

PRODUCTION AND OPERATIONS MANAGEMENT

Vol. 26, No. 9, September 2017, pp. 1760–1777
 ISSN 1059-1478 | EISSN 1937-5956 | 17 | 2609 | 1760

doi 10.1111/poms.12717
 © 2017 Production and Operations Management Society

Heuristics for Base-Stock Levels in Multi-Echelon Distribution Networks

Ying Rong

Antai College of Economics and Management, Shanghai Jiao Tong University, 1954 Huashan Road, Shanghai 200030, China,
 yrong@sjtu.edu.cn

Zümbül Atan

Department of Industrial Engineering and Innovation Sciences, Eindhoven University of Technology, P.O. Box 513, Pav.E8 NL-5600,
 MB Eindhoven, The Netherlands, z.atan@tue.nl

Lawrence V. Snyder

Department of Industrial and Systems Engineering, Lehigh University Harold S. Mohler Laboratory, 200 West Packer Avenue, Bethlehem,
 Pennsylvania 18015-1582, USA, larry.snyder@lehigh.edu

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 and 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. We show that both RO and DA are asymptotically optimal along multiple dimensions for two-echelon distribution systems. Finally, we show that, with minor changes, both RO and DA are applicable to the balanced allocation policy.

Key words: distribution network; multi-echelon supply chain; inventory; heuristics; asymptotic analysis

History: Received: July 2014; Accepted: March 2017 by Qi Feng, after 2 revisions.

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, that is, 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* (Axsäter 1990, Graves 1985), 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 contribute to the literature on multi-echelon distribution systems by introducing 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, that is, 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, that is,

the leaf locations. Similar to the nested decomposition method for solving serial systems (Chen and Zheng 1994, Clark and Scarf 1960, Gallego and 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. The analysis of these systems was initiated by Clark and 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 and Zipkin (1984a), Chen and Zheng (1994), and Huh and Janakiraman (2008) refined the findings of Clark and 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 and Janakiraman 2010), capacity constraints (Huh et al. 2010) and supply disruptions (DeCroix 2013). However, the solution procedures are typically cumbersome. Thus, studies by Shang and Song (2003) and Gallego and Özer (2005) propose newsvendor heuristics to provide closed-form expressions for the echelon base-stock levels in serial systems. In the decomposition step of DA, we use the heuristic introduced by Shang and Song (2003) to obtain near-optimal local base-stock levels for all locations in each serial system.

The second step of the DA heuristic is *aggregation*. 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 and Xiong (2008) propose a newsvendor heuristic that sets the warehouse's 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 and Xiong (2008) point out that their newsvendor heuristic produces high percentage errors compared with 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.

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 and 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. We call this method the projection method assuming unimodality (PMU). 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.

We contribute to the OWMR literature by introducing two new heuristics, RO and DA. 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 lead-time goes to infinity. For OWMR systems, we prove that DA is asymptotically optimal as the unit holding cost at the warehouse goes to zero. In addition, for general distribution networks we prove that DA is asymptotically optimal as the unit holding costs at non-leaf nodes go to zero. Our comprehensive numerical experiments indicate that both of our heuristics provide better solutions, on average, than the state-of-art 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.

Although the FCFS allocation policy is widely used in the literature and in practice for its simplicity and fairness, it does not fully take advantage of all the information in the system. When a location follows a policy known as “balanced allocation,” it considers the status of the system when allocating an item to one of its successors. In this paper, we also study the balanced allocation rule for OWMR systems. Under the balanced allocation rule, the warehouse tries to assign each item to the retailer that brings the largest cost reduction at the moment of allocation. Our numerical results indicate that shifting from FCFS to balanced allocation does not necessarily lead to lower cost.

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 for OWMR systems. We present the extensions of our heuristics to general distribution systems in section 4.1. The extension of the heuristics to OWMR systems with a balanced allocation rule is discussed in section 4.2. We demonstrate the performance of our heuristics through numerical analysis 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. We consider distribution networks with the FCFS allocation policy. In section 4.2, we extend our heuristics to the balanced allocation policy and point out a few changes in the notation.

Let $T = (V, E)$ represent a directed tree, that is, 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 $T(i) := \{j \in V : \exists \text{ a directed path from } i \text{ to } j\}$ 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 orders from successors of non-leaf location i constitute the demand for node i . The transportation leadtimes between all the locations are deterministic. L_i represents the leadtime between $\mathcal{P}(i)$ and i , and D_i represents the leadtime demand at location i . For each location i , D_i is a Poisson random variable with rate $\lambda_i L_i$. 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, holding costs are charged on items in transit. There is no fixed ordering cost at any location. We study local control, which means that each location observes its own inventory level and places orders with its predecessor.

2.1. Local and Echelon Base-Stock Policies

Each location employs a continuous-review base-stock policy to replenish its inventories. We define s_i as the local base-stock level at location i and

$\mathbf{s} := (s_i)_{i=0}^N$ as the vector of local base-stock levels. 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. Accordingly, $B_{\mathcal{P}(i)}(\mathbf{s})$ is the backorder level at $\mathcal{P}(i)$. We define $B_{\mathcal{P}(i),i}(\mathbf{s})$ as the backorders at $\mathcal{P}(i)$ generated by orders from location i . Given that $\mathcal{S}(\mathcal{P}(i))$ is the set of all successors of $\mathcal{P}(i)$, we have $B_{\mathcal{P}(i)}(\mathbf{s}) = \sum_{j \in \mathcal{S}(\mathcal{P}(i))} B_{\mathcal{P}(i),i}(\mathbf{s})$. The orders from all locations in $\mathcal{S}(\mathcal{P}(i))$ are independent Poisson processes and $\mathcal{P}(i)$ fills these orders according to the FCFS policy. Consequently, given a fixed $B_{\mathcal{P}(i)}(\mathbf{s})$, $B_{\mathcal{P}(i),i}(\mathbf{s})$ follows a binomial distribution with parameters $B_{\mathcal{P}(i)}(\mathbf{s})$ (total number of trials) and $\theta_i = \lambda_i / \lambda_{\mathcal{P}(i)}$ (probability of success). Let $\text{Bin}(K, p)$ denote a binomial random variable with parameters K (total number of trials) and p (probability of success). Then, we have $B_{\mathcal{P}(i),i}(\mathbf{s}) \stackrel{\text{Dist}}{=} \text{Bin}(B_{\mathcal{P}(i)}(\mathbf{s}), \theta_i)$. Here, $\stackrel{\text{Dist}}{=}$ denotes that two random variables have the same probability distribution. Note that $B_{\mathcal{P}(i),i}(\mathbf{s})$ does not depend on s_i or the base-stock levels of the other locations in $\mathcal{S}(\mathcal{P}(i))$. 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:

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

where $B_{\mathcal{P}(0),0} \equiv 0$. We refer the interested readers to Zipkin (2000) for the derivation of Equation (1). We denote IT_i as the inventory in-transit to location i . Under the setting analyzed, we have $E[IT_i] = E[D_i] = \lambda_i L_i$ 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:

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

We include the in-transit costs in Equation (2) because Equation (2) is equivalent to the cost evaluation under the echelon base-stock levels. We omit the in-transit costs when we perform numerical studies in section 5.

Next, we define echelon-based cost functions. We define S_i to be the echelon base-stock level at location i . We let $\mathbf{S} := (S_i)_{i=0}^N$ be the vector of echelon base-stock levels. 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 $\ell(\mathbf{S}) := (\ell_0(\mathbf{S}), \ell_1(\mathbf{S}), \dots, \ell_N(\mathbf{S}))$. 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 $C_{\mathcal{S}(i)}(x|\mathbf{S}) = (b_i + h_i)[x]^-$. Next, we define the following recursive functions:

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

Although the echelon base-stock policy represents merely a change of variables from the local base-stock policy, it facilitates a *recursive* evaluation procedure described in Equation (2). With the following proposition, we prove that the average cost of the echelon representing the entire distribution network is equal to the average cost in Equation (2).

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

All proofs are available in Appendix A.

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 $h_{\mathcal{P}(i)}$.

3. Heuristics

As mentioned above, Graves (1985) and Axäter (1990) develop the projection method to obtain the optimal base-stock levels for an OWMR system. This method can be extended to obtain the optimal base-stock levels for distribution systems with more than two echelons. However, its implementation is computationally time consuming since enumeration is required to find the optimal base-stock levels of all the non-leaf locations. Therefore, heuristic methods are required. In this section, we introduce two heuristics: the RO and the DA heuristics. The motivation for both heuristics comes from the recursive algorithm by Clark and Scarf (1960) used for evaluation and optimization of base-stock levels in serial systems. This algorithm finds the optimal base-stock levels by starting from the location closest to customers and solving a set of recursive equations. Inspired by this “bottom-up approach,” we first propose the RO heuristic. As a result, RO sequentially solves a

set of single-variable, convex minimization problems. Secondly, in the DA heuristic, we decompose the entire distribution network into serial systems. The recursive idea is embedded in obtaining base-stock levels in each serial system. Then, the serial systems are aggregated back into the distribution system by “matching” the backorders.

3.1. The Recursive Optimization Heuristic

We define S_i^{r0} as the *intermediate* echelon base-stock level of location $i \in \{0, 1, \dots, N\}$ 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 S), we develop the following RO heuristic to determine the values of s_i^r .

Recursive Optimization Heuristic

Step 1: Compute S_i^{r0} for the retailers, that is, $\forall i \in \{1, 2, \dots, N\}$, as follows:

$$\begin{aligned}\hat{C}_i(x) &= H_i x + (h_i + b_i)[x]^- \\ C_i(y) &= E[\hat{C}_i(y - D_i)] \\ S_i^{r0} &= \underset{y}{\operatorname{argmin}} \quad C_i(y) \\ \underline{C}_i(v) &= C_i(S_i^{r0} - v)\end{aligned}\tag{4}$$

Step 2: Compute S_0^{r0} for the warehouse as follows:

$$\begin{aligned}\hat{C}_0(x) &= H_0 x + \sum_{i=1}^N \underline{C}_i \left(\text{Bin} \left(\left[x - \sum_{i=1}^N S_i^{r0} \right]^- , \theta_i \right) \right) \\ C_0(y) &= E[\hat{C}_0(y - D_0)] \\ S_0^{r0} &= \underset{y}{\operatorname{argmin}} \quad C_0(y)\end{aligned}$$

Step 3: Calculate $\ell(\mathbf{S}^{r0})$. Using Equation (1), we define $I_i(s_i, \ell_0(\mathbf{S}^{r0}))$ and $B_i(s_i, \ell_0(\mathbf{S}^{r0}))$ to be the on-hand inventory level and backorder level at retailer i when its own local base-stock level is s_i and the local base-stock level of the warehouse is $\ell_0(\mathbf{S}^{r0})$. Set $s_0^r = \ell_0(\mathbf{S}^{r0})$ and calculate $s_i^r \forall i \in \{1, 2, \dots, N\}$ as follows:

$$s_i^r = \underset{s_i}{\operatorname{argmin}} \quad E[h_i I_i(s_i, \ell_0(\mathbf{S}^{r0})) + b_i B_i(s_i, \ell_0(\mathbf{S}^{r0}))]\tag{5}$$

In Step 1, the retailers' echelon base-stock levels, assuming independency from all other locations in the system, are obtained. In Step 2, the warehouse echelon base-stock level is obtained. Then, in Step 3, the retailers' base-stock levels are updated by

optimizing their costs with the warehouse base-stock level fixed at the value obtained in Step 2. With the following proposition, we prove the convexity of $C_i(\cdot)$ and $\underline{C}_i(\cdot)$ for all $i \in \{0, 1, \dots, N\}$. This property is the main reason for the efficiency of the RO heuristic.

PROPOSITION 2. *For OWMR systems, $C_i(\cdot)$ and $\underline{C}_i(\cdot)$ are convex $\forall i \in \{0, 1, \dots, N\}$ and $\forall i \in \{1, 2, \dots, N\}$, respectively.*

Note that the convexity of $C_i(y)$ is well known in the literature for the retailers. Proposition 2 shows that such convexity is maintained for the warehouse when the retailer base-stock levels are set using the first step of RO. (In contrast, the warehouse cost function is not convex in the projection method.) This property also holds for general distribution systems (refer to Proposition 3).

Unlike the method by Clark and Scarf (1960) for serial systems, the RO heuristic is not guaranteed to find the optimal solution for OWMR systems since it ignores the fact that the base-stock level of one retailer might be affected by the base-stock level of the other retailers. 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 1. *Consider an OWMR system with N retailers. 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 . Then we have

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

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

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

3. Let $c^r(L_0)$ and $c^*(L_0)$ denote the cost of RO and the optimal cost (respectively) 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

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

3.2. The Decomposition-Aggregation Heuristic

Under the DA heuristic, we decompose the OWMR system into N two-location serial systems, solve the base-stock levels of the locations in each serial system and aggregate the solutions utilizing a procedure we call “backorder matching.” We use s_i^d to denote the local base-stock level at location i based on the DA heuristic. Next, we provide a detailed description of the steps of the heuristic.

Decomposition-Aggregation Heuristic

Step 1: Decompose the system into N serial systems. Serial system i consists of the warehouse and retailer i . We use 0_i to refer to the warehouse in serial system i . Utilizing the procedure in Shang and Song (2003), we approximate the echelon base-stock levels of retailer i , S_i^{SS} , and the warehouse in serial system i , $S_{0_i}^{SS}$, as follows:

$$\begin{aligned} S_i^{SS} &= F_{D_i}^{-1}\left(\frac{b_i + h_0}{b_i + h_i}\right) \\ S_{0_i}^{SS} &= \frac{1}{2} \left[G_{\tilde{D}_i}^{-1}\left(\frac{b_i}{b_i + h_i}\right) + G_{\tilde{D}_i}^{-1}\left(\frac{b_i}{b_i + h_0}\right) \right] \end{aligned} \quad (6)$$

Here, \tilde{D}_i is the total leadtime demand in serial system i . \tilde{D}_i is a Poisson random variable with rate $\lambda_i(L_0 + L_i)$. F^{-1} is the inverse Poisson cumulative distribution function (CDF). For the warehouse, instead of using F^{-1} , we opt instead to use G_D^{-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:

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

Step 2: Calculate the local base-stock level for the warehouse in serial system i by $s_{0_i}^d = S_{0_i}^{SS} - S_i^{SS}$. For the retailers, we have $s_i^d = S_i^{SS}$, $\forall i \in \{1, 2, \dots, N\}$. We approximate the expected backorders of the warehouse in serial system i by

$$E[B_{0_i}] = E[(D_{0_i} - s_{0_i}^d)^+] = Q_{D_{0_i}}(s_{0_i}^d), \quad (7)$$

where D_{0_i} is a Poisson random variable with rate $\lambda_i L_0$, and $Q_D(x)$ is the loss function of the Poisson random variable D .

Step 3: Aggregate the serial systems back into the OWMR system utilizing a “backorder matching” procedure. We approximate the total expected

backorders at the warehouse by $E[B_0] \cong \sum_{i=1}^N E[B_{0_i}]$. Specifically, the backorder matching procedure sets s_0^a , the base-stock level at the warehouse, equal to the smallest integer s_0 such that

$$E[(D_0 - s_0)^+] \leq \sum_{i=1}^N E[B_{0_i}].$$

In other words,

$$s_0^a = Q_{D_0}^{-1}\left(\sum_{i=1}^N Q_{D_{0_i}}(s_{0_i}^d)\right) \quad (8)$$

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

Similar to Theorem 1 for RO, for OWMR systems, we now prove the asymptotic optimality of DA as h_0 goes to 0.

THEOREM 2. Suppose that $h_i > h_0$ for all $i > 0$. Let $c^a(h_0)$ and $c^*(h_0)$ denote the cost of DA and the optimal cost (respectively) for given h_0 . Then we have

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

4. Extensions

In section 3, we introduced the RO and DA heuristics for OWMR systems. Next, we consider two extensions. In section 4.1 we discuss the extension of the heuristics to general distribution systems, and in section 4.2 we consider the balanced allocation rule for OWMR systems.

4.1. Extension of the Heuristics for General Systems

Similar to section 3.1, 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.

Recursive Optimization Heuristic

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

$$\begin{aligned} \hat{C}_i(x) &= H_i x + \sum_{j \in \mathcal{S}(i)} \underline{C}_j \left(\text{Bin} \left(\left[x - \sum_{j \in \mathcal{S}(i)} S_j^{r0} \right]^-, \theta_j \right) \right) \\ C_i(y) &= E[\hat{C}_i(y - D_i)] \\ S_i^{r0} &= \underset{y}{\operatorname{argmin}} C_i(y) \\ \underline{C}_i(v) &= C_i(S_i^{r0} - v) \end{aligned} \quad (9)$$

Step 2: Calculate $\ell(\mathbf{S}^{r0})$. Define $\ell_{-i}(\mathbf{S})$ to be the vector $\ell(\mathbf{S})$ excluding the component $\ell_i(\mathbf{S})$. 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 $\ell_j(\mathbf{S}^{\text{r0}})$ for $j \neq i$. Calculate \mathbf{s}^r as follows:

$$s_i^r = \begin{cases} \underset{s_i}{\operatorname{argmin}} E[h_i I_i(s_i, \ell_{-i}(\mathbf{S}^{\text{r0}})) \\ + b_i B_i(s_i, \ell_{-i}(\mathbf{S}^{\text{r0}}))] & \forall i \in \mathcal{L} \\ \ell_i(\mathbf{S}^{\text{r0}}) & \forall i \notin \mathcal{L} \end{cases} \quad (10)$$

Note that, in Step 2, we update the base-stock levels at leaves *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 and Song 2006), and therefore the expected cost. Our numerical results suggest that by updating the base-stock levels at all nodes, we can improve the performance of RO slightly while the computational time increases significantly. Therefore, we suggest updating only at the leaves. With the following proposition, we prove that the convexity result for OWMR systems (refer to Proposition 2) holds for general distribution systems, too.

PROPOSITION 3. *For all i , $C_i(\cdot)$ and $\underline{C}_i(\cdot)$ are convex.*

Next, we provide the following 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 before, 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 1 for OWMR systems and the numerical studies for general distribution networks in section 5.1 indicate that the optimality loss due to this decoupling is limited.

Next, we extend the DA heuristic to general distribution systems. The main idea remains the same: decomposition of the system into serial systems, determination of base-stock levels in all the locations in all the serial systems, and aggregation through backorder matching.

Decomposition-Aggregation Heuristic

Step 1: Decompose the system into $|\mathcal{L}|$ serial systems. Given that k is the leaf location of a serial system w , serial system w consists of locations $\{0, \dots, \mathcal{P}(\mathcal{P}(k)), \mathcal{P}(k), k\}$. The demand rate at each location of this system is $\lambda_w := \lambda_k$. Define \mathcal{W} as the set

of serial systems. Utilizing the procedure in Shang and Song (2003), we approximate the echelon base-stock level of location i in serial system w as follows:

$$S_{i_w}^{\text{SS}} = \begin{cases} F_{\tilde{D}_{i_w}}^{-1} \left(\frac{b_i + \sum_{j \in \mathcal{A}(0, \mathcal{P}(i))} H_j}{b_i + \sum_{j \in \mathcal{A}(0, i)} H_j} \right), & i \in \mathcal{L} \\ \frac{1}{2} \left[G_{\tilde{D}_{i_w}}^{-1} \left(\frac{b_{\ell_w} + \sum_{j \in \mathcal{A}(0, \mathcal{P}(i))} H_j}{b_{\ell_w} + \sum_{j \in w} H_j} \right) \right. \\ \left. + G_{\tilde{D}_{i_w}}^{-1} \left(\frac{b_{\ell_w} + \sum_{j \in \mathcal{A}(0, \mathcal{P}(i))} H_j}{b_{\ell_w} + \sum_{j \in \mathcal{A}(0, i)} H_j} \right) \right], & i \notin \mathcal{L} \end{cases} \quad (11)$$

Here, $\mathcal{A}(i, j)$ is the set of locations in the directed path from i to j . \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 $\lambda_w \sum_{j \in \mathcal{A}(i, \ell_w)} L_j$.

Step 2: Calculate the local base-stock level for location i ($i \notin \mathcal{L}$) by $s_{i_w}^d = S_{i_w}^{\text{SS}} - S_{j_w}^{\text{SS}}$, where j_w is the successor of i_w in system w . For $i \in \mathcal{L}$, we have $s_{i_w}^d = S_{i_w}^{\text{SS}}$ and approximate the expected backorders at each location i_w in serial system w assuming that i_w has a supplier with infinite supply by

$$E[B_{i_w}] = E[(D_{i_w} - s_{i_w}^d)^+] = Q_{D_{i_w}}(s_{i_w}^d), \quad (12)$$

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 .

Step 3: Approximate the total expected backorders at location i in the distribution network by $E[B_i] \cong \sum_{w \in \mathcal{W}: i \in w} E[B_{i_w}]$. Set s_i^a , the base-stock level at location i , to the smallest integer s_i such that

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

In other words,

$$s_i^a = Q_{D_i}^{-1} \left(\sum_{w \in \mathcal{W}: i \in w} Q_{D_{i_w}}(s_{i_w}^d) \right) \quad (13)$$

The effectiveness of DA comes from the backorder matching procedure. Backorder matching is motivated by the recognition that backorders are the key drivers of the cost and, therefore, performance of the system. In particular, as is evident from Equation (1), only the backorders of location i affect the system state of its successors. Hence, it is logical to use backorders to derive the aggregated base-stock levels, rather than simply summing the serial-system base-stock levels as is proposed by Özer and Xiong (2008). With the next proposition we prove that the

aggregated local base-stock level under base-stock level matching is greater than that under backorder matching.

PROPOSITION 5. $s_i^a \leq \lceil \sum_{w \in \mathcal{W}: i \in w} s_{i_w}^d \rceil$

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 with 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 back-order matching.

We used the data set in section 5, consisting of 480 instances, to test the effectiveness of base-stock level matching. The average percentage error is 14.29, compared with 2.49 for backorder matching. Clearly, base-stock level matching performs significantly worse than backorder matching. Similarly, for OWMR systems with fill rate constraints, Özer and 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.

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 $j \notin T(i)$.

With the following theorem, we show that DA is asymptotically optimal when the unit holding costs at non-leaf nodes go to zero.

THEOREM 3. Let $\hat{h} = (h_i)_{i \notin \mathcal{L}}$, that is, the vector of unit holding costs for non-leaf nodes. Let $c^a(\hat{h})$ and $c^*(\hat{h})$ denote the cost of DA and the optimal cost of the distribution system (respectively) for given \hat{h} . Then we have

$$\lim_{\hat{h} \rightarrow 0} \frac{c^a(\hat{h})}{c^*(\hat{h})} = 1.$$

We note that the analysis of general distribution systems is more difficult than the analysis of OWMR systems. Although we provide multiple results on the asymptotic performance of the heuristics for OWMR systems, these do not extend easily to general distribution systems, which are inherently more difficult

both to analyze and to optimize. This also explains why the literature on general distribution systems is relatively sparse.

4.2. Balanced Allocation Policy

One of the shortcomings of the FCFS allocation policy is that it does not utilize all available information regarding the status of the system. Consider an OWMR system with two identical retailers and a local base-stock level of zero at the warehouse. Suppose that after retailer 1 places an order for a unit, retailer 2 incurs the next 100 units of demand before the next shipment arrives at the warehouse. Under the FCFS allocation policy, this unit is shipped to retailer 1 even though it will be more valuable to retailer 2, since it would clear a backorder for retailer 2 but may simply sit in inventory at retailer 1.

The optimal allocation policy for distribution networks is unknown. Zipkin (2000, section 8.6) suggests a balanced allocation policy for OWMR systems. Under the balanced allocation policy, the warehouse considers the status of the system when allocating an item to one of the retailers. The warehouse assigns each item to the retailer that brings the largest cost reduction at the moment of allocation. This requires determining the echelon base-stock level at the warehouse (S_0) and the aggregate base-stock level for all retailers (S_r). Let $ITP_i(t)$ denote the inventory-transit position at retailer i at time t . By definition, the inventory-transit position is the sum of the net-inventory (on-hand inventory minus backorders) and in-transit inventories. Suppose that the warehouse receives an item at time t . Let $ITP_i(t^-)$ be the inventory-transit position at retailer i after the arrival of this item to the warehouse but before its allocation to a retailer. Under the balanced allocation policy, if $\sum_{i>0} ITP_i(t^-) < S_r$, the warehouse solves the following problem to determine which retailer to ship to:

$$\begin{aligned} & \min \sum_{i>0} C_i(ITP_i(t)) \\ \text{s.t. } & \sum_{i>0} ITP_i(t) = 1 + \sum_{i>0} ITP_i(t^-) \\ & ITP_i(t) \geq ITP_i(t^-) \end{aligned}$$

where $C_i(y) = H_i E[(y - D_i)] + (b_i + h_i) E[(y - D_i)^-]$. When the retailers are identical, in the optimal solution, the difference between the largest and smallest inventory-transit positions among the retailers is always within one unit, that is, the retailers' inventory-transit positions are balanced.

The PMU, RD, RO, and DA heuristics can be modified to handle the balanced allocation policy. Suppose we have used heuristic k ($k \in \{\text{PMU}, \text{RD}, \text{RO}, \text{DA}\}$) to find approximate base-stock levels, denoted S_i^k , under

the FCFS policy. To convert this solution to a solution that is appropriate under the balanced allocation policy, we let $S_0^{bk} = S_0^k$ and $S_r^{bk} = \sum_{i>0} S_i^k$. Here, S_0^k and S_r^{bk} are the echelon base-stock levels of the warehouse and the aggregate retailer base-stock level under balanced allocation for heuristic k . We provide numerical evidence on the performance of these heuristics under balanced allocation in section 5.2.

5. Numerical Analysis

In section 5.1, we consider multiple network settings and report the results of our numerical analysis on the performance of our heuristics. In section 5.2, we study the impact of the balanced allocation rule.

5.1. Performance of the Heuristics under the FCFS Allocation Policy

In this section, we test the performance of RO and DA by comparing them with two other heuristics: the 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 and 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 PMU. The PMU is a heuristic rather than an exact algorithm; however, it nearly always finds the optimal solution (Zipkin 2000, p. 334).

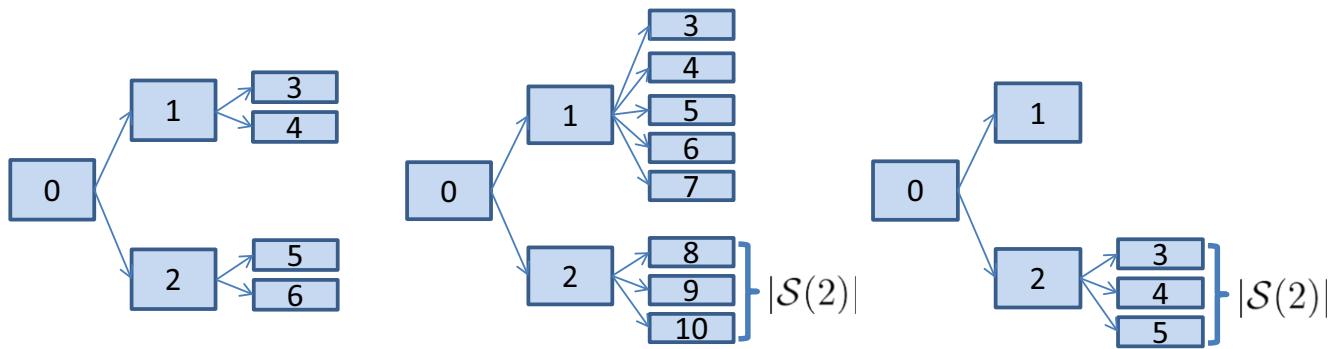
Given that c is the expected cost of the solution suggested by one of the methods and c^* is the expected cost of the solution obtained using PMU, we calculate the percentage cost error (or gap) by $\epsilon = 100 \frac{c - c^*}{c^*}$. Note that we exclude the in-transit inventory costs in our numerical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

In Table 1, 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 1a. 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

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

No. of echelons	2			3			4			5		
	RD	RO	DA	RD	RO	DA	RD	RO	DA	RD	RO	DA
Concave local holding cost coefficients												
$\bar{\epsilon}$	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 $_{\epsilon}$	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 $_{\epsilon}$	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												
$\bar{\epsilon}$	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 $_{\epsilon}$	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 $_{\epsilon}$	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												
$\bar{\epsilon}$	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 $_{\epsilon}$	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 $_{\epsilon}$	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		

Figure 1 Three-Echelon Distribution Networks [Color figure can be viewed at wileyonlinelibrary.com]



evenly distributed 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 1a 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.

Table 1 shows that the performance of RO is better than that of DA in all settings. This demonstrates the benefit of optimizing inventory levels for the distribution network as a whole. DA considers the connections only among the locations in the same echelon through backorder matching. On the other hand, backorder matching still ensures the good performance of DA. In addition, the calculations in the DA heuristic only involve newsvendor solutions and evaluations of Poisson inverse cumulative functions, which are computationally cheap and user-friendly to practitioners.

Table 1 also 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 deteriorates 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 Theorems 1 and 2 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 1, 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 with RD, DA runs much faster as the number of echelons increases.

In Table 2, we set $\lambda_i \sim U[1, 10]$ for $i \in \mathcal{L}$ and $H_i \sim U[0.1, 0.5]$ for all i . The other parameters are the same as those used in generating the results in Table 1. We generate 20 instances for each network topology. The results in Table 2 indicate that the trends in the performances of RO, DA, and RD remain the same as in Table 1.

Next, we investigate the impact of the network asymmetry on the performance of heuristics. In Table 3, the system with asymmetric leaves refers to a three-echelon distribution system with root node 0 and with nodes 1 and 2 in the second echelon. There are eight leaves. We vary the number of leaves attached to nodes 1 and 2. The system with

asymmetric echelons refers to a three-echelon distribution system with root node 0 and with nodes 1 and 2 in the second echelon. There are no leaves attached to node 1. We vary the number of leaves attached to node 2. Figure 1b and c demonstrate these two types of networks. Aside from the difference in the network structure, all other parameters in Table 3 are the same as those in Table 2. We generate 20 instances for each network topology.

In Table 3, we can see that DA performs better when leaves are more balanced under the system with asymmetric leaves. Other than that, the network structure does not have a systematic impact on the performance of RO or RD.

The results in this section are based on 480 different parameter combinations. We observe that the base-stock levels calculated through the RO heuristic are no higher than the base-stock levels suggested by the DA heuristic. In addition, PMU provides no higher base-stock levels at leaves than RO in 470 of the parameter combinations, while DA provides no higher base-stock levels at leaves than RD in 450 of the parameter combinations. Based on the performance of PMU, RO, DA,

Table 2 Performance of the Heuristics for Multi-Echelon Distribution Networks Under Random Demand and Random Holding Cost

No. of echelons	2			3			4			5		
	RD	RO	DA	RD	RO	DA	RD	RO	DA	RD	RO	DA
$\bar{\epsilon}$	7.32	0.17	0.38	9.31	0.55	2.01	9.10	0.98	3.00	9.20	1.41	5.71
med $_{\epsilon}$	7.51	0.00	0.00	8.63	0.35	1.44	9.02	0.83	2.17	8.79	1.42	5.42
max $_{\epsilon}$	12.14	1.46	1.64	14.76	1.52	5.91	13.59	2.35	8.41	12.85	2.54	10.66
σ_{ϵ}	2.71	0.43	0.56	3.19	0.56	1.71	2.17	0.57	1.88	1.70	0.54	2.17
\bar{t} (seconds)	0.012	0.054	0.011	0.043	0.219	0.022	0.232	1.078	0.055	1.012	4.911	0.122
		PMU			PMU		PMU			PMU		
\bar{t} (seconds)	0.081				0.862			37.122			1287.430	

Table 3 Performance of the Heuristics for Asymmetric Network Structure

$ S(2) $	1			2			3			4		
	RD	RO	DA									
Asymmetric leaves												
$\bar{\epsilon}$	8.57	1.10	5.32	9.08	0.78	4.34	9.18	0.60	3.79	9.39	0.80	3.46
med $_{\epsilon}$	8.04	0.94	5.15	9.22	0.65	4.17	9.34	0.58	3.31	8.85	0.72	3.13
max $_{\epsilon}$	14.01	3.23	11.66	14.44	2.28	10.78	14.25	1.54	12.44	13.86	2.11	10.12
σ_{ϵ}	2.84	0.92	2.09	2.63	0.64	2.19	2.73	0.45	2.37	2.88	0.66	2.08
\bar{t} (seconds)	0.115	0.482	0.038	0.113	0.469	0.039	0.104	0.459	0.036	0.108	0.448	0.038
		PMU			PMU		PMU			PMU		
\bar{t} (seconds)	2.784			2.721				2.601			2.633	
Asymmetric echelons												
$\bar{\epsilon}$	8.74	0.67	1.70	9.38	0.32	1.25	9.75	0.83	3.76	8.59	0.59	3.90
med $_{\epsilon}$	8.12	0.21	1.60	9.32	0.00	0.60	10.03	0.52	2.86	7.98	0.60	2.73
max $_{\epsilon}$	16.75	2.57	6.15	15.53	1.84	4.46	16.14	3.28	9.69	13.10	1.69	12.16
σ_{ϵ}	3.42	0.84	1.75	3.58	0.56	1.35	3.19	0.96	3.08	2.11	0.49	2.98
\bar{t} (seconds)	0.014	0.077	0.009	0.026	0.134	0.014	0.037	0.183	0.017	0.047	0.217	0.021
		PMU			PMU		PMU			PMU		
\bar{t} (seconds)	0.158			0.464				0.757			1.061	

and RD, we can conclude that the performance of a method/heuristic increases as the base-stock levels at the leaves decrease. This is partly due to better exploitation of the risk pooling effect.

5.2. Performance of the Heuristics under the Balanced Allocation Policy

One can only evaluate the system performance under the balanced allocation rule through simulation, since no expressions are available for the expected cost under this allocation rule. In order to find “the best” base-stock levels under balanced allocation, we evaluate the cost for different combinations of S_0 and S_r . Define S_0^{FCFS} as the optimal warehouse base-stock level under the FCFS allocation rule and S_r^{FCFS} as the sum of the optimal retailer base-stock levels under the same allocation rule. In our simulation study, we vary S_0 from $S_0^{FCFS} - R_0$ to $S_0^{FCFS} + R_0$ and similarly S_r from $S_r^{FCFS} - R_r$ to $S_r^{FCFS} + R_r$, where both R_0 and R_r are positive integers. The total number of solutions enumerated is equal to $(2R_0 + 1)(2R_r + 1)$. The higher R_0 and R_r are, the longer the computational time is.

We rely on our simulation study to evaluate the average costs of the heuristics under the balanced allocation rule and we compare these costs with the best average cost obtained by the enumeration of $(2R_0 + 1)(2R_r + 1)$ different S_0 and S_r combinations. We consider systems with 2, 3, and 4 retailers and use the same parameters as the ones used in obtaining the results in Table 2. We set the simulation time to $\frac{2500}{\sum_{i>0} i^2}$, that is, we simulate 10,000 demand arrivals, on average. We set both R_0 and R_r equal to 5. The results are summarized in Table 4. In the first set of rows in Table 4, we provide the statistics for the percentage cost savings achieved by switching from FCFS to the balanced allocation policy for both PMU, RD, RO and DA. In the second set of rows in Table 4, we compare the performance of PMU, RD, RO, and DA with the best cost found through simulation under the balanced allocation policy.

The results in Table 4 suggest that the balanced allocation rule does not necessarily lead to lower cost

compared with FCFS. Note that the average savings are negative, that is, switching from FCFS to the balanced allocation does not provide cost benefits on average. The maximum values imply that the percentage savings achieved by switching from FCFS to the balanced allocation can be as high as 12%, while the minimum values imply that switching to the balanced allocation policy can make the system worse off by almost 29%. One plausible reason, as discussed by Gallego et al. (2007) and Dogru et al. (2009), is due to the differences among the retailers. Compared to the balanced allocation policy, one needs to determine the individual base-stock levels for all retailers under FCFS. The wider decision space under FCFS can sometimes offset the shortcomings of this allocation policy. The second set of columns in Table 4 suggest that re-optimization of the base-stock levels improves the performance of the balanced allocation policy although it can still make the system worse off by 15%.

6. Conclusions

In this study, 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. We prove the asymptotic optimality of both RO and DA under multiple conditions. In addition, we consider the balanced allocation rule. We conclude that both RO and DA perform well under balanced allocation. The robustness of the heuristics allows us to extend them to the case with a fixed ordering cost at the warehouse. Our preliminary results confirm the good performance of the heuristics under this extension as well. Future work can further extend the heuristics to other variants of distribution networks with

Table 4 The Performance of Heuristics under Balanced Allocation Rule

No. of retailers	2				3				4			
	PMU	RD	RO	DA	PMU	RD	RO	DA	PMU	RD	RO	DA
<i>Cost savings from using balanced allocation</i>												
Average	-2.17	-0.59	-3.20	-1.97	-2.16	-0.94	-0.98	-1.44	-1.31	0.00	-1.85	-0.29
Max	9.58	7.48	6.07	7.68	7.99	11.54	8.58	6.94	7.74	12.18	8.57	8.11
Min	-28.87	-20.50	-24.90	-28.87	-14.10	-13.09	-17.68	-13.81	-23.72	-23.72	-15.98	-22.74
<i>Cost saving from using balanced allocation and base-stock re-optimization</i>												
Average	0.17	2.92	0.53	0.88	-0.16	2.81	0.12	0.99	0.53	3.34	0.94	2.02
Max	9.58	15.90	9.91	10.65	12.35	12.77	12.35	12.63	8.56	12.18	8.85	11.35
Min	-15.12	-13.45	-15.12	-15.12	-12.47	-10.72	-12.47	-12.47	-15.17	-15.17	-13.68	-12.63
\bar{t} (seconds)	224				312				421			

different cost structures and different replenishment policies.

Acknowledgments

The authors thank Qi Feng (Department Editor), the associate editor, and the reviewers for their detailed comments and many valuable suggestions that have significantly improved the quality of the article. The research of Ying Rong was supported by the National Natural Science Foundation of China (Grant 71522009, 71202068, 71632007, 71421002). This research was also supported in part by National Science Foundation grant CMMI-0726822.

Appendix A. Proofs

PROOF OF PROPOSITION 1. We first define $C^k(\mathbf{s})$ as the cost of a subtree $T(k)$ with the local base-stock level vector \mathbf{s} in the following manner:

$$\begin{aligned} C^k(\mathbf{s}) &= \sum_{i \in T(k) \setminus \mathcal{L}} (h_i - h_{\mathcal{P}(k)}) E \left[I_i^k(\mathbf{s}) + \sum_{j \in f(i)} IT_j \right] \\ &\quad + \sum_{i \in T(k) \cap \mathcal{L}} \left[(h_i - h_{\mathcal{P}(k)}) E[I_i^k(\mathbf{s})] \right. \\ &\quad \left. + (b_i + h_k) E[B_i^k(\mathbf{s})] \right], \end{aligned}$$

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

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

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 $C^0(\ell(\mathbf{s})) = C_0(S_0|\mathbf{s})$.

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

First, define a new vector $\gamma^n(\mathbf{s}|Z)$ 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 $B_{k,n}^k(\ell(\mathbf{s})) \stackrel{\text{Dist}}{=} \text{Bin}([S_k - \sum_{n \in \mathcal{S}(k)} S_n - D_n]^-, \theta_k)$. Then, one can verify that, for all $n \in \mathcal{S}(k)$ and $i \in T(k) \setminus \{k\}$, we have

$$\begin{aligned} E[I_i^k(\ell(\mathbf{s}))] &= E[I_i^n(\gamma^n(\mathbf{s}|Z))], \\ E[B_i^k(\ell(\mathbf{s}))] &= E[B_i^n(\gamma^n(\mathbf{s}|Z))] \end{aligned} \tag{A1}$$

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

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

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 (A1). 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. \square

PROOF OF PROPOSITION 2. Proposition 2 is a special case of Proposition 3, the proof of which is given below. \square

PROOF OF THEOREM 1. 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 $\sum_{i>0} c_i(s_i^u) \leq c^*(h_0)$. Therefore, $1 \leq \frac{c'(h_0)}{c^*(h_0)} \leq \frac{c'(h_0)}{\sum_{i>0} c_i(s_i^u)}$. Using Equations (4) and (5), we know that $c'(h_0) \leq \min_y C_0(y|S_1^{r0}(h_0), S_2^{r0}(h_0), \dots, S_N^{r0}(h_0))$ 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

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

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

$$E[\eta_i(S_i^{r0}(h_0) - B_{0i}|h_0) - \eta_i(S_i^{r0}(h_0)|h_0)] \leq (h_i + b_i)E[B_{0i}] \quad (\text{A2})$$

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

Then

$$\begin{aligned} c'(h_0) &\leq \min_y C_0(y|S_1^{r0}(h_0), S_2^{r0}(h_0), \dots, S_N^{r0}(h_0)) \\ &\leq h_0 \sum_{i>0} S_i^{r0}(h_0) + \sum_{i>0} \eta_i(S_i^{r0}(h_0)|h_0) \\ &\quad + \min_y \left(h_0 E \left[y - \sum_{i>0} S_i^{r0}(h_0) - D_0 \right] \right. \\ &\quad \left. + \sum_{i>0} (h_i + b_i)E \left[\left(y - \sum_{i>0} S_i^{r0}(h_0) - D_0 \right)^- \right] \right) \\ &\leq h_0 \sum_{i>0} S_i^{r0}(h_0) + \sum_{i>0} \eta_i(S_i^{r0}(h_0)|h_0) \\ &\quad + \sqrt{h_0 \left(\sum_{i>0} (h_i + b_i) - h_0 \right) L_0 \sum_{i>0} \lambda_i} \end{aligned}$$

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

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

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

Part 2:

From RO in Equation (4), we can see that S_i^{r0} 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 $\sum_{i=1}^N \eta_i(S_i^{r0}) \leq c^*(N)$ since $\sum_{i=1}^N \eta_i(S_i^{r0})$ ignores the echelon holding cost due to the warehouse's holding cost and the additional retailer costs due to warehouse backorders. Using Equations (4) and (5), we know that $c^*(N) \leq \min_y C_0(y|S_1^{r0}(h_0), S_2^{r0}(h_0), \dots, S_N^{r0}(h_0))$. Therefore, we only need to show $\frac{\min_y C_0(y|S_1^{r0}(h_0), S_2^{r0}(h_0), \dots, S_N^{r0}(h_0))}{\sum_{i=1}^N \eta_i(S_i^{r0})}$ approaches 1 as N increases.

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

$$\begin{aligned} C_0(y|S_1^{r0}, S_2^{r0}, \dots, S_N^{r0}) &\leq h_0 \sum_{i>0} S_i^{r0} + h_0 E \left[y - \sum_{i>0} S_i^{r0} - D_0 \right] \\ &\quad + \sum_{i>0} (h_i + b_i)E \left[\left(y - \sum_{i>0} S_i^{r0} - D_0 \right)^- \right] + \sum_{i>0} \eta_i(S_i^{r0}) \end{aligned}$$

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

$$\begin{aligned} & h_0 \sum_{i>0} S_i^{r0} + h_0 E \left[y - \sum_{i>0} S_i^{r0} - D_0 \right] \\ &\quad + \sum_{i>0} (h_i + b_i)E \left[\left(y - \sum_{i>0} S_i^{r0} - D_0 \right)^- \right] \\ &\leq \kappa \sum_{i>0} S_i^{r0} + \kappa E \left[\left(y - \sum_{i>0} S_i^{r0} - D_0 \right)^+ \right] \\ &\quad + \kappa E \left[\left(y - \sum_{i>0} S_i^{r0} - D_0 \right)^- \right] \end{aligned}$$

$$\begin{aligned} &\leq \kappa E[(y - D_0)^+] + \kappa E[(y - D_0)^-] \\ &+ \kappa E\left[\left(y - 2 \sum_{i>0} S_i^{r0} - D_0\right)^+\right] \\ &+ \kappa E\left[\left(y - 2 \sum_{i>0} S_i^{r0} - D_0\right)^-\right] \end{aligned}$$

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

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

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

Therefore,

$$\begin{aligned} &\min_y C_0(y|S_1^{r0}, S_2^{r0}, \dots, S_N^{r0}) \\ &\leq \sum_{i>0} \eta_i(S_i^{r0}) + 2\kappa \sqrt{L_0 \sum_{i>0} \lambda_i + E[R]} \\ &= \sum_{i>0} \eta_i(S_i^{r0}) + 2\kappa \sqrt{L_0 \sum_{i>0} \lambda_i + \sum_{i>0} S_i^{r0}}. \end{aligned}$$

Let $\underline{\varrho} = \min_i \eta_i(S_i^{r0})$, $\bar{\lambda} = \max_i \lambda_i$ 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 $\underline{\varrho}$ is strictly positive.

Then we have

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

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

Part 3:

First we show that

$$\begin{aligned} c^*(L_0) &= \min_s h_0 E[(s_0 - D_0)^+] \\ &+ \sum_{i>0} (h_i E[(s_i - D_i - B_{0i}(s_0))^+] \\ &+ b E[(s_i - D_i - B_{0i}(s_0))^-]) \end{aligned}$$

$$\begin{aligned} &\geq \min_s h_0 E[(s_0 - D_0)^+] \\ &+ \sum_{i>0} (h_i E[(s_i - \lambda_i L_i - B_{0i}(s_0))^+] \\ &+ b E[(s_i - \lambda_i L_i - B_{0i}(s_0))^-]) \\ &\geq \min_s h_0 E[(s_0 - D_0)^+] \\ &+ \sum_{i>0} (h_0 E[(s_i - B_{0i}(s_0))^+] \\ &+ b E[(s_i - B_{0i}(s_0))^-]) \\ &\geq \min_s h_0 E[(s_0 - D_0)^+] \\ &+ h_0 E\left[\left(\sum_{i>0} s_i - (s_0 - D_0)^-\right)^+\right] \\ &+ b E\left[\left(\sum_{i>0} s_i - (s_0 - D_0)^-\right)^-\right] \end{aligned}$$

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 $\text{Var}[\lambda_i + B_{0i}(s_0)] \leq \text{Var}[D_i + B_{0i}(s_0)]$ 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 $\min_s h_0 E[(s_0 - D_0)^+] + h_0 E[(\sum_{i>0} s_i - (s_0 - D_0)^-)^+] + b E[(\sum_{i>0} s_i - (s_0 - D_0)^-)^-]$. 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 $(\sum_{i>0} s_i^* - (s_0 - D_0)^-)^+ = \sum_{i>0} s_i^*$ and $(\sum_{i>0} s_i^* - (s_0 - D_0)^-)^- = 0$, 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

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

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

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

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

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

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

PROOF OF THEOREM 2. Theorem 2 is a special case of Theorem 3. The proof is shown under Theorem 3. \square

PROOF OF PROPOSITION 3. Our proof uses the following lemma from Shaked and Shanthikumar (1994, example 6.A.2). \square

LEMMA 10. 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 $i \in V \setminus \mathcal{L}$. 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

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

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

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

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

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

When $y \geq d_i + \sum_{j \in S(i)} S_j^{r0}$, $A_j(y, d_i) = \underline{C}_j(0)$ 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 $y < d_i + \sum_{j \in S(i)} S_j^{r0}$. Thus, we only need to show that $A_j(y, d_i)$ is decreasing convex in y when $y < d_i + \sum_{j \in S(i)} S_j^{r0}$. Indeed, by setting $n = d_i + \sum_{j \in S(i)} S_j^{r0} - y$, Lemma 10 suggests that

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

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 $y \geq d_i + \sum_{j \in S(i)} S_j^{r0}$, 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 $y = d_i + \sum_{j \in S(i)} S_j^{r0}$. Therefore, the convexity is also satisfied at $y = d_i + \sum_{j \in S(i)}$. Furthermore, $\underline{C}_i(\cdot)$ is convex since linear transformation preserves convexity.

PROOF OF PROPOSITION 4. The result follows directly from the construction of the RO heuristic, that is, the base-stock levels are calculated recursively from the leaf nodes to the root node. \square

PROOF OF PROPOSITION 5. The backorder matching procedure ensures that s_i^a is an optimal solution to the following problem.

$$\begin{aligned} s_i^a &= \underset{z \in \mathbb{Z}}{\operatorname{argmax}} E[(D_i - z)^+], \\ \text{s.t. } E[(D_i - z)^+] &\leq \sum_{w \in \mathcal{W}: i \in w} Q_{D_{i_w}}(s_{i_w}^d) \end{aligned} \quad (\text{A3})$$

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

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 $S_{i_w}^{SS}$ for any serial system w . In case 2, node j is not a leaf. Based on the form of Equation (11), changing the leadtime of node j does not affect $S_{i_w}^{SS}$, where node i is downstream from node j in serial system w . In both cases, it can be also shown that $S_{k_w}^{SS}$ remains the same, where k_w is the successor of i_w in serial system w . Therefore, $s_{i_w}^d$ 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 $T(i)$. \square

PROOF OF THEOREM 3. We first obtain an upper bound on $C(\mathbf{s})$ and then use it to derive an upper bound on $c^a(\hat{\mathbf{h}})$.

$$\begin{aligned} C(\mathbf{s}) &= \sum_{i \in V \setminus \mathcal{L}} h_i E \left[I_i(\mathbf{s}) + \sum_{j \in S(i)} IT_j \right] + \sum_{i \in \mathcal{L}} E[h_i I_i(\mathbf{s}) + b_i B_i(\mathbf{s})] \\ &\leq \sum_{i \in V \setminus \mathcal{L}} h_i E \left[s_i + \sum_{j \in S(i)} IT_j \right] \\ &\quad + \sum_{i \in \mathcal{L}} E[(h_i - h_{P(i)})(s_i - B_{P(i),i}(\mathbf{s}) - D_i)^+ \\ &\quad + (b_i + h_{P(i)})(B_{P(i),i}(\mathbf{s}) + D_i - s_i)^+] \\ &\quad + \sum_{i \in \mathcal{L}} h_{P(i)}(s_i - B_{P(i),i}(\mathbf{s}) - D_i) \\ &= \sum_{i \in V \setminus \mathcal{L}} E \left[H_i \left(S_i - \sum_{j \in T(i) \cap \mathcal{L}} S_j \right) + h_i \sum_{j \in S(i)} IT_j \right] \\ &\quad + \sum_{i \in \mathcal{L}} E[(h_i - h_{P(i)})(s_i - B_{P(i),i}(\mathbf{s}) - D_i)^+ \\ &\quad + (b_i + h_{P(i)})(B_{P(i),i}(\mathbf{s}) + D_i - s_i)^+] \\ &\quad + \sum_{i \in \mathcal{L}} h_{P(i)}(s_i - B_{P(i),i}(\mathbf{s}) - D_i) \\ &\leq \sum_{i \in V \setminus \mathcal{L}} E \left[H_i S_i + h_i \sum_{j \in S(i)} IT_j \right] \\ &\quad + \sum_{i \in \mathcal{L}} E[(h_i - h_{P(i)})(s_i - B_{P(i),i}(\mathbf{s}) - D_i)^+ \\ &\quad + (b_i + h_{P(i)})(B_{P(i),i}(\mathbf{s}) + D_i - s_i)^+] \\ &\quad + \sum_{i \in \mathcal{L}} h_{P(i)}(s_i - B_{P(i),i}(\mathbf{s}) - D_i) \end{aligned}$$

The first inequality follows from Equation (1) and the fact that $s_i \geq I_i(\mathbf{s})$. The second equality is due to the conversion between echelon and local base-stock levels and holding costs.

Note that this bound holds for any vector \mathbf{s} of base-stock levels; in particular, it holds for the vector returned by the DA heuristic, so:

$$\begin{aligned} c^a(\hat{\mathbf{h}}) &\leq \sum_{i \in V \setminus \mathcal{L}} E \left[H_i \sum_{w \in \mathcal{W}: i \in w} S_{i_w}^{SS}(\hat{\mathbf{h}}) + h_i \sum_{j \in S(i)} IT_j \right] \\ &\quad + \sum_{i \in \mathcal{L}} E[(h_i - h_{P(i)})(s_i^a(\hat{\mathbf{h}}) - B_{P(i),i}(\mathbf{s}^a) - D_i)^+ \\ &\quad + (b_i + h_{P(i)})(B_{P(i),i}(\mathbf{s}^a) + D_i - s_i^a(\hat{\mathbf{h}}))^+] \\ &\quad + \sum_{i \in \mathcal{L}} h_{P(i)}(s_i^a(\hat{\mathbf{h}}) - B_{P(i),i}(\mathbf{s}^a) - D_i) \\ &\leq \sum_{i \in V \setminus \mathcal{L}} E \left[H_i \sum_{w \in \mathcal{W}: i \in w} S_{i_w}^{SS}(\hat{\mathbf{h}}) + h_i \sum_{j \in S(i)} IT_j \right] \\ &\quad + \sum_{i \in \mathcal{L}} \left[(h_i + b_i) E[B_{P(i),i}(\mathbf{s}^a)] + \eta_i(s_i^a(\hat{\mathbf{h}})) + h_{P(i)} s_i^a(\hat{\mathbf{h}}) \right] \end{aligned}$$

$$\leq \sum_{i \in V \setminus \mathcal{L}} E \left[H_i \sum_{w \in \mathcal{W}: i \in w} S_{i_w}^{SS}(\hat{h}) + h_i \sum_{j \in \mathcal{S}(i)} IT_j \right] \\ + \sum_{i \in \mathcal{L}} \left[(h_i + b_i) E[B_{\mathcal{P}(i)}(\mathbf{s}^a)] + h_{\mathcal{P}(i)} s_i^a(\hat{h}) \right] + \sum_{i \in \mathcal{L}} \eta_i(s_i^u)$$

where the definition of $\eta_i(x)$ can be found in the proof of Theorem 1. The first inequality follows from Proposition 5, which indicates that the echelon base-stock level at location i after backorder matching is no larger than $\sum_{w \in \mathcal{W}: i \in w} S_{i_w}^{SS}(\hat{h})$. The second inequality follows from Equation (A2). The third inequality is due to the fact that $s_i^a(\hat{h})$ is the minimizer of $\eta_i(\cdot)$.

Note that $s_i^a(\hat{h})$ is finite for all $i \in \mathcal{L}$, as is $E[IT_j]$ for all j . Therefore, $\lim_{h_{\mathcal{P}(i)} \rightarrow 0} h_{\mathcal{P}(i)} s_i^a(\hat{h}) = 0$ for $i \in \mathcal{L}$ and $\lim_{h_i \rightarrow 0} h_i \sum_{j \in \mathcal{S}(i)} E[IT_j] = 0$ for all i . In addition, from the proof of Theorem 2, we know that $\lim_{h_{\mathcal{P}(i)} \rightarrow 0} s_{\mathcal{P}(i)}^a = \infty$ for $i \in \mathcal{L}$. Therefore, $\lim_{h_{\mathcal{P}(i)} \rightarrow 0} E[B_{\mathcal{P}(i)}(\mathbf{s}^a)] = 0$. Since $h_i \rightarrow 0$, we also have $H_i \rightarrow 0$. Following the same logic as in the proof of Theorem 2, we also have $\lim_{H_i \rightarrow 0} H_i S_{i_w}^{SS}(\hat{h}) = 0$ for $i \notin \mathcal{L}$. Therefore, we have

$$1 \leq \lim_{\hat{h} \rightarrow 0} \frac{c^a(\hat{h})}{c^*(\hat{h})} \leq \lim_{\hat{h} \rightarrow 0} \frac{c^a(\hat{h})}{\sum_{i \in \mathcal{L}} c_i(s_i^u)} = 1.$$

References

- Axsäter, S. 1990. Simple solution procedures for a class of two-echelon inventory problems. *Oper. Res.* **38**(1): 64–69.
- Axsäter, S. 1993. Continuous review policies for multi-level inventory systems with stochastic demand. *Handbooks Oper. Res. Manage. Sci.* **4**: 175–197.
- Chen, F., Y. Zheng. 1994. Lower bounds for multi-echelon stochastic inventory systems. *Management Sci.* **40**: 1426–1443.
- Clark, J., H. Scarf. 1960. Optimal policies for a multi-echelon inventory problem. *Management Sci.* **6**(4): 475–490.
- DeCroix, G. 2013. Inventory management for an assembly system subject to supply disruptions. *Management Sci.* **59**(9): 2079–2092.
- Dogru, M., A. de Kok, G. van Houtum. 2009. A numerical study on the effect of the balance assumption in one-warehouse multi-retailer inventory systems. *Flex. Serv. Manuf. J.* **21**(3): 114–147.
- Eppen, G. D. 1979. Effects of centralization on expected costs in a multi-location newsboy problem. *Management Sci.* **25**(5): 498–501.
- Federgruen, A., P. Zipkin. 1984a. Approximations of dynamic, multilocational production and inventory problems. *Management Sci.* **30**: 69–84.
- Gallego, G., I. Moon. 1993. The distribution-free newsboy problem: Review and extensions. *J. Oper. Res. Soc.* **44**(8): 825–834.
- Gallego, G., O. Özer. 2005. A new algorithm and a new heuristic for serial supply systems. *Oper. Res. Lett.* **33**: 349–362.
- Gallego, G., P. Zipkin. 1999. Stock positioning and performance estimation in serial productiontransportation systems. *Manuf. Serv. Oper. Manag.* **1**: 77–88.
- Gallego, G., O. Özer, P. Zipkin. 2007. Bounds, heuristics, and approximations for distribution systems. *Oper. Res.* **55**(3): 503–517.
- Graves, S. 1985. A multi-echelon inventory model for a repairable item with one-for-one replenishment. *Management Sci.* **31**: 1247–1256.
- Huh, W. T., G. Janakiraman. 2008. A sample-path approach to the optimality of echelon orderup- to policies in serial inventory systems. *Oper. Res. Lett.* **36**(5): 547–550.
- Huh, W., G. Janakiraman. 2010. On the optimal policy structure in serial inventory systems with lost sales. *Oper. Res.* **58**(2): 486–491.
- Huh, W., G. Janakiraman, M. Nagarajan. 2010. Technical note capacitated serial inventory systems: Sample path and stability properties under base-stock policies. *Oper. Res.* **58**(4-Part-1): 1017–1022.
- Özer, O., H. Xiong. 2008. Stock positioning and performance estimation for distribution systems with service constraints. *IIE Trans.* **40**(12): 1141–1157.
- Shaked, M., G. Shanthikumar. 1994. *Stochastic Orders and Their Applications*, Academic Press, San Diego, CA.
- Shang, K. 2012. Single-stage approximations for optimal policies in serial inventory systems with non-stationary demand. *Manuf. Serv. Oper. Manag.* **14**(3): 414–422.
- Shang, K., J.-S. Song. 2003. Newsvendor bounds and heuristic for optimal policies in serial supply chains. *Management Sci.* **49**(5): 618–638.
- Shang, K., J.-S. Song. 2006. A closed-form approximation for serial inventory systems and its application to system design. *Manuf. Serv. Oper. Manag.* **8**(4): 394–406.
- Sherbrooke, C. 1968. Metric: A multi-echelon technique for recoverable item control. *Oper. Res.* **16**: 122–141.
- Zipkin, P. 2000. *Foundations of Inventory Management*, McGraw-Hill, New York.