Stock rationing in a continuous review two-echelon inventory model

Sven Axsäter* Marcel Kleijn† Ton G. de Kok‡

JULY 1998
ECONOMETRIC INSTITUTE REPORT EI9827/A

Abstract

In this paper we consider a 1-warehouse, N-retailer inventory system where demand occurs at all locations. We introduce an inventory model which allows us to set different service levels for retailers and direct customer demand at the warehouse. For each retailer a critical level is defined, such that a retailer replenishment order is delivered from warehouse stock if and only if the stock level exceeds this critical level. It is assumed that retailer replenishment orders which are not satisfied from warehouse stock are delivered directly from the outside supplier, instead of being backlogged. We present an analytical upper bound on the total cost of the system, and develop a heuristic method to optimise the policy parameters. Numerical experiments indicate that our technique provides a very close approximation of the exact cost. Also, we show that differentiating among the retailers and direct customer demand can yield significant cost reductions.

1 Introduction

In the literature on multi-echelon inventory theory it is, in general, assumed that demand only occurs at end-stage stockpoints. However, in practice we often encounter the situation where higher-stage stockpoints also face direct customer demand. Consider for example a company having a factory in Finland, a central depot in The Netherlands and regional depots in France and Italy. A natural distribution scheme would be to ship all products to the central depot and from there to the regional depots. It is clear that it would be efficient to supply the customers in The Netherlands from the central depot. Hence, at the central depot three types of demand occur: direct customer demand, replenishment orders from the regional depot in France and replenishment orders from the regional depot in Italy. A problem at the central

*Lund University, Production Management, P.O. Box 118, S-221 00 Lund, Sweden.
†Erasmus University Rotterdam, Econometric Institute, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands.
‡Eindhoven University of Technology, Technology Management, P.O. Box 513, 5600 MB Eindhoven, The Netherlands.

depot is that these types of demand are not equally important and the inventory manager may want to set different service levels. For example, not satisfying a direct customer order from stock will, in general, be more costly than not satisfying a regional depot's replenishment order, and therefore it is desirable to set a higher service level for direct customer demand. But, the inventory manager may also want to set different service levels for different regional depots. For example, if France is a key market, the desired service level at the regional depot in France should be higher than the one in Italy. This may be obtained by holding more inventory at the regional depot in France and/or by requiring a higher service level for France at the central depot. The latter option may be more attractive if the unit holding cost at the central depot is less than at the regional depot.

Existing models for multi-echelon inventory systems can be adapted in various ways to incorporate direct customer demand at higher-stage stockpoints. To illustrate this, consider a 1-warehouse, N-retailer inventory system where the warehouse supplies both final customers and retailers. Introduce now an additional retailer with warehouse-retailer lead time zero, which supplies all direct customer demand. If the retailer holds no stock and each customer demand is immediately ordered at the warehouse, the system is equivalent to a system with direct customer demand at the warehouse. By holding inventory at the new retailer, we actually reserve some of the warehouse stock for direct customer demand. This way, a higher service level for direct customer demand can be obtained by holding more inventory.

In a periodic review multi-echelon inventory system, it is easy to set different service levels for different retailers at the warehouse, by defining different allocation functions. For an overview we refer to Diks, De Kok & Lagodimos [4]. However, nearly all continuous review multi-echelon inventory models apply a simplifying first come, first served assumption. In case of Poisson demand and one-for-one ordering retailers, the warehouse service level will then be the same for all retailers. In this paper, we will discuss a model which is able to set different service levels for the retailers and direct customer demand. A main difference between our model and the common two-echelon inventory models is the way the warehouse handles a stockout. Usually it is assumed that all demand at the warehouse is backlogged (see e.g. Axsäter [2]), but in our model only direct customer demand at the warehouse is backlogged. If the warehouse does not satisfy a replenishment order placed by a retailer from stock on hand, an (emergency) order is placed at the outside supplier which is shipped directly to the retailer. We shall assume that the emergency lead time to the retailer in this case exceeds the normal warehouse-retailer lead time. This assumption is a common one (see e.g. Muckstadt & Thomas [6]). On the other hand it is assumed that the emergency lead time does not exceed the total outside supplier-retailer lead time. An advantage of our model is that the lead time for a retailer replenishment order is less uncertain, since it can only attain two values. A disadvantage is that replenishment orders may overlap, which makes the mathematical analysis more difficult.

The allocation policy that we introduce in this paper is inspired by an inventory model for different demand classes (Dekker, Hill & Kleijn [3], Ha [5]). For every retailer we define a nonnegative *critical level*; a retailer replenishment order is delivered from warehouse stock if and only if the warehouse inventory level exceeds the critical level for this retailer. The critical level for direct customer demand is set equal to $-\infty$, implying that customer demand is always satisfied if there is stock on hand and backordered otherwise. By setting different critical levels for different retailers one can obtain different service levels.

In the next section we analyse the inventory model and derive expressions for the cost. In Section 3 the optimisation of the inventory policy is discussed, and in Section 4 the results of numerical experiments are presented. We conclude the paper by summarising the main results in Section 5.

2 The model

2.1 Definitions and assumptions

We analyse a 1-warehouse, N-retailer inventory system, where $(S-1, S)$ inventory policies are applied at all locations. Customer demand occurs at all stockpoints and is assumed to be Poisson distributed, with arrival rate λ_j for retailer j, $j = 1, \ldots, N$, and λ_0 for the warehouse. If a demand occurs at a retailer, a replenishment order for the warehouse is triggered, and, provided the order is accepted by the warehouse, the warehouse places a replenishment order at the outside supplier. The order-up-to level at retailer j is denoted by S_j, $j = 1, \ldots, N$, and the warehouse order-up-to level is S_0. The lead time between the outside supplier and the warehouse equals L_0 and the warehouse-retailer j lead time (transportation time) equals L_j, $j = 1, \ldots, N$. If the warehouse does not accept an order placed by retailer j, then an (emergency) order is placed at the outside supplier which is delivered directly to retailer j and arrives $L_j + \Delta_j$ time units later, with $0 < \Delta_j < L_0$. The cost of such an order is given by $\pi_j \geq 0$, $j = 1, \ldots, N$. Since we do not explicitly model the normal replenishment cost, the value of π_j can be interpreted as the additional cost of a direct delivery. Direct customer demand at the warehouse is always accepted. If there is no stock on hand the demand is backordered. The holding cost (per unit per unit time) at retailer j equals h_j and at the warehouse it equals h_0. Similarly, the shortage costs (per unit per unit time) are denoted by β_j and β_0. Finally, the warehouse allocation policy is characterised by the vector of critical levels $\boldsymbol{c} := (c_1, \ldots, c_N)$. A replenishment order from retailer j is delivered from warehouse stock if and only if the stock on hand strictly exceeds $c_j \geq 0$. Defining $\boldsymbol{S} := (S_0, S_1, \ldots, S_N)$ the inventory policy is completely characterised by $(\boldsymbol{c}, \boldsymbol{S})$. If $c_j \geq S_0$ no orders from retailer j are delivered from the warehouse stock. Consequently, there is no need to consider $c_j > S_0$. The inventory system is illustrated in Figure 1.

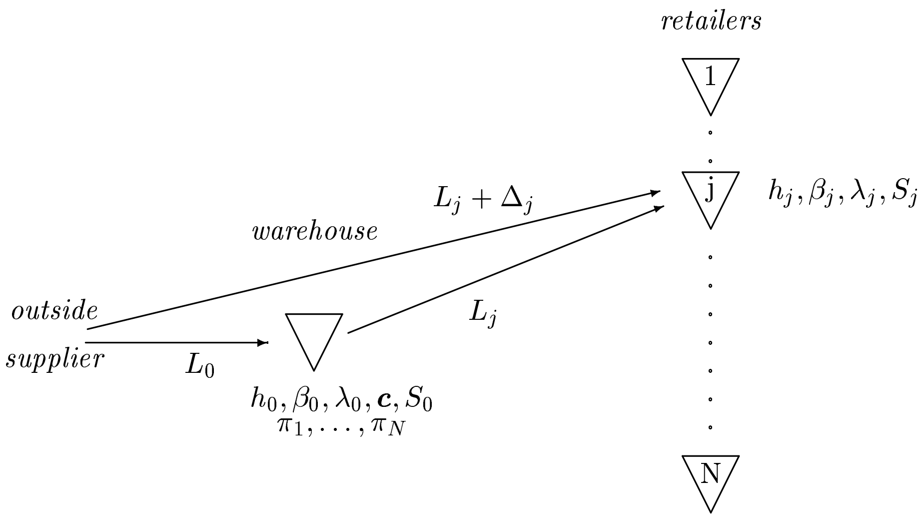

Figure 1: Illustration of the two-echelon inventory system.

2.2 Warehouse cost

For the warehouse cost the parameters $\boldsymbol{c}$ and S_0 are relevant. Given these values one can determine the steady-state probabilities of the number of outstanding orders at the warehouse. Let p_i, $i = 0, 1, \ldots$, denote the steady-state probability that i orders are outstanding or, equivalently, that there are $S_0 - i$ units in stock ($S_0 \geq i$) or $i - S_0$ units backordered ($S_0 < i$). Given the vector of critical levels $\boldsymbol{c}$ and S_0 we can determine the aggregate arrival rate μ_i in state i by

$$\mu_i = \lambda_0 + \sum_{j=1}^{N} \lambda_j 1_{\{S_0 - i > c_j\}}$$

since orders from retailer j are only satisfied if the stock level exceeds c_j. Direct customer demand is always delivered from stock on hand or backordered. It is easy to see that $\mu_i = \lambda_0$ for all $i \geq S_0$ because in this case there is no stock on hand and all retailer demand is satisfied from the outside supplier.

Using a similar analysis as Dekker, Hill & Kleijn [3] it follows that

$$p_i = \frac{\prod_{k=0}^{i-1} (\mu_k L_0)}{i!} p_0, \qquad i = 0, 1, 2, \ldots$$

with p_0 normalised in the usual way. The product term, here and elsewhere in the paper, is defined as 1 if the lower limit exceeds the upper limit by 1. Since $\mu_k = \lambda_0$ for $k \geq S_0$ we get that for $i \geq S_0$

$$p_i = \left\{ \prod_{k=0}^{S_0-1} \mu_k L_0 \right\} \frac{(\lambda_0 L_0)^{i-S_0}}{i!} p_0 = \left\{ \prod_{k=0}^{S_0-1} \frac{\mu_k}{\lambda_0} \right\} \frac{(\lambda_0 L_0)^{i}}{i!} p_0$$

With $\gamma := \prod_{k=0}^{S_0-1}(\mu_k/\lambda_0)$ we obtain that

$$p_0 = \left(\sum_{i=0}^{S_0-1} (\prod_{k=0}^{i-1} \mu_k - \gamma\lambda_0^i) \frac{L_0^i}{i!} + \gamma \exp(\lambda_0 L_0) \right)^{-1}$$

The probability that the warehouse can satisfy a replenishment order from retailer j, $j = 1, \ldots, N$, is given by

$$\alpha_j(\boldsymbol{c}, S_0) = \sum_{i=0}^{S_0-c_j-1} p_i \tag{1}$$

In the next result we derive the total warehouse cost.

Theorem 2.1 *The total warehouse cost for a* $(\boldsymbol{c}, S_0)$ *policy is given by*

$$\begin{aligned} C_0(\boldsymbol{c}, S_0) &= \sum_{i=0}^{S_0-1} \left[(h_0+\beta_0)(S_0-i) + \beta_0 L_0(\mu_i - \lambda_0)\right] p_i \\ &\quad + \beta_0(\lambda_0 L_0 - S_0) + \sum_{j=1}^{N} \lambda_j (1-\alpha_j(\boldsymbol{c}, S_0))\pi_j \end{aligned}$$

Proof: First, we observe that the total warehouse cost is given by

$$C_0(\boldsymbol{c}, S_0) = h_0 \sum_{i=0}^{S_0-1}(S_0-i)p_i + \beta_0 \sum_{i=S_0}^{\infty}(i-S_0)p_i + \sum_{j=1}^{N}\lambda_j(1-\alpha_j(\boldsymbol{c}, S_0))\pi_j$$

Because this expression involves an infinite sum, we wish to simplify the above formula. Observing that

$$\begin{aligned} \sum_{i=S_0}^{\infty}(i-S_0)p_i &= \sum_{i=0}^{\infty}(i-S_0)p_i - \sum_{i=0}^{S_0-1}(i-S_0)p_i \\ &= \sum_{i=0}^{\infty} ip_i - S_0 + \sum_{i=0}^{S_0-1}(S_0-i)p_i \end{aligned}$$

we get

$$C_0(\boldsymbol{c}, S_0) = (h_0+\beta_0)\sum_{i=0}^{S_0-1}(S_0-i)p_i + \beta_0(\sum_{i=0}^{\infty} ip_i - S_0) + \sum_{j=1}^{N}\lambda_j(1-\alpha_j(\boldsymbol{c}, S_0))\pi_j$$

Moreover, since the average number of outstanding orders must equal the average aggregate demand at the warehouse during a time interval of length L_0 we have

$$\begin{aligned} \sum_{i=0}^{\infty} ip_i &= L_0 \sum_{i=0}^{\infty}\mu_i p_i \\ &= \lambda_0 L_0 + L_0 \sum_{i=0}^{S_0-1}(\mu_i - \lambda_0)p_i \end{aligned}$$

and thus the desired result follows. □

2.3 Retailer cost

In this section we derive an upper bound as well as an 'approximate lower bound' on the retailer's inventory cost. Let $Q_j(t)$ denote the number of outstanding orders at retailer j, $j = 1, \ldots, N$, at time t. Furthermore, let $B_j(t, \tau)$ denote the number of replenishment orders from retailer j in time period $[t, \tau)$ which cannot be delivered from warehouse stock, and let $D_j(t, \tau)$ be the total customer demand in $[t, \tau)$. Then it follows that

$$Q_j(t + L_j) \stackrel{d}{=} B_j(t - \Delta_j, t) + D_j(t, t + L_j) \tag{2}$$

An intuitive explanation of this result is the following: at time $t + L_j$ all replenishment orders which were placed at the warehouse before t and could be satisfied from warehouse stock will have arrived at retailer j. Similary, all replenishment orders placed before $t - \Delta_j$ which could not be delivered from warehouse stock will have arrived at $t + L_j$. Replenishment orders which were placed in the time period $[t - \Delta_j, t)$ and could not be satisfied from warehouse stock will not yet have arrived at retailer j. Finally, no replenishment orders (or equivalently, no customer demand) placed in $[t, t + L_j)$ have yet arrived.
If we know the distributions of $B_j(t - \Delta_j, t)$ and $D_j(t, t + L_j)$ then we can convolute them to obtain the distribution of $Q_j(t + L_j)$ and hence we may calculate the average cost at retailer j, observing that the inventory level at time t equals $S_j - Q_j(t)$. Although it is very difficult to obtain the distribution of $B_j(t - \Delta_j, t)$, it can easily be verified that its expected value is given by

$$\mathbb{E}B_j(t - \Delta_j, t) = (1 - \alpha_j(\boldsymbol{c}, S_0))\lambda_j \Delta_j$$

with $1 - \alpha_j(\boldsymbol{c}, S_0)$ the probability that a replenishment order from retailer j cannot be delivered from warehouse stock. Since $D_j(t, t + L_j)$ is Poisson distributed with parameter $\lambda_j L_j$, we obtain that the expected number of outstanding orders at retailer j is given by

$$\mathbb{E}Q_j = \lambda_j(L_j + (1 - \alpha_j(\boldsymbol{c}, S_0))\Delta_j)$$

We shall approximate the steady-state distribution of Q_j by a Poisson distribution with parameter $\mathbb{E}Q_j$. This means that we are replacing the stochastic lead time by

$$\bar{L}_j(\boldsymbol{c}, S_0) := L_j + (1 - \alpha_j(\boldsymbol{c}, S_0))\Delta_j \tag{3}$$

i.e. the 'average' lead time for a replenishment order for a $(\boldsymbol{c}, S_0)$ policy at the warehouse. We are consequently using the well-known METRIC approximation, first introduced by Sherbrooke [7]. Let ℓ_j be the stochastic lead time for an order from retailer j, and let us first assume that ℓ_j is independent of the demand after the order. In that case the exact cost is given by

$$C_j(S_j | \boldsymbol{c}, S_0) = \mathbb{E}_{\ell_j} v_j(S_j | \ell_j)$$

where $v_j(S_j|y)$ is the cost for a given deterministic lead time y. It can be shown that $v_j(S_j|y)$ is given by

$$v_j(S_j|y) := (h_j + \beta_j) \sum_{k=0}^{S_j-1} (S_j - k) \frac{(\lambda_j y)^k}{k!} \exp(-\lambda_j y) + \beta_j(\lambda_j y - S_j) \tag{4}$$

The expected cost for the METRIC approach is given by

$$C_j^L(S_j|\boldsymbol{c}, S_0) := v_j(S_j|\bar{L}_j(\boldsymbol{c}, S_0)) \tag{5}$$

It is easy to show that the function $y \to v_j(S_j|y)$ is convex. Therefore, if ℓ_j were independent of the demand after the order, by Jensen's inequality $C_j^L(S_j|\boldsymbol{c}, S_0)$ is a lower bound for $C_j(S_j|\boldsymbol{c}, S_0)$. In our case, though, the demand after the order can affect the lead time, since replenishment orders may overlap. Therefore, we cannot prove that $C_j^L(S_j|\boldsymbol{c}, S_0)$ is a lower bound. From a practical point of view, however, $C_j^L(S_j|\boldsymbol{c}, S_0)$ can serve as an 'approximate lower bound'. In our simulation tests (see Section 4) we have not found a single case where $C_j^L(S_j|\boldsymbol{c}, S_0)$ significantly exceeds the exact cost.

We will now proceed to derive an upper bound on the retailer's inventory cost. Therefore, we will follow the approach of Axsäter [1] who made the following observation, which trivially holds in a two-echelon inventory system with backordering and first-come, first-served allocation policy at the warehouse: *any unit ordered by retailer j, $j = 1, \dots, N$, is used to fill the S_jth demand at this retailer following this order.* He derives the average cost by 'following' a unit through the system. In our situation, this observation does not hold because orders may overlap. Assume that at time 0 retailer j places an order at the warehouse and the warehouse cannot satisfy this order. Then an (emergency) shipment is initiated from the outside supplier which will arrive at retailer j at time $L_j + \Delta_j$. Let X_0 denote the time until the warehouse stock level is raised above c_j (so there is enough stock to satisfy an order from retailer j) and assume the next retailer j order occurs at X_1. If $X_0 < X_1 < \Delta_j$ then the replenishment order placed by retailer j at time X_1 will arrive earlier at this retailer than the previous replenishment order placed at time 0, i.e. respectively at $X_1 + L_j$ and $L_j + \Delta_j$.

If we assume that Axsäter's observation also holds for our system, then following the approach of Axsäter [1] and observing that the replenishment lead time for retailer j equals L_j with probability $\alpha_j(\boldsymbol{c}, S_0)$ and $L_j + \Delta_j$ with probability $1 - \alpha_j(\boldsymbol{c}, S_0)$, we obtain that the expected cost is given by

$$C_j^U(S_j|\boldsymbol{c}, S_0) = \alpha_j(\boldsymbol{c}, S_0) v_j(S_j|L_j) + (1 - \alpha_j(\boldsymbol{c}, S_0)) v_j(S_j|L_j + \Delta_j) \tag{6}$$

The superscript U is used because it can be shown that in fact this expression provides an upper bound on the real cost, which is proved in the next lemma.

Lemma 2.1 *The function $C_j^U(S_j|\boldsymbol{c}, S_0)$ is an upper bound on the real retailer j cost, if its order-up-to level equals S_j and the warehouse applies a $(\boldsymbol{c}, S_0)$ policy.*

Proof: Given the considered cost structure, it is easy to show that a first-come, first-served policy is optimal for the retailers. If a unit ordered by retailer j is used to fill the S_jth demand at retailer j following this order, we deviate from the first-come, first-served policy. Hence, we obtain that $C_j^U(S_j|\boldsymbol{c}, S_0)$ is an upper bound on the real cost. □

2.4 Total cost

From the above analysis we obtain that the total cost of a $(\boldsymbol{c}, \boldsymbol{S})$ policy is normally bounded from below by

$$C_0(\boldsymbol{c}, S_0) + \sum_{j=1}^{N} C_j^L(S_j|\boldsymbol{c}, S_0)$$

and bounded from above by

$$C_0(\boldsymbol{c}, S_0) + \sum_{j=1}^{N} C_j^U(S_j|\boldsymbol{c}, S_0)$$

3 Optimisation

In this section we discuss the optimisation of the $(\boldsymbol{c}, \boldsymbol{S})$ policy. First, we will show how to determine the order-up-to levels of the retailers, given the warehouse policy, and secondly, we discuss the determination of the warehouse parameters.

3.1 Retailer policy

We assume that the warehouse applies a $(\boldsymbol{c}, S_0)$ policy. The optimal retailer j order-up-to level, using the approximate lower bound on the retailer cost, is denoted by

$$S_j^{L*}(\boldsymbol{c}, S_0) := \arg\min\{C_j^L(x|\boldsymbol{c}, S_0) : x \geq 0\}$$

Similarly, the optimal order-up-to level using the upper bound on the cost is denoted by

$$S_j^{U*}(\boldsymbol{c}, S_0) := \arg\min\{C_j^U(x|\boldsymbol{c}, S_0) : x \geq 0\}$$

The following result is intuitively obvious. For completeness we give a formal proof.

Theorem 3.1 *It follows for $j = 1, \ldots, N$ that*

$$S_j^L \leq S_j^{L*}(\boldsymbol{c}, S_0) \leq S_j^U$$

and

$$S_j^L \leq S_j^{U*}(\boldsymbol{c}, S_0) \leq S_j^U$$

with

$$S_j^L := \arg\min\{v_j(x|L_j) : x \geq 0\}$$

and

$$S_j^U := \arg\min\{v_j(x|L_j+\Delta_j) : x \geq 0\}$$

Proof: We first prove the result for the lower bound. Observe that

$$C_j^L(x|\boldsymbol{c}, S_0) = v_j(x|\bar{L}_j(\boldsymbol{c}, S_0))$$

Then because

$$v_j(x+1|y) - v_j(x|y) = (\beta_j + h_j)\sum_{k=0}^{x} \frac{(\lambda_j y)^k}{k!}\exp(-\lambda_j y) - \beta_j \tag{7}$$

it follows that the function $x \to v_j(x+1|y) - v_j(x|y)$ is increasing for all $y > 0$. From (7) it can be verified that

$$\frac{d}{dy}(v_j(x+1|y) - v_j(x|y)) = -(\beta_j + h_j)\lambda_j \frac{(\lambda_j y)^x}{x!}\exp(-\lambda_j y) < 0$$

and hence, observing that due to $x \to v_j(x+1|y) - v_j(x|y)$ is increasing it follows that

$$\arg\min\{v_j(x|y) : x \geq 0\} = \min\{x \geq 0 : v_j(x+1|y) - v_j(x|y) \geq 0\}$$

we obtain that the function $y \to \arg\min\{v_j(x|y) : x \geq 0\}$ is increasing. Finally, we observe that $L_j \leq \bar{L}_j(\boldsymbol{c}, S_0) \leq L_j + \Delta_j$ and hence the first result follows. To prove the second part, we note that

$$C_j^U(x|\boldsymbol{c}, S_0) = \alpha_j(\boldsymbol{c}, S_0)v_j(x|L_j) + (1 - \alpha_j(\boldsymbol{c}, S_0))v_j(x|L_j + \Delta_j)$$

Since $x \to v_j(x|y)$ is convex for all $y > 0$ it follows that also $x \to \alpha_j(\boldsymbol{c}, S_0)v_j(x|L_j) + (1 - \alpha_j(\boldsymbol{c}, S_0))v_j(x|L_j + \Delta_j)$ is convex. The second result follows directly after observing that $0 \leq \alpha_j(\boldsymbol{c}, S_0) \leq 1$ and $y \to \arg\min\{v_j(x|y) : x \geq 0\}$ is increasing. □

We remark that the values of S_j^L and S_j^U are easy to determine using

$$\begin{aligned}\arg\min\{v_j(x|y) : x \geq 0\} &= \min\{x \geq 0 : v_j(x+1|y) - v_j(x|y) \geq 0\} \\ &= \min\{x \geq 0 : \sum_{k=0}^{x} \frac{(\lambda_j y)^k}{k!}\exp(-\lambda_j y) \geq \frac{\beta_j}{\beta_j + h_j}\}\end{aligned} \tag{8}$$

Given a warehouse policy $(\boldsymbol{c}, S_0)$ we may thus efficiently determine the value of $S_j^{L*}(\boldsymbol{c}, S_0)$ and $S_j^{U*}(\boldsymbol{c}, S_0)$ using e.g. a bisection algorithm. If we define

$$C^L(\boldsymbol{c}, S_0) = C_0(\boldsymbol{c}, S_0) + \sum_{j=1}^{N} C_j^L(S_j^{L*}(\boldsymbol{c}, S_0)|\boldsymbol{c}, S_0)$$

then it follows that $C^L(\boldsymbol{c}, S_0)$ is normally a lower bound on the total cost for a $(\boldsymbol{c}, S_0)$ policy at the warehouse. Equivalently,

$$C^U(\boldsymbol{c}, S_0) = C_0(\boldsymbol{c}, S_0) + \sum_{j=1}^{N} C_j^U(S_j^{U*}(\boldsymbol{c}, S_0)|\boldsymbol{c}, S_0)$$

is an upper bound on the total cost.

Assume now that we use $S_j^{L*}(\boldsymbol{c}, S_0)$, or $S_j^{U*}(\boldsymbol{c}, S_0)$, as our retailer policy. We are interested in the cost increase compared to using the (unknown) optimal policy

$$S_j^*(\boldsymbol{c}, S_0) := \arg\min\{C_j(x|\boldsymbol{c}, S_0) : x \geq 0\}$$

If we accept $C_j^L(S_j^{L*}(\boldsymbol{c}, S_0)|\boldsymbol{c}, S_0)$ as a lower bound we obtain the bounds

$$0 \leq C_j(S_j^{L*}(\boldsymbol{c}, S_0)|\boldsymbol{c}, S_0) - C_j(S_j^*(\boldsymbol{c}, S_0)|\boldsymbol{c}, S_0) \leq C_j^U(S_j^{L*}(\boldsymbol{c}, S_0)|\boldsymbol{c}, S_0) - C_j^L(S_j^{L*}(\boldsymbol{c}, S_0)|\boldsymbol{c}, S_0)$$

and

$$0 \leq C_j(S_j^{U*}(\boldsymbol{c}, S_0)|\boldsymbol{c}, S_0) - C_j(S_j^*(\boldsymbol{c}, S_0)|\boldsymbol{c}, S_0) \leq C_j^U(S_j^{U*}(\boldsymbol{c}, S_0)|\boldsymbol{c}, S_0) - C_j^L(S_j^{L*}(\boldsymbol{c}, S_0)|\boldsymbol{c}, S_0)$$

Note that the second bound is tighter.

3.2 Warehouse policy

In this section we discuss the optimisation of the warehouse policy $(\boldsymbol{c}, S_0)$. The determination of the (near) optimal values will be based on the heuristic approach introduced in Dekker, Hill & Kleijn [3]. For an $(S-1, S)$ lost sales inventory model with N demand classes, it was shown that the heuristic approach led to an optimal solution in 99.9% of the cases. In the inventory model of Dekker, Hill & Kleijn the priority ranking of the retailers was known a priori because it was solely based on the values of π_j, $j = 1, \ldots, N$. However, in our system the indirect effect of the warehouse policy on the retailer's cost also affects the priority ranking. We will determine the priority ranking of the retailers by means of the derivative of the total cost with respect to the expected number of emergency orders $\lambda_j(1 - \alpha_j(\boldsymbol{c}, S_0))$ and we assume that a higher value of this derivative corresponds to a higher priority.

Lemma 3.1 *The derivative of the cost associated with retailer j, $j = 1, \ldots, N$, with respect to the expected number of emergency orders $\lambda_j(1-\alpha_j(\boldsymbol{c}, S_0))$ and using the approximate lower bound on the retailer cost, is given by*

$$\delta_j^L(\boldsymbol{c}, S_0) = -(\beta_j + h_j)\Delta_j \sum_{k=0}^{S_j^{L*}(\boldsymbol{c}, S_0)-1} \frac{(\lambda_j \bar{L}_j(\boldsymbol{c}, S_0))^k}{k!} \exp(-\lambda_j \bar{L}_j(\boldsymbol{c}, S_0)) + \Delta_j\beta_j + \pi_j$$

Proof: First, we observe that the cost associated with retailer j is given by

1. the cost for shipments from the outside supplier to the retailer in case the warehouse does not deliver the retailer order from stock on hand

2. the retailer inventory holding and shortage cost

Regarding the first cost component it is easy to verify that the derivative with respect to the stockout probability equals π_j.
For the second cost component we observe that

$$C_j^L(S_j^{L*}(\boldsymbol{c}, S_0)|\boldsymbol{c}, S_0) = v_j(S_j^{L*}(\boldsymbol{c}, S_0)|L_j + z\Delta_j/\lambda_j)$$

with $z = \lambda_j(1 - \alpha_j(\boldsymbol{c}, S_0))$. Since by (4) it can be shown that

$$\begin{aligned}\frac{\partial}{\partial y}v_j(x|y) &= (h_j + \beta_j)\sum_{k=1}^{x-1}(x-k)k\lambda_j\frac{(\lambda_j y)^{k-1}}{k!}\exp(-\lambda_j y) \\ &\quad -(h_j + \beta_j)\lambda_j\sum_{k=0}^{x-1}(x-k)\frac{(\lambda_j y)^k}{k!}\exp(-\lambda_j y) + \beta_j\lambda_j \\ &= -(h_j + \beta_j)\lambda_j\sum_{k=0}^{x-1}\frac{(\lambda_j y)^k}{k!}\exp(-\lambda_j y) + \beta_j\lambda_j\end{aligned}$$

the desired result follows by setting $y := L_j + z\Delta_j/\lambda_j$ and $x := S_j^{L*}(\boldsymbol{c}, S_0)$, and observing that $\partial y/\partial z = \Delta_j/\lambda_j$. □

A similar result can be derived for the upper bound on the total cost.

Lemma 3.2 *The derivative of the cost associated with retailer j, $j = 1, \ldots, N$, with respect to the expected number of emergency orders $\lambda_j(1 - \alpha_j(\boldsymbol{c}, S_0))$ and using the upper bound on the retailer cost, is given by*

$$\delta_j^U(\boldsymbol{c}, S_0) = \frac{1}{\lambda_j}\Big(v_j(S_j^{U*}(\boldsymbol{c}, S_0)|L_j + \Delta_j) - v_j(S_j^{U*}(\boldsymbol{c}, S_0)|L_j)\Big) + \pi_j$$

Proof: Following the first part of the proof of Lemma 3.1 we obtain that π_j is the derivative with respect to the emergency shipment cost. Observing that

$$\begin{aligned}C_j^U(S_j^{U*}(\boldsymbol{c}, S_0)|\boldsymbol{c}, S_0) &= v_j(S_j^{U*}(\boldsymbol{c}, S_0)|L_j) + \lambda_j(1 - \alpha_j(\boldsymbol{c}, S_0)) \times \\ &\quad \frac{1}{\lambda_j}\Big(v_j(S_j^{U*}(\boldsymbol{c}, S_0)|L_j + \Delta_j) - v_j(S_j^{U*}(\boldsymbol{c}, S_0)|L_j)\Big)\end{aligned}$$

the result follows directly. □

With these observations, we now present the heuristic approach to determine the warehouse policy. Before doing so, we define $\boldsymbol{e}_k$ as an N-dimensional vector containing all zeros, except for the kth position where it equals 1. Furthermore, we define $|\mathcal{X}|$ as the number of elements contained in the set $\mathcal{X}$. Finally, for the functions $C(\boldsymbol{c}, S_0)$ and $\delta_j(\boldsymbol{c}, S_0)$ which are used in the heuristic, we may use either $C^L(\boldsymbol{c}, S_0)$ and $\delta_j^L(\boldsymbol{c}, S_0)$ or $C^U(\boldsymbol{c}, S_0)$ and $\delta_j^U(\boldsymbol{c}, S_0)$.
The heuristic approach starts in Step 0 by determining the optimal simple policy where all critical levels are zero. To determine this policy it is assumed that the total cost is unimodal

Step 0:
Set $\boldsymbol{c}^{(0)} := \mathbf{0}$, $S_0^{(0)} = \arg\min\{x \geq 0 : C(\boldsymbol{c}^{(0)}, x)\}$,
$C^{(0)} = C(\boldsymbol{c}^{(0)}, S_0^{(0)})$, $k = 0$ and $\mathcal{I} := \{1, 2, \dots, N\}$.
Step 1:
Let $i^* = \arg\min\{i \in \mathcal{I} : \delta_i(\boldsymbol{c}^{(k)}, S_0^{(k)})\}$. Then we determine
$\boldsymbol{c}^{(k+1)} := \boldsymbol{c}^{(k)} + \boldsymbol{e}_{i^*}$, $S_0^{(k+1)} = \arg\min\{x \geq 0 : C(\boldsymbol{c}^{(k+1)}, x)\}$,
$C^{(k+1)} = C(\boldsymbol{c}^{(k+1)}, S_0^{(k+1)})$ and we set $\mathcal{I} := \mathcal{I} - \{i^*\}$.
Step 2:
If $C^{(k+1)} < C^{(k)}$ and $\mathcal{I} \neq \emptyset$, then set $k := k + 1$ and goto Step 1.
If $C^{(k+1)} < C^{(k)}$ and $\mathcal{I} = \emptyset$, then set $k := k + 1$, $\mathcal{I} := \{1, 2, \dots, N\}$ and goto Step 1.
If $C^{(k+1)} \geq C^{(k)}$ and $|\mathcal{I}| < N - 1$, then set $\mathcal{I} := \{1, 2, \dots, N\}$ and goto Step 1.
If $C^{(k+1)} \geq C^{(k)}$ and $|\mathcal{I}| = N - 1$, then STOP and take $(\boldsymbol{c}^{(k)}, S_0^{(k)})$ as a solution.

Algorithm 3.1: Heuristic approach.

in the warehouse order-up-to level. We performed a number of numerical tests and this assumption was always satisfied. Then in Step 1 we determine the retailer with the smallest value of δ_j and raise its critical level with one unit. After re-optimising the warehouse order-up-to level and the retailer order-up-to levels, we check in Step 2 if there is a cost improvement. If there is, we determine again the retailer with the largest value of δ_j, excluding the previous retailer(s) of which the critical level was raised, and raise its critical level by one unit. If excluding these retailers of which the critical levels were previously raised implies that there are none left, we consider again all retailers as candidates. If there is no cost improvement, we do not raise the critical level and we consider again all retailers and determine the one with the lowest value of δ_j. If raising the critical level of this one does not lead to a cost reduction, the algorithm is terminated.

4 Numerical results

In this section we present the results of a number of numerical experiments. We are interested in the quality of the approximate cost expressions, the quality of the heuristic approach, and the profitability of using a critical level, order-up-to policy at the warehouse. For $N = 3$ retailers we generated 500 data sets in the following way: first, we defined two groups of parameter values, which are presented in Table 1. For each group we randomly generated 250 data sets. All parameter distributions were uniform on the given intervals.

group 1		group 2	
$\lambda_0 \in [0.2, 1]$	$L_0 \in [0.5, 3]$	$\lambda_0 = 1$	$L_0 = 2$
$h_0 \in [1, 5]$	$\beta_0 \in [10, 500]$	$h_0 = 1$	$\beta_0 = 500$
$\lambda_j \in [0.1, 1]$	$L_j \in [0.2, 2]$	$\lambda_j \in [0.1, 1]$	$L_j = 1$
$h_j \in [1, 5]$	$\beta_j \in [10, 500]$	$h_j = 1$	$\beta_j \in [10, 1000]$
$\Delta_j \in [0.5, \min\{3, L_0\}]$	$\pi_j = 0$	$\Delta_j \in [0.5, 2]$	$\pi_j = 0$
$c_j \in [0, \min\{4, S_0\}]$	$S_0 \in [2, 15]$	$c_j \in [0, \min\{4, S_0\}]$	$S_0 \in [2, 15]$

Table 1: Determination of data sets.

4.1 Quality of approximate cost expressions

First of all, we are interested in the quality of our approximate cost expressions. Therefore, we evaluated for the 500 randomly generated data sets the value of the approximate lower bound and the upper bound, and compared them with the exact cost obtained by simulation. Whereas the warehouse policy is prespecified for each case, the retailer policy was determined by minimising the retailer cost using the approximate lower bound. The exact cost was calculated using the average of 5 simulation runs of 20000 time units each (all starting with full stock and using a warming up period of 5000 time units). In Table 2 the minimum, maximum, average and standard deviation of the relative differences between the approximations and the simulated cost are presented. We have defined *lowsim* as the relative difference between the approximate lower bound and simulation, i.e.

$$lowsim := \frac{\text{approximate lower bound} - \text{simulated cost}}{\text{simulated cost}}$$

and we defined *uppsim* as the relative difference between the upper bound and simulation, i.e.

$$uppsim := \frac{\text{upper bound} - \text{simulated cost}}{\text{simulated cost}}$$

		minimum	maximum	average	standard deviation
group 1	*lowsim*	-0.14	0.02	-0.014	0.021
	uppsim	-0.01	0.32	0.020	0.035
group 2	*lowsim*	-0.11	0.02	-0.022	0.024
	uppsim	-0.02	0.31	0.035	0.050

Table 2: Relative differences between approximations and the simulated cost.

In 81.6% of the cases for group 1 and 75.6% of the cases for group 2 the approximate lower bound was closer to the simulated cost than the upper bound. From Table 2 we see that the average deviation of the approximate lower bound is only 1.4% (group 1) and 2.2% (group 2) and therefore we conclude that it is a good approximation for the real cost. Due to the fact that the maximum and standard deviation of the relative difference between simulation and the upper bound are larger than the corresponding values for the approximate lower bound, it follows that the approximate lower bound also is a more reliable approximation than the upper bound.

Another interesting aspect to analyse is the quality of using the average value of the approximate lower bound and the upper bound as an approximation for the real cost. It turns out that this gives a very good approximation. For group 1 the average deviation was only 0.3% and for group 2 it was 0.7%. The maximum deviation was for both groups about 10%. In 90.6% (group 1) and 60.7% (group 2) of the cases the (absolute) deviation was less than 1%, and in only 0.8% respectively 3% the deviation exceeded 5%. The results are summarised in Table 3.

	minimum	maximum	average	standard deviation
group 1	-0.011	0.092	0.003	0.010
group 2	-0.042	0.105	0.007	0.018

Table 3: Relative difference between the average value of the approximations and the simulated cost.

Figure 2 illustrates the quality of the approximate lower bound and the upper bound for two special cases of the group 1 data set. The figure shows that the bounds can be tight (special case 1) but also much different from the exact cost (special case 2). The parameter values of the two special cases are presented in Table 4. The critical levels were the values which minimised the upper bound on the total cost. To compare the approximate lower bound, the upper bound and the simulated cost, we have used for each value of S_0 the retailer order-up-to levels which minimised the approximate lower bound on the cost. This explains why in special case 2 the upper bound is not a smooth function of S_0.

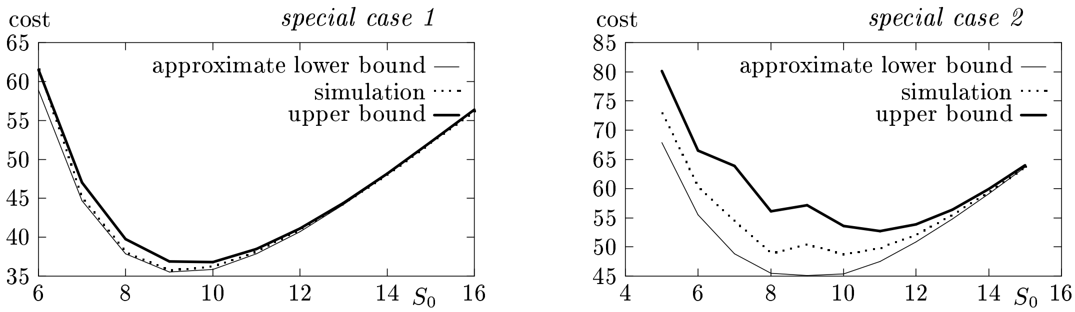

Figure 2: Approximate lower bound, upper bound and simulated cost for the special cases.

	special case 1						special case 2					
warehouse	λ_0	L_0	β_0	h_0			λ_0	L_0	β_0	h_0		
	0.98	2.51	421.86	4.27			0.44	2.79	410.72	4.72		
retailer	λ_1	L_1	Δ_1	β_1	h_1	π_1	λ_1	L_1	Δ_1	β_1	h_1	π_1
1	0.75	0.82	0.59	412.48	1.43	0	0.39	0.24	0.56	185.09	4.75	0
retailer	λ_2	L_2	Δ_2	β_2	h_2	π_2	λ_2	L_2	Δ_2	β_2	h_2	π_2
2	0.46	1.84	1.85	220.87	2.16	0	0.75	0.83	2.40	284.17	3.94	0
retailer	λ_3	L_3	Δ_3	β_3	h_3	π_3	λ_3	L_3	Δ_3	β_3	h_3	π_3
3	0.29	1.37	2.32	16.74	1.59	0	0.68	0.27	0.27	332.10	3.54	0
critical	c_1	c_2	c_3				c_1	c_2	c_3			
levels	1	1	2				0	0	2			

Table 4: Parameter values for two special cases.

4.2 Quality of heuristic approach

Now that we have analysed the quality of the approximate lower bound and the upper bound, we proceed to evaluate the performance of the heuristic approach. As mentioned in Section 3.2, this approach is based on the method of Dekker, Hill & Kleijn [3]. The numerical experiments in [3] showed that in almost all cases the heuristic approach led to an optimal solution. However, in their model the priority ranking of the demand classes was prespecified, whereas in our model this ranking is not known beforehand and we need to determine it using the derivative approach. Hence, we need to evaluate again the performance of the heuristic approach.

For all 500 data sets we have determined the optimal critical level, order-up-to level policy using enumeration, both for the approximate lower bound and the upper bound. First, we used the heuristic approach to find an initial value S_0^H for the order-up-to level. Then we enumerated over all values of $0 \leq c_j \leq S_0$, $j = 1, \ldots, N$, and $0 \leq S_0 \leq \bar{S}_0$, with $\bar{S}_0 := 1.25 S_0^H + 1$, to determine the optimal policy. The numerical results showed that the average cost error of the heuristic approach is 2.1% for the approximate lower bound, and 1.0% for the upper bound. The maximum cost error for all 500 cases was 10% for the approximate lower bound and 8% for the upper bound. In Table 5 the minimum, average, and maximum cost error of the heuristic approach compared to the optimal cost are presented. The last column *optimal* denotes the percentage of cases in which the heuristic approach recovered the optimal policy.

4.3 Profitability of rationing

We will now analyse the profitability of applying a critical level, order-up-to policy at the warehouse instead of using a simple policy (where all critical levels are equal to zero). However,

		minimum	maximum	average	optimal (%)
group 1	*lower*	0	0.10	0.013	30.8
	upper	0	0.08	0.009	32.4
group 2	*lower*	0	0.10	0.029	10.0
	upper	0	0.06	0.011	22.4

Table 5: Cost error of heuristic approach.

based on only the approximate lower bound and the upper bound on the cost and the heuristic approach to determine an optimal policy it is impossible to calculate the exact cost reduction. Nevertheless, we are able to identify a sufficient condition such that applying a critical level policy is cost efficient. If the cost of the optimal critical level policy using the upper bound is less than the cost of the optimal simple policy using the approximate lower bound, then applying a critical level policy leads to a cost reduction. We have calculated for all 500 data sets the following values: *upper bound clp* = the cost of the optimal critical level policy using the upper bound on the cost and *lower bound sp* = the cost of the optimal simple policy using the approximate lower bound on the cost. The guaranteed cost reduction was defined as

$$gcr := \max\{\frac{\text{lower bound sp} - \text{upper bound clp}}{\text{lower bound sp}}, 0\}$$

The results are presented in Table 6.

	$gcr > 0$	average ($gcr > 0$)	maximum	$gcr > 0$ and $c_j \neq c$
group 1	22.8%	0.022	0.062	10.8%
group 2	48.8%	0.026	0.071	25.6%

Table 6: Guaranteed cost reduction due to applying a critical level policy.

In 22.8% (group 1) and 48.8% (group 2) of the cases the cost can be reduced by applying a critical level policy. It is also interesting to see how often the optimal critical levels are not equal. The last column of Table 6 indicates that in about half of the cases where the guaranteed cost reduction was strictly positive, the optimal critical levels were all equal. If the guaranteed cost reduction is strictly positive, the average value is about 0.024. However, using the simulated cost of the optimal critical level policy using the upper bound, we are able to calculate a better guaranteed reduction. It turned out that for the cases where the guaranteed cost reduction is strictly positive, the average simulated guaranteed cost reduction equals 0.035 (group 1) and 0.042 (group 2), and the maximum values were 0.08 respectively 0.09. For the special cases discussed above we present the results in detail in Table 7.

special case 1	S_1	S_2	S_3	c_1	c_2	c_3	S_0	cost
lower bound sp	4	4	1	0	0	0	11	39.23
upper bound clp	4	4	2	1	1	1	10	36.81
simulated clp								36.10
special case 2	S_1	S_2	S_3	c_1	c_2	c_3	S_0	cost
lower bound sp	1	3	2	0	0	0	9	44.98
upper bound clp	1	3	3	0	0	2	10	51.36
simulated clp								48.58

Table 7: Optimal policies and costs for special cases.

5 Summary and conclusions

In this paper we have analysed a 1-warehouse, N-retailer inventory system where demand occurs at all stockpoints. Hence, the warehouse faces both direct customer demand and replenishment orders from retailers. Since not satisfying direct customer demand from stock is usually more expensive than not satisfying a retailer replenishment order, it is desirable to set a higher service level for direct customer demand. Therefore, we have introduced a critical level policy at the warehouse; for every retailer we set a critical level such that a retailer replenishment order is satisfied from warehouse stock if and only if the stock level is above the retailer's critical level. This policy not only allows us to distinguish between direct customer demand and retailer orders, but also between different retailers.

We have derived an approximate lower bound and an upper bound on the total cost and presented a heuristic approach to determine (near) optimal policies. The numerical results indicate that the average difference between the bounds and the exact cost is about 2.5%. Approximating the exact cost by taking the average value of the lower and upper bound gives very good results; on average a deviation of only 0.5%. The average cost error of the heuristic approach, compared to the optimal policy determined using enumeration, was about 1.5%.

Applying a critical level policy at the warehouse, instead of using a simple policy where all demands are handled in the same way, leads in many cases to a cost reduction. Our numerical experiments have indicated that cost reductions of 5% and more are quite often possible.

References

[1] S. Axsäter. Simple solution procedures for a class of two–echelon inventory problems. *Operations Research*, 38:64–69, 1990.

[2] S. Axsäter. Continuous review policies for multi–level inventory systems with stochastic demand. In S.C. Graves, A.H.G. Rinnooy Kan, and P. Zipkin, editors, *Handbooks in OR*

& MS, vol. 4. Elseviers Science Publishers, North–Holland, 1993.

[3] R. Dekker, R.M. Hill, and M.J. Kleijn. On the $(S-1, S)$ lost sales inventory model with priority demand classes. Technical Report 9743/A, Econometric Institute, Erasmus University Rotterdam, The Netherlands, 1997.

[4] E.B. Diks, A.G. de Kok, and A.G. Lagodimos. Multi-echelon systems: A service measure perspective. *European Journal of Operational Research*, 95:241–263, 1996.

[5] A.Y. Ha. Inventory rationing in a make-to-stock production system with several demand classes and lost sales. *Management Science*, 43:1093–1103, 1997.

[6] J.A. Muckstadt and L.J. Thomas. Are multi-echelon inventory methods worth implementing in systems with low-demand rate items? *Management Science*, 26:483–494, 1980.

[7] C.C. Sherbrooke. METRIC: A multi-echelon technique for recoverable item control. *Operations Research*, 16:122–141, 1968.