Heuristics for Base-Stock Levels in Multi-Echelon Distribution Networks with First-Come First Served Policies

Ying Rong*, Zümbül Atan† and Lawrence V. Snyder‡

Abstract

We study inventory optimization for locally controlled, continuous-review distribution systems with stochastic customer demands. Each node follows a base-stock policy and a first-come, first-served allocation policy. We develop two heuristics, the *recursive optimization* (RO) heuristic and the *decomposition-aggregation* (DA) heuristic, to approximate the optimal base-stock levels of all the locations in the system. The RO heuristic applies a bottom-up approach that sequentially solves single-variable, convex problems for each location. The DA heuristic decomposes the distribution system into multiple serial systems, solves for the base-stock levels of these systems using the newsvendor heuristic of Shang & Song (2003), and then aggregates the serial systems back into the distribution system using a procedure we call "backorder matching." A key advantage of the DA heuristic is that it does not require any evaluation of the cost function (a computationally costly operation that requires numerical convolution). We show that, for both RO and DA, changing some of the parameters, such as leadtime, unit backordering cost, and demand rate, of a location has an impact only on its own local base-stock level and its upstream locations' local base-stock levels. An extensive numerical study shows that both heuristics perform well, with the RO heuristic providing more accurate results and the DA heuristic consuming less computation time. Finally, we show that both RO and DA are asymptotically optimal along multiple dimensions for two-echelon distribution systems.

Keywords: Distribution Network; Heuristics; Asymptotic Analysis

*Dept. of Industrial Engineering and Operations Research , University of California, Berkeley, CA 94720 USA.
†Dept. of Industrial Engineering and Innovation Sciences, Eindhoven University of Technology, The Netherlands.
‡Industrial and Systems Engineering Dept., Lehigh University, Bethlehem, PA 18015 USA.

1 Introduction

From an inventory-optimization perspective, distribution systems are among the most difficult network topologies to analyze and solve. Their optimal ordering and allocation policies under stochastic demands are still unknown (Zipkin 2000). Many studies focus on two-echelon distribution networks, i.e., one-warehouse, multiple-retailer systems (OWMR), with base-stock policies employed at all locations and a first-come, first-served (FCFS) allocation policy employed at the warehouse. However, even for this restricted scenario, the best known exact algorithm, called *the projection method* (Graves 1985, Axsäter 1990), involves an exhaustive search over the warehouse base-stock levels.

In this paper, we consider an infinite-horizon problem for multi-echelon, continuous-review distribution systems facing Poisson demands. The systems are locally controlled in the sense that each location monitors its own inventory and order information. We assume that each location uses a base-stock ordering policy and a FCFS allocation policy. We introduce two heuristics to obtain the base-stock levels at all locations with the objective of minimizing the expected cost per unit time.

The projection method and early approximate solution methods, including METRIC (Sherbrooke 1968) and the two-moment approximation by Graves (1985), are top-down methods. That is, they determine the base-stock levels by starting from the upstream part of the supply chain, i.e., the root location. In contrast, in our first heuristic, the *recursive optimization* (RO) heuristic, we apply a bottom-up method, which starts from the most downstream locations, i.e., the leaf locations. Similar to the nested decomposition method for solving serial systems (Clark & Scarf 1960, Chen & Zheng 1994, Gallego & Zipkin 1999, etc.), we show that the RO heuristic only needs to solve single-variable, convex minimization problems sequentially to obtain the base-stock levels for all locations. Due to this convexity property, it is much more computationally efficient than the projection method, METRIC and the two-moment approximation, especially when the number of echelons is more than two. However, unlike its counterpart for serial systems, the RO heuristic does not guarantee the optimal solution for distribution systems. In order to further improve its performance, in the last step of the RO heuristic, we re-optimize the base-stock levels of the leaf locations, keeping the base-stock levels at the non-leaf locations fixed to their values found by the earlier steps of the heuristic.

In our second heuristic, the *decomposition-aggregation* (DA) heuristic, we initially decompose the distribution network into serial systems. We use the heuristic introduced by Shang & Song (2003) to obtain near-optimal local base-stock levels for all locations in each serial system. Then we propose a "backorder matching" procedure to aggregate the serial systems back into the distribution network.

The backorder matching procedure sets the local base-stock level of a given location in the distribution network so that its expected backorder level is equal to the sum of the expected backorder levels in all of its counterparts in the decomposed serial systems. One key advantage of the DA heuristic is that it does not require any evaluation of the cost function (a computationally costly operation) when determining the base-stock levels. The cost function only needs to be calculated after the heuristic completes its operation if the user wants to know the expected cost of the solution.

Two heuristics for OWMR systems have used the same decomposition procedure as ours (but different aggregation procedures). It was first used in the direct search method by Gallego et al. (2007) to set the maximal echelon base-stock level at the warehouse for OWMR systems with a central control scheme. For OWMR systems with service level constraints, Özer & Xiong (2008) propose a newsvendor heuristic that sets the warehouses base-stock level equal to the sum of the warehouse's base-stock levels in all of the decomposed serial systems. We call this aggregation procedure "base-stock level matching". Özer & Xiong (2008) point out that their newsvendor heuristic produces high percentage errors compared to the optimal solution. Our DA heuristic addresses this problem by recognizing that backorders are the main bottleneck connecting different echelons. By matching the backorders, our DA heuristic generates sufficient protection for downstream locations to receive goods on time and reduces inventory holding costs at the same time. (See Section 3.2 for further discussion.)

In our numerical study, we compare our two heuristics with two other approaches. The first is the current state-of-the-art OWMR heuristic, namely, the restriction-decomposition (RD) heuristic proposed by Gallego et al. (2007), or more specifically, its extension to distribution networks as discussed by Özer & Xiong (2008). The second is a modification of the projection method that assumes unimodality of the cost function at non-leaf locations, and thus executes more quickly (but possibly with some loss of optimality) than the classical projection method. The former approach represents the best heuristic, balancing accuracy and computation time, for distribution networks that is currently available, while the second is a reasonable heuristic based on an exact method, and therefore is worthwhile as a benchmark. We refer the reader to Axsäter (1993) and Gallego et al. (2007) for extensive reviews of algorithms introduced before the RD heuristic. The RD heuristic consists of three subheuristics: cross-docking, zero-safety-stock and stock-pooling. It selects the solution with minimal cost among the three subheuristics. The cross-docking and stock-pooling solutions are two extremes, with minimal and maximal base-stock levels at non-leaf locations, respectively, while the zero-safety-stock solution sets the corresponding base-stock levels to the average demand during the leadtime. In our RO heuristic, the base-stock level of each node i is optimal for node i in the subgraph rooted at i, given that the base-stock levels of all of its downstream locations are fixed. The DA heuristic aggregates the information from the decomposed serial systems to set the base-stock levels in the distribution network. Both the RO and DA heuristics make explicit use of the cost parameters to set base-stock levels at non-leaf locations. In contrast, although RD as a whole accounts for the cost parameters by switching among the subheuristics, the base-stock levels of non-leaf locations are independent of the cost parameters in the cross-docking and zero-safety-stock subheuristics, and set to their maximal levels in the stock-pooling subheuristic.

For OWMR systems, we prove that RO is asymptotically optimal as i) the unit holding cost at the warehouse goes to zero, ii) the number of retailers goes to infinity, or iii) the warehouse leadtime goes to infinity. In addition, we prove that DA is asymptotically optimal as the unit holding cost at the warehouse goes to zero. Moreover, our comprehensive numerical experiments indicate that both of our heuristics provide better solutions, on average, than the RD heuristic for general distribution networks. Both RO and DA are faster than the modified projection method, and DA is faster than the RD heuristic.

Our heuristics also have the potential to be extended to even more general supply chain settings. To give an example of this, we consider OWMR systems with a fixed ordering cost at the warehouse, leading to an (r, q) inventory policy at the warehouse. Extensions of the RO and DA heuristics for this case still provide compelling performance in terms of both computation times and optimality gaps. Thus, both methods shed new light on the inventory mechanics of distribution systems by suggesting that solution methods for serial supply chains can be translated to distribution networks, although not as neatly or as accurately as can be for serial and assembly systems (Rosling 1989). The connection between distribution networks and serial supply chains may help researchers to improve their understanding of more generalized systems, such as acyclic supply chains (e.g. Billington et al. 2004, Graves & Willems 2003, Shi & Zhao 2010).

The structure of this paper is as follows. We introduce our notation in Section 2. In Section 3, we discuss both the RO and DA heuristics. We present computation results in Section 4. The extension of our heuristics to OWMR systems with a fixed ordering cost at the warehouse is discussed in Section 5. Conclusions and directions for future work are provided in Section 6.

2 Preliminaries

Before discussing the details of our heuristics, we introduce the notation used throughout the paper. Initially, we provide the notation required to define the distribution network itself. Then, we provide the inventory-related notation. This section ends with the expressions that clarify the relationships among the state variables and the expression for the expected cost.

Let $T = (V, E)$ represent a directed tree, i.e., a distribution network. V is the set of all locations and E is the set of all directed edges. There are $N + 1$ locations in V, with 0 representing the root location. Representing the edge from location i to location j by $\langle i, j \rangle$, we define $\mathcal{P}(i) := \{j \in V : \langle j, i \rangle \in E\}$ and $\mathcal{S}(i) := \{j \in V : \langle i, j \rangle \in E\}$ to be the *single* predecessor of location i and the *set* of successors of location i, respectively. We define $\mathcal{P}(0)$ as the external supplier, which is assumed to have infinite supply. We can always arrange the indices of nodes such that $\mathcal{P}(i) < i$ for all $i \in V$. In addition, we define $\mathcal{L} := \{i \in V : \mathcal{S}(i) = \emptyset\}$ as the set of all leaves and $T(i) := \{j \in V : \exists$ a directed path from i to $j\}$ as the subtree rooted at location i consisting of location i and all downstream locations.

We assume that all items initially proceed from the external supplier to the root location. The root location supplies its successors, who in turn supply their successors, and so on until the items reach the leaf locations, where the customer demands occur. Demands are Poisson processes with rate λ_i at leaf location i and rate $\lambda_i := \sum_{j \in \mathcal{S}(i)} \lambda_j$ at non-leaf location i. The transportation leadtimes between all the locations are deterministic, though waiting times are stochastic due to stockouts. L_i represents the leadtime between $\mathcal{P}(i)$ and i, and D_i represents the leadtime demand at location i. Unsatisfied demands at each location are backordered, but only the leaf locations pay penalty costs for unsatisfied demands, with the unit backorder cost at location i given by $b_i > 0$. All locations are allowed to carry inventory, and they incur holding costs for doing so. The unit local holding cost at location i is h_i. The echelon holding cost is calculated by $H_i = h_i - h_{\mathcal{P}(i)}$, with $H_0 = h_0$. Since downstream holding costs are typically greater than upstream ones, we assume that $H_i \geq 0$ for all i. Moreover, we assume that holding costs are charged on items in transit. We assume local control, which means that each location observes its own inventory level and places orders with its predecessor. The assumptions in this paragraph are consistent with much of the literature on multi-echelon inventory theory, e.g., Axsäter (1990) and Zipkin (2000), though some of the notation is different.

In Section 3 and 4, we assume that there is no fixed cost at any location and that a continuous-review base-stock policy is employed by all locations. The decision variables are $\mathbf{s} := (s_i)_{i=0}^N$, where s_i is the local base-stock level at location i. We can obtain the echelon base-stock level at location i by $S_i = \sum_{j \in T(i)} s_j$. We let $\mathbf{S} := (S_i)_{i=0}^N$ be the vector of echelon base-stock levels. (In Section 5, we consider OWMR systems with a fixed ordering cost at the warehouse.)

We define $I_i(\mathbf{s})$ and $B_i(\mathbf{s})$ as the on-hand inventory level and backorder level at location i for a given $\mathbf{s}$, respectively. Let $Bin(K, p)$ denote a binomial random variable with parameters K (total number) and p (probability). $B_{\mathcal{P}(i)}(\mathbf{s})$ (the total backorder level at $\mathcal{P}(i)$) and $B_{\mathcal{P}(i),i}(\mathbf{s})$ (the backorders at $\mathcal{P}(i)$ due to demands from i) are related as $B_{\mathcal{P}(i),i}(\mathbf{s}) \overset{Dist}{=} Bin(B_{\mathcal{P}(i)}(\mathbf{s}), \theta_i)$. Here, $\overset{Dist}{=}$ denotes that two random variables have the same probability distribution and $\theta_i := \lambda_i/\lambda_{\mathcal{P}(i)}$ is the demand rate proportion of location i at its predecessor $\mathcal{P}(i)$. Moreover, $B_{\mathcal{P}(i),i}(\mathbf{s})$ and D_i are independent (Zipkin 2000).

Let $x^+ = \max\{0, x\}$ and $x^- = \max\{0, -x\}$. For any s_i, $i \in V$, the relationships among the limiting probabilities of the state and decision variables can be described by the following expressions (Zipkin 2000):

$$\begin{aligned}
B_i(\mathbf{s}) &= [B_{\mathcal{P}(i),i}(\mathbf{s}) + D_i - s_i]^+ \\
I_i(\mathbf{s}) &= [B_{\mathcal{P}(i),i}(\mathbf{s}) + D_i - s_i]^- \\
B_{i,j}(\mathbf{s}) &\overset{Dist}{=} Bin(B_i(\mathbf{s}), \theta_j), \ \forall j \in \mathcal{S}(i)
\end{aligned} \tag{1}$$

where $B_{\mathcal{P}(0),0} \equiv 0$. We denote IT_i as the inventory in-transit to location i. Under both local and echelon base-stock policies, we have $E[IT_i] = E[D_i]$, where $E[\cdot]$ denotes the expectation of a random variable. Using the system dynamic equations (1), we can evaluate the long-run expected cost per unit time for any given $\mathbf{s}$ as follows:

$$\mathscr{C}(\mathbf{s}) = \sum_{i \in V \setminus \mathcal{L}} h_i E\left[I_i(\mathbf{s}) + \sum_{j \in \mathcal{S}(i)} IT_j\right] + \sum_{i \in \mathcal{L}} E[h_i I_i(\mathbf{s}) + b_i B_i(\mathbf{s})]. \tag{2}$$

The first sum is the expected cost of holding inventory at the non-leaf locations and in transit. The second sum is the total holding and penalty cost at the leaf locations. Note that the in-transit costs are not affected by the local base-stock levels. We include the in-transit costs in (2) because (2) is equivalent to the cost evaluation under the echelon base-stock levels introduced in Section 3.1. We omit the in-transit costs when we perform numerical studies in Section 4. A summary of all notation is provided in Appendix A.

3 Heuristics

As mentioned above, Graves (1985) and Axsäter (1990) develop the projection method to obtain the optimal base-stock levels for an OWMR system. For a given warehouse base-stock level, this method finds the optimal base-stock levels at the retailers and evaluates the corresponding total cost. An exhaustive search needs to be performed to find the optimal warehouse base-stock level because the cost is not a convex function of the warehouse base-stock level (and the corresponding optimal retailer base-stock levels). This is computationally time consuming, specifically because each objective function evaluation requires numerical convolutions. Using the same idea, the projection method can be extended to solve distribution systems with more than two echelons. However, its implementation is even harder in this case since enumeration is required to find the optimal base-stock levels of all the non-leaf locations. This is why we propose the following heuristic procedures to approximate the base-stock levels of all locations in a distribution system.

3.1 The Recursive Optimization (RO) Heuristic

The recursive algorithm by Clark & Scarf (1960), with subsequent studies by Chen & Zheng (1994), Gallego & Zipkin (1999), is used for evaluation and optimization of base-stock levels in serial systems. They find the optimal base-stock levels by starting from the location closest to customers and solving a set of recursive equations. This approach is called a "bottom-up approach." Inspired by this algorithm, in this section, we propose a similar approach for distribution systems.

The key enabler of the bottom-up optimization approach for serial systems is the existence of two equivalent cost evaluation procedures, one using local and the other using echelon base-stock levels. Using the echelon base-stock policy, the cost function of serial supply chains is constructed recursively. This recursion enables sequential optimization of the cost by solving only single-variable, convex minimization problems. The development of the RO heuristic relies on the same idea. Below, we define cost functions based on echelon base-stock levels that are equivalent to the local-quantity cost function (2) and which enable a recursive evaluation of the cost in distribution networks.

For a given echelon base-stock level vector $\mathbf{S}$, we define $\ell_i(\mathbf{S})$ to be the corresponding local base-stock level at location i. That is, $\ell_i(\mathbf{S}) = S_i - \sum_{j \in \mathcal{S}(i)} S_j$. Let $\ell(\mathbf{S}) := \{\ell_0(\mathbf{S}), \ell_1(\mathbf{S}), ..., \ell_N(\mathbf{S})\}$. Next we define the following functions to facilitate the cost calculations based on the echelon base-stock levels. First, for $i \in \mathcal{L}$, we define an auxiliary location attached to leaf node i as $\mathcal{S}(i)$. Then, for

$i \in \mathcal{L}$, we define $\underline{C}_{\mathcal{S}(i)}(x|\mathbf{S}) = (b_i + h_i)[x]^-$. Next, we define the following recursive functions:

$$\begin{aligned} \hat{C}_i(x|\mathbf{S}) &= H_i x + \sum_{j \in \mathcal{S}(i)} E\left[\underline{C}_j\left(Bin\left(\left[x - \sum_{j \in \mathcal{S}(i)} S_j\right]^-, \theta_j\right) \middle| \mathbf{S}\right)\right] \\ C_i(y|\mathbf{S}) &= E\left[\hat{C}_i(y - D_i|\mathbf{S})\right] \\ \underline{C}_i(v|\mathbf{S}) &= C_i(S_i - v|\mathbf{S}) \end{aligned} \tag{3}$$

The existence of two equivalent cost evaluations, one based on local and the other on echelon base-stock levels, is a well known result (Zipkin 2000, page 340). We further extend this result by providing a *recursive* evaluation procedure using echelon base-stock levels and, with the following proposition, we prove that resulting cost is equal to the cost based on local base-stock levels.

Proposition 1. *For any echelon base-stock vector* $\mathbf{S}$*, we have* $\mathscr{C}(\ell(\mathbf{S})) = C_0(S_0|\mathbf{S})$.[1]

From the proof of Proposition 1, one can treat $C_i(S_i|\mathbf{S})$ as the expected cost of subtree $T(i)$ when the vector of echelon base-stock levels is $\mathbf{S}$, with the local holding cost of all locations in $T(i)$ decreased by $h_{\mathcal{P}(i)}$ and the backorder cost of locations in $T(i) \cap \mathcal{L}$ increased by $h_{\mathcal{P}(i)}$. These adjustments are similar to the truncation of the serial system implied by Proposition 2 in Shang & Song (2003). $\hat{C}_i(x|\mathbf{S})$ and $\underline{C}_i(v|\mathbf{S})$ are auxiliary functions which help in simplifying the expression of $C_i(y|\mathbf{S})$.

We define S_i^{r0} as the *intermediate* echelon base-stock level of location i and s_i^r as the *final* local base-stock level of location i suggested by the RO heuristic, respectively. With slight abuse of the notation (dropping the dependency of $\hat{C}$, C and $\underline{C}$ on $\mathbf{S}$), we develop the following RO heuristic to determine the values of s_i^r.

Recursive Optimization (RO) Heuristic

Step 1: Compute S_i^{r0} *as follows:*

$$\begin{aligned} \hat{C}_i(x) &= H_i x + \sum_{j \in \mathcal{S}(i)} \underline{C}_j\left(Bin\left(\left[x - \sum_{j \in \mathcal{S}(i)} S_j^{r0}\right]^-, \theta_j\right)\right) \\ C_i(y) &= E[\hat{C}_i(y - D_i)] \\ S_i^{r0} &= \underset{y}{\operatorname{argmin}} C_i(y) \\ \underline{C}_i(v) &= C_i(S_i^{r0} - v) \end{aligned} \tag{4}$$

[1] All proofs are in Appendix B.

Step 2: Calculate $\ell(\mathbf{S^{r0}})$. Define $\ell_{-i}(\mathbf{S})$ to be the vector $\ell(\mathbf{S})$ excluding the component $\ell_i(\mathbf{S})$. Using (1), we define $I_i(s_i, \ell_{-i}(\mathbf{S^{r0}}))$ and $B_i(s_i, \ell_{-i}(\mathbf{S^{r0}}))$ to be the on-hand inventory level and backorder level at location i when its own local base-stock level is s_i and the local base-stock level of node j is $\ell_j(\mathbf{S^{r0}})$ for $j \neq i$. Calculate $\mathbf{s}^r$ as follows:

$$s_i^r = \begin{cases} \underset{s_i}{\operatorname{argmin}} E\left[h_i I_i(s_i, \ell_{-i}(\mathbf{S^{r0}})) + b_i B_i(s_i, \ell_{-i}(\mathbf{S^{r0}}))\right] & \forall i \in \mathcal{L} \\ \ell_i(\mathbf{S^{r0}}) & \forall i \notin \mathcal{L} \end{cases} \tag{5}$$

The RO heuristic optimizes the base-stock levels recursively, starting from downstream. In order to obtain S_i^{r0}, it first decomposes the subgraph downstream from location i into $|\mathcal{S}(i)|$ independent subtrees. In this sense, the RO heuristic shares the same bottom-up idea as the approaches for distribution systems proposed by Clark & Scarf (1960). After the base-stock levels for all locations in those systems have been determined, this procedure searches for S_i^{r0} by finding the y that minimizes $C_i(y)$ (step 1). Then, in step 2, it updates the leaf locations' base-stock levels by optimizing their costs when all the non-leaf locations' base-stock levels are fixed to the values obtained in step 1. We update the base-stock levels for leaf locations *only*, since these locations provide direct protection against costly customer backorders, and since inventories at these locations have a large impact on improving the fill rate (Shang & Song 2006), and therefore the expected cost. With the following proposition, we prove the convexity of $C_i(\cdot)$ and $\underline{C}_i(\cdot)$ for all i. This property is the main reason for the efficiency of the RO heuristic.

Proposition 2. *For all i, $C_i(\cdot)$ and $\underline{C}_i(\cdot)$ are convex.*

Note that the convexity of $C_i(y)$ is well known in the literature for $i \in \mathcal{L}$. Proposition 2 shows that such convexity is maintained in the upstream portion of the distribution network, too, when the downstream base-stock levels are set using the first step of RO. (In contrast, the upstream cost function is not convex in the projection method.)

Unlike the method by Clark & Scarf (1960) for serial systems, the RO heuristic is not guaranteed to find the optimal solution for distribution systems since we solve for the base-stock levels of the successors of a location assuming the successors are completely independent. This ignores the fact that the base-stock level of one successor of i is affected by the base-stock level of the other successors of i. For example, increasing base-stock level at one successor of i reduces the need to hold inventory at node i. As a result, the base-stock levels at the other successors of i may also need to increase.

In other words, the distribution system cannot be perfectly decomposed into subsystems without considering their interactions.

Although RO is not guaranteed to find the optimal solution, it is asymptotically optimal in multiple ways. The following theorem summarizes our asymptotic results.

Theorem 3. *Suppose that $h_i > h_0$ for all $i > 0$.*

1. *Let $c^r(h_0)$ and $c^*(h_0)$ denote the cost of RO and the optimal cost (respectively) for given h_0 under the OWMR system. Then we have*

$$\lim_{h_0 \to 0} \frac{c^r(h_0)}{c^*(h_0)} = 1.$$

2. *Let $c^r(N)$ and $c^*(N)$ denote the cost of RO and the optimal cost (respectively) for the OWMR system with N retailers. Suppose that there exists $\delta > 0$ such that $h_0 < h_i - \delta$ for all i. Then we have*

$$\lim_{N \to \infty} \frac{c^r(N)}{c^*(N)} = 1.$$

3. *Let $c^r(L_0)$ and $c^*(L_0)$ denote the cost of RO and the optimal cost (respectively) for the OWMR system with warehouse leadtime L_0. Suppose that there exists $\delta > 0$ such that $h_0 < h_i - \delta$ for all i. Then, when $b_i = b$ for all i, we have*

$$\lim_{L_0 \to \infty} \frac{c^r(L_0)}{c^*(L_0)} = 1.$$

Next, we provide another result which states that the impact of the unit backordering cost, demand rate and leadtime of node j is only limited to nodes that have direct paths to node j.

Proposition 4. *S_i^{r0} remains unchanged if b_j and λ_j vary for $j \notin T(i) \cap \mathcal{L}$, or if L_j varies for $j \notin T(i)$.*

As discussed above, RO is not guaranteed to provide the optimal base-stock levels, because the decomposition is not exact. Proposition 4 shows that the decomposition procedure of RO decouples certain connections in the distribution network. However, Theorem 3 and the numerical studies for general distribution networks in Section 4 indicate that the optimality loss due to this decoupling is limited.

3.2 The Decomposition-Aggregation Heuristic

As can be seen in the numerical study in Section 4, RO performs very well, indicating that the bottom-up decomposition approach can be effective for distribution systems. However, it still involves an optimization step that can be cumbersome in practice. Therefore, in this section we propose another heuristic procedure in which, instead of decomposing the system at each location one at a time (as in the RO heuristic), we go a bit further and use a single decomposition step to decompose the entire distribution system into serial systems. Then, we approximate the base-stock levels of all the locations in all the decomposed serial systems with the newsvendor heuristic introduced by Shang & Song (2003). The question is then how to set the base-stock levels when we re-aggregate the serial systems back into the original distribution system so that we can use the results for serial systems and approximate the true optimal base-stock levels of the distribution system as closely as possible. We propose to aggregate the serial systems back into the distribution system using an approach we call "backorder matching." The backorder matching procedure is motivated by the recognition that backorders (or, equivalently, service levels) are the key driver in the cost and, therefore, performance of the system. In particular, as is evident from (1), only the backorders of location i affect the system state of its successors. Hence, it is logical to use backorders to drive the aggregated base-stock levels, rather than simply summing the serial-system base-stock levels. After discussing the details of the decomposition-aggregation (DA) heuristic, we elaborate on other possibilities for the aggregation step.

3.2.1 Description of DA

Initially, we decompose the distribution system into $|\mathcal{L}|$ individual serial systems. For each leaf, we obtain a serial system that consists of the unique path from the root location to this leaf. Hence, given that k is the leaf location of a given serial system w, serial system w consists of locations $\{0, \ldots, \mathcal{P}(\mathcal{P}(k)), \mathcal{P}(k), k\}$. The demand rate at each location of this system is $\lambda_w := \lambda_k$. The cost parameters and the leadtimes between the locations are exactly the same as in the distribution system. Define $\mathcal{W}$ as the set of serial systems. Then we have $|\mathcal{W}| = |\mathcal{L}|$, and after the decomposition, there are $|T(i) \cap \mathcal{L}|$ serial systems containing location i. If $i \in w$, then we refer to the copy of location i in serial system w as location i_w.

The analysis of serial systems was initiated by Clark & Scarf (1960). They prove that echelon base-stock policies are optimal and show that the optimal base-stock levels can be obtained by a recursive minimization of one-dimensional convex cost functions. Subsequently, Federgruen & Zipkin (1984a), Chen & Zheng (1994), and Huh & Janakiraman (2008) refined the findings of Clark & Scarf (1960). The research on serial systems has also been extended in multiple ways, including such factors as non-stationary demand (Shang 2012), lost sales (Huh & Janakiraman 2010), capacity constraints (Huh et al. 2010) and supply disruptions (DeCroix 2013). However, the solution procedures are typically cumbersome. Thus, studies initiated by Shang & Song (2003) and followed by Gallego & Özer (2005) propose newsvendor heuristics to provide closed-form expressions for the echelon base-stock levels in serial systems. In this paper, we adapt the main newsvendor heuristics proposed in Shang & Song (2003). In particular, we use s_i^a to denote the local base-stock level at location i based on the DA heuristic, and use $s_{i_w}^d$ [$S_{i_w}^d$] to denote the local [echelon] base-stock level at location i in the decomposed serial system w. The echelon base-stock levels obtained by the heuristic by Shang & Song (2003) are

$$S_{i_w}^{SS} = \begin{cases} F_{\tilde{D}_{i_w}}^{-1}\left(\dfrac{b_i + \sum_{j\in\mathcal{A}(0,\mathcal{P}(i))} H_j}{b_i + \sum_{j\in\mathcal{A}(0,i)} H_j}\right), & i \in \mathcal{L} \\ \frac{1}{2}\left[G_{\tilde{D}_{i_w}}^{-1}\left(\dfrac{b_{\ell_w} + \sum_{j\in\mathcal{A}(0,\mathcal{P}(i))} H_j}{b_{\ell_w} + \sum_{j\in w} H_j}\right) + G_{\tilde{D}_{i_w}}^{-1}\left(\dfrac{b_{\ell_w} + \sum_{j\in\mathcal{A}(0,\mathcal{P}(i))} H_j}{b_{\ell_w} + \sum_{j\in\mathcal{A}(0,i)} H_j}\right)\right], & i \notin \mathcal{L} \end{cases} \tag{6}$$

Here, $\mathcal{A}(i,j)$ is the set of locations in the directed path from i to j. Adapting Shang & Song's notation, $\tilde{D}_{i_w}$ is the total leadtime demand in the subsystem of serial system w consisting of locations i to the leaf node of w, which we denote ℓ_w. $\tilde{D}_{i_w}$ is a Poisson random variable with rate $\lambda_w \sum_{j\in\mathcal{A}(i,\ell_w)} L_j$. F^{-1} is the inverse Poisson cumulative distribution function. For non-leaf locations, instead of using F^{-1}, we opt instead to use G^{-1}, the inverse function of an approximate Poisson CDF. Let $\lceil x \rceil$ and $\lfloor x \rfloor$ be the smallest integer no less than x and the largest integer no greater than x, respectively. We define this approximate CDF of a Poisson random variable D to be the following continuous, piecewise linear function:

$$G_D(x) = \begin{cases} 2F_D(0)x, & \text{if } x \le 0.5 \\ F_D(\lfloor x-0.5 \rfloor) + \left[F_D(\lceil x-0.5 \rceil) - F_D(\lfloor x-0.5 \rfloor)\right](x - 0.5 - \lfloor x-0.5 \rfloor), & \text{if } x \ge 0.5 \end{cases}$$

The primary advantage of using this revised CDF is that it is continuous and not a step function, and therefore it allows the resulting base-stock levels for the decomposed serial systems to be non-integer. To see why this is useful, consider an OWMR system. (The explanation for general distribution systems is similar.) In some cases, the base-stock level determined by (6) using F^{-1}

instead of G^{-1} at the warehouse may equal 0 for every copy of the warehouse in the decomposed serial systems. This may happen, for example, if the demand rates are small, the leadtimes are short, or the echelon holding costs at the retailers are small. These zero base-stock levels provide little information about inventory requirements at the warehouse to the aggregation step. On the other hand, if the base-stock levels may be non-integer, then the warehouse base-stock levels may be set to something positive (but small), which provides more information to the aggregation step. (Note, however, that in the aggregated system, all base-stock levels will be integer.)

Next, we obtain the local base-stock level for location i ($i \notin \mathcal{L}$) by $s^d_{i_w} = S^{SS}_{i_w} - S^{SS}_{j_w}$, where j_w is the successor of i_w in system w. For $i \in \mathcal{L}$, we have $s^d_{i_w} = S^d_{i_w}$. Given the local base-stock levels of all the locations in all serial systems, we can approximate the expected backorders at each location i_w in serial system w assuming that i_w has a supplier with infinite supply.

$$E[B_{i_w}] = E[(D_{i_w} - s^d_{i_w})^+] = Q_{D_{i_w}}(s^d_{i_w}), \tag{7}$$

where D_{i_w} is the leadtime demand rate of location i, a Poisson random variable with rate $\lambda_w L_i$, and $Q_D(x)$ is the loss function of the Poisson random variable D. This expression is approximate because it ignores the portion of the serial system that is upstream from i.

We now discuss how to aggregate the serial systems back into the distribution system. S-ince we approximate the expected backorders at location i due to the demand of its successors in serial system w by (7), we can approximate the total expected backorders at location i by $E[B_i] \cong \sum_{w\in\mathcal{W}:i\in w} E[B_{i_w}]$. Then, for each $i \in V$, we search for a local base-stock level s_i that generates an expected backorder level which is smaller than but as close as possible to $E[B_i]$. We call this procedure "backorder matching."

Specifically, the backorder matching procedure sets s^a_i, the base-stock level at location i equal to the smallest integer s_i such that

$$E[(D_i - s_i)^+] \le \sum_{w\in\mathcal{W}:i\in w} E[B_{i_w}].$$

In other words,

$$s^a_i = Q^{-1}_{D_i}\left(\sum_{w\in\mathcal{W}:i\in w} Q_{D_{i_w}}(s^d_{i_w})\right) \tag{8}$$

where $Q^{-1}_D(y)$ is defined as $\min\{s \in \mathbb{Z} : Q_D(s) \le y\}$.

Similar to Theorem 3 for RO, we now prove the asymptotic optimality of DA as h_0 goes to 0.

Theorem 5. *Suppose that $h_i > h_0$ for all $i > 0$. Let $c^a(h_0)$ denote the cost of DA for given h_0 under the OWMR system. Then we have*

$$\lim_{h_0 \to 0} \frac{c^a(h_0)}{c^*(h_0)} = 1.$$

In addition, similar to Proposition 4 for RO, the impact of the unit backordering cost, demand rate and leadtime of node j is only limited to nodes that have direct paths to node j.

Proposition 6. *s_i^a remains unchanged if b_j and λ_j vary for $j \notin T(i) \cap \mathcal{L}$, or if L_j varies for $j \notin T(i)$.*

3.2.2 Other Aggregation Procedures

The DA heuristic can accommodate aggregation procedures other than backorder matching. For example, matching can be performed in terms of the base-stock and on-hand inventory.

Base-stock level matching is proposed by Özer & Xiong (2008) by approximating the base-stock level of a location with the sum of the base-stock levels in all of its counterparts in the decomposed serial systems. As the following proposition indicates, the aggregated local base-stock level under base-stock level matching is greater than that under backorder matching.

Proposition 7. $s_i^a \leq \lceil \sum_{w \in \mathcal{W}: i \in w} s_{i_w}^d \rceil$

Since a high demand at one of the successors of a location is sometimes compensated by a low demand at another successor of the same location, backorder matching can maintain the same expected backorders as the sum of the decomposed systems with less investment in inventory compared to base-stock level matching—in other words, it can take advantage of the risk-pooling effect (Eppen 1979). Thus, base-stock level matching leads to higher costs than backorder matching.[2]

On-hand inventory matching implies finding the largest integer s_i such that the inequality

$$E[(s_i - D_i)^+] \leq \sum_{w \in \mathcal{W}: i \in w} E[(s_{i_w}^d - D_{i_w})^+]$$

holds. In other words, the purpose of on-hand inventory matching is to equate the inventory level of a location to the sum of the inventory levels in all of its counterparts in the decomposed serial systems. However, the true benefit of aggregating serial systems lies in reducing the inventory rather

[2]For OWMR systems with fill rate constraints, Özer & Xiong (2008) report that base-stock level matching results in an average optimality gap of 18.66%. Their base-stock level matching procedure is customized for their model with fill rate constraints, but the similarity of fill rate-constrained models and backorder cost models implies that similar observations would hold for the problem studied in this paper.

than in keeping it constant. On the other hand, by utilizing the identity $x^+ - x^- = x$, the on-hand inventory matching procedure is equivalent to finding the largest integer s_i such that the inequality

$$s_i + E[(D_i - s_i)^+] \leq \sum_{w \in \mathcal{W}: i \in w} s^d_{i_w} + \sum_{w \in \mathcal{W}: i \in w} E[(D_{i_w} - s^d_{i_w})^+]$$

holds. This inequality suggests that on-hand inventory matching can be treated as a hybrid of base-stock level matching ($s_i \leq \sum_{w \in \mathcal{W}: i \in w} s^d_{i_w}$) and backorder matching ($E[(D_i - s_i)^+] \leq \sum_{w \in \mathcal{W}: i \in w} E[(D_{i_w} - s^d_{i_w})^+]$). Since base-stock level matching does not provide better performance than backorder matching in general, this hybrid of the two does not offer better performance than backorder matching alone.

We use the data set in Section 4, consisting of 240 instances, to test the effectiveness of the three matching methods discussed here. The average percentage errors are 16.24, 13.70 and 1.76 for base-stock level, on-hand inventory and backorder matching, respectively. Clearly, base-stock level matching and on-hand inventory matching perform significantly worse than backorder matching.

4 Numerical Analysis

In this section, we test the performance of RO and DA by comparing them with two other heuristics: the restriction-decomposition (RD) heuristic and the projection method assuming unimodality (PMU). The RD heuristic consists of three subheuristics: cross-docking, zero-safety-stock and stock-pooling. It selects the solution with minimal cost among the three subheuristics. We refer the interested reader to Gallego et al. (2007) and Özer & Xiong (2008) for the details of RD. The former paper studies OWMR systems only, while the latter discusses RD's extension to distribution networks. The expected cost of the subtree rooted at a non-leaf location i is not unimodal in the local base-stock level s_i when its predecessors' local base-stock levels (i.e., locations in $\mathcal{A}(0, \mathcal{P}(i))$) are fixed. However, one can still *assume* that it is unimodal and apply a bisection search on s_i rather than enumerating all possible values of s_i as is done in the projection method (PM). We call this method the *projection method assuming unimodality* (PMU). The PMU is a heuristic rather than an exact algorithm; however, it nearly always finds the optimal solution (Zipkin 2000, page 334).

Before testing the performance in terms of cost, we compare the approximate computation times of RO, DA, RD, PM, and PMU. Let α be the time required to evaluate the expected cost and β the time required to evaluate the approximate Poisson inverse function G_D^{-1}. When the network consists

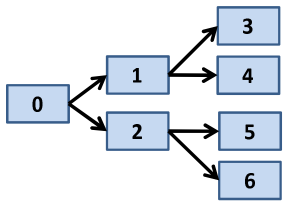

Figure 1: Three-Echelon Distribution Network

of many echelons, α is typically much greater than β. In addition, α is also affected by the size of the network (i.e., N). For RO, since $C_i(\cdot)$ is convex, S_i^{r0} can be found using line-search techniques (e.g., bisection search). Thus, the computation time is $O(\alpha \sum_{i \notin \mathcal{L}} \log_2 \hat{s}_i)$, where $\hat{s}_i$ is the upper bound on the local base-stock level of node i. The DA heuristic does not need cost evaluations. The depth and width of a distribution network with $N+1$ nodes are each at most $N+1$. Hence, the decomposition step generates at most $N+1$ serial systems, and the maximum length of each is $N+1$. Thus, the worst-case computation time of DA is $O(\beta N^2)$. However, for many networks, the DA heuristic generates far fewer locations in the decomposed serial systems. For example, for an OWMR system, the DA heuristic only generates N two-echelon serial systems so that the computation time is $O(\beta N)$.

Since the base-stock levels must be found through exhaustive search in PM, its computation time is $O(\alpha \prod_{i \notin \mathcal{L}} \hat{s}_i)$. For PMU, performing the bisection search recursively reduces the computation time to $O(\alpha \prod_{i \notin \mathcal{L}} log_2(\hat{s}_i))$. RD needs to evaluate the average cost three times. Thus, its computation time is $O(\alpha)$. Therefore, for $\alpha \gg \beta$, we can sort the computation times as: $t_{PM} \geq t_{PMU} \geq t_{RO} \geq t_{RD} \geq t_{DA}$.

Given that c is the expected cost of the solution suggested by one of the methods and c^* is the expected cost of the solution obtained using PMU, we calculate the percentage cost error (or gap) by $\epsilon = 100\frac{c-c^*}{c^*}$. Note that we exclude the in-transit inventory costs in our numerical analysis so that the percentage errors reflect the controllable portions of the expected cost.

We consider systems with 2, 3, 4 and 5 echelons. In all the systems, each location except the leaves has two successors. An example three-echelon distribution network is shown in Figure 1. For $i \in \mathcal{L}$, we set $\lambda_i = 8$, $L_i \sim U[0.1, 0.25]$ and $b_i \sim U[9, 39]$. For $i \notin \mathcal{L}$, we set $L_i \sim U[0.1, 0.5]$. We consider three different scenarios that describe how the local holding cost coefficients increase as one moves downstream in the system. In particular, the scenarios assume the costs increase in a linear, concave, and convex manner—essentially, whether the total holding cost is evenly distributed

Concave Local Holding Cost Coefficients												
# of echelons		2			3			4			5	
	RD	RO	DA	RD	RO	DA	RD	RO	DA	RD	RO	DA
$\bar{\epsilon}$	3.23	0.49	0.82	4.88	1.55	2.61	5.38	1.52	3.27	5.82	1.82	4.69
med_ϵ	2.82	0.00	0.66	4.70	1.35	2.47	5.20	1.43	3.38	5.81	1.81	4.62
max_ϵ	7.42	1.83	3.56	7.26	3.25	4.23	6.88	3.05	5.37	6.56	2.49	6.24
σ_ϵ	2.07	0.61	0.99	1.14	0.69	1.08	0.68	0.65	0.98	0.42	0.41	0.71
$\bar{t}$ (seconds)	0.006	0.027	0.003	0.034	0.148	0.005	0.218	0.920	0.008	1.204	5.328	0.015
		PMU			PMU			PMU			PMU	
$\bar{t}$ (seconds)		0.024			0.683			32.032			1696.347	
Linear Local Holding Cost Coefficients												
# of echelons		2			3			4			5	
	RD	RO	DA	RD	RO	DA	RD	RO	DA	RD	RO	DA
$\bar{\epsilon}$	6.37	0.14	0.19	8.43	0.47	1.27	8.20	0.60	1.67	8.01	0.80	2.43
med_ϵ	5.98	0.00	0.00	8.23	0.32	1.03	8.00	0.59	1.75	7.97	0.74	2.44
max_ϵ	11.69	1.02	1.05	11.56	1.87	3.93	10.28	1.23	2.67	8.85	1.53	3.38
σ_ϵ	2.60	0.31	0.34	1.62	0.52	1.10	1.11	0.34	0.65	0.51	0.27	0.53
$\bar{t}$ (seconds)	0.006	0.027	0.003	0.034	0.145	0.005	0.218	0.880	0.008	1.195	5.010	0.015
		PMU			PMU			PMU			PMU	
$\bar{t}$ (seconds)		0.025			0.687			31.161			1606.932	
Convex Local Holding Cost Coefficients												
# of echelons		2			3			4			5	
	RD	RO	DA	RD	RO	DA	RD	RO	DA	RD	RO	DA
$\bar{\epsilon}$	6.37	0.14	0.19	12.92	0.19	0.38	16.29	0.32	1.27	18.87	0.18	2.38
med_ϵ	5.98	0.00	0.00	13.28	0.07	0.22	16.32	0.24	1.29	18.81	0.16	2.34
max_ϵ	11.69	1.02	1.05	16.64	0.65	2.21	20.27	0.70	2.42	20.46	0.39	3.71
σ_ϵ	2.60	0.31	0.34	2.17	0.21	0.54	2.31	0.18	0.62	1.04	0.09	0.49
$\bar{t}$ (seconds)	0.006	0.027	0.003	0.034	0.140	0.006	0.216	0.832	0.009	1.174	4.566	0.015
		PMU			PMU			PMU			PMU	
$\bar{t}$ (seconds)		0.025			0.672			30.965			1541.090	

Table 1: Performance of the Heuristics for Multi-Echelon Distribution Networks

among the echelons, weighted more toward the upstream portion, or weighted more toward the downstream portion. (Note that the holding costs are still linear functions of the on-hand inventory. These functional shapes describe how the *unit* local holding costs change with the echelon number.) For the linear case, we set the local holding cost of location i equal to $h_i = \frac{\lceil \log_2(i+2) \rceil}{J}$, where J is the total number of echelons. Here, $\lceil \log_2(i+2) \rceil$ identifies the echelon that location i belongs to. Using Figure 1 as an example, location 0 belongs to the first echelon, whose local holding cost is $\frac{1}{3}$; locations 1 and 2 belong to the second echelon, whose local holding cost is $\frac{2}{3}$; and the remaining locations belong to the third echelon, whose local holding cost is 1. The corresponding formulas for the concave and convex holding costs are $h_i = \sqrt{\frac{\lceil \log_2(i+2) \rceil}{J}}$ and $h_i = 2^{\lceil \log_2(i+2) \rceil - J}$, respectively. In all three scenarios, we generate 20 instances for each network topology. The results are summarized in Table 1.

Running PM for distribution networks with many echelons is very time consuming. Therefore, we use PMU as a benchmark. However, it is evident from Table 1 that even PMU's computation time is sharply increasing in the number of echelons. Thus, the computation time difference between PMU and the other heuristics increases in the number of echelons.

Table 1 suggests that the performance of the heuristics depends on the structure of the unit local holding costs. RO and DA perform the best for convex unit local holding costs and their performance deteriorate as the unit holding costs become linear and then concave. The way we set the unit local holding costs suggests that as we move from the convex structure to the concave structure, the relative value added by the leaves decreases. Both RO and RD perform the best when the value added by the leaves is relatively more than the value added by the non-leaf locations. In fact, we prove with Theorem 3 and Theorem 5 that, for OWMR systems, both RO and DA are asymptotically optimal for small values of h_0. Our numerical observations are in line with these results. In contrast to RO and DA, the performance of RD improves when the unit local holding costs move from the convex to the concave structure.

When the number of echelons increases, the performance of RO, DA and RD deteriorates, in general, since the number of upstream locations increases in the number of echelons. RO's performance under the convex holding cost structure is the least affected by the number of echelons. This is because, in our experimental setting, the ratios of the leaves' echelon holding costs to their local holding costs (under the convex local holding cost structure) remain 0.5 regardless of the number of echelons. Note that, with Theorem 3, we prove that RO is asymptotically optimal for OWMR systems with a large number of leaf locations. Both the ratio and asymptotic optimality contribute to the good performance of RO under the convex holding cost structure regardless of the number of echelons. On the other hand, the ratios of the leaves' echelon holding costs to their local holding costs decrease under the linear and concave local holding cost structures. Therefore, the relative value added by the leaf locations decreases. Although the number of leaves does not have a significant impact on RO, the performance of RO is affected by the decreasing value added at leaves as the number of echelons increases under both the linear and concave holding cost settings.

Table 1 suggests, as expected, that DA has the shortest computation times and is the least affected by the scale of the distribution networks. RD has the second shortest computation time, followed by RO. As the number of echelons increases, the ratio of the computation times for RD and RO stays more or less the same, while the ratio of the computation times for RD and DA increases. RD requires calculation of the expected cost for each subheuristic, while no cost calculations are necessary for DA. Since the time required to calculate the expected cost for distribution networks is more sensitive to the number of echelons than the time to calculate the approximate Poisson inverse function, compared to RD, DA runs much faster as the number of echelons increases.

5 Extension: Fixed Ordering Cost at Warehouse

The RO and DA methods can serve as a starting point for heuristics for other types of distribution systems than those considered in this paper. In this section, we give one example of this by considering an extension of RO and DA for an OWMR system with a fixed ordering cost, ξ, at the warehouse. We assume that the warehouse follows an (r, q) ordering policy and that the retailers follow a base-stock policy. Making a slight change from the notation used in Section 2, we define $\mathbf{s} = [s_1, s_2, ..., s_N]$ to be the vector of retailers' base-stock levels (that is, we exclude s_0, the base-stock level of the warehouse, from the vector). We use $\pi(r, q, \mathbf{s})$ to denote the cost of this policy. Then $\pi(r, q, \mathbf{s})$ can be evaluated as follows (Zipkin 2000, pg. 338):

$$\pi(r, q, \mathbf{s}) = \frac{\xi \sum_{i=1}^{N} \lambda_i + \sum_{s_0=r+1}^{r+q} \mathscr{C}(s_0, \mathbf{s})}{q}. \tag{9}$$

$\pi(r, q, \mathbf{s})$ is a convex function in $\mathbf{s}$. Therefore, in order to obtain the optimal values of r, q and $\mathbf{s}$, PM can be used by finding the best $\mathbf{s}$ for fixed r and q and then enumerating all possible r and q combinations. This method is computationally intensive but we will use it to provide a benchmark for our heuristics.

Next, we introduce the revised RO heuristic. For $i \in \mathcal{L}$, we define an auxiliary location attached to retailer i as $\mathcal{S}(i)$. Then, for $i \in \mathcal{L}$, we define $\underline{\Pi}_{\mathcal{S}(i)}(x) = (b_i + h_i)[x]^-$ and $S^{r0}_{\mathcal{S}(i)} = 0$.

Recursive Optimization (RO) Heuristic for OWMR with Fixed Cost at Warehouse

Step 1: Compute S_i^{r0} for retailer i as follows:

$$\begin{aligned} \hat{\Pi}_i(x) &= H_i x + \sum_{j \in \mathcal{S}(i)} \underline{\Pi}_j \left(Bin \left(\left[x - \sum_{k \in \mathcal{S}(i)} S_k^{r0} \right]^-, \theta_j \right) \right) \\ \Pi_i(y) &= E[\hat{\Pi}_i(y - D_i)] \\ S_i^{r0} &= \arg\min_y \Pi_i(y) \\ \underline{\Pi}_i(v) &= \Pi_i(S_i^{r0} - v) \end{aligned} \tag{10}$$

Step 2: Compute (r^r, q^r) for the warehouse as follows:

$$\begin{aligned}
\hat{\Pi}_0(x) &= H_0 x + \sum_{j \in \mathcal{S}(0)} \underline{\Pi}_j \left(Bin \left(\left[x - \sum_{k \in \mathcal{S}(0)} S_k^{r0} \right]^- , \theta_j \right) \right) \\
\Pi_0(r,q) &= \frac{\xi \sum_{i=1}^N \lambda_i + \sum_{s_0=r+1}^{r+q} E[\hat{\Pi}_0(s_0 + \sum_{i=1}^N S_i^{r0} - D_0)]}{q} \qquad (11) \\
(r^r, q^r) &= \arg\min_{r,q} \Pi_0(r,q)
\end{aligned}$$

Step 3: Calculate $\mathbf{s}^r$ for the retailers as follows:

$$\mathbf{s}^r = \arg\min_{\mathbf{s}} \Pi(r^r, q^r, \mathbf{s}) \qquad (12)$$

The properties suggested by the following proposition make the modified RO heuristic computationally efficient.

Proposition 8. *1. $\Pi_0(r,q)$ is convex in r;*

2. Let $r^*(q)$ be the smallest optimal solution to $\min_r \Pi_0(r,q)$. Then $r^*(q)$ is a decreasing function.

Next, we explain how to customize DA for the case with a fixed ordering cost in the warehouse. The decomposition procedure from the OWMR system to serial supply chains remains the same. We use the procedure in Shang (2008) to solve for s_i^d (the base-stock level) for retailer i (> 0), $r_{0_i}^d$ (the reorder point) and $q_{0_i}^d$ (the order quantity) for the warehouse in the decomposed serial supply chain containing retailer i. In particular,

$$s_i^d = F_{D_i}^{-1}\left(\frac{b_i + h_0}{b_i + h_i}\right)$$

and $q_{0_i}^d$ is obtained by solving the following problem

$$q_{0_i}^d = \arg\min_q \frac{\xi \lambda_i + \sum_{x=1}^q \chi_i((R(q) + x)}{q},$$

where $\chi_i(y) = E[h_0(y - \tilde{D}_{0_i}) + (b_i + h_0)(y - \tilde{D}_{0_i})^-]$ and $R_i(q) = \arg\min_y \sum_{x=1}^q \chi_i(y + x)$. Then

$$r_{0_i}^d = R_i(q_{0_i}^d).$$

Given these, the expected backorder at the warehouse can be calculated as

$$E[B_0] = \sum_{i=1}^N \frac{1}{q_{0_i}^d} \sum_{x=r_{0_i}^d+1}^{r_{0_i}^d + q_{0_i}^d} E[(D_{0_i} - x)^+]. \qquad (13)$$

	ξ							
	5		10		15		20	
	RO	DA	RO	DA	RO	DA	RO	DA
$\bar{\epsilon}$	0.01	0.75	0.01	2.02	0.02	2.64	0.0	3.04
med_ϵ	0.00	0.70	0.00	1.98	0.00	2.43	0.00	2.98
max_ϵ	0.13	1.41	0.14	2.78	0.22	3.96	0.08	4.01
σ_ϵ	0.03	0.45	0.03	0.60	0.06	0.64	0.02	0.59
$\bar{t}$	204.55	0.13	404.34	0.20	598.82	0.28	802.63	0.35
	PM		PM		PM		PM	
$\bar{t}$	632.091		1256.49		1868.631		2497.646	

Table 2: Impact of Fixed Ordering Cost on Performance of the Heuristics

We approximate the order quantity at the warehouse using the EOQ model with backorders (e.g., Zipkin 2000). In particular:

$$\begin{aligned} s_i^a &= s_i^d \\ q^a &= \sqrt{\frac{\xi \sum_{i=1}^N \lambda_i (h_0 + b_0)}{h_0 b_0}} \\ r^a &= \min \left\{ y \in \mathbb{Z} : \frac{1}{q^a} \sum_{x=y+1}^{y+q^a} E[(D_0 - x)^+] \le \sum_{i=1}^N \frac{1}{q_{0_i}^d} \sum_{x=r_{0_i}^d+1}^{r_{0_i}^d + q_{0_i}^d} E[(D_{0_i} - x)^+] \right\} \end{aligned}$$

In Table 2, we provide numerical results to examine the impact of a fixed ordering cost on the performance of the heuristics. The results suggest that the cost gap of RO is negligible. We use $[\frac{q^a}{3}, 3q^a]$ as the search space for the order quantity under PM and RO. Due to Proposition 8, the computation time of RO is reduced to one-third of PM's computation time. However, no further reduction in computation time is obtained. The unimodality of $\Pi_0(r^*(q), q)$ in q is not established when r and q can only take integer values (Zipkin 2000, Section 6.5.4). One can, of course, reduce the computation time of both PM and RO by assuming unimodality, as PMU does.

Zheng (1992) and Axsäter (1996) show that the EOQ with backorders is a good approximation for the order quantity of a continuous (r, q) policy at a single location. As the results in Table 2 related to DA suggest, the EOQ with backorders is a good approximation for our problem as well. DA is more than 1000 times faster than both PM and RO.

6 Conclusions

In this paper, we introduce the RO and DA heuristics to approximate the base-stock levels for all locations in a distribution system with a standard setting. By comparing our heuristic with the RD heuristic (Gallego et al. 2007), we find that RO performs the best, on average, followed by DA, and then RD. At the same time, DA is the least computationally intensive heuristic, followed by RD, and then RO.

Our RO and DA heuristics have the potential to be extended to other problems related to distribution systems. We study one such example, consisting of an OWMR system with a fixed order cost at the warehouse. Our extensions of RO and DA are able to extend to an (r, q) policy in warehouse. Both heuristics generate compelling cost gaps, with a much shorter computation time than PM. Future work can extend the DA and RO heuristics to other variants of distribution networks.

References

Axsäter, S. (1990), 'Simple solution procedures for a class of two-echelon inventory problems', *Operations Research* **38**(1), 64–69.

Axsäter, S. (1993), 'Continuous review policies for multi-level inventory systems with stochastic demand', *Handbooks in Operations Research and Management Science* **4**, 175–197.

Axsäter, S. (1996), 'Using the deterministic eoq formula in stochastic inventory control', *Management Science* **42**(6), 830–834.

Billington, C., Callioni, G., Crane, B., Ruark, J., Rapp, J., White, T. & Willems, S. (2004), 'Accelerating the profitability of hewlett-packard's supply chains', *Interfaces* **34**(1), 59–72.

Chen, F. & Zheng, Y. (1994), 'Lower bounds for multi-echelon stochastic inventory systems', *Management Science* **40**, 1426–1443.

Clark, J. & Scarf, H. (1960), 'Optimal policies for a multi-echelon inventory problem', *Management Science* **6**(4), 475–490.

DeCroix, G. (2013), 'Inventory management for an assembly system subject to supply disruptions', *Management Science* **59**(9), 2079–2092.

Eppen, G. D. (1979), 'Effects of centralization on expected costs in a multi-location newsboy problem', *Management Science* **25**(5), 498–501.

Federgruen, A. & Zipkin, P. (1984a), 'Approximations of dynamic, multilocation production and inventory problems', *Management Science* **30**, 69–84.

Gallego, G. & Moon, I. (1993), 'The distribution-free newsboy problem: Review and extensions', *Journal of Operations Research Society* **44**(8), 825–834.

Gallego, G. & Özer, O. (2005), 'A new algorithm and a new heuristic for serial supply systems', *Operations Research Letters* **33**, 349–362.

Gallego, G., Özer, O. & Zipkin, P. (2007), 'Bounds, heuristics, and approximations for distribution systems', *Operations Research* **55**(3), 503–517.

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

Graves, S. (1985), 'A multi-echelon inventory model for a repairable item with one-for-one replenishment', *Management Science* **31**, 1247–1256.

Graves, S. & Willems, S. (2003), 'Supply chain design: safety stock placement and supply chain configuration', *Handbooks in Operations Research and Management Science* **11**, 95–132.

Huh, W. & Janakiraman, G. (2010), 'On the optimal policy structure in serial inventory systems with lost sales', *Operations Research* **58**(2), 486–491.

Huh, W., Janakiraman, G. & Nagarajan, M. (2010), 'Technical notecapacitated serial inventory systems: Sample path and stability properties under base-stock policies', *Operations Research* **58**(4-Part-1), 1017–1022.

Huh, W. T. & Janakiraman, G. (2008), 'A sample-path approach to the optimality of echelon order-up-to policies in serial inventory systems', *Operations Research Letters* **36**(5), 547–550.

Özer, O. & Xiong, H. (2008), 'Stock positioning and performance estimation for distribution systems with service constraints', *IIE Transactions* **40**(12), 1141–1157.

Rosling, K. (1989), 'Optimal inventory policies for assembly systems under random demand', *Operations Research* **37**, 565–579.

Shaked, M. & Shanthikumar, G. (1994), *Stochastic orders and their applications*, Academic Press, San Diego, California.

Shang, K. (2008), 'Note: A simple heuristic for serial inventory systems with fixed order costs', *Operations Research* **56**(4), 1039–1043.

Shang, K. (2012), 'Single-stage approximations for optimal policies in serial inventory systems with non-stationary demand', *Manufacturing & Service Operations Management* **14**(3), 414–422.

Shang, K. & Song, J.-S. (2003), 'Newsvendor bounds and heuristic for optimal policies in serial supply chains', *Management Science* **49**(5), 618–638.

Shang, K. & Song, J.-S. (2006), 'A Closed-Form Approximation for Serial Inventory Systems and Its Application to System Design', *Manufacturing & Service Operations Management* **8**(4), 394–406.

Sherbrooke, C. (1968), 'Metric: A multi-echelon technique for recoverable item control', *Operations Research* **16**, 122–141.

Shi, J. & Zhao, Y. (2010), 'Technical note: some structural results on acyclic supply chains', *Naval Research Logistics* **57**(6), 605–613.

Zheng, Y.-S. (1992), 'On properties of stochastic inventory systems', *Management Science* **38**(1), 87–103.

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

Appendix to
Heuristics for Base-Stock Levels in Multi-Echelon Distribution Networks with First-Come First Served Policies

Appendix A: Definitions and Notation

	Network Topology
0	root location
N	number of non-root locations
V	set of all locations
E	set of all directed edges
T	directed tree for a distribution system with location set V and edge set E
$\langle i, j \rangle$	edge from location i to location j
$\mathcal{P}(i)$	predecessor of location i, $\{j \in V : \langle j, i \rangle \in E\}$
$\mathcal{S}(i)$	set of successors of location i, $\{j \in V : \langle i, j \rangle \in E\}$
$\mathcal{P}(0)$	external supplier
$\mathcal{L}$	set of all leaves, $\{i \in V : \mathcal{S}(i) = \emptyset\}$
$T(i)$	subtree rooted at location i and containing all downstream locations
$\mathcal{A}(i, j)$	set of locations in the directed path from i to j
w	index for serial systems obtained after decomposing T
	Parameters
λ_i	demand rate at location i, $\forall i \in V$
L_i	leadtime between $\mathcal{P}(i)$ and i, $\forall i \in V$
D_i	leadtime demand at location i, $\forall i \in V$
$\tilde{D}_{i_w}$	total leadtime demand for locations i to the leaf node of serial system w
b_i	unit backordering cost at location i, $\forall i \in \mathcal{L}$
h_i	unit local holding cost at location i, $\forall i \in V$
H_i	unit echelon holding cost at location i, $H_i = h_i - h_{\mathcal{P}(i)}$, with $H_0 = h_0$
ξ	fixed ordering cost at the warehouse, $\xi = 0$(except Section 5)

Table 3: Definitions and Notation for Network Topology and Parameters

I_i	local on-hand inventory at location i, $\forall i \in V$
B_i	local backorders at location i, $\forall i \in V$
$B_{i,j}$	backorders at location i due to demands from location j, $\forall j \in \mathcal{S}(i)$
IT_i	inventory in transit to location i, $\forall i \in V$

Table 4: Definitions and Notation for Local State Variables

$\mathbf{S}$	vector of echelon base-stock levels, $\mathbf{S} := (S_i)_{i=0}^N$
$\mathbf{s}$	vector of local base-stock levels, $\mathbf{s} := (s_i)_{i=0}^N$
$\ell(\mathbf{S})$	$:= \{\ell_0(\mathbf{S}), \ell_1(\mathbf{S}), ..., \ell_N(\mathbf{S})\}$, with $\ell_i(\mathbf{S}) = S_i - \sum_{j\in\mathcal{S}(i)} S_j$
S_i^{r0}	intermediate echelon base-stock level at location i suggested by the RO heuristic
s_i^r	final local base-stock level at location i suggested by the RO heuristic
$S_{i_w}^{SS}$	echelon base-stock level (Shang & Song (2003)) at location i in serial system w
$s_{i_w}^{SS}$	local base-stock level (Shang & Song (2003)) at location i in serial system w
$S_{i_w}^d$	$\min_{k_w\in\mathcal{A}(0,i_w)} \{S_{k_w}^{SS}\}$
$s_{i_w}^d$	$S_{i_w}^d - S_{j_w}^d$, where j_w is the successor of i_w in system w
s_i^a	local base-stock level at location i suggested by the DA heuristic
$\mathscr{C}(\mathbf{s})$	long-run expected cost per unit time for any given $\mathbf{s}$
$C_0(S_0\|\mathbf{S})$	long-run expected cost per unit time for any given $\mathbf{S}$

Table 5: Definitions and Notation for Decision Variables and Cost Functions

Appendix B: Proofs

Proof of Proposition 1

We first define $\mathscr{C}^k(\mathbf{s})$ as the cost of a subtree $T(k)$ with the local base-stock level vector $\mathbf{s}$ in the following manner:

$$\mathscr{C}^k(\mathbf{s}) = \sum_{i\in T(k)\setminus\mathcal{L}} (h_i - h_{\mathcal{P}(k)})E\left[I_i^k(\mathbf{s}) + \sum_{j\in\int(i)} IT_j\right] + \sum_{i\in T(k)\cap\mathcal{L}} \left[(h_i - h_{\mathcal{P}(k)})E[I_i^k(\mathbf{s})] + (b_i + h_{\mathcal{P}(k)})E[B_i^k(\mathbf{s})]\right],$$

where $I_i^k(\mathbf{s})$ and $B_i^k(\mathbf{s})$ are defined as

$$\begin{aligned} B_i^k(\mathbf{s}) &= [B_{\mathcal{P}(i),i}^k(\mathbf{s}) + D_i - s_i]^+ \qquad \forall i \in T(k) \\ I_i^k(\mathbf{s}) &= [B_{\mathcal{P}(i),i}^k(\mathbf{s}) + D_i - s_i]^- \qquad \forall i \in T(k) \\ B_{i,j}^k(\mathbf{s}) &\overset{Dist}{=} Bin(B_i^k(\mathbf{s}), \theta_j), \qquad \forall i \in T(k) \quad \& \quad \forall j \in \mathcal{S}(i), \end{aligned}$$

with $B_{\mathcal{P}(k),k}^k(\mathbf{s}) \equiv 0$. Compared to expressions (1), it is easy to see that $I_i^0(\mathbf{s}) = I_i(\mathbf{s})$ and $B_i^0(\mathbf{s}) = B_i(\mathbf{s})$. Therefore, we only need to show $\mathscr{C}^0(\ell(\mathbf{S})) = C_0(S_0|\mathbf{S})$.

We use mathematical induction to prove this proposition. For all $n \in \mathcal{L}$, we have $\mathscr{C}^n(\ell(\mathbf{S})) = C_n(S_n|\mathbf{S})$. Next, we assume that, for all $n \in \mathcal{S}(k)$, $\mathscr{C}^n(\ell(\mathbf{S})) = C_n(S_n|\mathbf{S})$ holds. Then we need to show that $\mathscr{C}^k(\ell(\mathbf{S})) = C_k(S_k|\mathbf{S})$.

First, define a new vector $\boldsymbol{\gamma}^n(\mathbf{S}|Z)$ as $\gamma_i^n(\mathbf{S}|Z) = \ell_i(\mathbf{S})$ for $i \neq n$ and $\gamma_n^n(\mathbf{S}|Z) = \ell_n(\mathbf{S}) - Z$. By the definition of $B_{k,n}^k(\ell(\mathbf{S}))$ for $n \in \mathcal{S}(k)$, we have $B_{k,n}^k(\ell(\mathbf{S})) \overset{Dist}{=} Bin\left(\left[S_k - \sum_{n\in\mathcal{S}(k)} S_n - D_n\right]^-, \theta_k\right)$. Then, one can verify that, for all $n \in \mathcal{S}(k)$ and $i \in T(k) \setminus \{k\}$, we have

$$E[I_i^k(\ell(\mathbf{S}))] = E[I_i^n(\boldsymbol{\gamma}^n(\mathbf{S}|Z))], \quad E[B_i^k(\ell(\mathbf{S}))] = E[B_i^n(\boldsymbol{\gamma}^n(\mathbf{S}|Z))] \tag{14}$$

for $Z \overset{Dist}{=} B_{k,n}^k(\ell(\mathbf{S}))$. Next we have,

$$\begin{aligned}
C_k(S_k|\mathbf{S}) &= H_k E[S_k - D_k] + \sum_{n\in\mathcal{S}(k)} E\left[C_n\left(S_n - B_{k,n}^k(\ell(\mathbf{S}))\middle|\mathbf{S}\right)\right] \\
&= H_k E[S_k - D_k] + \sum_{n\in\mathcal{S}(k)} \sum_{i\in T(n)\setminus\mathcal{L}} (h_i - h_k) E\left[I_i^n(\gamma^n(\mathbf{S}|B_{k,n}^k(\ell(\mathbf{S})))) + \sum_{j\in\mathcal{S}(i)} IT_j\right] \\
&\quad + \sum_{n\in\mathcal{S}(k)} \sum_{i\in T(n)\cap\mathcal{L}} \Big((h_i - h_k)E[I_i^n(\gamma^n(\mathbf{S}|B_{k,n}^k(\ell(\mathbf{S}))))] + (b_i + h_k)E[B_i^n(\gamma^n(\mathbf{S}|B_{k,n}^k(\ell(\mathbf{S}))))]\Big) \\
&= H_k E[S_k - D_k] + \sum_{n\in\mathcal{S}(k)} \sum_{i\in T(n)\setminus\mathcal{L}} (h_i - h_k) E\left[I_i^k(\ell(\mathbf{S})) + \sum_{j\in\mathcal{S}(i)} IT_j\right] \\
&\quad + \sum_{n\in\mathcal{S}(k)} \sum_{i\in T(n)\cap\mathcal{L}} \Big((h_i - h_k)E[I_i^k(\ell(\mathbf{S}))] + (b_i + h_k)E[B_i^k(\ell(\mathbf{S}))]\Big) \\
&= H_k \left(\sum_{i\in T(k)\setminus\mathcal{L}} E\left[I_i^k(\ell(\mathbf{S})) + \sum_{j\in\mathcal{S}(i)} IT_j\right] + \sum_{i\in T(k)\cap\mathcal{L}} E[I_i^k(\ell(\mathbf{S})) - B_i^k(\ell(\mathbf{S}))]\right) \\
&\quad + \sum_{n\in\mathcal{S}(k)} \sum_{i\in T(n)\setminus\mathcal{L}} (h_i - h_k) E\left[I_i^k(\ell(\mathbf{S})) + \sum_{j\in\mathcal{S}(i)} IT_j\right] \\
&\quad + \sum_{n\in\mathcal{S}(k)} \sum_{i\in T(n)\cap\mathcal{L}} ((h_i - h_k)E[I_i^k(\ell(\mathbf{S}))] + (b_i + h_k)E[B_i^k(\ell(\mathbf{S}))]) \\
&= \mathscr{C}^k(\ell(\mathbf{S}))
\end{aligned}$$

The first equality can directly be obtained using the recursive functions (3). The second equality is due to the induction hypothesis. The third equality utilizes expressions (14). The fourth equality follows from the use of the echelon base-stock policy. Finally, the last equality is due to the definition of the unit echelon holding cost.

Proof of Proposition 2

Our proof uses the following lemma from Shaked & Shanthikumar (1994, Example 6.A.2).

Lemma 9. *Let $\beta(i,n,p)$ represent the binomial probability mass function of i units drawn from n units with each unit picked with probability p. Let $G(n) = \sum_{i=0}^{n} f(i)\beta(i,n,p)$ for $n \geq 0$.*

1. *If $f(i)$ is convex in i, then $G(n)$ is convex in n.*

2. *If $f(i)$ is increasing in i, then $G(n)$ is increasing in n.*

We apply mathematical induction on the echelon, starting from the leaves. Certainly for $i \in \mathcal{L}$, $C_i(\cdot)$ and $\underline{C}_i(\cdot)$ are convex. Now let $i \in V \backslash \mathcal{L}$. Suppose that $\underline{C}_j(\cdot)$ is convex for all $j \in T(i) \setminus \{i\}$. For arbitrary d_i, a realization of D_i, we have

$$C_i(y|D_i = d_i) = H_i(y - d_i) + \sum_{j\in\mathcal{S}(i)} E\left[\underline{C}_j\left(Bin\left(\left(d_i - \left(y - \sum_{j\in\mathcal{S}(i)} S_j^{r0}\right)\right)^+, \theta_j\right)\right)\right]$$

Next, we show that $C_i(y|D_i = d_i)$ is convex in y. Then by the preservation of convexity under expectation, we can infer the convexity of $C_i(\cdot)$. Define

$$A_j(y,d) = E\left[\underline{C}_j\left(Bin\left(\left(d_i - \left(y - \sum_{j\in\mathcal{S}(i)} S_j^{r0}\right)\right)^+, \theta_j\right)\right)\right].$$

In order to show that $C_i(y|D_i = d_i)$ is convex in y, it suffices to show that, for each $j \in \mathcal{S}(i)$, $A_j(y, d_i)$ is convex in y. We have

$$A_j(y,d_i) = \begin{cases} \underline{C}_j(0), & y \geq d_i + \sum_{j\in\mathcal{S}(i)} S_j^{r0} \\ E\left[\underline{C}_j\left(Bin\left(d_i + \sum_{j\in\mathcal{S}(i)} S_j^{r0} - y, \theta_j\right)\right)\right], & y < d_i + \sum_{j\in\mathcal{S}(i)} S_j^{r0} \end{cases}$$

When $y \geq d_i + \sum_{j\in\mathcal{S}(i)} S_j^{r0}$, $A_j(y,d_i) = \underline{C}_j(0)$ is constant. Moreover, $\underline{C}_j(0) = \min_v \underline{C}_j(v)$, since S_j^{r0} is the minimizer of $C_j(\cdot)$. Therefore, $\underline{C}_j(0) \leq A_j(y,d_i)$ for $y < d_i + \sum_{j\in\mathcal{S}(i)} S_j^{r0}$. Thus, we only need to show that $A_j(y,d_i)$ is decreasing convex in y when $y < d_i + \sum_{j\in\mathcal{S}(i)} S_j^{r0}$. Indeed, by setting $n = d_i + \sum_{j\in\mathcal{S}(i)} S_j^{r0} - y$, Lemma 9 suggests that

$$E\left[\underline{C}_j\left(Bin\left(d_i + \sum_{j\in\mathcal{S}(i)} S_j^{r0} - y, \theta_j\right)\right)\right]$$

is decreasing convex in y since $\underline{C}_j(n)$ is increasing convex for $n \geq 0$ by the inductive hypothesis. Since $A_j(y,d)$ is a decreasing function and $A_j(y,d_i) - A_j(y-1,d_i) = 0$ for $y \geq d_i + \sum_{j\in\mathcal{S}(i)} S_j^{r0}$, then we have $A_j(y+1,d_i) - A_j(y,d_i) \geq 0 = A_j(y,d_i) - A_j(y-1,d_i)$ for $y = d_i + \sum_{j\in\mathcal{S}(i)} S_j^{r0}$.

Therefore, the convexity is also satisfied at $y = d_i + \sum_{j\in\mathcal{S}(i)}$. Furthermore, $\underline{C}_i(\cdot)$ is convex since linear transformation preserves convexity.

Proof of Theorem 3

Part 1:

We first show the asymptotic result for RO. Let $c_i(s_i) = h_i E[(s_i - D_i)^+] + b_i E[(D_i - s_i)^+]$ be the single-stage cost function for location i. Let s_i^u be the minimizer of $c_i(s_i)$. From Proposition 1 in Gallego et al. (2007), we know that $\sum_{i>0} c_i(s_i^u) \le c^*(h_0)$. Therefore, $1 \le \frac{c^r(h_0)}{c^*(h_0)} \le \frac{c^r(h_0)}{\sum_{i>0} c_i(s_i^u)}$.

Using (4) and (5), we know that $c^r(h_0) \le \min_y C_0(y|S_1^{r0}(h_0), S_2^{r0}(h_0), ..., S_N^{r0}(h_0))$ for a given h_0, where $S_i^{r0}(h_0)$ is the minimizer of $\eta_i(x|h_0) = (h_i - h_0)E[x - D_i] + (h_i + b_i)E[(x - D_i)^-]$. Since $h_i > h_0$, $S_i^{r0}(h_0)$ is a finite number. Let B_{0i} be the backorder level at the warehouse due to demands from retailer i. Then we have

$$\begin{aligned}
& C_0(y|S_1^{r0}(h_0), S_2^{r0}(h_0), ..., S_N^{r0}(h_0)) \\
= \; & h_0 E\left[y - D_0\right] + \sum_{i>0} \left[(h_i - h_0)E\left[S_i^{r0}(h_0) - (D_i + B_{0i})\right] + (h_i + b_i)E[(S_i^{r0}(h_0) - (D_i + B_{0i}))^-]\right] \\
\le \; & h_0 E\left[y - D_0\right] + \sum_{i>0}(h_i + b_i)E[B_{0i}] + \sum_{i>0} \eta_i(S_i^{r0}(h_0)|h_0) \\
\le \; & h_0 \sum_{i>0} S_i^{r0}(h_0) + h_0 E\left[y - \sum_{i>0} S_i^{r0}(h_0) - D_0\right] \\
& + \sum_{i>0}(h_i + b_i)E\left[\left(y - \sum_{i>0} S_i^{r0}(h_0) - D_0\right)^-\right] + \sum_{i>0} \eta_i(S_i^u|h_0)
\end{aligned}$$

The first inequality is due to the fact that $S_i^{r0}(h_0)$ is the minimizer of $\eta_i(x|h_0)$. Since $|\eta_i'(x|h_0)| \le h_i + b_i$, we have $\eta_i(S_i^{r0}(h_0) - y|h_0) - \eta_i(S_i^{r0}(h_0)|h_0) \le (h_i + b_i)|y|$. Therefore,

$$E[\eta_i(S_i^{r0}(h_0) - B_{0i}|h_0) - \eta_i(S_i^{r0}(h_0)|h_0)] \le (h_i + b_i)E[B_{0i}]$$

since $B_{0i} \ge 0$. The second inequality is due to the fact that $E[B_{0i}] \le E[B_0] = E\left[\left(y - \sum_{i>0} S_i^{r0}(h_0) - D_0\right)^-\right]$. Moreover, $S_i^{r0}(h_0)$ is the minimizer of $\eta_i(x|h_0)$.

Then

$$\begin{aligned}
c^r(h_0) \; & \le \; \min_y C_0(y|S_1^{r0}(h_0), S_2^{r0}(h_0), ..., S_N^{r0}(h_0)) \\
& \le \; h_0 \sum_{i>0} S_i^{r0}(h_0) + \sum_{i>0} \eta_i(S_i^u|h_0)
\end{aligned}$$

$$+\min_y \left(h_0 E\left[y - \sum_{i>0} S_i^{r0}(h_0) - D_0 \right] + \sum_{i>0}(h_i+b_i) E\left[\left(y - \sum_{i>0} S_i^{r0}(h_0) - D_0 \right)^- \right] \right)$$

$$\le \quad h_0 \sum_{i>0} S_i^{r0}(h_0) + \sum_{i>0} \eta_i(S_i^u|h_0) + \sqrt{h_0\left(\sum_{i>0}(h_i+b_i) - h_0\right) L_0 \sum_{i>0} \lambda_i}$$

The last inequality is due to the distribution-free upper bound developed by Gallego & Moon (1993).

It is easy to see that $\lim_{h_0\to 0} h_0 \sum_{i>0} S_i^{r0}(h_0) = 0$ and $\lim_{h_0\to 0} \sqrt{h_0\left(\sum_{i>0}(h_i+b_i) - h_0\right) L_0 \sum_{i>0} \lambda_i} = 0$. Moreover, $\lim_{h_0\to 0} \eta_i(S_i^u|h_0) = c_i(s_i^u)$. Therefore,

$$\begin{aligned} 1 \quad &\le \quad \lim_{h_0\to 0} \frac{c^r(h_0)}{c^*(h_0)} \le \lim_{h_0\to 0} \frac{\min_y C_0(y|S_1^{r0}(h_0), S_2^{r0}(h_0), ..., S_N^{r0}(h_0))}{\sum_{i>0} c_i(s_i^u)} \\ &\le \quad \lim_{h_0\to 0} \frac{h_0 \sum_{i>0} S_i^{r0}(h_0) + \sum_{i>0} \eta_i(S_i^u|h_0) + \sqrt{h_0\left(\sum_{i>0}(h_i+b_i) - h_0\right) L_0 \sum_{i>0} \lambda_i}}{\sum_{i>0} c_i(s_i^u)} = 1. \end{aligned}$$

Now we show the asymptotic result for DA. For sufficiently small h_0, we have $s_i^a(h_0) = S_i^{r0}(h_0)$ for $i > 0$ and $s_{0_i}^d(h_0) = S_{0_i}^{SS}(h_0) - s_i^a(h_0)$. From the previous part of the proof, we know that

$$\begin{aligned} & C_0(s_0^a(h_0)|s_1^a(h_0), s_2^a(h_0), ..., s_N^a(h_0)) \\ \le \quad & h_0 \sum_{i>0} s_i^a(h_0) + \sum_{i>0} \eta_i(S_i^u|h_0) \\ & + h_0 E\left[s_0^a(h_0) - \sum_{i>0} s_i^a(h_0) - D_0 \right] + \sum_{i>0}(h_i+b_i) E\left[\left(s_0^a(h_0) - \sum_{i>0} s_i^a(h_0) - D_0 \right)^- \right] \end{aligned}$$

To show the asymptotic result for DA, we only need to show that

$$h_0 E\left[s_0^a(h_0) - \sum_{i>0} s_i^a(h_0) - D_0 \right] + \sum_{i>0}(h_i+b_i) E\left[\left(s_0^a(h_0) - \sum_{i>0} s_i^a(h_0) - D_0 \right)^- \right]$$

goes to 0 when h_0 goes to zero.

To show that $\lim_{h_0\to 0}(h_i+b_i) E\left[\left(s_0^a(h_0) - \sum_{i>0} s_i^a(h_0) - D_0 \right)^- \right] = 0$ for each i, we only need to show that $\lim_{h_0\to 0} s_0^a(h_0) = \infty$. From (6), we can see that $\lim_{h_0\to 0} S_{0_i}^{SS}(h_0) = \infty$ and $\lim_{h_0\to 0} s_i^a(h_0)$ approaches a finite number since $h_i > h_0$. Therefore, $\lim_{h_0\to 0} Q_{D_{0_i}}(s_{0_i}^d(h_0)) = 0$. As a result, from (8), we can see that $\lim_{h_0\to 0} s_0^a(h_0)$ goes to ∞.

Since both $s_i^a(h_0)$ and $E[D_0]$ have finite upper bounds, we only need to show that $\lim_{h_0\to 0} h_0 s_0^a(h_0) = 0$. From Theorem 7, we know that $s_0^a(h_0) \le \sum_i s_{0_i}^d(h_0)$. For each i, we have $s_{0_i}^d(h_0) \le S_{0_i}^{SS}(h_0) \le G_{\tilde{D}_{0_i}}^{-1}\left(\frac{h_0}{b_i+h_0}\right)$ where the second inequality follows (6). Assume that the CDF of $\tilde{D}_{0_i}$ is the $G(\cdot)$ function defined in Section 3.2. Then $G_{\tilde{D}_{0_i}}^{-1}\left(\frac{b_i}{b_i+h_0}\right)$ is the optimal solution to the problem

$$\min_y E\left[h_0(y - \tilde{D}_{0_i}) + (h_0+b_i)(y - \tilde{D}_{0_i})^- \right].$$

By the distribution-free upper bound developed by Gallego & Moon (1993), we know that

$$E\left[h_0\left(G^{-1}_{\tilde{D}_{0_i}}\left(\frac{h_0}{b_i+h_0}\right)-\tilde{D}_{0_i}\right)+(h_0+b_i)\left(G^{-1}_{\tilde{D}_{0_i}}\left(\frac{h_0}{b_i+h_0}\right)-\tilde{D}_{0_i}\right)^-\right]\le\sqrt{h_0b_iE[\tilde{D}_{0_i}]}.$$

Therefore, $\lim_{h_0\to 0} h_0 G^{-1}_{\tilde{D}_{0_i}}\left(\frac{h_0}{b_i+h_0}\right)=0$. As a result, $\lim_{h_0\to 0} h_0 s^a_0(h_0)=0$.

Part 2:

From RO in (4), we can see that S^{r0}_i is the minimizer of $\eta_i(x)=(h_i-h_0)E[x-D_i]+(h_i+b_i)E[(x-D_i)^-]$. Moreover, it is evident that $\sum_{i=1}^N\eta_i(S^{r0}_i)\le c^*(N)$ since $\sum_{i=1}^N\eta_i(S^{r0}_i)$ ignores the echelon holding cost due to the warehouse's holding cost and the additional retailer costs due to warehouse backorders. Using (4) and (5), we know that $c^r(N)\le\min_y C_0(y|S^{r0}_1(h_0),S^{r0}_2(h_0),...,S^{r0}_N(h_0))$ Therefore, we only need to show $\frac{\min_y C_0(y|S^{r0}_1(h_0),S^{r0}_2(h_0),...,S^{r0}_N(h_0))}{\sum_{i=1}^N\eta_i(S^{r0}_i)}$ approaches 1 as N increases.

From the proof of asymptotic optimality of RO as $h_0\to 0$, we know that

$$\begin{aligned}&C_0(y|S^{r0}_1,S^{r0}_2,...,S^{r0}_N)\\ \le\;& h_0\sum_{i>0}S^{r0}_i+h_0E\left[y-\sum_{i>0}S^{r0}_i-D_0\right]+\sum_{i>0}(h_i+b_i)E\left[\left(y-\sum_{i>0}S^{r0}_i-D_0\right)^-\right]+\sum_{i>0}\eta_i(S^{r0}_i)\end{aligned}$$

Let $\kappa=\max\left(h_0,\sum_{i>0}(h_i+b_i)-h_0\right)$. Then we can get the following relationship:

$$\begin{aligned}&h_0\sum_{i>0}S^{r0}_i+h_0E\left[y-\sum_{i>0}S^{r0}_i-D_0\right]+\sum_{i>0}(h_i+b_i)E\left[\left(y-\sum_{i>0}S^{r0}_i-D_0\right)^-\right]\\ \le\;&\kappa\sum_{i>0}S^{r0}_i+\kappa E\left[\left(y-\sum_{i>0}S^{r0}_i-D_0\right)^+\right]+\kappa E\left[\left(y-\sum_{i>0}S^{r0}_i-D_0\right)^-\right]\\ \le\;&\kappa E\left[(y-D_0)^+\right]+\kappa E\left[(y-D_0)^-\right]+\kappa E\left[\left(y-2\sum_{i>0}S^{r0}_i-D_0\right)^+\right]+\kappa E\left[\left(y-2\sum_{i>0}S^{r0}_i-D_0\right)^-\right]\end{aligned}$$

The last inequality is due to the fact that $\omega+\psi^++\psi^-\le(\psi+\omega)^++(\psi+\omega)^-+(\psi-\omega)^++(\psi-\omega)^-$ for all ω and ψ.

Then we define a random variable R where $R=0$ with probability 0.5 and $R=2\sum_{i>0}S^{r0}_i$ with probability 0.5. Then we have

$$\begin{aligned}&\kappa E\left[(y-D_0)^+\right]+\kappa E\left[(y-D_0)^-\right]+\kappa E\left[\left(y-2\sum_{i>0}S^{r0}_i-D_0\right)^+\right]+\kappa E\left[\left(y-2\sum_{i>0}S^{r0}_i-D_0\right)^-\right]\\ =\;&2\kappa E\left[(y-D_0+R)^+\right]+2\kappa E\left[(y-D_0+R)^-\right].\end{aligned}$$

Therefore,

$$\min_y C_0(y|S^{r0}_1,S^{r0}_2,...,S^{r0}_N)\le\sum_{i>0}\eta_i(S^{r0}_i)+2\kappa\sqrt{L_0\sum_{i>0}\lambda_i+E[R]}=\sum_{i>0}\eta_i(S^{r0}_i)+2\kappa\sqrt{L_0\sum_{i>0}\lambda_i+\sum_{i>0}S^{r0}_i}.$$

Let $\underline{c} = \min_i \eta_i(S_i^{r0})$, $\bar{\lambda} = \max_i \lambda_i$ and $\bar{S}_i^{r0} = \max_i S_i^{r0}$. Since there exists $\delta > 0$ such that $h_0 < h_i - \delta$ for all i and $b_i, \lambda_i < \infty$, it follows that the S_i^{r0} are bounded above by a finite number and $\underline{c}$ is strictly positive. Then we have

$$\frac{\min_y C_0(y|S_1^{r0}(h_0), S_2^{r0}(h_0), ..., S_N^{r0}(h_0))}{\sum_{i=1}^N \eta_i(S_i^{r0})} \leq 1 + \frac{2\kappa\sqrt{N\left(L_0\bar{\lambda} + \bar{S}_i^{r0}\right)}}{N\underline{c}}.$$

It is clear that, when N approaches ∞, the method is asymptotically optimal.

Part 3:

First we show that

$$\begin{aligned}
c^*(L_0) &= \min_{\mathbf{s}} h_0 E[(s_0 - D_0)^+] + \sum_{i>0} \left(h_i E[(s_i - D_i - B_{0i}(s_0))^+] + bE[(s_i - D_i - B_{0i}(s_0))^-]\right) \\
&\geq \min_{\mathbf{s}} h_0 E[(s_0 - D_0)^+] + \sum_{i>0} \left(h_i E[(s_i - \lambda_i L_i - B_{0i}(s_0))^+] + bE[(s_i - \lambda_i L_i - B_{0i}(s_0))^-]\right) \\
&\geq \min_{\mathbf{s}} h_0 E[(s_0 - D_0)^+] + \sum_{i>0} \left(h_0 E[(s_i - B_{0i}(s_0))^+] + bE[(s_i - B_{0i}(s_0))^-]\right) \\
&\geq \min_{\mathbf{s}} h_0 E[(s_0 - D_0)^+] + h_0 E\left[\left(\sum_{i>0} s_i - (s_0 - D_0)^-\right)^+\right] + bE\left[\left(\sum_{i>0} s_i - (s_0 - D_0)^-\right)^-\right]
\end{aligned}$$

The first inequality is due to the equality $E[\lambda_i L_i + B_{0i}(s_0)] = E[D_i + B_{0i}(s_0)]$, the inequality $Var[\lambda_i + B_{0i}(s_0)] \leq Var[D_i + B_{0i}(s_0)]$ and the convexity of $h_i(s_i - x)^+ + b(s_i - x)^-$ in x. Redefining s_i as $s_i - \lambda_i L_i$ for all i, the second inequality holds since $h_i \geq h_0$. Finally, the last inequality is due to $x^+ + y^+ \geq (x + y)^+$.

Since $(s_0 - D_0)^-$ is a nonnegative random variable for any s_0, $\sum_{i>0} s_i^*$ is nonnegative where s_i^* is the optimal solution to $\min_{\mathbf{s}} h_0 E[(s_0 - D_0)^+] + h_0 E[(\sum_{i>0} s_i - (s_0 - D_0)^-)^+] + bE[(\sum_{i>0} s_i - (s_0 - D_0)^-)^-]$. For $D_0 \leq s_0$, we have

$$\begin{aligned}
&h_0(s_0 - D_0)^+ + h_0\left(\sum_{i>0} s_i^* - (s_0 - D_0)^-\right)^+ + b\left(\sum_{i>0} s_i^* - (s_0 - D_0)^-\right)^- \\
&= h_0\left(s_0 + \sum_{i>0} s_i^* - D_0\right) \\
&= h_0\left(s_0 + \sum_{i>0} s_i^* - D_0\right)^+ + b\left(s_0 + \sum_{i>0} s_i^* - D_0\right)^-
\end{aligned}$$

Since $(s_0 - D_0)^- = 0$ for $D_0 \leq s_0$ and $\sum_{i>0} s_i^* \geq 0$, we have $\left(\sum_{i>0} s_i^* - (s_0 - D_0)^-\right)^+ = \sum_{i>0} s_i^*$ and $\left(\sum_{i>0} s_i^* - (s_0 - D_0)^-\right)^- = 0$, which leads to the first equality. The second equality is due to the fact that $s_0 + \sum_{i>0} s_i^* - D_0 \geq 0$ when $D_0 \leq s_0$. For $D_0 > s_0$, we have

$$
\begin{aligned}
& h_0(s_0 - D_0)^+ + h_0 \left(\sum_{i>0} s_i^* - (s_0 - D_0)^- \right)^+ + b \left(\sum_{i>0} s_i^* - (s_0 - D_0)^- \right)^- \\
= \;& h_0 \left(s_0 + \sum_{i>0} s_i^* - D_0 \right)^+ + b \left(s_0 + \sum_{i>0} s_i^* - D_0 \right)^-
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
& \min_{\mathbf{s}} h_0 E[(s_0 - D_0)^+] + h_0 E\left[\left(\sum_{i>0} s_i - (s_0 - D_0)^- \right)^+ \right] + bE\left[\left(\sum_{i>0} s_i - (s_0 - D_0)^- \right)^- \right] \\
= \;& \min_{s_0} E\left[h_0(s_0 - D_0)^+ + h_0 \left(\sum_{i>0} s_i^* - (s_0 - D_0)^- \right)^+ + b \left(\sum_{i>0} s_i^* - (s_0 - D_0)^- \right)^- \right] \\
= \;& \min_{s_0} h_0 E\left[\left(s_0 + \sum_{i>0} s_i^* - D_0 \right)^+ \right] + bE\left[\left(s_0 + \sum_{i>0} s_i^* - D_0 \right)^- \right]
\end{aligned}
$$

By redefining $s_0 := s_0 + \sum_{i>0} s_i^*$, we have

$$
c^*(L_0) \geq \min_{s_0} h_0 E[(s_0 - D_0)^+] + bE[(s_0 - D_0)^-].
$$

Next, we provide a bound for $c^r(L_0)$.

$$
\begin{aligned}
& c^r(L_0) \\
\leq \;& \min_{s_0} C_0 \left(s_0 + \sum_{i>0} S_i^{r0} \right) - h_0 \sum_{i>0} E[D_i] \\
= \;& \min_{s_0} h_0 E\left[s_0 + \sum_{i>0} S_i^{r0} - D_0 \right] - h_0 \sum_{i>0} E[D_i] \\
& + \sum_{i>0} \left(H_i E[S_i^{r0} - D_i - B_{0i}(s_0)] + (b + h_i) E[(S_i^{r0} - D_i - B_{0i}(s_0))^-] \right) \\
= \;& \min_{s_0} h_0 E[(s_0 - D_0)^+] - h_0 \sum_{i>0} E[B_{0i}(s_0)] + h_0 \sum_{i>0} \left(E[(S_i^{r0} - D_i)^+] - E[(S_i^{r0} - D_i)^-] \right) \\
& + \sum_{i>0} \left(H_i E[(S_i^{r0} - D_i - B_{0i}(s_0))^+] + (b + h_0) E[(S_i^{r0} - D_i - B_{0i}(s_0))^-] \right) \\
\leq \;& \min_{s_0} h_0 E[(s_0 - D_0)^+] \\
& + \sum_{i>0} \left(h_0 E[(S_i^{r0} - D_i)^+] + H_i E[(S_i^{r0} - D_i - B_{0i}(s_0))^+] + bE[(S_i^{r0} - D_i - B_{0i}(s_0))^-] \right) \\
\leq \;& \sum_{i>0} \left(h_0 E[(S_i^{r0} - D_i)^+] + H_i E[(S_i^{r0} - D_i)^+] + bE[(S_i^{r0} - D_i)^-] \right) \\
& + \min_{s_0} h_0 E[(s_0 - D_0)^+] + bE[(s_0 - D_0)^-]
\end{aligned}
$$

The first inequality is due to the fact that step 2 of RO has not been implemented and that in-transit inventory has been excluded. The second inequality is due to the fact that $B_{0i} + (S_i^{r0} - D_i)^- \geq (S_i^{r0} - D_i - B_{0i})^-$. The last inequality is due to $(S_i^{r0} - D_i - B_{0i}(s_0))^- \leq (S_i^{r0} - D_i)^- + B_{0i}(s_0)$ and $(S_i^{r0} - D_i - B_{0i}(s_0))^+ \leq (S_i^{r0} - D_i)^+$.

When L_0 goes to ∞, $\min_{s_0} h_0 E[(s_0 - D_0)^+] + bE[(s_0 - D_0)^-]$ goes to ∞, while $\sum_{i>0} h_0 E[(S_i^{r0} - D_i)^+] + H_i E[(S_i^{r0} - D_i)^+] + bE[(S_i^{r0} - D_i)^-]$ stays constant. Thus, we have part 3.

Proof of Proposition 4

The result follows directly from the construction of the RO heuristic, i.e., the base-stock levels are calculated recursively from the leaf nodes to the root node.

Proof of Theorem 5

Part 1:

For sufficiently small h_0, we have $s_i^a(h_0) = S_i^{r0}(h_0)$ for $i > 0$ and $s_{0_i}^d(h_0) = S_{0_i}^{SS}(h_0) - s_i^a(h_0)$. From the previous part of the proof, we know that

$$\begin{aligned} & C_0(s_0^a(h_0)|s_1^a(h_0), s_2^a(h_0), ..., s_N^a(h_0)) \\ \leq\ & h_0 \sum_{i>0} s_i^a(h_0) + \sum_{i>0} \eta_i(S_i^u|h_0) \\ & + h_0 E\left[s_0^a(h_0) - \sum_{i>0} s_i^a(h_0) - D_0\right] + \sum_{i>0}(h_i + b_i)E\left[\left(s_0^a(h_0) - \sum_{i>0} s_i^a(h_0) - D_0\right)^-\right] \end{aligned}$$

To show the asymptotic result for DA, we only need to show that

$$h_0 E\left[s_0^a(h_0) - \sum_{i>0} s_i^a(h_0) - D_0\right] + \sum_{i>0}(h_i + b_i)E\left[\left(s_0^a(h_0) - \sum_{i>0} s_i^a(h_0) - D_0\right)^-\right]$$

goes to 0 when h_0 goes to zero.

To show that $\lim_{h_0 \to 0}(h_i + b_i)E\left[\left(s_0^a(h_0) - \sum_{i>0} s_i^a(h_0) - D_0\right)^-\right] = 0$ for each i, we only need to show that $\lim_{h_0 \to 0} s_0^a(h_0) = \infty$. From (6), we can see that $\lim_{h_0 \to 0} S_{0_i}^{SS}(h_0) = \infty$ and $\lim_{h_0 \to 0} s_i^a(h_0)$ approaches a finite number since $h_i > h_0$. Therefore, $\lim_{h_0 \to 0} Q_{D_{0_i}}(s_{0_i}^d(h_0)) = 0$. As a result, from (8), we can see that $\lim_{h_0 \to 0} s_0^a(h_0)$ goes to ∞.

Since both $s_i^a(h_0)$ and $E[D_0]$ have finite upper bounds, we only need to show that $\lim_{h_0 \to 0} h_0 s_0^a(h_0) = 0$. From Theorem 7, we know that $s_0^a(h_0) \leq \sum_i s_{0_i}^d(h_0)$. For each i, we have $s_{0_i}^d(h_0) \leq S_{0_i}^{SS}(h_0) \leq$

$G^{-1}_{\tilde{D}_{0_i}}\left(\frac{h_0}{b_i+h_0}\right)$ where the second inequality follows (6). Assume that the CDF of $\tilde{D}_{0_i}$ is the $G(\cdot)$ function defined in Section 3.2. Then $G^{-1}_{\tilde{D}_{0_i}}\left(\frac{b_i}{b_i+h_0}\right)$ is the optimal solution to the problem

$$\min_y E\left[h_0(y - \tilde{D}_{0_i}) + (h_0 + b_i)(y - \tilde{D}_{0_i})^-\right].$$

By the distribution-free upper bound developed by Gallego & Moon (1993), we know that

$$E\left[h_0\left(G^{-1}_{\tilde{D}_{0_i}}\left(\frac{h_0}{b_i+h_0}\right) - \tilde{D}_{0_i}\right) + (h_0+b_i)\left(G^{-1}_{\tilde{D}_{0_i}}\left(\frac{h_0}{b_i+h_0}\right) - \tilde{D}_{0_i}\right)^-\right] \leq \sqrt{h_0 b_i E[\tilde{D}_{0_i}]}.$$

Therefore, $\lim_{h_0\to 0} h_0 G^{-1}_{\tilde{D}_{0_i}}\left(\frac{h_0}{b_i+h_0}\right) = 0$. As a result, $\lim_{h_0\to 0} h_0 s_0^a(h_0) = 0$.

Proof of Proposition 6

For $j \notin T(i)$, either there is no direct path between node i and node j (case 1), or node i is in $T(j) \setminus \{j\}$ (case 2). In case 1, changes in the parameters of node j do not affect $S^{SS}_{i_w}$ for any serial system w. In case 2, node j is not a leaf. Based on the form of (6), changing the leadtime of node j does not affect $S^{SS}_{i_w}$, where node i is downstream from node j in serial system w. In both cases, it can be also shown that $S^{SS}_{k_w}$ remains the same, where k_w is the successor of i_w in serial system w. Therefore, $s^d_{i_w}$ is not affected for each i_w. Finally, the backorder matching procedure at node i is only affected by its own leadtime and the demand rate of each leaf in $T(i)$.

Proof of Proposition 7

The backorder matching procedure ensures that s_i^a is an optimal solution to the following problem.

$$\begin{aligned} s_i^a &= \operatorname*{argmax}_{z\in\mathbb{Z}} E[(D_i - z)^+], \\ & \quad s.t.\ E[(D_i - z)^+] \leq \sum_{w\in\mathcal{W}:i\in w} Q_{D_{i_w}}(s^d_{i_w}) \end{aligned} \tag{15}$$

It is clear that $z = \lceil \sum_{w\in\mathcal{W}:i\in w} s^d_{i_w} \rceil$ is a feasible solution. Since $E[(D_i - z)^+]$ is a decreasing function in z, $s_i^a \leq \lceil \sum_{w\in\mathcal{W}:i\in w} s^d_{i_w} \rceil$.

Proof of Proposition 8

Part 1: From the proof of Proposition 2, one can see that $\hat{\Pi}_0(x)$, equal to $\hat{C}_0(x)$, is a convex function. Since convexity is preserved under summation and expectation, $\Pi_0(r, q)$ is convex in r.

Part 2: Let $f(x) = E[\hat{\Pi}_0(x + \sum_{i=1}^{N} S_i^{r0} - D_0)]$. It is clear that $f(x)$ is a convex function. The overall cost function is $\Pi_0(r,q) = \frac{\xi \sum_{i=1}^{N} \lambda_i + \sum_{s_0=r+1}^{r+q} f(s_0)}{q}$. By the first order difference, $r^*(q)$ needs to satisfy the following equations

$$f(r^*(q)+1) \quad \leq \quad f(r^*(q)+q+1) \tag{16}$$

$$f(r^*(q)) \quad > \quad f(r^*(q)+q) \tag{17}$$

(17) is strict since $r^*(q)$ is the smallest optimal solution to $\min_r \Pi_0(r,q)$. Now, suppose that $r^*(q+1) \geq r^*(q)+1$. By (17), we must have $f(r^*(q+1)) > f(r^*(q+1)+q+1)$. However,

$$f(r^*(q+1)+q+1) - f(r^*(q+1)) \geq f(r^*(q)+q+2) - f(r^*(q)+1) \geq 0$$

The first inequality is due to the convexity of $f(x)$ and $r^*(q+1) \geq r^*(q)+1$. Due to the convexity of $f(x)$ and equation (16), we have $f(r^*(q)+1) \leq f(r^*(q)+q+1) \leq f(r^*(q)+q+2)$, which leads to the second inequality. However, this causes a contradiction. Therefore the proposition holds.