

## AN INVENTORY MODEL FOR OPTIMIZATION OF A TWO-ECHELON PERIODIC-REVIEW SYSTEM

Denilson Ricardo de Lucena Nunes<sup>a</sup>; Antônio Fernando de Castro Vieira<sup>a</sup>

<sup>a</sup> Pontifical Catholic University of Rio de Janeiro (PUC-RJ), Rio de Janeiro, RJ, Brasil - Department of Industrial Engineering

### Abstract

We propose a single-item inventory model for a two-echelon periodic-review system made up of a warehouse and  $N$  retailers. The warehouse adopts an interval between reviews that is an integer multiple of the interval used in the retailers. Unlike in other models in the literature, the retailers carry out reviews at equal, synchronized intervals, and orders may be partially met by the warehouse. A methodology for breaking the system down into  $N + 1$  facilities that results in a univariate convex cost function is proposed. This allows a golden-section search algorithm to be developed that minimizes total system cost while meeting the retail service level. The effectiveness of the proposed model is confirmed in a system with three retailers and a warehouse in which the movement of one item with normal retail demand is simulated in different situations. The retail service level achieved with the model is satisfactory. However, the greater the differences between the demand variance at each retailer, the more the service level at those retailers with greater variance fall short of the target while the service level at retailers with smaller variance exceeds the target.

**Keywords:** inventory model, warehouse, service level, retail.

### INTRODUCTION

Multi-echelon inventory systems have been the subject of study for the past four decades. Interest in this type of system has arisen from, among other factors, the need to make the supply-chain network more efficient. To achieve this objective, it is essential to reduce inventory costs while maintaining an acceptable service level. It is against this background that this article describes a proposal for optimizing a two-echelon inventory model in a system like that shown in Figure 1.

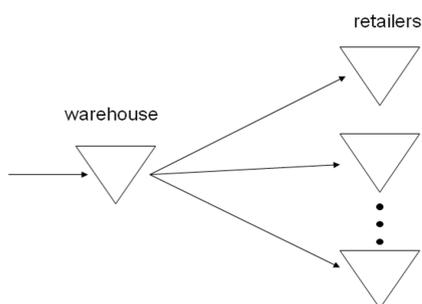


Figure 1 – Two-echelon system

Figure 1 represents the movement of an item in a system made up of a warehouse supplying  $N$  retailers, where all the stockpoints (warehouse and retailers) adopt a policy of periodic inventory reviews. The review interval is the same in all the retailers, while in the warehouse the review interval used is a whole multiple of the interval in the retailers. The replenishment policy adopted at the stockpoints involves placing an order at each review interval  $T$  to reach the order-up-to level  $S$ . The review intervals in the retailers are equal and synchronized. Consequently, orders placed by the retailers reach the warehouse at the same time.

The system is centralized, i.e., all the information about demand and inventory levels in the retailers is shared with the warehouse. This system differs from others described in the literature in that the review intervals for the retailers and warehouse differ, the orders arrive at the warehouse simultaneously and orders may be only partially met when there is insufficient inventory in the warehouse.

Heijden (1992) proposed a system similar to that described here but which used different review interval at retailers so that orders arrived at the warehouse at different times and were served on a FIFO (first-in-first-out) basis. Heijden (1992) estimated mean inventory, service level

and order delay, all as a function of pre-defined maximum inventory levels  $S$  at each stockpoint. These estimates were validated by simulation. In the model proposed by Heijden (1992), various values of  $S$  were used until the estimated variables reached acceptable levels. The solution is therefore based on trial and error, which does not guarantee an optimal result.

Although Heijden's (1992) approach was a general one, it is not applicable to the problem studied here: firstly, using Heijden's approach, orders can only be met in full, unlike what happens in practice, where partial deliveries are common; and secondly, Heijden assumes that all orders are met by the warehouse on a FIFO basis, which does not apply when orders arrive simultaneously.

Another similar system found in the literature was studied by Heijden (2000). In this case the stockpoints also used a periodic review, but the system could have multiple echelons, i.e., it constituted a supply chain. Heijden (2000) gives approximate equations that are a function of order-up-to-levels adopted at each stockpoint so that mean inventory levels and the retail service level can be estimated. The objective is to minimize system cost and guarantee the retail service level. To do this, Heijden (2000) gives a search algorithm. Despite the similarity with the system studied here, this solution is also not applicable here as for Heijden (2000) all the stockpoints in the network have the same review interval and orders are synchronized in all the stockpoints.

Recently, Chu *et al.* (2010) described an approximate solution for safety stock at all the stockpoints in a two-echelon periodic-review system. In their model, the ratio between review intervals at the warehouse and at retailers is limited to a power of two, orders are not synchronized and the service level for all retailers (and the warehouse) is defined in advance.

Advances in information technology have made possible the development of multiechelon inventory models that take into account the availability of real-time information [Cachon *et al.* (2000); Özer (2003); Axsäter *et al.* (2008); Bakal, *et al.* (2011); Howard *et al.* (2011)]. In line with this trend, Wang (2013) pointed a solution for the system described in this article that assumes that the warehouse allocates stock to retailers as a function of the real-time stock level at each retailer. Wang assumes that the retailers face a Poisson demand and describes an algorithm to jointly optimize the review interval and replenishment levels at all stockpoints. The algorithm requires considerable computational effort except when the average demand and number of retailers are small.

Hence, to our knowledge there is no other solution in the literature for the model proposed here. This is an important gap in multi-echelon inventory theory.

The aim of the model we describe is to define the order-up-to-level in all the stockpoints in accordance with a policy to be described so that total system inventory cost is minimized and the service level objective is met at the retailers. The service level used here is the fill rate (the fraction of orders met from inventory on hand). This article makes two main contributions: firstly, we develop a model and describe an optimization algorithm for this model; and secondly, we develop a methodology for determining effective lead time—defined as the sum of lead time plus mean order delay caused by shortages at the warehouse—so that each retailer can be treated as an isolated problem.

A simulation was performed to observe the service level for an item with normal demand in a system with three retailers using the solution to the proposed model as order-up-to-levels. The resulting service levels were close to the targets that had been set. However, as the differences in retail demand variances increase significantly, the service level at retailers with greater demand variance tends to fall below target.

The rest of this article is organized as follows: The section RESEARCH METHODOLOGY discusses the methodology used. The section DESCRIPTION OF THE MODEL describes the inventory policy and the overall formulation of the model. The section FORMULATION FOR NORMAL DEMAND gives the formulation for a system with normal demand in the retailers and warehouse. This section also addresses a number of issues related to the cost function and the methodology for optimizing this. The section NUMERICAL ILLUSTRATION gives a numerical example to illustrate how the model is applied. The section SIMULATION gives the results of the simulations to test the solution to the algorithm in three different situations. The conclusions are presented in the section CONCLUSIONS.

## RESEARCH METHODOLOGY

The aim of this article is to describe an inventory model for a two-echelon system (warehouse and retailers), a common type of inventory system in the retail sector. To this end, a literature review was carried out to find solutions for the proposed problem. As this failed to identify any solutions, a solution had to be developed based on the theoretical framework investigated.

The complexity of multi-echelon inventory systems lies in the direct influence that inventory decisions (replenishment parameters) in one echelon have on others, suggesting that the solution should be defined jointly for all the stockpoints in the system, which is very often impracticable. An alternative to reduce the complexity of these models is to define an element that decomposes the system into isolated stockpoints.

The principle of decomposition of the system was used here and involved the development of a methodology to determine the effective time between an order being placed with the warehouse and the time the order is received at the retailer. This interval was called the effective lead time and is directly affected by the order-up-to-level adopted at the warehouse. Once the expected effective lead time is determined, the system can be decomposed into isolated stockpoints.

To obtain an optimal solution for the model proposed here a golden section search algorithm was used. This was possible because the cost function for inventory in the system is convex and is solely a function of the order-up-to-level in the warehouse. The effectiveness of the proposed formulation was tested in a simulation with one item with normal demand in a two-echelon system consisting of a warehouse and three retailers.

## DESCRIPTION OF THE MODEL

In the system shown in Figure 1 the lead times at the warehouse and retailers ( $L_0$  and  $L_i$  respectively) are considered constant. The external supplier always has sufficient capacity to supply the warehouse, a condition

known as infinite capacity. Items can be stocked at all the stockpoints, and there are no constraints on inventory capacity or transportation. Demands in the retailers are random, identically distributed, independent variables. Retailer  $i$  must serve the end customer with a service level  $\beta_i$  (a management target), which in this case represents the fraction of demand promptly met from in-retailer inventory.

At the end of each review interval  $T_0$  the warehouse places an order with the external supplier at time  $t_0$  to reach the order-up-to level  $S_0$ . The interval  $T$  is the same and synchronized in all the retailers, and at each review (at time  $t$ ) retailer  $i$  places an order with the warehouse to reach its order-up-to level  $S_i$ . The orders placed with the warehouse by the retailers can be partially met and the items that could not be supplied are sent as soon as inventory is available. The holding cost in the system is charged at the end of each interval  $T$ .

Figure 2 illustrates the changes in inventory levels for an item in the warehouse and in a particular retailer  $i$ . In this case  $T_0 = 2T$ .

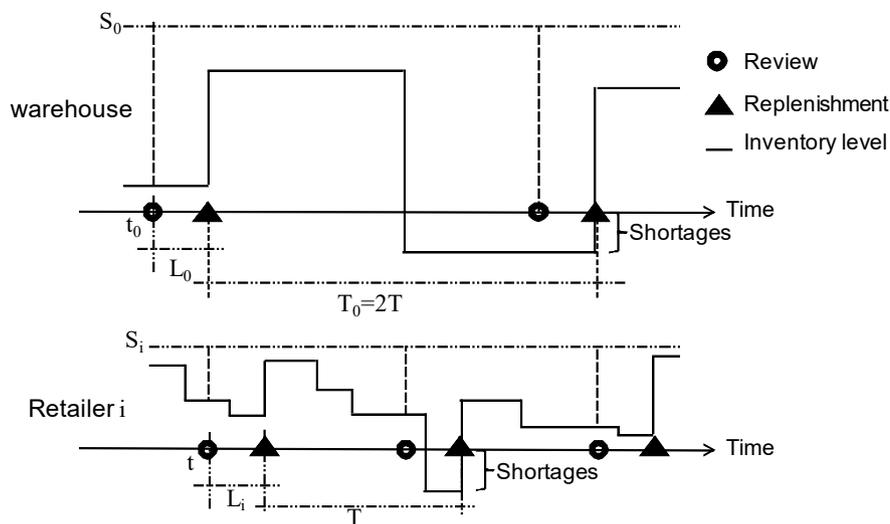


Figure 2 – Inventory level at the warehouse and at retailer  $i$  for  $T_0 = 2T$

### Basic Notation

$N$  - number of retailers;

$T$  - review interval at retailers. The same at all the retailers;

$m$  - total number of reviews in a warehouse replenishment cycle;

$T_0$  - Review interval at the warehouse, restricted to  $T_0 = mT$ , where  $m$  is a positive integer;

$\mu_i$  - mean demand at retailer  $i$  per unit time;

$\sigma_i^2$  - mean demand variance at retailer  $i$  per unit time;

$h_i, h_0$  - cost of holding an item in inventory per unit time at retailers and at the warehouse during the interval  $T$ , respectively;

#### Determining Mean Inventory in the warehouse and in each Retailer

The mean inventory level in the warehouse, as well as in each retailer  $i$ , is used to calculate the expected holding cost in each of these stockpoints. The mean desired inventory level is the inventory present in a replenishment cycle, which is defined as the interval between the time an order arrives and the instant immediately before the arrival of the next order. Inventory levels in retailer  $i$  at these two instants are, respectively:  $(S_i - D_i(L_i))$  and  $(S_i - D_i(L_i + T))$ . Because inventory level can be either positive (inventory available) or negative (inventory not available), only the positive part of these levels should be used to calculate the mean available inventory in retailer  $i$ , which is given by

$$\bar{I}_i = \left( E(S_i - D_i(L_i))^+ + E(S_i - D_i(L_i + T))^+ \right) / 2 \quad (1)$$

where  $E(S_i - D_i(L_i))^+$  and  $E(S_i - D_i(L_i + T))^+$  are the expected values of available inventory at the beginning and end of the retailer replenishment cycle.

Mean available inventory level at the warehouse is calculated in the same way as mean inventory level at retailers. However, it should be noted that a replenishment cycle  $T_0$  contains  $m$  cycles  $T$  and that because of the policy adopted, the inventory level at the warehouse only changes when there are reviews at the retailers, i.e., at intervals of  $T$  (see Figure 2). Consequently, the expected available inventory before the next replenishment at the

warehouse is given by  $E(S_0 - D_0(L_0 + (m-1)T))^+$ , which is constant until the next replenishment. The expected available inventory immediately after replenishment is given by  $E(S_0 - D_0(L_0))^+$ . Hence, mean inventory at the warehouse is given by

$L_i, L_0$  - lead times at retailer  $i$  and at the warehouse, respectively;

$S_i, S_0$  - order-up-to levels at retailer  $i$  and the warehouse, respectively;

$D_i(a)$  - demand at retailer  $i$  during an interval  $[0, a)$ ;

$D_0(a)$  - demand at the warehouse during an interval  $[0, a)$ ;

$(x)^+ = \max\{x, 0\}$ ;  $(x)^- = \max\{-x, 0\}$ .

$$\bar{I}_0 = \left( E(S_0 - D_0(L_0))^+ + E(S_0 - D_0(L_0 + (m-1)T))^+ \right) / 2 \quad (2)$$

#### Allocation of Shortages to Retailers

Shortages at the warehouse cause delays in deliveries to retailers, which in this case have to wait longer than  $L_i$  to receive the items they ordered. Hence, shortages at the warehouse allocated to retailer  $i$  are directly related to

the effective lead time  $\bar{l}_i$  (discussed in detail in the next section), making it necessary to know the expected number of shortages allocated to each retailer  $i$  at each review.

As previously mentioned, changes in inventory levels at the warehouse only occur when reviews are performed at retailers, so shortages at the warehouse are reflected in the negative part of the inventory level (inventory not available) on those occasions. Therefore, the expected number of shortages at the warehouse up to the  $j^{\text{th}}$  review at retailer inventory points is given by

$$E(S_0 - D_0(L_0 + jT))^- \quad j \in \{1, \dots, m-1\} \quad (3)$$

where  $j = 0$  corresponds to the first review at retailer stockpoints in a warehouse replenishment cycle and  $j = m - 1$  corresponds to the last review.

For a given interval between two reviews at retailer stockpoints, the expected number of shortages at the CD at the beginning and end of the interval are, respectively,  $E(S_0 - D_0(L_0 + jT))^-$  and

$E(S_0 - D_0(L_0 + (j-1)T))^-$ . The expected number of shortages at the  $j^{\text{th}}$  review at retailer stockpoints is the difference between these expected values. The only exception is for the first review at retailer stockpoints ( $j = 0$ ), which coincides with replenishment at the warehouse; in this case the expected number of shortages at this review is

given by  $E(S_0 - D_0(L_0))^-$ . Hence, the shortages at each review at retailer stockpoints are given by

$$B_{0j} = E(S_0 - D_0(L_0 + jT))^- - E(S_0 - D_0(L_0 + (j-1)T))^- \quad j \in \{1, \dots, m-1\}$$

$$B_{00} = E(S_0 - D_0(L_0))^- \quad \text{for } j=0$$

we know that for any non-negative value of  $a$ ,  $E(S_0 - D_0(a))^- = E(D_0(a) - S_0)^+$ . Therefore

$$B_{0j} = E(D_0(L_0 + jT) - S_0)^+ - E(D_0(L_0 + (j-1)T) - S_0)^+ \quad j \in \{1, \dots, m-1\}$$

$$B_{00} = E(D_0(L_0) - S_0)^+ \quad \text{for } j=0$$

(4)

Thus, the expected number of shortages allocated to retailer  $i$  at the  $j^{\text{th}}$  review is given by

$$B_{ij} = p_i B_{0j} \quad \text{for } i > 0$$

(5)

where  $p_i$  is the fraction of shortages allocated to retailer  $i$  when there is no inventory available at the warehouse. In such situations the number of items in each order that will be supplied must be defined (a procedure known as rationing). Rationing can also be interpreted as the act of allocating shortages at the warehouse to retailers, i.e., it corresponds to the fraction  $p_i$  ( $0 \leq p_i \leq 1$ ) of each order that will not be met at retailer  $i$ , such that  $\sum_{i=1}^N p_i = 1$ .

Determining  $p_i$  is very important for system performance as it affects not only the probability of shortages at retailer  $i$ , but also the level of service at the retailer. In addition, rationing may sometimes be performed in such a way that the number of shortages allocated to retailer  $i$  is greater than the order the retailer placed with the warehouse, a phenomenon known as imbalance, or negative inventory allocation.

Some of the better-known proposals for determining the fraction  $p_i$  include *Consistent Appropriate Share* (CAS) (De Kok, 1990), *Fair Share* (FS) (Eppen et Schrage, 1981) and *Balanced Rationing* (BR) (Heijden, 1997). These approaches differ in their objectives: in CAS the aim is to maintain the same probability of a shortage at all the retailers, while in FS the procedure prioritizes the service-level target and in BR negative allocation is minimized. Many other proposals based on these approaches have since appeared. Heijden et

al., (1997) proposed a solution for  $p_i$  that was independent of  $S_i$ , unlike other proposals, which determine  $p_i$  and  $S_i$  simultaneously. In their approach, only the variance in retailer demand influences rationing, so that:

$$p_i = \frac{1}{2N} + \frac{\sigma_i^2}{2 \sum_{j=1}^N \sigma_j^2}.$$

(6)

According to Heijden et al., (2000), the heuristic in Equation 6 was tested in a simulation together with CAS, FS and BR and had the best performance in terms of meeting the service level in retailers. Because of this, the fractions obtained using Equation 6 are used in this article for the rationing rule; in other words, they are used to allocated shortages at the warehouse to retailers.

### Effective Retail Lead Time and Decomposition of the System

In a two-echelon inventory system the movement of inventory at retailers is influenced by the order-up-to-levels used at the warehouse. The basic principle when modeling a two-echelon inventory system involves finding a way of breaking the system down into  $N$  stockpoints that can be treated separately. One way of doing this is to determine the effective lead time  $\bar{\ell}_i$ . The effective lead time is influenced by the order-up-to-level at the warehouse through the delay in deliveries caused by shortages at the warehouse. Hence, once  $\bar{\ell}_i$  has been determined, retailer  $i$  can be treated separately, as the influence of the order-up-to-level at the warehouse is being taken into account implicitly.

As shortages at the warehouse are a random variable, the effective lead time is also random and its expected value is calculated by adding lead time to mean order delay:

$$\bar{\ell}_i = L_i + \bar{w}_i,$$

(7)

where  $\bar{w}_i$  is the expected order delay at retailer  $i$ .

Because of the supply policy adopted, replenishments occur at fixed points in time. It follows that order delay is a discrete random variable.

Hence, if there are shortages in the first review at retailer  $i$  at time  $t$  (see Figure 2), the items that are out of inventory will take a total time  $L_i + 2T$  to be delivered, i.e., they

will arrive with a delay of  $2T$ . For shortages in the second review ( $t + T$ ), this time is reduced to  $L_i + T$ , i.e., a delay of  $T$ . It can be concluded that the possible delay values are given by  $(m - j)T$ , where  $j \in [0, 1, \dots, m - 1]$ , which represents the delay for the items that are out of inventory in review  $j$  at retailer  $i$ .

Delays only occur when there are shortages at the warehouse. When this happens, of all the items requested from the warehouse in a cycle  $T_0$  only  $B_{ij}$  will be out of inventory at retailer  $i$  in review  $j$ . Hence, the fraction of demand that cannot be met at the  $j^{\text{th}}$  review is given by the fraction  $B_{ij} / \mu_i m T$ , where  $B_{ij} = p_i B_{0j}$  as in Equation 5 and  $\mu_i$  is the expected retail demand per unit time.

The expected delay  $\bar{w}_i$  is therefore given by

$$\bar{w}_i = \sum_{j=0}^{m-1} [(m - j)T] B_{ij} / \mu_i m T \quad (8)$$

Once  $\bar{\ell}_i$  is known (Equation 7), the system can be treated as  $N + 1$  independent systems by using  $\bar{\ell}_i$  instead of  $L_i$ . The mean inventory at retailer  $i$  calculated in Equation 1 can then be rewritten as

$$\bar{I}_i = (E(S_i - D_i(\bar{\ell}_i))^+ + E(S_i - D_i(\bar{\ell}_i + T))^+) / 2.$$

We know that for any non-negative value  $a$ ,  $E(S_i - D_i(a))^+ = E(D_i(a) - S_i)^-$ . Therefore

$$\bar{I}_i = (E(D_i(\bar{\ell}_i) - S_i)^- + E(D_i(\bar{\ell}_i + T) - S_i)^-) / 2,$$

Furthermore,

$$E(D_i(a) - S_i)^- = E(D_i(a) - S_i)^+ - E(D_i(a) - S_i),$$

which when substituted in the previous expression becomes

$$\bar{I}_i = (E(D_i(\bar{\ell}_i) - S_i)^+ - E(D_i(\bar{\ell}_i) - S_i) + E(D_i(\bar{\ell}_i + T) - S_i)^+ - E(D_i(\bar{\ell}_i + T) - S_i)) / 2$$

taking into account the result  $E(D_i(a) - S_i) = \mu_i a - S_i$  in the previous expression we have

$$\bar{I}_i = (E(D_i(\bar{\ell}_i) - S_i)^+ + E(D_i(\bar{\ell}_i + T) - S_i)^+ + 2S_i - \mu_i(2\bar{\ell}_i + T)) / 2 \quad (9)$$

Likewise, the same approach can be used to rewrite the mean inventory at the warehouse (Equation 2), giving

$$\bar{I}_0 = (E(D_0(L_0) - S_0)^+ + E(D_0(L_0 + (m - 1)T) - S_0)^+ + 2S_0 - \mu_0(2L_0 + (m - 1)T)) / 2 \quad (10)$$

The retail service level, here represented by  $\beta_i$ , is the fraction of demand that is promptly met (fill rate).

$\beta_i$  should also take into account the effective lead time  $\bar{\ell}_i$ . To calculate the fill rate, the shortages in a retail replenishment cycle must be known, which in this case correspond to the difference between the expected number of shortages at the beginning and end of the cycle, i.e.,

$$E(S_i - D_i(\bar{\ell}_i))^- \text{ and } E(S_i - D_i(\bar{\ell}_i + T))^- , \text{ respectively.}$$

The expected demand in a cycle is given by  $\mu_i T$ . Hence, the fraction of demand that is not met is given by  $(E(S_i - D_i(\bar{\ell}_i + T))^- - E(S_i - D_i(\bar{\ell}_i))^-) / \mu_i T$  and the fraction that is promptly met is given by

$$\beta_i = 1 - (E(S_i - D_i(\bar{\ell}_i + T))^- - E(S_i - D_i(\bar{\ell}_i))^-) / \mu_i T$$

which, as  $E(S_i - D_i(a))^- = E(D_i(a) - S_i)^+$ , can be rewritten

$$\beta_i = 1 - (E(D_i(\bar{\ell}_i + T) - S_i)^+ - E(D_i(\bar{\ell}_i) - S_i)^+) / \mu_i T \quad (11)$$

## FORMULATION FOR NORMAL DEMAND

A normal distribution is a common choice for modeling demands, particularly when the demand is high. Demanding for an item is expressed per interval of time (e.g., per week) and is in turn made up of demands expressed per another, smaller time (e.g., days). According to the central limit theorem, the sum of various independent random variables has an approximately normal distribution, justifying the use of this distribution in many situations.

According to Axsäter (2006), for demand to approximate to a normal distribution, the coefficient of variation

$CV = \sigma / \mu$  must be reasonably less than one, which reduces the probability of the distribution containing negative demand values. Silver *et al.*, (1998) recommend

that a normal distribution should be used when  $cv \leq 0.5$ .

### Determining Inventory Level for Normal Demand

Let  $X$  be a random variable with mean  $\mu_x$ , standard deviation  $\sigma_x$  and density function  $f(x)$ . Zipkin (2000) defines the first-order loss for a particular value  $S$  to be

$E(X - S)^+ = \int_S^\infty (y - S) f(y) dy = F^1(S)$  According to Zipkin (2000) the loss  $F^1(S)$  for a standardized normal distribution can be expressed as

$$F^1(S) = \sigma_x \Phi^1(z), \quad (12)$$

where  $\Phi^1(z)$  is given by

$$\Phi^1(z) = \phi(z) - z[1 - \Phi(z)], \quad (13)$$

and  $\Phi^1(z)$  is the first-order loss function,  $\phi(z)$  the density function and  $\Phi(z)$  the cumulative distribution function of the standardized normal distribution, all defined in  $z = (S - \mu_x) / \sigma_x$ .

Hence, we can rewrite Equations, 4, 9, 10 and 11 taking into consideration the loss function given by  $E(D(a) - S)^+ = \sigma \sqrt{a} \Phi^1(z)$  for any non-negative value of  $a$ , where  $z = (S - \mu a) / \sigma \sqrt{a}$ .

Thus, Equation 4 becomes

$$B_{0j} = \Phi^1(z) \sigma_0 \sqrt{L_0 + jT} - \Phi^1(z) \sigma_0 \sqrt{L_0 + (j-1)T} \quad j \in \{1, \dots, m-1\}$$

$$B_{00} = \Phi^1(z) \sigma_0 \sqrt{L_0} \quad \text{for } j = 0$$

or

$$B_{0j} = \Phi^1(z) \sigma_0 \left( \sqrt{L_0 + jT} - \sqrt{L_0 + (j-1)T} \right) \quad j \in \{1, \dots, m-1\}$$

$$B_{00} = \Phi^1(z) \sigma_0 \sqrt{L_0} \quad \text{for } j = 0 \quad (14)$$

Equation 9 becomes

$$\bar{I}_i = \left( \sigma_i \sqrt{\bar{\ell}_i} \Phi^1(z) + \sigma_i \sqrt{\bar{\ell}_i + T} \Phi^1(z) + 2S_i - \mu_i (2\bar{\ell}_i + T) \right) / 2$$

or

$$\bar{I}_i = \left( \Phi^1(z) \sigma_i \left( \sqrt{\bar{\ell}_i} + \sqrt{\bar{\ell}_i + T} \right) + 2S_i - \mu_i (2\bar{\ell}_i + T) \right) / 2 \quad (15)$$

Equation 10 becomes

$$\bar{I}_0 = \left( \sigma_0 \sqrt{L_0} \Phi^1(z) + \sigma_0 \sqrt{L_0 + (m-1)T} \Phi^1(z) + 2S_0 - \mu_0 (2L_0 + (m-1)T) \right) / 2$$

or

$$\bar{I}_0 = \left( \Phi^1(z) \sigma_0 \left( \sqrt{L_0} + \sqrt{L_0 + (m-1)T} \right) + 2S_0 - \mu_0 (2L_0 + (m-1)T) \right) / 2 \quad (16)$$

Equation 11 becomes

$$\beta_i = 1 - \left( \Phi^1(z) \sigma_i \sqrt{\bar{\ell}_i + T} - \Phi^1(z) \sigma_i \sqrt{\bar{\ell}_i} \right) / \mu_i T$$

or

$$\beta_i = 1 - \left( \Phi^1(z) \sigma_i \left( \sqrt{\bar{\ell}_i + T} - \sqrt{\bar{\ell}_i} \right) \right) / \mu_i T. \quad (17)$$

In the previous expressions,  $\mu_i$  and  $\sigma_i$  represent the mean and standard deviation of the demand at retailer  $i$  per unit time. Similarly,  $\mu_0$  and  $\sigma_0$  are the mean and standard deviation of the demand at the warehouse, so that

$$\mu_0 = \sum_{i=1}^N \mu_i \quad \text{and} \quad \sigma_0^2 = \sum_{i=1}^N \sigma_i^2.$$

### Cost Function and Optimization

The only cost considered in the cost function is the cost of holding inventory, which is charged at the end of each retail replenishment cycle, and the only other constraint is the retail service level, which must reach a defined target

$\beta_i$ . Let  $h_1$  and  $h_0$  be the cost of holding an item in inventory at a retailer and the warehouse, respectively, for the period  $T$ . The total cost of holding inventory in the system during

this period is given by  $c = h_0 \bar{I}_0 + \sum_{i=1}^N h_i \bar{I}_i$ , where  $\bar{I}_i$  and  $\bar{I}_0$  are given by Equations 15 and 16.

We wish then to determine the order-up-to levels  $S_i$  and  $S_0$  that meet the target retail service level at the lowest system cost. This constitutes a minimization problem. ,

$$(18)$$

$$\begin{aligned} \min \quad & c = h_0 \bar{I}_0 + \sum_{i=1}^N h_i \bar{I}_i \\ \text{subject to} \quad & \\ & 1 - \left( \Phi^1(z) \sigma_i \left( \sqrt{\bar{\ell}_i + T} - \sqrt{\bar{\ell}_i} \right) \right) / \mu_i T = \beta_i \quad i \in \{1, 2, \dots, N\} \end{aligned} \quad (18)$$

The cost  $c$  is a linear combination of mean available inventory  $I_0$  and  $I_i$ . These available inventories are obtained from the expected available inventory values, which in this case are first-order loss functions of the normalized distribution. According to Axsäter (2006), loss functions are continuous, decreasing and convex. Hence, the objective function of Problem 18 is a linear combination of convex functions, which according to Bazaraa *et al.*, (2006) results in a convex function.

The expression used to calculate the service level in Equation 17 is also convex as it uses the expected number of shortages obtained from the loss function. Thus for each pair  $(\bar{\ell}_i, \beta_i)$  there is only one  $S_i$  that satisfies the constraint on Problem 18. This value of  $S_i$  can be found by using a search method with Equation 17, an example being the secant method.

In addition, the effective lead time  $\bar{\ell}_i$ , which influences  $S_i$ , is related to  $S_0$  through the shortages  $B_{0j}$ . We conclude, therefore, that for every  $S_0$  there is only one  $S_i$  that satisfies the constraint on the problem. Hence, the cost  $c$  is a function of only  $S_0$ , as the values of  $S_i$  are a consequence of the constraint on the problem.

The cost  $c$  is therefore convex and has only one argument. Although there is a number of constraints on Problem 18, the characteristics of the cost function allowed an optimization algorithm based on the golden section method (first developed for problems without any constraints) to be used.

### Optimization Algorithm

The algorithm developed here is based on the golden section algorithm (Bazaraa *et al.*, 2006). In this method a search interval is defined and convergence to the optimum occurs when the difference between two points being tested is less than a stopping criterion  $\delta$ . In practice  $S_0$  is a discrete variable, so it is reasonable that the smallest interval

tested should have a minimum length equal to a unit, i.e.,

$$\delta = S_{0(1)}^k - S_{0(2)}^k < 1 \text{ where } k \text{ is the iteration number and } S_{0(1)}^k \text{ and } S_{0(2)}^k \text{ the points tested in each iteration.}$$

In addition, the golden section method is started from the extreme bounds of the search for  $S_0$ , which are described in Appendix A and given below:

$$\mu_0(L_0 - T) < S_0 < 5\sqrt{L_0 + (m-1)T}\sigma_0 + \mu_0[L_0 + (m-1)T] \quad (19)$$

Following is the algorithm based on the golden section method:

Data:  $\mu_i, \sigma_i, L_i, T, h_i, \beta_i, \mu_0, \sigma_0, L_0, T_0, h_0, m$ .

Do

$$A^1 = \mu_0(L_0 - T);$$

$$B^1 = 5\sqrt{L_0 + (m-1)T}\sigma_0 + \mu_0[L_0 + (m-1)T];$$

$$\delta = A^1 - B^1; k = 1 \text{ (iteration number);}$$

Calculate the initial test points  $S_{0(1)}^k$  and  $S_{0(2)}^k$  for  $k = 1$

$$S_{0(1)}^1 = A^1 + (1 - 0.618)(B^1 - A^1) \text{ and } S_{0(2)}^1 = A^1 + 0.618(B^1 - A^1)$$

While  $\delta > 1$  do:

For each point  $\{S_{0(1)}^k, S_{0(2)}^k\}$  do:

Calculate the mean inventory at the warehouse  $I_{0(1)}^k$  and

$I_{0(2)}^k$  as in Equation 16

For each review  $j \in \{0, 1, \dots, m-1\}$  do:

Calculate the shortages at the warehouse  $B_{0j(1)}^k$  and

$B_{0j(2)}^k$  as in Equations 14

end

for each retailer  $i \in \{1, 2, \dots, N\}$  do:

Determine the effective lead times  $\bar{\ell}_{i(1)}^k$  and  $\bar{\ell}_{i(2)}^k$  as in Equation 7

With a search method, define  $S_{i(1)}^k$  and  $S_{i(2)}^k$  for  $\beta_i$  in Equation 17

Calculate the mean inventory levels  $I_{i(1)}^k$  and  $I_{i(2)}^k$  as in Equation 15

end

Calculate the costs  $c_{(1)}^k$  and  $c_{(2)}^k$  as in Problem 18

If ( $c_{(1)}^k > c_{(2)}^k$ ) then

$$A^{k+1} = S_{0(1)}^k; \quad B^{k+1} = B^k; \quad S_{0(1)}^{k+1} = S_{0(2)}^k;$$

$$S_{0(2)}^{k+1} = A^{k+1} + 0,618(B^{k+1} - A^{k+1});$$

$$c \text{ min} = c_{(2)}^k, \quad S_0^* = S_{0(2)}^k \text{ and } S_i^* = S_{i(2)}^k.$$

Or if ( $c_{(1)}^k \leq c_{(2)}^k$ ) then

$$A^{k+1} = A^k; \quad B^{k+1} = S_{0(2)}^k; \quad S_{0(2)}^{k+1} = S_{0(1)}^k;$$

$$S_{0(1)}^{k+1} = A^{k+1} + (1 - 0,618)(B^{k+1} - A^{k+1});$$

$$c \text{ min} = c_{(1)}^k, \quad S_0^* = S_{0(1)}^k \text{ and } S_i^* = S_{i(1)}^k.$$

end

$$\delta = A^k - B^k; \quad k = k + 1.$$

end

### NUMERICAL ILLUSTRATION

In this example the aim is to determine the order-up-to levels for a particular product in a system with three retailers ( $N = 3$ ) that carry out inventory reviews every day ( $T = 1$ ). The warehouse checks inventory every three

days ( $T_0 = 3$ ) so that orders can be placed with the external supplier. The variance and mean for daily demand in each

retailer are  $\mu_1 = 27; \sigma_1^2 = 23 \quad \mu_2 = 81; \sigma_2^2 = 39$

$\mu_3 = 54; \sigma_3^2 = 31$ . Other data for the problem:

$L_0 = 1; h_0 = 1; L_i = 1; h_i = 4; \beta_i = 0.9$ .

In the proposed algorithm, a new set of values obtained from the points  $S_{0(1)}^k$  and  $S_{0(2)}^k$  is generated at each iteration  $k$ . The values obtained in some of the iterations are shown in Table 1, where the reduction in cost  $c^*$  with

each iteration can be observed. The order-up-to level  $S_0^*$  also decreases, causing the shortages at the warehouse to increase, which in turn leads to larger effective lead times  $\bar{\ell}_i$ . Consequently,  $S_1^*$ ,  $S_2^*$  and  $S_3^*$  increase to meet the target service level  $\beta_i = 0.9$ .

At the end of the iterations a solution is found with a minimum cost  $c^* = R\$ 329.79$  for order-up-to levels  $S_0^* = 153, S_1^* = 106, S_2^* = 220$  and  $S_3^* = 162$ .

Table 1- Results of the iterations using the proposed algorithm

	$A$	$B$	$S_0^*$	$S_1^*$	$S_2^*$	$S_3^*$	$c^*$
lt. 1	1,335.84	2,161.12	1,335.84	54.01	154.96	104.20	1,324.96
lt. 4	315.29	510.08	315.29	72.43	177.97	124.75	395.64
lt. 7	74.41	120.39	120.39	115.84	232.38	173.33	332.25

Figure 3 shows the graphs of variables  $S_1, S_2, S_3$  and  $c$  for a variation  $S_0[1 \quad 729]$  using the data from the example described here. As expected, the increase in  $S_0$  causes a reduction in order-up-to levels  $S_i$  as a consequence of the reduction in effective lead time. However, the reduction

in  $\bar{\ell}_i$  is limited to the lead time  $L_i$  itself, and from this point on  $S_i$  remains constant even when  $S_0$  continues to increase. The convex behavior of the cost function can also be observed.

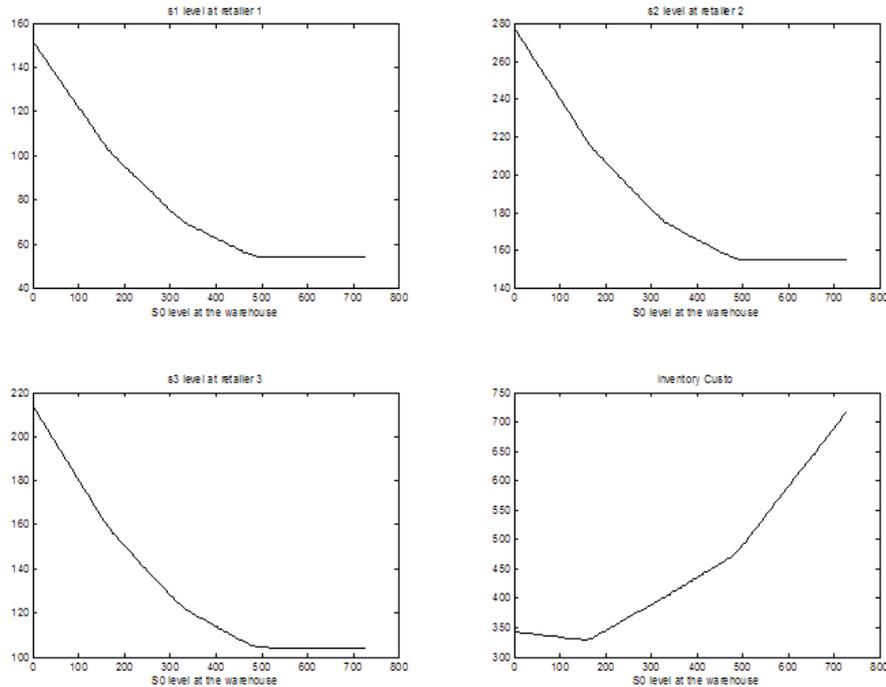


Figure 3 – Values of  $S_1, S_2, S_3$  and  $C$  for  $S_0 \in [1 \quad 729]$ .

**SIMULATION**

A simulation was performed to observe how the values  $S_i^*$  and  $S_0^*$  supplied by the proposed algorithm influence the behavior of the fill rate in relation to the defined target  $\beta_i$ . The MATLAB standard number generator was used for the simulation.

Three cases with different demands in the system described above were simulated. In each case 10,000 daily demands were generated, and the change in inventory levels in the system was a consequence of the order-up-to levels  $S_i^*$  and  $S_0^*$  produced by the proposed algorithm. Lastly, the service level was observed at each retailer. The results are shown in Table 2.

As can be seen from this table, the solution proposed by the algorithm produces a service level (fill rate) close to the defined target, particularly when the variances of the retail demands are similar. However, as the variances differ, the target is no longer reached (values in bold in Table 2). Because of the rationing rule in Equation 6, the retailers with greater demand variance have more shortages allocated by the warehouse, with a consequent reduction in fill rate compared with the defined target. For those stockpoints with smaller demand variance, the opposite happens and the service level is exceeded.

Table 2 – Results of the simulation

Retail demand		Order-up-to levels				Target	Observed $\beta$ (%)		
Case 1		$S_0^*$	$S_1^*$	$S_2^*$	$S_3^*$	$\beta$	$\beta_1$	$\beta_2$	$\beta_3$
$\mu_1 = 27$	$\sigma_1^2 = 22$	75	79	80	80	80%	85.13	85.38	86.14
$\mu_1 = 27$	$\sigma_1^2 = 23$	78	83	84	84	90%	91.24	91.56	91.97
$\mu_1 = 27$	$\sigma_1^2 = 24$	80	86	87	88	95%	94.97	95.30	95.94
		84	94	94	94	99%	98.83	98.94	99.01

Case 2		$S_0^*$	$S_1^*$	$S_2^*$	$S_3^*$	$\beta$	$\beta_1$	$\beta_2$	$\beta_3$
$\mu_1 = 81$	$\sigma_1^2 = 39$	185	223	167	194	80%	80.25	88.59	82.86
$\mu_2 = 54$	$\sigma_2^2 = 31$	190	231	173	201	90%	<b>85.33</b>	96.52	89.68
$\mu_3 = 67$	$\sigma_3^2 = 35$	193	237	177	206	95%	<b>89.37</b>	98.44	94.17
		196	247	186	215	99%	<b>94.18</b>	99.83	98.36
Case 3		$S_0^*$	$S_1^*$	$S_2^*$	$S_3^*$	$\beta$	$\beta_1$	$\beta_2$	$\beta_3$
$\mu_1 = 27$	$\sigma_1^2 = 23$	149	102	212	156	80%	99.32	78.21	83.68
$\mu_2 = 81$	$\sigma_2^2 = 39$	153	106	220	162	90%	99.84	<b>81.37</b>	<b>88.28</b>
$\mu_3 = 54$	$\sigma_3^2 = 31$	156	110	226	167	95%	99.97	<b>84.58</b>	<b>92.13</b>
		160	116	235	175	99%	99.99	<b>88.74</b>	<b>97.40</b>

## CONCLUSIONS

The simulation carried out here showed the effectiveness of the proposed model, which is an alternative solution for a common retail situation. Unlike other solutions in the literature, the model proposed here assumes that the review intervals at retailers are equal and synchronized and that retail orders can be partially fulfilled by the warehouse. The formulation proposed in this article assumes that the demand distributions in the retailers and warehouse are known. While a normal distribution was used here, the formulation can be used with other distributions, such as a standardized gamma distribution or a Poisson distribution, by writing the expected value  $E(D(a) - S)^+$  as was done for the normal distribution.

The cost function formulated for the system studied here is convex and for each order-up-to level used at the warehouse there is only one corresponding order-up-to level at each retailer that satisfies the service level target. Hence, in addition to being convex, the cost function has a single argument. These two characteristics allow a golden section search algorithm to be used to minimize the total system cost.

The effective lead time obtained using the proposed methodology allows the system being studied to be decomposed into  $N$  single-retailer problems by substituting the lead time by the effective lead time, thereby making calculation of  $S_i$  an isolated problem for each retailer.

For the cases simulated here, the results in Table 2 show that when the solution produced by the proposed algorithm is used, the service level is close to the target value and in some cases exceeds it. However, as the differences in

retail demand variances increase significantly, the service level at stockpoints with greater demand variance tends to fall below target. In contrast, in stockpoints with smaller demand variance the targets are exceeded.

## REFERENCES

- Axsäter, S. (2006) Inventory Control. 2th ed. New York: Springer.
- Bazaraa, M.S. (2006) Sherali, H. D.; Shetty, C. M. Nonlinear Programming: Theory and Algorithms. 3th ed. New Jersey: Wiley-Interscience.
- Bakal, I.S.; Erkip, N.; Güllü, R. (2011) Value of supplier's capacity information in a two-echelon supply chain. Ann Oper Res, vol. 191, pp.115–135.
- Cachon, G.P.; Fisher M. (2000) Supply chain inventory management and the value of shared Information. Management Science, vol. 46, 8, pp. 1032–1048.
- Chu, L.Y.; Shen, Z.M. (2010) A Power-of-Two ordering policy for one-warehouse multiretailer systems with stochastic demand. Operations Research, vol. 58, 2, pp. 492–502.
- De Kok, A.G. (1990) Hierarchical production planning for consumer goods. European Journal of Operational Research, vol.45, pp. 55-69.
- Eppen, G.; Schrage, L. (1981) Centralized ordering policies in a multi-warehouse system with lead times and random demand. TIMS Studies in the Management Sciences, vol. 16, pp. 51-67.
- Heijden, M.C. (1992) Analysing divergent logistic networks with local (R, S) inventory control. International Journal of Production Economics, vol. 27, pp. 187-219.



Heijden, M.C.; Diks, E.B.; De Kok, A.G. (1997) Inventory allocation in general multi-echelon distribution systems with (R, S) order-up-to-policies. *International Journal of Production Economics*. vol. 49, pp.157-174.

Heijden, M.C. (1997) Supply rationing in multi-echelon divergent systems. *European Journal of Operational Research*. vol. 101, pp. 532-549.

Heijden, M.C. (2000) Near cost-optimal inventory control policies for divergent networks under fill rate constraints. *Int. J. Production Economics*. vol. 63, pp. 161-179.

Hpward, C.; Marklund, J. (2011) Evaluation of stock allocation policies in a divergent inventory system with shipment consolidation. *European Journal of Operational Research*. Vol. 211, pp. 298–309.

Özer, Ö. (2003) Replenishment strategies for distribution systems under advance demand information. *Management Science*, vol. 49, 3, pp. 255–272.

Silver, E. A.; Pyke, D. F.; Peterson, R. (1998) *Decision Systems for Inventory Management and Production Planning and Scheduling*. 3th ed. New Jersey: John Wiley.

Wang Q. (2013) A periodic-review inventory control policy for a two-level supply chain with multiple retailers and stochastic demand. *European Journal of Operational Research*. vol. 230, pp. 53–62.

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

## APPENDIX A

As was seen in the optimization algorithm, an interval is defined for the variable  $S_0$ . To do this the upper and lower bounds for  $S_0$  must be found. The metric chosen for this purpose was the fraction of demand not met by the warehouse in the interval  $T_0 = mT$ .

The expected total number of shortages at the warehouse (the unmet demand) in an interval  $mT$  is given by  $E(D_0(L_0 + (m-1)T) - S_0)^+$ , where the expected demand at the warehouse in the same interval is  $\mu_0 mT$ . The fraction of demand not promptly met by the warehouse is therefore given by:

$$f = \frac{E(D_0(L_0 + (m-1)T) - S_0)^+}{\mu_0 mT}.$$

Clearly,  $0 \leq f \leq 1$ , and since  $E(D_0(a) - S_0)^+ = \sigma_0 \sqrt{a} \Phi^1(z)$  we have:

$$0 \leq \frac{\sqrt{L_0 + (m-1)T} \sigma_0 \Phi^1 \left\{ \left[ S_0 - \mu_0 (L_0 + (m-1)T) \right] / \sqrt{L_0 + (m-1)T} \sigma_0 \right\}}{\mu_0 mT} \leq 1 \quad (\text{A.1})$$

Considering only the upper bound in (A.1) gives:

$$\frac{\sqrt{L_0 + (m-1)T} \sigma_0 \Phi^1 \left\{ \left[ S_0 - \mu_0 (L_0 + (m-1)T) \right] / \sqrt{L_0 + (m-1)T} \sigma_0 \right\}}{\mu_0 mT} \leq 1,$$

or:

$$\Phi^1 \left\{ \left[ S_0 - \mu_0 (L_0 + (m-1)T) \right] / \sqrt{L_0 + (m-1)T} \sigma_0 \right\} \leq \frac{\mu_0 mT}{\sqrt{L_0 + (m-1)T} \sigma_0} \quad (\text{A.2})$$

In Equation B.2,  $\mu_0 mT$  and  $\sqrt{L_0 + (m-1)T} \sigma_0$  are the mean and standard deviation of the demand at the warehouse in an interval  $mT$ . This demand has a normal distribution, and according to Axsäter (2006), for demand to approximate to a normal distribution the quotient  $\sigma/\mu$  must be less than 1. Silver, *et al.*, (1998) recommend that  $\sigma/\mu < 0.5$ , i.e.,  $\mu/\sigma > 2$ . Hence, the right-hand side of Inequality A.2 is always greater than or equal to 2. For  $z \geq 2$ , the approximation  $\Phi^1(-z) = z$  can be used. Hence, for Inequality A.2 to be satisfied:

$$\left[ S_0 - \mu_0 (L_0 + (m-1)T) \right] / \sqrt{L_0 + (m-1)T} \sigma_0 \geq - \frac{\mu_0 mT}{\sqrt{L_0 + (m-1)T} \sigma_0}$$

After some algebraic manipulation we get:

$$S_0 \geq \mu_0 (L_0 - T) \quad (\text{A.3})$$



Considering the lower bound in Expression A.1 we get:

$$0 \leq \frac{\sqrt{L_0 + (m-1)T} \sigma_0 \Phi^1 \left\{ \left[ S_0 - \mu_0 (L_0 + (m-1)T) \right] / \sqrt{L_0 + (m-1)T} \sigma_0 \right\}}{\mu_0 m T}$$

or:

$$\Phi^1 \left\{ \left[ S_0 - \mu_0 (L_0 + (m-1)T) \right] / \sqrt{L_0 + (m-1)T} \sigma_0 \right\} \geq 0. \quad (\text{A.4})$$

For  $z > 5$ , the approximation  $\Phi^1(z) = 0$  can be used, giving:

$$\left[ S_0 - \mu_0 (L_0 + (m-1)T) \right] / \sqrt{L_0 + (m-1)T} \sigma_0 \leq 5$$

$$\text{or: } S_0 \leq 5 \sqrt{L_0 + (m-1)T} \sigma_0 + \mu_0 [L_0 + (m-1)T] \quad (\text{A.5})$$