability ($n=120$)	Binomial (10, 0.1) (%)	Binomial (2, 0.5) (%)	
Average	0.89	0.99	
Median	0.75	0.89	
90th percentile	1.72	2.09	
Maximum	4.21	5.85	
Maximum lead time ($n=120$)	$L_j^{\max}=5$ (%)	$L_j^{\max}=11$ (%)	
Average	0.80	1.08	
Median	0.73	0.88	
90th percentile	1.69	2.24	
Maximum	3.87	5.85	
Lead time type ($n=80$)	Centered (%)	Uniform (%)	Dispersed (%)
Average	0.78	1.45	0.59
Median	0.75	1.30	0.44
90th percentile	1.37	2.67	1.36
Maximum	2.40	5.85	1.83
Holding cost increment ($n=120$)	$h_1-h_2=1$ (%)	$h_1-h_2=4$ (%)	
Average	0.88	1.00	
Median	0.77	0.83	
90th percentile	1.85	1.87	
Maximum	4.13	5.85	
Backlog/holding cost ($n=120$)	$b/h_1=2$ (%)	$b/h_1=10$ (%)	
Average	0.75	1.13	
Median	0.66	1.00	
90th percentile	1.46	2.40	
Maximum	2.76	5.85	
Total ($n=240$)	(%)		
Average	0.94		
Median	0.79		
90th percentile	1.87		
Maximum	5.85		

The effect of the lead time on the accuracy of the cost estimate is twofold. First, the support of the lead time distribution is either 5, 11, 101, 201, or 301 periods for two-stage systems, and 5 or 11 periods for five-stage systems. We observe that the accuracy of the method does not deteriorate in cases with long lead times, which are highly variable. Changing the support is one way of varying the variability of the lead time distribution. Another way is to change the shape of the distribution, which we have done by using three types of distributions—centered, uniform, and dispersed (see Figure 3). Interestingly, the uniform type has the highest average error, even though it is not the most variable lead time type. This suggests that the variance of the distribution is not the sole determinant of the accuracy of the method; rather, the shape of the distribution is also an important factor. In fact, for five-stage problems, the centered type has a higher average error than the dispersed type. Hence, there seems to be an intricate way in which the shape of the distribution affects the accuracy of the method. The good news is that the average errors are quite small across the board. The effect of the demand variability on the accuracy of the method is not significant in the numerical experiments. The error increases with a higher holding cost increment or a higher backlog to holding cost ratio, but the differences are very small.

For two-stage systems, we checked the accuracy of the single-unit cost estimation against the difference in base-stock levels of the two stages, inspired by Proposition 3.12. Proposition 3.12 states that the single-unit method is exact when this difference is large enough. Figure 4 illustrates this effect. The cost estimation error is plotted against the

Figure 4. The error in cost estimation vs. the difference in base-stock levels of the two stages.



difference in base-stock levels for all 264 instances of two-stage problems with a maximum lead time $L_j^{\max}$ of 11 plus 8 additional instances to make sure that small values of the difference in the base-stock levels are captured as well. As the difference in base-stock levels increases, the error goes to zero. In fact, the error is close to zero far before the value prescribed by Proposition 3.12, which is $L_2^{\max} \cdot D^{\max}$, equaling 110 for half of the instances and 22 for the other half. We see that for values of $s_2 - s_1$ larger than 11, the error is very close to zero.

Finally, we ran a set of experiments to test the cost of ignoring order crossing. In particular, we determined base-stock levels by finding the "lead time demand" distribution and using it to approximate the distribution of pipeline inventory. In systems with i.i.d. lead times, the lead time demand is defined as the total demand over the lead time distribution. This is a widely used method, initially proposed by Hadley and Whitin (1963). We also obtained base-stock levels using the single-unit method. We simulated the systems under both base-stock levels to obtain the corresponding costs. Tables 8 and 9 report the cost of ignoring order crossing for numerical examples with maximum lead time of 5 or 11 and for two- and five-stage systems, respectively. The cost increase is defined as

$$\text{cost increase} = \frac{\text{HW cost} - \text{single-unit cost}}{\text{single-unit cost}},$$

where HW cost is the cost under the Hadley and Whitin approach. The cost increase was positive for all numerical examples that we considered except for one, where the difference was -0.04%.

As the tables demonstrate, ignoring order crossing can be quite costly. The average cost increase was 9.5% for two-stage systems and 6.11% for five-stage systems, and can be as high as 52.78% for two-stage systems and 36.60% for five-stage systems. One interesting thing to note is that for the centered type, the cost of ignoring order crossing was quite small. This is intuitive because the prevalence of order crossing is expected to be much lower in these problems. For the uniform lead time type, the average cost increase is 5.78% and 3.64%, respectively. For the dispersed lead time type, the average cost increase is 21.75%

Table 8. Cost of ignoring order crossing for two-stage systems.

Demand variability ($n = 24$)	Binomial $(10, 0.1)$ (%)	Binomial $(2, 0.5)$ (%)	
Average	6.93	12.07	
Maximum	36.02	52.78	
Maximum lead time ($n = 24$)	5 (%)	11 (%)	
Average	4.23	14.77	
Maximum	19.43	52.78	
Lead time type ($n = 24$)	Centered (%)	Uniform (%)	Dispersed (%)
Average	0.97	5.78	21.75
Maximum	6.54	27.71	52.78
Holding cost increment ($n = 24$)	$h_1 - h_2 = 1$ (%)	$h_1 - h_2 = 4$ (%)	
Average	7.83	11.17	
Maximum	45.98	52.78	
Backlog/holding cost ($n = 24$)	$b/h_1 = 2$ (%)	$b/h_1 = 10$ (%)	
Average	6.16	12.84	
Maximum	52.78	45.98	
Total ($n = 48$)	(%)		
Average	9.50		
Maximum	52.78		

and 14.04%, respectively. Similarly, there is a higher level of cost increase when the support of the lead time distribution is larger, suggesting that the increase would be even more for maximum lead time values of 101, 201, and 301. Overall, we find that the single-unit method offers a substantially better alternative when the system experiences nonnegligible order crossing.

The numerical experiments demonstrate that across a wide range of parameter combinations, the single-unit method is an effective way of determining base-stock levels in multistage systems with stochastic lead times. The base-stock levels are near optimal. The method provides a way of accurately estimating the cost of any given base-stock level. Finally, the cost of ignoring order crossing can be quite substantial.

6. Conclusions

In this paper, we study inventory systems with exogenous stochastic lead times operating under base-stock policies. The class of exogenous lead times is a broad class that includes all previously studied lead time models, and it can also capture phenomena such as history or congestion-dependent lead times. We related the cost of the inventory system with the cost of a corresponding single-unit single-customer problem. For single-stage problems, the relationship enables one to easily optimize the base-stock level or

Table 9. Cost of ignoring order crossing for five-stage systems.

Demand variability ($n=24$)	Binomial (10, 0.1) (%)	Binomial (2, 0.5) (%)	
Average	4.88	7.34	
Maximum	27.15	36.60	
Maximum lead time ($n=24$)	$L_j^{\max}=5$ (%)	$L_j^{\max}=11$ (%)	
Average	2.56	9.67	
Maximum	11.92	36.60	
Lead time type ($n=16$)	Centered (%)	Uniform (%)	Dispersed (%)
Average	0.66	3.64	14.04
Maximum	2.12	12.32	36.60
Holding cost increment ($n=24$)	$h_1-h_2=1$ (%)	$h_1-h_2=4$ (%)	
Average	5.94	6.29	
Maximum	33.46	36.60	
Backlog/holding cost ($n=24$)	$b/h_1=2$ (%)	$b/h_1=10$ (%)	
Average	3.23	9.00	
Maximum	18.26	36.60	
Total ($n=48$)	(%)		
Average	6.11		
Maximum	36.60		

compute the cost of a given base-stock policy by simply solving a single-unit problem. The same is true for multi-stage problems under certain conditions. If those conditions are not satisfied, then the single-unit method is an approximation that yields near-optimal base-stock levels.

The analysis involves the notion of an order-based ordered lead time process $\hat{L}$. The order-based ordered lead time process represents the stochastic durations between the ith order release and the ith order arrival at a given stage for all i. The idea is that the ith order release and the ith order arrival may not correspond to the same physical shipment because order crossing may have taken place. The only relevant information we need about the (possibly very complicated) lead time process is a single random vector (which has one component for every stage): the steady-state distribution of the order-based ordered lead time process $\hat{L}$. This random vector is used as the lead time distribution in the associated single-unit problem. One important implication is that one does not even need to have a model for the overall lead time process as long as one can estimate the distribution of the steady-state random variable $\hat{\mathbf{L}}^{ss}$. This can be done by observing the release and arrival epochs of orders and by fitting a distribution to the observed order-based ordered lead time process. The release and arrival epochs of orders are readily available in most companies. Alternatively, one can also analytically derive the distribution of $\hat{\mathbf{L}}^{ss}$ if the original lead time process is i.i.d. and its distribution is readily available. Another useful fact about the steady-state random variable $\hat{\mathbf{L}}^{ss}$ of the order-based ordered lead time process $\hat{L}$ is that it has the same steady-state distribution as the outstanding order process V.

We used a single-unit approach as the main tool in our analysis. An alternative approach is the conventional analysis based on counting the number of units in various parts of the system by keeping track of inventory positions, net inventories, backlog levels, etc. Many of the results we obtained can also be obtained using this kind of *echelon-based* approach, again by going through the same surrogate problem as we did in our analysis. For systems with deterministic lead times, it is well known that this echelon-based approach uses the distribution of lead time demand in a stage-by-stage recursive algorithm to optimize base stock levels (see §8.3.3 in Zipkin 2000). Svoronos and Zipkin (1991) and Gallego and Zipkin (1999) show that using the lead time demand distribution in such a recursion yields the optimal base-stock levels for systems with exogenous sequential lead times as well. Our single-unit method can be seen as analogous to using the echelon-based recursion with the lead time demand defined as the demand over the steady-state random variable $\hat{\mathbf{L}}^{ss}$ of the order-based ordered lead time process. In fact, using the single-unit algorithm with any given lead time distribution is equivalent to using the echelon-based recursion with the lead time demand over the same lead time distribution.

We believe that extending the methods developed in this paper to more general supply chain configurations with stochastic lead times is a promising future research direction. For example, there is a substantial body of literature on assembly systems and assemble-to-order systems with stochastic lead times (see Song and Zipkin 2003) where order crossing may be a complicating factor.

7. Electronic Companion

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

References

Axsäter, S., K. Rosling. 1993. Installation vs. echelon stock policies for multilevel inventory control. *Management Sci.* **39**(10) 1274–1280.

Bradley, J. R., L. W. Robinson. 2005. Improved base-stock approximations for independent stochastic lead times with order crossover. *Manufacturing Service Oper. Management* **7**(4) 319–329.

Chen, F., B. Yu. 2004. Qualifying the value of lead time information in a single-location inventory system. Working paper, Columbia Business School, New York.

Ehrhardt, R. 1984. (s, S) policies for a dynamic inventory model with stochastic lead times. *Oper. Res.* **32**(1) 121–132.

Gallego, G., P. Zipkin. 1999. Stock positioning and performance estimation in serial production-transportation systems. *Manufacturing Service Oper. Management* **1**(1) 77–88.

Hadley, G., T. M. Whitin. 1963. *Analysis of Inventory Systems*. Prentice Hall, Englewood Cliffs, NJ.

Kaplan, R. S. 1970. A dynamic inventory model with stochastic lead times. *Management Sci.* **16**(7) 491–507.

Karlin, S., H. M. Taylor. 1975. *A First Course in Stochastic Processes*. Academic Press, New York.

Muharremoglu, A., J. N. Tsitsiklis. 2008. A single-unit decomposition approach to multiechelon inventory systems. *Oper. Res.* **56**(5) 1089–1103.

Nahmias, S. 1979. Simple approximations for a variety of dynamic lead time lost-sales inventory models. *Oper. Res.* **27**(5) 904–924.

Robinson, L. W., J. R. Bradley. 2008. Further improvements on base-stock approximations for independent stochastic lead times with order crossover. *Manufacturing Service Oper. Management* **10**(2) 325–327.

Robinson, L. W., J. R. Bradley, L. J. Thomas. 2001. Consequences of order crossover under order-up-to inventory policies. *Manufacturing Service Oper. Management* **3**(3) 175–188.

Song, J. S., P. H. Zipkin. 1996a. Inventory control with information about supply conditions. *Management Sci.* **42**(10) 1409–1419.

Song, J., P. Zipkin. 1996b. The joint effect of lead time variance and lot size in a parallel processing environment. *Management Sci.* **42**(9) 1352–1363.

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

Svoronos, A., P. Zipkin. 1991. Evaluation of one-for-one replenishment policies for multiechelon inventory systems. *Management Sci.* **37**(1) 68–83.

Zalkind, D. 1978. Order-level inventory systems with independent stochastic lead times. *Management Sci.* **24**(13) 1384–1392.

Zipkin, P. 1986. Stochastic lead time in continuous-time inventory models. *Naval Res. Logist.* **33**(4) 763–774.

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