

# Inventory Management With an Exogenous Supply Process

**Alp Muharremoglu**

Graduate School of Business

Columbia University

New York, NY, 10027

e-mail: alp2101@columbia.edu

**Nan Yang**

Johnson School of Management

Cornell University

Ithaca, NY, 14853

e-mail: ny38@cornell.edu

November 24, 2008

## Abstract

We study single and multi-stage inventory systems with stochastic leadtimes. We study a class of stochastic leadtime processes, which we refer to as *exogenous* leadtimes. This class of leadtime processes includes as special cases all leadtime models from existing literature (such as Kaplan's leadtimes with no order crossing, see Kaplan (1970), or i.i.d. leadtimes with order crossing, among others) but is a substantially broader class. For a system with an exogenous leadtime 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 multi-stage systems under certain conditions. If the conditions are not satisfied, the method obtains near optimal base stock levels and accurate approximations of cost for multi-stage systems.

# 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, leadtime uncertainty and capacity uncertainty. These different ways of thinking about supply uncertainty are related to each other, as for example, one can think about capacity shortage at a supplier leading to a longer leadtime for the delivery of orders. In this paper, we address the issue of supply uncertainty in the form of uncertainty in leadtimes. We study *exogenous* leadtimes, which is a broad class of stochastic leadtimes, that includes sequential and order-crossing leadtime processes. We study both single and serial multi-stage systems.

There are numerous papers in the inventory management literature that deal with the issue of uncertain leadtimes. An early reference is the book of Hadley & Whitin (1963) (Chapter 5.14), who study an inventory model with independent and identically distributed leadtimes. 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. leadtime, i.e., the leadtime demand. As Hadley & Whitin (1963) point out, the two assumptions of leadtimes being independent and no order-crossing can be satisfied at the same time only under specific conditions. For example, if leadtimes are smaller than the review period, then the analysis is exact. Another example is the case when the smallest possible leadtime and the largest possible leadtime 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. leadtimes can give rise to complicated optimal policies that depend on the state of the pipeline inventory vector. Zalkind (1978) gives an example demonstrating the non-optimality of base stock policies when leadtimes are i.i.d. Still, the predominant approach for dealing with leadtime uncertainty, both in the literature and in practice has been to use the method of Hadley & Whitin (1963), i.e. to assume i.i.d. leadtimes and to use the leadtime 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 leadtime demand. In many cases, leadtime demand is a good approximation for pipeline inventory, but in many other cases, using such an approximation leads to substantial suboptimality, as demonstrated in Section 5.

There are some notable exceptions to Hadley & Whitin (1963)’s paradigm, that fall into two main categories. The first set of papers assume a certain type of non-order crossing stochastic leadtime process, initially proposed by Kaplan (1970). Kaplan (1970) shows the optimality of base stock policies in a single stage, periodic review setting under his proposed leadtime 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 leadtime process. Zipkin (1986) extends Kaplan’s leadtime model to include a more general class of non-order crossing stochastic leadtime processes, which he denotes as “exogenous, sequential leadtimes” (Zipkin 2000, see Chapter 7.4). Svoronos & Zipkin (1991) evaluate one-for-one replenishment policies in serial systems with such leadtimes. Muharremoglu & Tsitsiklis (2008) show that echelon base stock policies are optimal for serial systems with Kaplan’s non-order crossing stochastic leadtime processes. All these papers assume that orders do not cross.

The second set of papers that diverge from Hadley & Whitin (1963)’s paradigm are Zalkind (1978), Robinson, Bradley & Thomas (2001), Bradley & Robinson (2005) and Robinson & 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 leadtimes are taken to be independently and identically distributed. The class of i.i.d. leadtimes 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. leadtimes. 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 & Robinson (2005) and Robinson & Bradley (2008) improve upon this approximation with tighter upper bounds for the variance of the pipeline inventory. Song & Zipkin (1996*b*) study the effect of leadtime variance in an  $(r, q)$  system with i.i.d. leadtimes and discuss the relationship of i.i.d. leadtimes with exogenous sequential leadtimes. Note that all these papers assume i.i.d. leadtimes in a single stage setting.

We analyze a set of leadtime processes that is more general than previously studied leadtime models. We refer to this class as *exogenous* leadtime processes. In fact, to our knowledge, all leadtime models from existing literature are in the class of exogenous processes. For example, both Kaplan’s leadtimes with no order crossing and i.i.d. leadtimes with order crossing are special cases of exogenous leadtime processes, as are the Markov modulated leadtime models of Song & Zipkin (1996*a*) and Chen & Yu (2004), among many others. Leadtimes 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, *iii*) when applied to multi-stage systems, (echelon) base stock policies can be implemented in a decentralized fashion as shown by Axsäter & 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* leadtime process. For single stage systems, the method finds the optimal base stock levels and provides the exact cost. For multi-stage 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 & Whitin (1963) and Zalkind (1978). The idea is to assume that orders are being released at every period, and a stochastic leadtime 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 leadtime 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 ‘leadtime process of stage  $j$ ’. The joint process of all stages constitutes the ‘leadtime process’. A leadtime process is *exogenous*, if the leadtime process of any stage is independent from the leadtime processes of all other stages and independent of the demand process. We also assume that the leadtime process is ergodic throughout the paper. (For a formal definition of ergodicity see Section 2). The class of *exogenous* leadtime processes is an extension of the *exogenous-sequential leadtime* class that Zipkin (1986) introduced, in that we relax the sequential restriction. The ergodicity assumption is reasonable, if the conditions underlying the leadtime process remain stable over a relatively long period, as compared to the frequency of the shipment decisions. The assumption of the leadtimes 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* leadtimes. For example, if a third party logistics provider adopts a congestion dependent dispatch policy, where if more than a certain 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 leadtimes. The method is based on relating the original problem to a single-unit, single-customer problem (Proposition 3.3). In contrast to Muharremoglu & 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 leadtimes, by removing the sequential leadtimes condition. It also extends Zalkind (1978), who studies single stage systems with i.i.d. leadtimes, by allowing more general stochastic leadtime processes.
- (c) For multi-stage 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 leadtime, when the difference between base-stock levels is zero (Proposition 3.13). This extends Svoronos & Zipkin (1991), which studies multi-stage systems with exogenous, sequential leadtimes, by removing the sequential leadtimes condition.
- (d) For multi-stage 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. (Section 5).

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

The rest of the paper is organized as follows. Section 2 formulates the problem. Section 3 contains the main results. This is followed by Section 4 which discusses estimation issues. Section 5 provides

	Single Stage	Multi Stage
Kaplan's leadtimes	Kaplan (1970)	Muharremoglu & Tsitsiklis (2008)
Exogenous and sequential leadtimes	Zipkin (1986)	Svoronos & Zipkin (1991)
i.i.d. leadtimes	Zalkind (1978), Robinson et al. (2001)	This paper <sup>1</sup>
Exogenous leadtimes	This paper	This paper <sup>1</sup>

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

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 Section 6.

## 2 Problem Formulation

We consider an  $M$ -stage serial system with stochastic leadtimes. 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  $j = 1, \dots, M - 1$ ) 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 leadtime 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, 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  $\{X_j(i)\}, i = 1, \dots, \infty$  and  $j = 1, \dots, M - 1$  be a stochastic process. We denote vectors using bold letters, i.e.,  $\mathbf{X}(i) = (X_1(i), \dots, X_{M-1}(i))$  for all  $i$ . We refer to the process as stochastic process  $X$ . The realization of this process under sample path  $\omega$  is  $\mathbf{x}(i, \omega), i = 1, \dots, \infty$ . We use the notation  $\mathbf{X}^{ss}$  to denote the steady state random vector of the process  $X$ .

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

**Definition 2.2.** (a) **The order-based ordered leadtime process**  $\hat{L}_j(t)$  is the time between the  $t^{\text{th}}$  order release from stage  $j + 1$  and the  $t^{\text{th}}$  order arrival at stage  $j$ . Note that these orders

do not have to be the same order, since when orders cross, orders do not arrive in the same sequence they were released.

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

Figure 1 illustrates the order-based ordered leadtime process, as well as the unit-based ordered leadtime process. Part a) of the figure depicts the leadtimes of three successive orders placed at  $t = 1, 2$  and  $3$ . Note that order 2 has a size of 0. Part b) of the figure depicts the order-based order leadtime 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 leadtime goes from period 1 to period 9 in Figure 1 (b), and thus equals 8. Part c) depicts the leadtimes that the actual units experience. Units 1 and 2 are part of Order 1, and therefore have the same leadtime as Order 1. Unit 3 is part of Order 3, and has the same leadtime as Order 3. Part d) depicts the unit-based ordered leadtime 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 leadtime goes from period 3 to period 11 in Figure 1 (d), and thus equals 8.

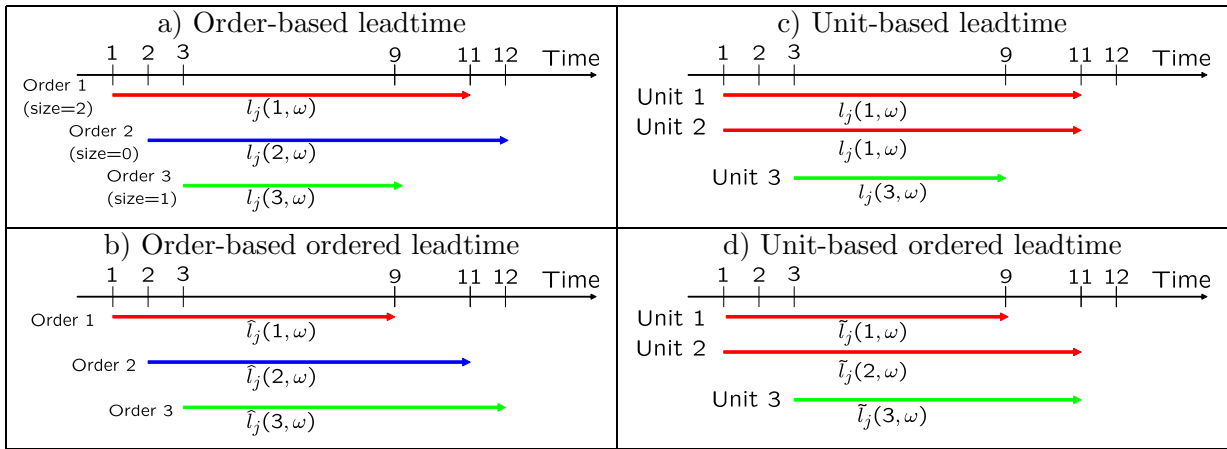


Figure 1: A realization of the leadtime processes and their ordered versions (under sample path  $\omega$ ). Each arrow starts at the departure time and ends at the arrival time of an order/unit.

We assume the ergodicity of the original leadtime process, as well as the ordered leadtime 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 & Taylor 1975, see Theorem 5.6):*

(a)

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

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

(b)

$$\lim_{i \rightarrow \infty} \Pr(\mathbf{X}(i) = \mathbf{x}) = f^X(\mathbf{x})$$

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 leadtimes. Leadtimes are sequential, if  $t + L_j(t)$  is non-decreasing 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 cost for both single and multi-echelon problems. Finally, we analyze the case of non-sequential leadtimes. In this case, the single-unit method is exact for single stage problems and for multi-stage problems under certain conditions, but is an approximation otherwise. In Section 5, we study the performance of the algorithm for multi-stage problems under non-sequential leadtimes 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 & Tsitsiklis (2008). We use the concepts of *the location of a unit* and *the position of a customer* as defined in Muharremoglu & Tsitsiklis (2008). For convenience, we include these definitions in the Appendix. 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 leadtime 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  $y^{\text{th}}$  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, and 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 & 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 & 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 re-labeled 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.

Since 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  $i^{\text{th}}$  unit and  $i^{\text{th}}$  customer, i.e., the  $i^{\text{th}}$  unit-customer pair. From now on, whenever we refer to the  $i^{\text{th}}$  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 non-negative 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

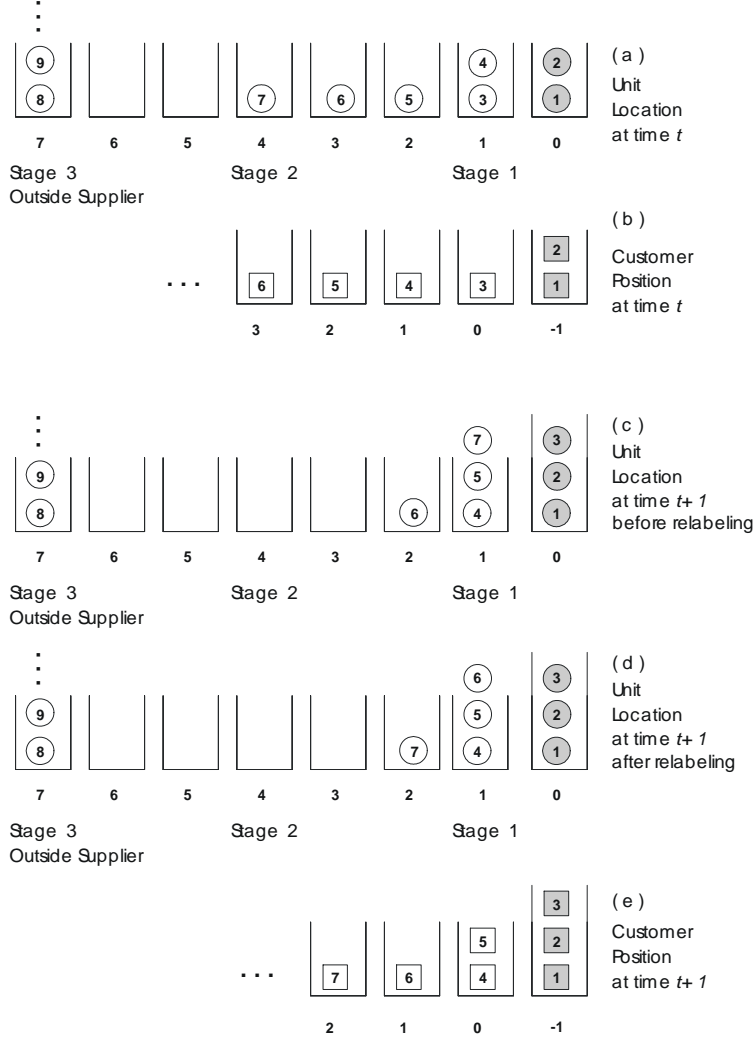


Figure 2: Illustration of the unit positions and customer locations and the re-labeling of units after order crossing.

on inventory positions translates into release/don't release decisions for single unit, single customer pairs. The timing of these threshold crossings and the leadtimes 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  $R_j(i) = \min\{t|y_t^i < s_j\}$ , 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

$$\mathbf{W}(i) = (W_1(i), W_2(i), \dots, W_{M-1}(i)).$$

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 re-labeling of the units in every period and the definition of the unit-based ordered leadtime 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

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

for all  $\omega$ . Note that  $f^{\tilde{L}, W}(\mathbf{l}, \mathbf{w})$  represents the long-run fraction of unit-customer pairs for which  $(\tilde{\mathbf{I}}(i, \omega) = \mathbf{l}, \mathbf{w}(i, \omega) = \mathbf{w})$ . 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  $\omega$ . We can write the total infinite horizon average cost for base stock policy  $\mathbf{s}$  as

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

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)*

$$C(\mathbf{s}) = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=\lceil \epsilon T + 1 \rceil}^{\lceil (\bar{d} - \epsilon) T \rceil} g(\tilde{\mathbf{I}}(i, \omega), \mathbf{w}(i, \omega)) + h(\epsilon), \quad \text{where } \lim_{\epsilon \rightarrow 0} h(\epsilon) = 0.$$

*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 since the cost for pair  $i$  is determined by  $(\tilde{\mathbf{I}}(i, \omega), \mathbf{w}(i, \omega))$ , 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  $\omega$  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 & 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 leadtimes have a special structure, in order 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 & 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, since the process  $W$  represents the durations between threshold crossings, i.e., 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 Section 3.1, we study the case where orders do not cross. In Section 3.2, we study the case where orders are allowed to cross, and give an exact method for single stage problems. For multi-stage problems, the method is exact under some conditions. If the conditions are not satisfied, the method is an approximation.

### 3.1 Sequential Leadtime Processes

In this subsection, we develop an exact method for optimizing the base stock levels in a multi-echelon system where orders do not cross, i.e. systems with sequential leadtime 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 leadtime process, which can be very cumbersome. Rather,

it requires a single leadtime distribution for a single unit, hence one needs to collect data and fit a distribution for a single random variable for the leadtime of each stage of the supply chain and a single random variable for the demand. The leadtime distribution needed by the method is the steady state distribution  $\mathbf{L}^{ss}$  of the original leadtime process. Systems with ergodic and sequential leadtimes were studied by Zipkin (1986) and Svoronos & Zipkin (1991) as well. We develop an alternative method for problems with ergodic and sequential leadtimes 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 Section 3.2.

Proposition 3.4 a) below states that the order-based ordered leadtime process and the unit-based ordered leadtime process have the same steady state distribution as the original leadtime process, if the leadtimes are sequential. The order-based leadtime  $L$  and the order-based ordered leadtime  $\hat{L}$  are of course identical in this case. The sequential nature of the leadtimes, coupled with the fact that the order sizes for a particular stage are independent of the leadtimes for that stage allow us to show that the steady state distribution  $f^{\hat{L}}(\mathbf{l})$  of the order-based ordered leadtime and the steady state distribution  $f^{\tilde{L}}(\mathbf{l})$  of the unit-based ordered leadtime 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 multi-stage problem and the cost of a single-unit problem.

**Proposition 3.4.** *If the leadtime 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 leadtime 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 Section 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 15<sup>th</sup> customer to walk through the door, counting from time 0. Note that we don’t know when ‘the customer’ will show up and who she is, we just know that she is the 15<sup>th</sup> 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 (to which we refer as ‘the unit’). Letting  $z_t$  be the location of the unit, we have  $z_0 = M$ . Let  $\mathbf{X} = (X_1, \dots, X_{M-1})$  be the vector of leadtimes 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 leadtime vector  $\mathbf{X}$ . Let the sequence of functions  $(\mu_1(y), \mu_2(y), \dots, \mu_{M-1}(y))$  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  $\mathbf{s} = (s_1, \dots, s_{M-1})$  such that*

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

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

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

Consider the version of the single unit problem where the leadtimes have the steady state distribution of the original leadtime 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  $C^S(\mathbf{s}, \mathbf{L}^{ss}) = \lim_{y_0 \rightarrow \infty} J(y_0, \mathbf{s}, \mathbf{L}^{ss})$ .  $C^S(\mathbf{s}, \mathbf{L}^{ss})$  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 leadtime process for the original problem is sequential, then*

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

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 leadtime 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:*

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

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, since 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 leadtimes are sequential. Note that this result was achieved, even though the original problem (even if leadtimes are sequential) does not decompose into single unit, single customer problems in the sense of the separability of optimal controls as in Muharremoglu & 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 Section 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 leadtime process, which can be quite complicated. Next on the agenda is to study the case where the leadtime process is not sequential.

### 3.2 Non-sequential Leadtime Processes

In this section, we study systems where orders may cross. The leadtime process is assumed to be exogenous and ergodic. Note that Proposition 3.3 still holds even if leadtimes 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 multi-stage 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 leadtime process  $\hat{L}$ .
- (b) The infinite horizon average cost of a single stage problem depends on the leadtime 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 leadtime process.

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

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

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

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  $\lambda$  is equal to 1. 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  $i^{th}$  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 leadtime process. In particular, the leadtime process for the surrogate problem is defined to be stochastically identical to the order-based ordered leadtime process  $\hat{L}$  of the original problem. In particular, given a sample path of the leadtime process for the original problem, there is a corresponding sample path of the leadtime 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 leadtime dynamics of the

	Time	3	4	5	6	7	8	9	10	11	12	13	14
Original Problem	Release Time	2	1	3		5 & 6	4	7	9			8, 11 & 12	10
	Order Size	10	3	7		8 & 6	5	1	0			2, 15 & 4	12
Surrogate Problem	Release Time	1	2	3		4 & 5	6	7	8			9, 10 & 11	12
	Order Size	3	10	7		5 & 8	6	1	2			0, 12 & 15	4

Table 2: An example illustrating the leadtime dynamics of the original problem and its surrogate problem. 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.

original problem and the surrogate problem. In period 7, two orders arrive in the original problem. The first order was released at time 5, the second 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 one was released at time 4 and the second 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 leadtime  $L(1) = 3$ . The order-based ordered leadtime  $\hat{L}(1)$  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 leadtime  $L^{Surrogate}(1) = 2 = \hat{L}(1)$ . 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 leadtime process was not sequential, the leadtime 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 multi-stage 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 leadtime process only through its effect on  $\hat{L}^{ss}$ ; i.e., if two systems differ only in their leadtime processes yet have the same steady state distribution  $\hat{L}^{ss}$  of the order-based ordered leadtime 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.*

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

Part a) of Proposition 3.10 relies on the fact that the order sizes in a single stage system under a base stock policy are realizations of the i.i.d. demand distribution and that the steady state distribution of the order-based ordered leadtime 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 distribution 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 leadtime 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  $\hat{L}^{ss}$  in order 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 Multi-Stage Systems

Note that Proposition 3.9 is valid for multi-stage systems. Similar to the single stage case, one can talk about a surrogate problem for a multi-stage system as well. The surrogate problem has the same steady state distribution ( $\hat{\mathbf{L}}^{ss}$ ) of the order-based ordered leadtime process, and furthermore, the surrogate problem can be optimized by solving its related single unit problem  $S1(\hat{\mathbf{L}}^{ss})$ . In this subsection, we show that the cost of the surrogate problem is equal to the cost of the original problem for multi-stage 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 multi-stage 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, since the steady state distribution of the order-based ordered leadtime  $\hat{\mathbf{L}}^{ss}$  is equal to  $\mathbf{V}^{ss}$  (by Proposition 3.9) and the distribution of the order-based ordered leadtime 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. Since 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  $M - 2, \dots, 1$ . Therefore, the sum of expected costs incurred at stages  $M - 2, \dots, 1$  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 multi-stage 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 multi-stage 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 multi-stage 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 Section 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  $\mathbf{s} = (s_1, \dots, s_{M-1})$ . Suppose that  $s_j - s_{j-1} \geq L_j^{\max} \cdot D^{\max}$  for all  $j = 2, \dots, M - 1$ , where  $D^{\max}$  is the maximum level of demand in any period and  $L_j^{\max}$  is the maximum leadtime 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 leadtime 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 leadtimes, 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 leadtime 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 0, the method is exact. This means that for both small and large values of the difference, the method is expected to do well. In Section 5, we indeed show that the performance of the approximation is quite good over a diverse set of instances.

## 4 Estimation Issues

In Sections 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 leadtimes of the unit are distributed as the random variable  $\hat{\mathbf{L}}^{ss}$ . The detailed definition of this single unit problem was given in Section 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 multi-stage problems under certain conditions. Otherwise, they are not guaranteed to be optimal, but are near-optimal, as demonstrated in Section 5.

In order 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 leadtime 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 leadtime process and fitting parameters to that model. One then needs to derive the distribution of  $\hat{\mathbf{L}}^{ss}$  from this estimated original leadtime process. Such a derivation is already available for the case of i.i.d. leadtimes. Zalkind (1978) assumes that the original leadtime 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 leadtime 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 leadtime process and to fit a distribution to the observed realizations. The realizations are obtained by subtracting the time of the  $i^{th}$  order release from the time of the  $i^{th}$  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 leadtime process, which is all that is needed for the solution procedure. The leadtime process  $L$  and the order-based ordered leadtime 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. In order 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 leadtime 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 leadtime process  $L_j^+(i)$  is the time between*

the  $i^{\text{th}}$  positive-sized order release from stage  $j + 1$  and the  $i^{\text{th}}$  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  $\mathbf{V}^{+ss}$ . In order to optimize base stock levels, we need the steady state distribution  $\hat{\mathbf{L}}^{ss}$  of the order-based ordered leadtime 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 leadtime 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 leadtime value recorded during the observation phase of the estimation procedure. Let

$$\mathbf{f}_j = \left( \Pr(V_j^{+ss} = 1), \Pr(V_j^{+ss} = 2), \dots, \Pr(V_j^{+ss} = L_j^{\max}) \right)'$$

and

$$\mathbf{g}_j = \left( \Pr(\hat{L}_j^{ss} = 1), \Pr(\hat{L}_j^{ss} = 2), \dots, \Pr(\hat{L}_j^{ss} = L_j^{\max}) \right)'$$

**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 leadtime 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

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

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 leadtime processes are equal, when leadtimes are sequential between two stages. When leadtimes 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 order 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 leadtimes 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 leadtimes 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 leadtimes, i.e., Proposition 4.3(a) applies to stages  $1, \dots, M-2$ .
- 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 leadtime 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 multi-stage 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 leadtimes 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 leadtime process  $L^+$  and/or the the number of positive-sized 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 leadtime process and/or  $\mathbf{V}^{+ss}$ , the steady state distribution of the number of positive-sized 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 leadtime 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 leadtime 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 multi-stage 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 leadtime and demand distributions to test a wide variety of problem instances. The leadtime process for a given stage is assumed to be an independent and identically distributed process where the leadtime 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 leadtime distributions, which are illustrated in Figure 3. All leadtime 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 leadtime distribution is used for all stages of the system.

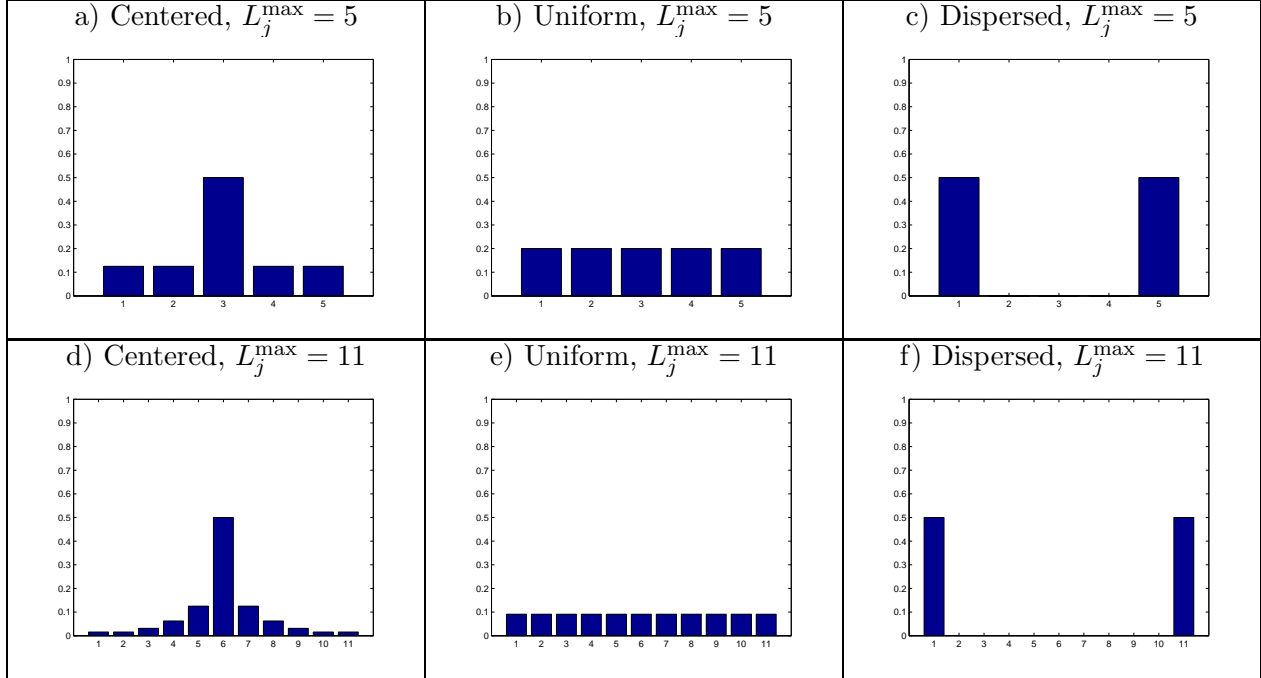


Figure 3: The types of leadtime distributions.

The demand is assumed to have a binomial distribution with mean 1. In half of the experiments, we used a binomial distribution with 2 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 1 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 1 or 4. For example, when the holding cost increment is 4,  $h_j - h_{j+1} = 4$  for all stages  $j$ . Finally, the backlog cost rate is assumed to be proportional 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:

$$\text{Loss of Optimality} = \frac{\text{Simulated Cost of } \mathbf{s}_u - \text{Simulated Cost of } \mathbf{s}^*}{\text{Simulated Cost of } \mathbf{s}^*}$$

Table 3: The effectiveness of the single-unit method for two stage systems. 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  $L_j^{\max} = 5$ ).

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. Leadtime ( $n = 24$ )	$L_j^{\max} = 5$	$L_j^{\max} = 11$	$L_j^{\max} = 101$	$L_j^{\max} = 201$	$L_j^{\max} = 301$
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
Leadtime 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				

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 leadtimes, due to the discreteness of inventory levels. For example, with a maximum leadtime of 5, the optimal base stock levels are on the order of (5, 10),

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 Leadtime ( $n = 24$ )	$L_j^{\max} = 5$	$L_j^{\max} = 11$	
Average Loss	0.0768%	0.1021%	
Maximum Loss	0.4394%	0.2976%	
Total optimal	11	5	
Leadtime 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		

whereas with a maximum leadtime 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 leadtimes. 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 leadtimes. The average loss of optimality in the five non-optimal cases with  $L_j^{\max}$  values of 5 or 11 is 0.1458%, whereas the average loss of optimality in the 42 non-optimal cases with  $L_j^{\max}$  values of 101, 201 or 301 is 0.0821%. Table 10 in Appendix C 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 leadtime distribution, we observe that the effectiveness of the method is best for the *centered* leadtime type for both two-stage and five-stage systems. This is intuitive, since 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 either to find a set of base stock levels or we can use it 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 non-optimal 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:

$$\text{Error} = \frac{|\text{Estimated Cost} - \text{Simulated Cost}|}{\text{Simulated Cost}} \quad (6)$$

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

Table 5: Accuracy of Cost Estimation for Two-Stage Systems, Instances with Maximum Leadtime of 5 or 11. (The percentages reported are the error values as defined in Equation (6)).

Demand Variability ( $n = 264$ )	Binomial (10, 0.1)	Binomial (2, 0.5)	
Average	0.75%	0.78%	
Median	0.67%	0.70%	
90 <sup>th</sup> Percentile	1.41%	1.51%	
Maximum	2.04%	2.31%	
Maximum Leadtime ( $n = 264$ )	$L_j^{\max} = 5$	$L_j^{\max} = 11$	
Average	0.70%	0.82%	
Median	0.66%	0.74%	
90 <sup>th</sup> Percentile	1.31%	1.60%	
Maximum	2.30%	2.31%	
Leadtime Type ( $n = 176$ )	Centered	Uniform	Dispersed
Average	0.52%	1.04%	0.72%
Median	0.51%	1.08%	0.72%
90 <sup>th</sup> 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%	
90 <sup>th</sup> 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%	
90 <sup>th</sup> Percentile	1.34%	1.55%	
Maximum	2.31%	2.30%	
Total ( $n = 528$ )			
Average	0.76%		
Median	0.68%		
90 <sup>th</sup> Percentile	1.47%		
Maximum	2.31%		

five-stage systems. The maximum error over all instances is 2.31% and 5.85%, respectively.

The effect of the leadtime on the accuracy of the cost estimate is two-fold. First, the support of the leadtime 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 leadtimes, which are highly variable. Changing the support is one way of varying the variability of the leadtime 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 leadtime type. This suggests that the variance of the distribution is not the sole determinant of the

Table 6: Accuracy of Cost Estimation for Two-Stage Systems, Instances with Maximum Leadtime 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%	
90 <sup>th</sup> Percentile	1.42%	1.43%	
Maximum	1.82%	2.00%	
Maximum Leadtime ( $n = 48$ )	$L_j^{\max} = 101$	$L_j^{\max} = 201$	$L_j^{\max} = 301$
Average	0.85%	0.74%	0.68%
Median	1.06%	0.90%	0.81%
90 <sup>th</sup> Percentile	1.60%	1.42%	1.24%
Maximum	2.00%	1.61%	1.61%
Leadtime Type ( $n = 48$ )	Centered	Uniform	Dispersed
Average	0.08%	1.26%	0.92%
Median	0.06%	1.24%	0.97%
90 <sup>th</sup> 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%	
90 <sup>th</sup> 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%	
90 <sup>th</sup> Percentile	1.35%	1.53%	
Maximum	1.66%	2.00%	
Total ( $n = 144$ )			
Average	0.76%		
Median	0.88%		
90 <sup>th</sup> 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%	
90 <sup>th</sup> Percentile	1.72%	2.09%	
Maximum	4.21%	5.85%	
Maximum Leadtime ( $n = 120$ )	$L_j^{\max} = 5$	$L_j^{\max} = 11$	
Average	0.80%	1.08%	
Median	0.73%	0.88%	
90 <sup>th</sup> Percentile	1.69%	2.24%	
Maximum	3.87%	5.85%	
Leadtime Type ( $n = 80$ )	Centered	Uniform	Dispersed
Average	0.78%	1.45%	0.59%
Median	0.75%	1.30%	0.44%
90 <sup>th</sup> 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%	
90 <sup>th</sup> 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%	
90 <sup>th</sup> Percentile	1.46%	2.40%	
Maximum	2.76%	5.85%	
Total ( $n = 240$ )			
Average	0.94%		
Median	0.79%		
90 <sup>th</sup> Percentile	1.87%		
Maximum	5.85%		

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 difference in base stock levels for all 264 instances of two stage problems with a maximum leadtime  $L_j^{\max}$  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 0.

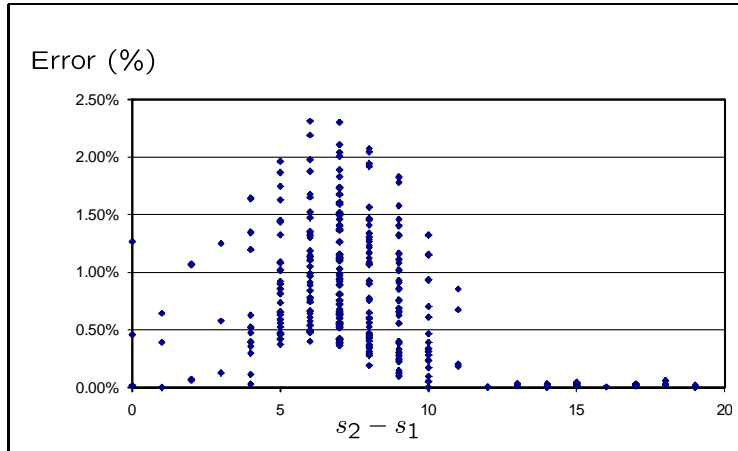


Figure 4: The error in cost estimation vs. the difference in base stock levels of the two stages.

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 ‘leadtime demand’ distribution and using it to approximate the distribution of pipeline inventory. In systems with i.i.d. leadtimes, the leadtime demand is defined as the total demand over the leadtime distribution. This is a widely used method, initially proposed by Hadley & 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 leadtime 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, since the prevalence of order crossing is expected to be much lower in these problems. For the uniform leadtime type, the

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 Leadtime ( $n = 24$ )	5	11	
Average	4.23%	14.77%	
Maximum	19.43%	52.78%	
Leadtime 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%		

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 Leadtime ( $n = 24$ )	$L_j^{\max} = 5$	$L_j^{\max} = 11$	
Average	2.56%	9.67%	
Maximum	11.92%	36.60%	
Leadtime 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%		

average cost increase is 5.78% and 3.64%, respectively. For the dispersed leadtime type, the average cost increase is 21.75% and 14.04% respectively. Similarly, there is a higher level of cost increase when the support of the leadtime distribution is larger, suggesting that the increase would be even more for maximum leadtime values of 101, 201 and 301. Overall, we find that the single-unit method offers a

substantially better alternative when the system experiences non-negligible 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 multi-stage systems with stochastic leadtimes. 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

We studied inventory systems with exogenous stochastic leadtimes operating under base stock policies. The class of exogenous leadtimes is a broad class that includes all previously studied leadtime models, and it can also capture phenomena such as history or congestion dependent leadtimes. 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 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 leadtime process  $\hat{L}$ . The order-based ordered leadtime process represents the stochastic durations between the  $i^{th}$  order release and the  $i^{th}$  order arrival at a given stage, for all  $i$ . The idea is that the  $i^{th}$  order release and the  $i^{th}$  order arrival may not correspond to the same physical shipment, as order crossing may have taken place. The only relevant information we need about the -possibly very complicated- leadtime process is a single random vector (which has one component for every stage): the steady state distribution of the order-based ordered leadtime process  $\hat{L}$ . This random vector is used as the leadtime distribution in the associated single-unit problem. One important implication is that one does not even need to have a model for the overall leadtime 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 leadtime 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 leadtime 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 leadtime 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 leadtimes, it is well known that this echelon based approach uses the distribution of leadtime demand in a stage-by-stage recursive algorithm to optimize base stock levels (see Section 8.3.3 in Zipkin (2000)). Svoronos & Zipkin (1991) and Gallego & Zipkin (1999) show that using the leadtime demand distribution in such a recursion yields the optimal base stock levels for systems with exogenous, sequential leadtimes as well. Our single-unit method can be seen as analogous to using the echelon based recursion with the leadtime demand defined as the demand over the steady state random variable  $\hat{\mathbf{L}}^{ss}$  of the order-based ordered leadtime process. In fact, using the single-unit algorithm with any given leadtime distribution is equivalent to using the echelon based recursion with the leadtime demand over the same leadtime distribution.

We believe that extending the methods developed in this paper to more general supply chain configurations with stochastic leadtimes 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 leadtimes (see Song & Zipkin (2003)), where order-crossing may be a complicating factor.

## Endnote

<sup>1</sup> The multi-stage analysis is exact in the four cases outlined in item (c) of the list of contributions given in the Introduction section, approximate otherwise.

## References

- Axsäter, S. & Rosling, K. (1993), ‘Installation vs. Echelon Stock Policies for Multi-Level Inventory Control’, *Management Science* **39**, 1274–1280.
- Bradley, J. R. & Robinson, L. W. (2005), ‘Improved Base-Stock Approximations for Independent Stochastic Lead Times with Order Crossover’, *Manufacturing & Service Operations Management* **7**(4), 319–329.
- Chen, F. & Yu, B. (2004), ‘Qualifying the Value of Leadtime Information in a Single-Location Inventory System’, Working paper.
- Ehrhardt, R. (1984), ‘ $(s, S)$  Policies for A Dynamic Inventory Model with Stochastic Leadtimes’, *Operations Research* **32**(1), 121–132.
- Gallego, G. & Zipkin, P. (1999), ‘Stock Positioning and Performance Estimation in Serial Production-Transportation Systems’, *Manufacturing and Service Operations Management* **1**, 77–88.

- Hadley, G. & Whitin, T. M. (1963), *Analysis of Inventory Systems*, Prentice Hall.
- Kaplan, R. S. (1970), ‘A Dynamic Inventory Model with Stochastic Lead Times’, *Management Science* **16**(7), 491–507.
- Karlin, S. & Taylor, H. M. (1975), *A First Course in Stochastic Processes*, Academic Press.
- Muharremoglu, A. & Tsitsiklis, J. N. (2008), ‘A Single-Unit Decomposition Approach to Multi-Echelon Inventory Systems’, *Operations Research* **56**(5), 1089–1103.
- Nahmias, S. (1979), ‘Simple Approximations for A Variety of Dynamic Leadtime Lost-sales Inventory Models’, *Operations Research* **27**(5), 904–924.
- Robinson, L. W. & Bradley, J. R. (2008), ‘Further Improvements on Base-Stock Approximations for Independent Stochastic Lead Times with Order Crossover’, *Manufacturing & Service Operations Management* **10**(2), 325–327.
- Robinson, L. W., Bradley, J. R. & Thomas, L. J. (2001), ‘Consequences of Order Crossover Under Order-up-to Inventory Policies’, *Manufacturing & Service Operations Management* **3**(3), 175–188.
- Song, J. S. & Zipkin, P. H. (1996a), ‘Inventory Control with Information about Supply Conditions’, *Management Science* **42**(10), 1409–1419.
- Song, J. S. & Zipkin, P. H. (2003), *Supply chain operations: assemble-to-order systems*, in *Handbooks in Operations Research and Management Science, Vol. XXX: Supply Chain Management*, North-Holland, chapter 11.
- Song, J. & Zipkin, P. (1996b), ‘The Joint Effect of Leadtime Variance and Lot Size in a Parallel Processing Environment’, *Management Science* **42**(9), 1352–1363.
- Svoronos, A. & Zipkin, P. (1991), ‘Evaluation of One-for-one Replenishment Policies for Multi-echelon Inventory Systems’, *Management Science* **37**(1), 68–83.
- Zalkind, D. (1978), ‘Order-level Inventory Systems with Independent Stochastic Leadtimes’, *Management Science* **24**(13), 1384–1392.
- Zipkin, P. (1986), ‘Stochastic Leadtime in Continuous-time Inventory Models’, *Naval Research Logistics* **33**(4), 763–774.
- Zipkin, P. H. (2000), *Foundations of Inventory Management*, McGraw-Hill Book Co.

# Appendix for “Inventory Management With an Exogenous Supply Process” by Alp Muharremoglu and Nan Yang

## A Definitions

Next, we formally define the concept of a unit-customer pair. Towards this goal, we start by associating a unique label with each unit and customer. At any given time, there will be a number of units at each stage or on order between two given stages. In addition, conceptually, we have a countably infinite number of units at the outside supplier, which we call stage  $M$ . We will now introduce a set of conceptual locations in the system that can be used to describe where a unit is found and, if it is part of an outstanding order, how long ago it was ordered.

**Definition A.1.** The location of a unit: *First, each of the actual stages in the system will constitute a location. Next, we insert  $L_i^{\max} - 1$  artificial locations between the locations corresponding to stages  $i$  and  $i+1$ , for  $i = 1, \dots, M-1$ , in order to model the units in transit between these two stages.<sup>1</sup> If a unit is part of an order between stages  $i$  and  $i+1$  that has been outstanding for  $k$  periods,  $1 \leq k \leq L_i^{\max} - 1$ , then it will be in the  $k^{\text{th}}$  location between stages  $i+1$  and  $i$ . Finally, for any unit that has been given to a customer, we define its location to be 0. Thus, the set of possible locations is  $\{0, 1, \dots, N+1\}$ , where  $N = \sum_{i=1}^{M-1} L_i^{\max}$ . We index the locations starting from location 0. Location 1 corresponds to stage 1. Location 2 corresponds to units that have been released  $L_1^{\max} - 1$  times ago from stage 2. Location  $L_1^{\max}$  corresponds to units that have been released from stage 2 one period earlier. Location  $L_1^{\max} + 1$  corresponds to stage 2, etc. Location  $N+1$  corresponds to the outside supplier (stage  $M$ ). For any unit, we define its location as the index of the location at which the unit can be found. For example, in Figure 2(a), unit 5 is in location 2 at time  $t$ , which means that this unit has been released from stage 2 (location 4) at time  $t - 2$ .*

Let us now turn to the customer side of the model, which we describe using a countably infinite pool of past and potential future customers, with each such customer treated as a distinct object. At any given time, there is a finite number of customers that have arrived and whose demand is either satisfied or backlogged. In addition, conceptually, there is a countably infinite number of potential customers that may arrive to the system at a future period. Consider the system at time 0. Let  $k$  be the number of customers that have arrived, whose demand is already satisfied. We index them as customers  $1, \dots, k$ , in any arbitrary order. Let  $l$  be the number of customer that have arrived, whose demand is backlogged. We index them as customers  $k+1, k+2, \dots, k+l$  in any order. The remaining (countably infinite) customers are assigned indices  $k+l+1, k+l+2, \dots$ , in order of their

---

<sup>1</sup>For ease of exposition, we are assuming here that the leadtime process is bounded. It is possible to model cases with infinite leadtime support by considering a two dimensional location variable.

arrival times to the system, breaking ties arbitrarily, starting with the earliest arrival time. Of course, we do not know the exact arrival times of future customers, but we can conceptually talk about a “next customer,” a “second to next customer,” etc. This way, we index the past and potential future customers. We now define a quantity that we call “the position of a customer.”

**Definition A.2.** The position of a customer: *Suppose that at time  $t$  a customer  $i$  has arrived and its demand is satisfied. We define the position of such a customer to be  $-1$ . Suppose that the customer has arrived but its demand is backlogged. Then, we define the position of the customer to be  $0$ . If on the other hand, customer  $i$  has not yet arrived but customers  $1, 2, \dots, m$  have arrived, then the position of customer  $i$  at time  $t$  is defined to be  $i - m$ . In particular, a customer whose position at time  $t$  is  $k$ , will have arrived by the end of the current period if and only if  $d_t \geq k$ .*

## B Proofs

### Proof of Proposition 3.3:

(a) The infinite horizon average cost for base stock policy  $\mathbf{s}$ ,  $C(\mathbf{s})$ , is defined in (1) as:

$$C(\mathbf{s}) = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=1}^{\infty} g_T(\tilde{\mathbf{I}}(i, \omega), \mathbf{w}(i, \omega)) \quad (\text{A1})$$

Let  $\epsilon > 0$  be a constant less than  $\bar{d}/2$ . We decompose the above expression as follows:

$$\begin{aligned} & \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=1}^{\infty} g_T(\tilde{\mathbf{I}}(i, \omega), \mathbf{w}(i, \omega)) \\ = & \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=1}^{\lceil \epsilon \cdot T \rceil} g_T(\tilde{\mathbf{I}}(i, \omega), \mathbf{w}(i, \omega)) + \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=\lceil \epsilon \cdot T \rceil + 1}^{\lceil (\bar{d} - \epsilon) \cdot T \rceil} g_T(\tilde{\mathbf{I}}(i, \omega), \mathbf{w}(i, \omega)) \\ & + \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=\lceil (\bar{d} - \epsilon) \cdot T \rceil + 1}^{\lceil (\bar{d} + \epsilon) \cdot T \rceil} g_T(\tilde{\mathbf{I}}(i, \omega), \mathbf{w}(i, \omega)) + \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=\lceil (\bar{d} + \epsilon) \cdot T \rceil + 1}^{\infty} g_T(\tilde{\mathbf{I}}(i, \omega), \mathbf{w}(i, \omega)) \end{aligned} \quad (\text{A2})$$

We will show that the first and the third terms in the above sum go to zero as  $\epsilon \rightarrow 0$ , and that the fourth term is equal to zero.

Under a given non-negative base stock policy  $\mathbf{s}$ , on any given sample path  $\omega$ , the maximum length of time during which a positive cost is incurred for the  $i^{\text{th}}$  unit and customer pair is  $\left( \sum_{j=1}^{M-1} (\tilde{l}_j(i, \omega) + w_j(i, \omega)) \right)$ , and the maximum per period cost is  $(h^{\max} + b)$ , where  $h^{\max} =$

$\max_j \{h_j\}$ . Thus,

$$g_T(\tilde{\mathbf{I}}(i, \omega), \mathbf{w}(i, \omega)) \leq (h^{\max} + b) \left( \sum_{j=1}^{M-1} (\tilde{l}_j(i, \omega) + w_j(i, \omega)) \right), \quad \text{for all } i \text{ and } T. \quad (\text{A3})$$

Consequently, the first term satisfies:

$$\begin{aligned} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=1}^{\lceil \epsilon T \rceil} g_T(\tilde{\mathbf{I}}(i, \omega), \mathbf{w}(i, \omega)) &\leq \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=1}^{\lceil \epsilon T \rceil} (h^{\max} + b) \left( \sum_{j=1}^{M-1} (\tilde{l}_j(i, \omega) + w_j(i, \omega)) \right) \\ &= \lim_{T \rightarrow \infty} \epsilon \frac{1}{\epsilon T} \sum_{i=1}^{\lceil \epsilon T \rceil} (h^{\max} + b) \left( \sum_{j=1}^{M-1} (\tilde{l}_j(i, \omega) + w_j(i, \omega)) \right) \\ &= \epsilon (h^{\max} + b) \lim_{T \rightarrow \infty} \frac{1}{\epsilon T} \sum_{i=1}^{\lceil \epsilon T \rceil} \left( \sum_{j=1}^{M-1} (\tilde{l}_j(i, \omega) + w_j(i, \omega)) \right) \\ &= \epsilon (h^{\max} + b) \left( \sum_{(\mathbf{l}, \mathbf{w}) \in A} f^{\tilde{L}, W}(\mathbf{l}, \mathbf{w})(\mathbf{e}' \cdot (\mathbf{1} + \mathbf{w})) \right) \xrightarrow{\epsilon \rightarrow 0} 0 \end{aligned}$$

where the inequality follows from (A3) and the last equality follows from the ergodicity assumption. Note that, also by the ergodicity assumption,  $\left( \sum_{(\mathbf{l}, \mathbf{w}) \in A} f^{\tilde{L}, W}(\mathbf{l}, \mathbf{w})(\mathbf{e}' \cdot (\mathbf{1} + \mathbf{w})) \right) = \left( \sum_{j=1}^{M-1} (E[\tilde{L}_j^{ss}] + E[W_j^{ss}]) \right)$ , which is a finite number.

Similarly, for the third term,

$$\begin{aligned} &\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=\lceil (\bar{d}-\epsilon) \cdot T+1 \rceil}^{\lceil (\bar{d}+\epsilon) \cdot T \rceil} g_T(\tilde{\mathbf{I}}(i, \omega), \mathbf{w}(i, \omega)) \\ &\leq \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=\lceil (\bar{d}-\epsilon) \cdot T+1 \rceil}^{\lceil (\bar{d}+\epsilon) \cdot T \rceil} (h^{\max} + b) \left( \sum_{j=1}^{M-1} (\tilde{l}_j(i, \omega) + w_j(i, \omega)) \right) \\ &= 2\epsilon (h^{\max} + b) \left( \sum_{(\mathbf{l}, \mathbf{w}) \in A} f^{\tilde{L}, W}(\mathbf{l}, \mathbf{w})(\mathbf{e}' \cdot (\mathbf{1} + \mathbf{w})) \right) \xrightarrow{\epsilon \rightarrow 0} 0 \end{aligned}$$

To get the result for the fourth term, consider a unit  $i$  and its corresponding customer. After an interval of  $T$  periods, the position of this customer will be  $(i - d^T(\omega))^+$ , where  $d^T(\omega)$  is the realization of the sum of demands in  $T$  periods, on the sample path  $\omega$ . Under a given base stock policy  $\mathbf{s}$ , if  $(i - d^T(\omega))^+ > s_{M-1}$ , unit  $i$  will not be released from the outside supplier at location

$M$  and this unit customer pair will have a cost of 0 during the  $T$ -step horizon. Therefore,

$$\begin{aligned}
& \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=\lceil(\bar{d}+\epsilon) \cdot T\rceil+1}^{\infty} g_T(\tilde{\mathbf{l}}(i, \omega), \mathbf{w}(i, \omega)) \\
&= \lim_{T \rightarrow \infty} \frac{1}{T} \left\{ \sum_{i=\lceil(\bar{d}+\epsilon) \cdot T\rceil+1}^{s_{M-1}+d^T(\omega)-1} g_T(\tilde{\mathbf{l}}(i, \omega), \mathbf{w}(i, \omega)) + \sum_{s_{M-1}+d^T(\omega)}^{\infty} g_T(\tilde{\mathbf{l}}(i, \omega), \mathbf{w}(i, \omega)) \right\} \\
&= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=\lceil(\bar{d}+\epsilon) \cdot T\rceil+1}^{s_{M-1}+d^T(\omega)-1} g_T(\tilde{\mathbf{l}}(i, \omega), \mathbf{w}(i, \omega)) \tag{A4}
\end{aligned}$$

Since the demand process is ergodic, we have:

$$\lim_{T \rightarrow \infty} \frac{d^T(\omega) + S_{M-1}}{T} = \lim_{T \rightarrow \infty} \frac{d^T(\omega)}{T} = \sum_d dF^D(d) = \bar{d}, \tag{A5}$$

on every sample path  $\omega$ . Therefore, on a given sample path  $\omega$ , for any given  $\epsilon > 0$ , there exists  $T(\epsilon, \omega)$ , such that  $|(d^T(\omega) + S_{M-1})/T - \bar{d}| \leq \epsilon$  for all  $T \geq T(\epsilon, \omega)$ , i.e.,  $d^T(\omega) + S_{M-1} - 1 < (d^T(\omega) + S_{M-1}) \leq (\bar{d} + \epsilon)T < \lceil(\bar{d} + \epsilon) \cdot T\rceil + 1$  for all  $T \geq T(\epsilon, \omega)$ , meaning that there is no term in the summation in (A4). Therefore, the fourth term equals 0.

This means that only the second term in the right hand side of (A2) remains positive as  $\epsilon \downarrow 0$ , and

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

for some function  $h$  that satisfies  $\lim_{\epsilon \downarrow 0} h(\epsilon) = 0$ .

Again by the ergodicity of the demand process, as stated in (A5), on every sample path  $\omega$ , the number of customers who would have arrived in the first  $T$  periods would converge to  $\bar{d}T$  as  $T \rightarrow \infty$ . Therefore, for every fixed  $\epsilon > 0$ , there exists a  $T(\epsilon, \omega)$  such that the first  $\lceil(\bar{d} - \epsilon) \cdot T\rceil$  units would have been given to their corresponding customers by time  $T$  for any  $T > T(\epsilon, \omega)$ . In other words, on a given sample path  $\omega$ , for every fixed  $\epsilon > 0$ , there exists a  $T(\epsilon, \omega)$  such that, for any  $T > T(\epsilon, \omega)$ ,  $g_T(\tilde{\mathbf{l}}(i, \omega), \mathbf{w}(i, \omega)) = g(\tilde{\mathbf{l}}(i, \omega), \mathbf{w}(i, \omega))$  for  $i = 1, \dots, \lceil(\bar{d} - \epsilon) \cdot T\rceil$ .

(b) Immediate from part (a) and the assumption that the joint process  $(\tilde{L}, W)$  is ergodic. ■

### Proof of Proposition 3.4:

(a) If the leadtime process is sequential, then orders always arrive in the same order in which they

were placed. By the definition of the processes  $L$  and  $\hat{L}$ , the two processes are identical on every sample path, i.e.,  $f^L(\mathbf{1}) = f^{\hat{L}}(\mathbf{1})$  for all  $\mathbf{1}$ .

Consider the  $i^{\text{th}}$  unit released from stage  $j+1$  to stage  $j$ . Since the leadtime process is sequential, the  $i^{\text{th}}$  unit released from stage  $j+1$  is also the  $i^{\text{th}}$  unit that arrives in stage  $j$ . Similarly, if the  $i^{\text{th}}$  unit is part of the  $k^{\text{th}}$  order released from stage  $j+1$ , it is also part of the  $k^{\text{th}}$  order that arrives in stage  $j$ . Let  $OS_j(k, \omega)$  denote the order size of the  $k^{\text{th}}$  order between stages  $j+1$  and  $j$  on a sample path  $\omega$ . Consider the units in the first  $K$  orders on a given sample path  $\omega$ , the total number of units that have a leadtime of  $l$  is  $\sum_{k=1}^K \mathbb{I}\{\hat{l}_j(k, \omega) = l\} OS_j(k, \omega)$ , while the total number of units in the first  $K$  orders is  $\sum_{k=1}^K OS_j(k, \omega)$ . By the ergodicity assumption, we have:

$$\begin{aligned}
f^{\tilde{L}_j}(l) &= \lim_{K \rightarrow \infty} \frac{\sum_{k=1}^K \mathbb{I}\{\hat{l}_j(k, \omega) = l\} OS_j(k, \omega)}{\sum_{k=1}^K OS_j(k, \omega)} \\
&= \lim_{K \rightarrow \infty} \frac{\left(\sum_{k=1}^K \mathbb{I}\{\hat{l}_j(k, \omega) = l\} OS_j(k, \omega)\right) / K}{\left(\sum_{k=1}^K OS_j(k, \omega)\right) / K} \\
&= \frac{\lim_{K \rightarrow \infty} \left\{ \left(\sum_{k=1}^K \mathbb{I}\{\hat{l}_j(k, \omega) = l\} OS_j(k, \omega)\right) / K \right\}}{\lim_{K \rightarrow \infty} \left\{ \left(\sum_{k=1}^K OS_j(k, \omega)\right) / K \right\}} \\
&= \frac{\bar{d} f^{\hat{L}_j}(l)}{\bar{d}} = f^{\hat{L}_j}(l), \tag{A6}
\end{aligned}$$

where the first equality follows directly from the definition of the ergodicity of the  $\tilde{L}_j$  process, and the fourth equality follows from the ergodicity of the  $\hat{L}_j$  process and the fact that the order-based ordered leadtime process between stages  $j+1$  and  $j$ ,  $\hat{L}_j$ , is independent of the sizes of the orders between these two stages on any sample path  $\omega$ . To show the latter fact, note that on a given sample path  $\omega$ ,  $OS_j(k, \omega)$  is uniquely determined by the demand process  $D$  and the leadtime processes in the upstream stages  $L_{j+1}, \dots, L_{M-1}$  through the  $k^{\text{th}}$  time period, and the base stock levels. By the assumption that the leadtime processes are *exogenous*, we have:  $L_j$ , and thus  $\hat{L}_j$ , is independent of the demand process  $D$  and the leadtime processes at all other stages. It follows that  $\hat{L}_j$  is independent of the order sizes  $OS_j$ . Finally, since (A6) is true for all  $j$ , we have  $f^{\tilde{L}}(\mathbf{1}) = f^{\hat{L}}(\mathbf{1})$  for all  $l$ .

- (b) Under a base stock policy, the unit  $i$  will be released from stage  $j+1$  when the position  $y_t^i$  of the corresponding customer reaches or exceeds the base stock level  $s_j$ . Let  $t^*$  be the period when the unit  $i$  is released from stage  $j+1$ , i.e., such that  $y_{t^*}^i > s_j$  and  $y_{t^*+1}^i \leq s_j$ . Define  $Q_j(i)$  to be the overshoot of customer  $i$  at stage  $j$ , where  $Q_j(i) = s_j - y_{t^*+1}^i$ , the amount by which the position of the customer is less than the threshold that it just crossed.

Consider the random joint process  $(\tilde{L}_j, W_j)$  and the joint process  $(\tilde{L}_j, W_j, Q_j)$  and their steady state distributions  $f^{\tilde{L}_j, W_j}(l, w)$  and  $f^{\tilde{L}_j, W_j, Q_j}(l, w, q)$ , for stage  $j$ . We have:

$$\begin{aligned}
f^{\tilde{L}_j, W_j}(l, w) &= \sum_{q=1}^{\infty} f^{\tilde{L}_j, W_j, Q_j}(l, w, q) \\
&= \sum_{q=1}^{\infty} \left[ f^{\tilde{L}_j, Q_j}(l, q) \cdot f^{W_j, Q_j}(w, q) \right] \\
&= \sum_{q=1}^{\infty} \left[ f^{\tilde{L}_j}(l) \cdot f^{Q_j}(q) \cdot f^{W_j, Q_j}(w, q) \right] \\
&= f^{\tilde{L}_j}(l) \cdot \sum_{q=1}^{\infty} \left[ f^{Q_j}(q) \cdot f^{W_j, Q_j}(w, q) \right] \\
&= f^{\tilde{L}_j}(l) \cdot f^{W_j}(w).
\end{aligned}$$

The first equality above is just algebra. The second equality follows from the fact that given the overshoot  $Q_j(i)$  of a particular customer  $i$ , the unit-based ordered leadtime  $\tilde{L}_j(i)$  for the corresponding unit  $i$  and the time between threshold crossings  $W_j(i)$  for customer  $i$  are independent, because the leadtimes are sequential. Moreover, the overshoot  $Q_j(i)$  of customer  $i$  and the unit-based ordered leadtime  $\tilde{L}_j(i)$  for the corresponding unit  $i$  are independent as well, again as a consequence of sequential leadtimes, leading to the third equality. ■

**Proof of Proposition 3.6:** The proof is based on showing a property of the value function of the single unit problem  $\mathbb{S}1(\mathbf{X})$ . Let  $\hat{J}(z, y)$  denote the value function of the single-unit problem, defined as in Muharremoglu & Tsitsiklis (2008), except that it uses the leadtime distribution  $\mathbf{X}$  for the single unit. Here, we give an overview of the proof. Given that the unit is at location  $z$  right now, let  $r_z$  be the probability that the unit will arrive at the next stage  $v(z)$  in the next period. Let

$$\Delta(z, y) = r_z \hat{J}(v(z), y) + (1 - r_z) \hat{J}(z - 1, y) - \hat{J}(z, y).$$

The proof relies on showing that  $\Delta(z, y)$  is non-decreasing in  $y$  for every  $z > 1$ . Using this fact, it is easy to show that if it is optimal to release a unit from location  $z$ , when the customer position is  $y$ , it is also optimal to release the unit from stage  $z$  when the customer position is  $y - 1$ , resulting in a threshold policy. The monotonicity of the threshold levels  $s_j$  is a consequence of the monotonicity of the holding cost rates  $h_j$ . ■

**Proof of Proposition 3.7:**

(a) Consider the single unit problem  $\mathbb{S}1(\mathbf{L}^{ss})$ , where the leadtime random vector of the single unit

is  $\mathbf{L}^{ss}$ , and the threshold crossing random vector of the single customer is  $\mathbf{W}(y_0)$ , where  $y_0$  is the initial position of the customer. Then the cost of the single unit problem can be written as:

$$\begin{aligned}
C^S(\mathbf{s}, \mathbf{L}^{ss}) &= \lim_{y_0 \rightarrow \infty} J(y_0, \mathbf{s}, \mathbf{L}^{ss}) \\
&= \lim_{y_0 \rightarrow \infty} E [g(\mathbf{L}^{ss}, \mathbf{W}(y_0))] \\
&= \lim_{y_0 \rightarrow \infty} \sum_{(\mathbf{l}, \mathbf{w}) \in A} g(\mathbf{l}, \mathbf{w}) \Pr(\mathbf{L}^{ss} = \mathbf{l}) \Pr(\mathbf{W}(y_0) = \mathbf{w}) \\
&= \sum_{(\mathbf{l}, \mathbf{w}) \in A} g(\mathbf{l}, \mathbf{w}) \Pr(\mathbf{L}^{ss} = \mathbf{l}) \left[ \lim_{y_0 \rightarrow \infty} \Pr(\mathbf{W}(y_0) = \mathbf{w}) \right] \\
&= \sum_{(\mathbf{l}, \mathbf{w}) \in A} g(\mathbf{l}, \mathbf{w}) f^{\tilde{L}}(\mathbf{l}) f^W(\mathbf{w}) \\
&= \sum_{(\mathbf{l}, \mathbf{w}) \in A} g(\mathbf{l}, \mathbf{w}) f^{\tilde{L}, W}(\mathbf{l}, \mathbf{w}),
\end{aligned}$$

where the next to last equality follows from the fact that the leadtime is sequential, by Proposition 3.4 (a) and by the ergodicity of the  $W$  process (which is implied by the ergodicity of the demand process). The last equality follows from Proposition 3.4(b).

(b) Immediate from part (a) and Proposition 3.3 (b). ■

**Proof of Proposition 3.9:**

(a) Consider a given time  $t + k$ . The event that there are at most  $k$  outstanding orders between stage  $j + 1$  and stage  $j$  implies the event that at least  $t$  orders have arrived at stage  $j$  by time  $t + k$ , which is equivalent to the event that the  $t^{\text{th}}$  order arrival time at stage  $j$  is no later than  $t + k$ . By the definition of the order-based ordered leadtime process, this implies the event that the order-based ordered leadtime between the  $t^{\text{th}}$  order release from stage  $j + 1$  and the  $t^{\text{th}}$  order arrival is at most  $k$ . In other words,

$$(V_j(t + k) \leq k) \Rightarrow (\hat{L}_j(t) \leq k). \tag{A7}$$

On the other hand, also consider the time  $t + k$ , the event that there are more than  $k$  outstanding orders between stage  $j + 1$  and stage  $j$  implies the event that less than  $t$  orders have arrived at stage  $j$  by time  $t + k$ , which is equivalent to the event that the  $t^{\text{th}}$  order arrival time at stage  $j$  is later than  $t + k$ . By the definition of the order-based ordered leadtime process, this implies the event that the order-based ordered leadtime between the  $t^{\text{th}}$  order release from stage  $j + 1$

and the  $t^{\text{th}}$  order arrival is more than  $k$ . In other words,

$$(V_j(t+k) > k) \Rightarrow (\hat{L}_j(t) > k). \quad (\text{A8})$$

By (A7) and (A8), we have:

$$(V_j(t+k) \leq k) \Leftrightarrow (\hat{L}_j(t) \leq k), \quad (\text{A9})$$

which implies

$$\Pr(\hat{L}_j(t) \leq k) = \Pr(V_j(t+k) \leq k) \quad \text{for all } j, t \text{ and } k.$$

- (b) For any fixed vector  $\mathbf{k}$ ,  $\Pr(\hat{\mathbf{L}}^{ss} \leq \mathbf{k}) = \Pr(\mathbf{V}^{ss} \leq \mathbf{k})$ , by part (a) and the assumption that the leadtime processes at different stages are independent of each other. Since this is true for all vectors  $\mathbf{k}$ , part (b) follows. ■

**Proof of Proposition 3.10:**

- (a) First, notice that by Proposition 3.9 (b), the original problem and the surrogate problem have the same steady state distribution of the number of outstanding orders. Denote the two single-stage inventory systems as  $\mathbb{P}1 = (h, b, L1, D)$  and  $\mathbb{P}2 = (h, b, L2, D)$ . Let  $PI1^{ss}$  and  $PI2^{ss}$  denote the steady state distribution of pipeline inventory, i.e. the number of outstanding units in system 1 and system 2, respectively. Then at time  $t$  on a given sample path  $\omega$ ,  $PI1(t, \omega)$  is the sum of  $V1(t, \omega)$  independent demands and  $PI2(t, \omega)$  is the sum of  $V2(t, \omega)$  independent demands. Since  $V1^{ss} \stackrel{d}{=} V2^{ss}$  and the two single-stage inventory systems have the same demand processes, we have:  $PI1^{ss} \stackrel{d}{=} \sum_{i=1}^{V1^{ss}} D(i) \stackrel{d}{=} \sum_{i=1}^{V2^{ss}} D(i) \stackrel{d}{=} PI2^{ss}$ , and consequently,

$$\begin{aligned} \text{Cost of } \mathbb{P}1 &= hE[s - PI1^{ss}]^+ + bE[s - PI1^{ss}]^- \\ &= hE[s - PI2^{ss}]^+ + bE[s - PI2^{ss}]^- \\ &= \text{Cost of } \mathbb{P}2, \quad \text{under any base stock level } s. \end{aligned}$$

- (b) By the construction of the surrogate problem, the original problem and the surrogate problem have the same realization of the number of outstanding orders on any sample path, so the two single-stage problems differ only in their leadtime processes but have the same steady state distribution of the number of outstanding orders. The first equality in part (b), i.e.,  $C(s) = C'(s)$ , follows from part (a) of this Proposition. By the construction of the surrogate problem, the leadtime process of the surrogate problem is stochastically identical to the order-based ordered

leadtime process  $\hat{L}$  of the original problem, which is sequential. Therefore, the second equality in part (b), i.e.,  $C'(s) = \bar{d}C^S(s, \hat{L}^{ss})$ , follows from Proposition 3.7(b).

(c) Since part (b) is true for any base stock level  $s$ , part (c) follows immediately. ■

**Proof of Proposition 3.11:** Let **PI1** and **PI2** denote the steady state distribution of pipeline inventory, i.e. the number of outstanding units in the original problem and the surrogate problem, respectively. Then at time  $t$  on a given sample path  $\omega$ ,  $PI1_{M-1}(t, \omega)$  is the sum of  $V1_{M-1}(t, \omega)$  independent demands and  $PI2_{M-1}(t, \omega)$  is the sum of  $V2_{M-1}(t, \omega)$  independent demands. By the construction of the surrogate problem and Proposition 3.9,  $V1_{M-1}^{ss} \stackrel{d}{=} V2_{M-1}^{ss}$ . The two multi-stage inventory systems have the same demand processes, so we have:  $PI1_{M-1}^{ss} \stackrel{d}{=} \sum_{i=1}^{V1_{M-1}^{ss}} D(i) \stackrel{d}{=} \sum_{i=1}^{V2_{M-1}^{ss}} D(i) \stackrel{d}{=} PI2_{M-1}^{ss}$ . Consequently,  $IN1_{M-1}^{ss} \stackrel{d}{=} s_{M-1} - PI1_{M-1}^{ss} \stackrel{d}{=} s_{M-1} - PI2_{M-1}^{ss} \stackrel{d}{=} IN2_{M-1}^{ss}$ , i.e., under any given base stock level  $s_{M-1}$ , the steady state distribution of the net inventory at stage  $M-1$  in the original problem is identical to the steady state distribution of the net inventory at stage  $M-1$  in the surrogate problem. Therefore,  $h_{M-1}E[IN1_{M-1}^{ss}]^+ = h_{M-1}E[IN2_{M-1}^{ss}]^+$ , i.e., the expected holding cost at stage  $M-1$  is the same in both systems. On the other hand, consider all the downstream stages from stage  $M-1$  to stage 1. It is easy to verify that the expected costs from stage  $M-1$  down to stage 1 are the same in both systems, because: (i) the net inventory at stage  $M-1$  has the same steady state distribution in both systems; (ii) this steady state distribution is independent of the leadtime process in the downstream stages; (iii) the leadtime processes in the downstream stages of the two systems are identical, since there is no order crossing in the downstream stages in both systems; and (iv) all the cost parameters are identical in both systems. This implies that the total expected holding costs from stage 1 to stage  $M-1$  and the expected backlogging costs are the same in both systems. This proves that  $C(\mathbf{s}) = C'(\mathbf{s})$ . Since the leadtime process in the surrogate problem is sequential, we can apply Proposition 3.7(b) to the surrogate problem and obtain  $C'(\mathbf{s}) = \bar{d}C^S(\mathbf{s}, \hat{\mathbf{L}}^{ss})$ . Since the above two equations hold for any given base stock levels  $\mathbf{s}$ , it follows that the set of minimizers of the cost functions are equivalent. ■

**Proof of Proposition 3.12:** We first show that under the echelon base stock policy  $\mathbf{s}$  such that  $s_j - s_{j-1} \geq L_j^{\max} \cdot D^{\max}$  for all  $j = 2, \dots, M-1$ , there are no internal backlogs in the system. Let  $IP_j$  and  $IN_j$  denote the echelon inventory position and the net inventory at stage  $j$ , respectively. Without loss of generality, assume that the system starts from an initial state such that  $IP_j = s_j$  for all  $j$  and nothing is in the pipeline. Consider stage  $(M-1)$  first. Since stage  $M$  provides infinite supply to stage  $(M-1)$ , at any time period  $t$ , stage  $(M-1)$  places an order of the same size as the demand  $D(t)$ ,

which gets shipped immediately from stage  $M$ . In any time period  $t$ , there are at most  $(L_{M-1}^{\max} - 1)$  outstanding orders between stage  $M$  and stage  $(M - 1)$ . This implies that

$$IN_{M-1} \geq s_{M-2} + D^{\max}.$$

This in turn implies that there are never any backlogs between stage  $M - 1$  and  $M - 2$ . Continuing the same argument for every stage, we prove that there is no internal backlog at any stage in the system. Therefore, each order size at each stage is a realization of single period demand.

Let **PI1** and **PI2** denote the steady state distribution of pipeline inventory, i.e. the number of outstanding units in the original problem and the surrogate problem, respectively. By the construction of the surrogate problem, we have:  $\mathbf{V1}^{ss} \stackrel{d}{=} \mathbf{V2}^{ss}$ , which, together with the fact that there is no internal backlog in the system, implies that: (i)  $IP_j = s_j$ ; (ii)  $PI1_j^{ss} \stackrel{d}{=} \sum_{i=1}^{V1_j^{ss}} D(i) \stackrel{d}{=} \sum_{i=1}^{V2_j^{ss}} D(i) \stackrel{d}{=} PI2_j^{ss}$ ; and (iii)  $IN1_j^{ss} \stackrel{d}{=} s_j - PI1_j^{ss} \stackrel{d}{=} s_j - PI2_j^{ss} \stackrel{d}{=} IN2_j^{ss}$ , for all  $j = 1, \dots, M - 1$ . As a result,

$$C(\mathbf{s}) = E \left[ (b + h'_1) (IN1_1^{ss})^- + \sum_{j=1}^{M-1} h_j IN1_j^{ss} \right] = E \left[ (b + h'_1) (IN2_1^{ss})^- + \sum_{j=1}^{M-1} h_j IN2_j^{ss} \right] = C'(\mathbf{s}),$$

where  $h_j$  is the echelon holding cost at stage  $j$  and  $h'_1$  is the local holding cost at stage 1. Since the leadtime process in the surrogate problem is sequential, we can apply Proposition 3.7(b) to the surrogate problem and obtain:  $C'(\mathbf{s}) = \bar{d}C^S(\mathbf{s}, \hat{\mathbf{L}}^{ss})$ .  $\blacksquare$

**Proof of Proposition 3.13:** Since  $s_2 = s_1$ , no inventory is kept at stage 2 and the two-stage system can be viewed as a single-stage system with an exogenous leadtime process satisfying all the assumptions of the class. (Note that unless the upstream stage has deterministic leadtimes, the single stage analog may not satisfy the assumption that the leadtimes across stages are independent). The leadtime process of the single stage analog is  $L'_1(t) = L_1(t) + l_2$ , where  $L_1(t)$  is the leadtime process between stage 2 and stage 1, and  $l_2$  is the deterministic leadtime between the outside supplier and stage 2. The result of this Proposition follows immediately from Proposition 3.10.  $\blacksquare$

**Proof of Proposition 4.3:**

- (a) If orders do not cross between stage  $j + 1$  and  $j$ , the first two equalities follow from Definition 2.3(a) and Proposition 3.4(a). Since orders do not cross between stage  $j + 1$  and  $j$ , the order-based ordered positive leadtime process  $L_j^+$  is the leadtime process of the positive-sized orders. As the leadtime process between stage  $j + 1$  and stage  $j$  is independent of the order sizes of the orders between stage  $j + 1$  and stage  $j$ , the last equality follows.

- (b) Consider the situation that there are exactly  $k$  positive-sized outstanding orders between stage  $j + 1$  and stage  $j$ . This situation can happen if and only if there are  $v$  ( $v \geq k$ ) outstanding orders between stage  $j + 1$  and stage  $j$ , and of these  $v$  outstanding orders, exactly  $k$  orders have positive size. In other words,

$$\begin{aligned}
\mathbf{f}_j(k) = \Pr(V_j^{+ss} = k) &= \sum_{v=k}^{L_j^{\max}} \Pr(V_j^{ss} = v) \Pr(k \text{ of } v \text{ orders have positive sizes}) \\
&= \sum_{v=k}^{L_j^{\max}} \Pr(V_j^{ss} = v) \binom{v}{k} (p_0^j)^{v-k} (1 - p_0^j)^k \\
&= \sum_{v=k}^{L_j^{\max}} \mathbf{g}_j(k) \binom{v}{k} (p_0^j)^{v-k} (1 - p_0^j)^k,
\end{aligned}$$

where the last equality follows, since  $\mathbf{V}^{ss} \stackrel{d}{=} \hat{\mathbf{L}}^{ss}$  by Proposition 3.9(b). Writing in matrix form, we have:  $\mathbf{f}_j = B_j \mathbf{g}_j$ . Except the trivial cases where  $p_0^j = 0$  (so all orders are of positive size and we can obtain  $\hat{\mathbf{L}}^{ss}$  directly by observing the  $\hat{L}^+$  process) or  $p_0^j = 1$  (so all orders are of zero size and the base stock levels are clearly optimized at  $\mathbf{s} = \mathbf{0}$ ),  $B_j$  is an upper triangular matrix whose diagonal entries are all positive, so  $B_j$  is clearly an invertible matrix, and  $\mathbf{g}_j = B_j^{-1} \mathbf{f}_j$ .

■

## C Additional Numerical Results

Here, we provide a detailed table of the two-stage numerical examples in which the single-unit method did not produce optimal base stock levels.

Table 10: The results of the 47 two-stage numerical examples (out of 120) where the single unit method did not produce optimal base stock levels.

Demand	$L_j^{\max}$	Leadt. Type	$h_1 - h_2$	$\frac{b}{h_1}$	$s_u$	$s^*$	Sim. cost $s_u$	Sim. cost $s^*$	Loss of Opt.
B(10,0.1)	5	Uni.	1	10	(6,10)	(7,10)	13.0616	13.0468	0.1134%
B(10,0.1)	11	Disp.	4	10	(10,19)	(10,20)	39.3315	39.1978	0.3410%
B(10,0.1)	11	Uni.	4	2	(8,15)	(7,15)	25.0098	24.9597	0.2006%
B(10,0.1)	11	Cent.	1	10	(10,17)	(11,17)	18.8788	18.8715	0.0388%
B(2,0.5)	11	Cent.	1	2	(8,13)	(8,14)	11.9059	11.9018	0.0351%
B(10,0.1)	101	Disp.	1	2	(59,108)	(60,108)	76.6505	76.6139	0.0478%
B(10,0.1)	101	Disp.	1	10	(66,120)	(67,120)	93.166	93.1592	0.0073%
B(10,0.1)	101	Disp.	4	10	(64,123)	(63,125)	145.3989	145.0814	0.2188%
B(2,0.5)	101	Disp.	1	2	(58,107)	(59,107)	72.5483	72.5271	0.0292%
B(2,0.5)	101	Disp.	1	10	(63,117)	(64,117)	86.1155	86.0535	0.0720%
B(2,0.5)	101	Disp.	4	2	(55,110)	(55,111)	99.796	99.7324	0.0638%
B(2,0.5)	101	Disp.	4	10	(61,120)	(61,121)	129.2598	129.1439	0.0898%
B(10,0.1)	101	Uni.	1	10	(65,119)	(65,120)	91.0078	90.994	0.0152%
B(10,0.1)	101	Uni.	4	2	(56,111)	(55,112)	106.0286	105.925	0.0978%
B(10,0.1)	101	Uni.	4	10	(63,122)	(63,124)	139.9276	139.6338	0.2105%
B(2,0.5)	101	Uni.	1	10	(62,116)	(62,117)	83.6196	83.5966	0.0275%
B(2,0.5)	101	Uni.	4	2	(55,110)	(55,111)	96.0239	96.0147	0.0096%
B(2,0.5)	101	Uni.	4	10	(61,118)	(60,120)	123.3377	122.829	0.4142%
B(10,0.1)	201	Disp.	1	2	(112,211)	(114,211)	137.1009	137.0283	0.0530%
B(10,0.1)	201	Disp.	1	10	(122,227)	(123,227)	160.0348	160.0036	0.0195%
B(10,0.1)	201	Disp.	4	2	(108,216)	(108,217)	182.5833	182.5633	0.0110%
B(10,0.1)	201	Disp.	4	10	(119,232)	(118,234)	232.7175	232.4676	0.1075%
B(2,0.5)	201	Disp.	1	2	(111,210)	(112,209)	131.338	131.3087	0.0223%
B(2,0.5)	201	Disp.	1	10	(118,223)	(119,223)	150.2543	150.1817	0.0484%
B(2,0.5)	201	Disp.	4	10	(116,227)	(116,229)	210.9255	210.6746	0.1191%
B(10,0.1)	201	Uni.	1	2	(112,210)	(112,211)	135.4271	135.4066	0.0151%
B(10,0.1)	201	Uni.	1	10	(120,226)	(121,227)	156.9732	156.9303	0.0273%
B(10,0.1)	201	Uni.	4	2	(108,215)	(107,216)	178.2336	178.0999	0.0751%
B(10,0.1)	201	Uni.	4	10	(117,230)	(117,232)	225.1259	224.6653	0.2050%
B(2,0.5)	201	Uni.	1	10	(117,221)	(117,223)	146.9465	146.7619	0.1258%
B(2,0.5)	201	Uni.	4	2	(107,213)	(106,214)	164.5178	164.3259	0.1168%
B(2,0.5)	201	Uni.	4	10	(114,225)	(114,227)	202.0782	201.6386	0.2180%
B(10,0.1)	301	Disp.	1	2	(165,313)	(167,313)	195.0665	195.0124	0.0278%
B(10,0.1)	301	Disp.	1	10	(176,333)	(177,333)	222.941	222.8822	0.0264%
B(10,0.1)	301	Disp.	4	2	(160,319)	(160,320)	250.6043	250.5749	0.0117%
B(10,0.1)	301	Disp.	4	10	(172,339)	(172,341)	311.3267	311.0999	0.0729%
B(2,0.5)	301	Disp.	1	2	(163,311)	(165,311)	188.0793	188.041	0.0204%
B(2,0.5)	301	Disp.	1	10	(172,328)	(173,328)	211.0284	210.9773	0.0242%
B(2,0.5)	301	Disp.	4	2	(159,316)	(158,317)	235.0093	234.8879	0.0517%
B(2,0.5)	301	Disp.	4	10	(169,333)	(169,335)	284.7837	284.6159	0.0590%
B(10,0.1)	301	Uni.	1	2	(164,312)	(165,313)	193.0142	192.993	0.0110%
B(10,0.1)	301	Uni.	1	10	(174,331)	(175,332)	219.2662	219.1464	0.0547%
B(10,0.1)	301	Uni.	4	2	(159,318)	(159,320)	245.2409	245.1281	0.0460%
B(10,0.1)	301	Uni.	4	10	(171,336)	(171,339)	302.507	301.8358	0.2224%
B(2,0.5)	301	Uni.	1	10	(170,326)	(170,327)	206.8846	206.811	0.0356%
B(2,0.5)	301	Uni.	4	2	(158,315)	(157,317)	228.6957	228.4719	0.0980%
B(2,0.5)	301	Uni.	4	10	(167,330)	(167,333)	274.374	273.778	0.2177%