

Article

A Two-Echelon Inventory System with a Minimum Order Quantity Requirement

Huaxiao Shen ¹, Tian Tian ² and Han Zhu ^{2,*} 

¹ Business School, Sun Yat-sen University, Guangzhou 510275, China; shenhx3@mail.sysu.edu.cn

² School of Management Science and Engineering, Dongbei University of Finance and Economics, Dalian 116025, China; tiantian@dufe.edu.cn

* Correspondence: hanzhu@dufe.edu.cn

Received: 17 July 2019; Accepted: 11 September 2019; Published: 16 September 2019



Abstract: In this paper, we study a two-echelon inventory system with one warehouse and multiple retailers, under the setting of periodic review and infinite horizon. In each period, retailers replenish their stocks from the warehouse, and the warehouse in turn replenishes from an external supplier. Particularly, as stipulated by the supplier, there is a minimum order quantity (MOQ) requirement for the warehouse. That is, the warehouse must order either none or at least as much as the MOQ. To investigate this system analytically, we assume retailers adopt the base-stock policy, and we design for the warehouse a new heuristic ordering policy, called refined base-stock policy, which conforms to the MOQ requirement. Moreover, in the case of shortages, we assume the warehouse adopts a virtual allocation policy, and therefore the orders for individual units are filled in the same order as the original demands at the retailers. To evaluate the long-run average system cost exactly, we present a position-based cost-accounting scheme, in which the cost associated with each unit is assigned to its first position at the warehouse. We also derive lower and upper bounds of the inventory parameters, facilitating the search for the optimal policy that minimizes the long-run average system cost.

Keywords: inventory management; two-echelon inventory system; warehouse; minimum order quantity; analytical model; cost optimization

1. Introduction

We study a two-echelon inventory system with one warehouse and multiple retailers (OWMR) in this paper. As shown in Figure 1, the OWMR model consists of one warehouse and N retailers. Inventory is reviewed periodically, and the horizon is infinite. In each period, retailers replenish their stocks from the warehouse, and the warehouse in turn replenishes from an external supplier. The supply chain is operated under an installation-stock based control. Linear purchasing cost is considered, but the warehouse must order either none or at least as much as a *minimum order quantity* (MOQ).

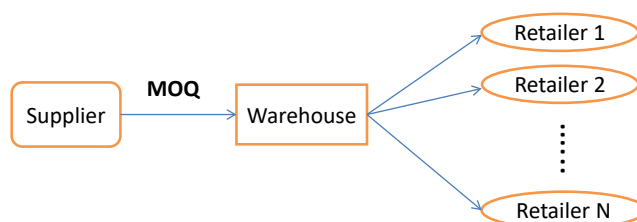


Figure 1. A two-echelon inventory system.

The MOQ requirement is enforced by the external supplier (manufacturing facility) to achieve economies of scale, and it has been widely applied by the firms manufacturing apparel, consumer packaged goods, and chemical products. The application of a MOQ is common in practice. With the prevalence of e-commerce, MOQs have been used more and more widely, especially in online business-to-business (B2B) sourcing portals such as alibaba.com, where suppliers often impose such requirements. MOQs are also applied in manufacturing industries for products that have short lifetimes or long leadtimes. A well-known example is Sport Obermeyer, a fashion sport ski-wear manufacturer, which has a minimum production level of 600 garments in Hong Kong and 1200 garments in China per order (Zhao and Katehakis [1]). In fact, the MOQ requirement is quite common in China and other low cost manufacturing countries, where low profit margins force manufacturers to pursue large production quantities to break even.

Retailers face independent Poisson demand processes, and they do not need to consider the MOQ requirement. Unfulfilled demands at each retailer are backlogged, incurring a backlogging cost (also called penalty cost). At the replenishment occasion of each period, the orders of all retailers are sent together to the warehouse. Orders that cannot be satisfied at the warehouse are also backlogged, but they will not incur any backlogging cost. At the end of each period, holding costs are incurred for the inventory left at both the warehouse and retailers. The objective is to minimize the long-run average system cost over an infinite horizon, achieving a sustainable inventory system. Facing such a complicated system, natural questions regard how to effectively manage this complicated inventory system: When should each installation place an order? If an order is placed, what is the optimal order quantity?

Multi-echelon inventory systems are more complex than single-echelon inventory systems. Single-echelon inventory systems only focus on the inventory position and the cost incurred at a single installation. Multi-echelon inventory systems, however, need to coordinate all installations. For example, in a serial two-echelon system, one upstream installation with a large inventory will incur a large holding cost. However, if one upstream installation does not have enough inventory, orders from the downstream installation will be delayed, and this delay may result in backlogging cost at the downstream installation. In fact, such delay at the upstream installation stands as a substantial challenge for the management of multi-echelon supply chains, and it is the biggest difference from single-echelon inventory systems.

To answer these questions, we need to first build a solid understanding of the structure of the inventory system. The OWMR model studied in this paper belongs to a *pure distribution* two-echelon inventory system, which is more challenging than the serial systems for the following two reasons. First, the demand process faced by the upstream installation is a superposition of the order processes of all its consecutive downstream installations. Second, in the case of shortage, the upstream installation needs to allocate current available inventory among its consecutive downstream installations, and different allocation policies will lead to different costs of the system (even with the same inventory policy and parameters).

Therefore, the allocation policy at the warehouse in the case of shortages is a key issue in the OWMR model. As suggested by Graves [2], we assume the warehouse adopts a *virtual allocation policy*, which means the orders for individual units are filled in the same order as the original demands at the retailers. Therefore, retailers' orders are filled in the order in which they are created. This policy gives rise to a significant tractability and a near-optimal performance. With the prevalence of information science and technology, the virtual allocation policy is realistic and achievable in practice. For example, the warehouse can observe retailers' demands in real time, and the demand information together with occurrence time are reported to the warehouse promptly.

Apart from the above, the MOQ requirement brings new challenges for the management of two-echelon inventory systems, because it reduces the inventory system's flexibility in responding to demand, and eventually increases inventory costs. In our model, the external supplier imposes the MOQ requirement on the warehouse, whereas retailers do not need to consider the MOQ. For such a

model, it is not clear how to evaluate the cost and optimize the system, and our efforts are therefore devoted to work them out.

The contributions of our paper are three-fold.

- First, to the best of our knowledge, we fill a gap in the literature in the sense that we are the first to *analytically* study a two-echelon inventory system with a MOQ requirement over an infinite horizon. The MOQ requirement is quite common in low cost manufacturing countries. Our work enriches the literature in the areas of inventory management and MOQ requirement.
- Second, we propose a new refined base stock policy for the inventory system with a MOQ requirement, and also provide optimization methods such that the optimal parameters can be easily obtained. In addition, to evaluate the system cost, we present a new *position-based cost-accounting* scheme, in which the cost of each unit is assigned to its first position at the warehouse.
- Third, we derive tight lower and upper bounds of the inventory parameters, which not only facilitate the search for the optimal policy that minimizes the long-run average system cost, but also shed light on inventory control in reality with some useful managerial insights. Our analytical results can be referred to as a good theoretical foundation for future related research.

The remainder of this paper is organized as follows. The literature related to our paper is discussed in Section 2. In Section 3, the model description is presented. Section 4 introduces a new heuristic policy developed for the inventory system with a MOQ requirement. In Section 5, we propose a new cost-accounting scheme to evaluate the system. The optimization method of our model is discussed in Section 6, and the paper is concluded in Section 7.

2. Literature Review

Our paper is related with a large body of multi-echelon inventory literature that studies cost evaluation and parameter optimization. Basically, multi-echelon inventory systems include serial systems, distribution systems, assembly systems, and general networks (see [3] for a good survey). The best-known early work for multi-echelon inventory systems may be the one by Clark and Scarf [4], and their decomposition technique is exact for serial systems. For a classical serial system, the echelon base-stock policy is optimal for both finite horizon ([4]) and infinite horizon ([5]) criteria. Chen [6] first established the optimality of (R, nQ) policy for a serial system with batch ordering. More recently, Shang and Zhou [7] studied the optimal policy parameters of an echelon (R, nQ, T) policy and heuristics for a serial system with fixed costs. Since a serial system is a special case of distribution inventory systems, our work, focusing on distribution networks, can be easily extended to the serial systems. In the remainder of this section, we only review the studies on distribution systems, with a particular focus on OWMR models.

2.1. Allocation Policy at the Warehouse

A key issue in distribution inventory systems, especially in periodic-review study, is the allocation policy at the warehouse in case of shortages. Up to now, the balance assumption and the virtual allocation policy are two commonly used policies.

The balance assumption, first suggested by Clark and Scarf [4], means that an upstream installation can make negative allocations to its downstream neighbors, or equivalently, lateral transshipment is allowed among all retailers. The balance assumption allows complete and costless reallocation of all stock in the system in every period. This implies that the warehouse stock can be used more efficiently than what is possible in reality, and therefore the balance assumption provides a lower bound for the real cost. The balance assumption has been widely applied in inventory study. Literature that uses or relies on the balance assumption includes the works in [5,8–16], and references therein.

The virtual allocation is another well-known allocation policy. It assumes that the upstream warehouse can observe the demand processes at all its downstream installations in real time. At the time when a demand event occurs, the warehouse commits or reserves a unit of its inventory

(if available) to replenish the downstream installation; however, the actual shipment of this unit does not start until the shipment occasion of that period arrives. Therefore, under the virtual allocation policy, orders for individual units at warehouse are filled in the same sequence as the origin demands at the retailers, which is consistent with the first-come-first-served basis in the continuous-review model. The virtual allocation policy was first proposed by Graves [2], and it is shown to be near optimal in many cases. After that, it was extensively used by the authors of [17–19]. Other allocation policies include the “ship-up-to- S ” policy ([20]), the “rationing-fraction- p_i ” policy ([21]), the α -policy ([22]), the two-period allocation ([23]), the two-interval allocation policy ([24]), and the asymptotically optimal allocation ([16]).

2.2. Methodologies for Distribution Inventory Systems

In general, as Simchi-Levi and Zhao [3] summarized, there are three generic methods for distribution inventory systems: the queuing-inventory method, the leadtime demand method, and flow-unit method.

The queuing-inventory method, just as the name indicates, models the inventory systems by a queuing system. In the OWMR model, the warehouse can be interpreted as a queuing system, where the waiting customers are backorders. Most of the previous studies using queuing-inventory method consider only continuous-review systems and one-for-one replenishment. The best-known work is the METRIC (Multi-Echelon Technique for Recoverable Item Control) approximation proposed by Sherbrooke [25]. In the OWMR model, the retailers’ leadtime may be stochastic due to possible delays at the warehouse. METRIC approximation means to replace this stochastic delay with its expectation.

The METRIC method is simple and computationally efficient because it decomposes the multi-echelon system into single-echelon systems, and therefore all the installation can be easily handed. Due to this advantage, several variations of METRIC are also proposed. Sherbrooke [26] proposed a VARI-METRIC framework, which performs better than METRIC in a numerical study. Recently, in [27], the authors applied the VARI-METRIC in designing an innovative model of spare parts allocation problem, aimed at minimizing backorders and ensuring system availability. Based on METRIC, Patriarca et al. [28] developed a novel system-approach model to address the stock allocation within a complex network, where unidirectional lateral transshipment has been concretely considered. Patriarca et al. [29] extended METRIC by considering a specific performance based contract (called PBC-METRIC), whose performance is demonstrated in a case study on a European airline that aims at reducing the ownership cost for the customer airline, while ensuring a target system performance. In [30], the authors extended METRIC by using the zero-inflated Poisson distribution to describe the demand pattern of the high-cost spare items (called ZIP-METRIC). In an extensive case study, the ZIP-METRIC outperforms the traditional Poisson-based approach. However, due to the MOQ requirement, we cannot apply the METRIC and the extensions mentioned above in our inventory system. Specifically, the common assumption in METRIC that the $(S-1, S)$ policy is appropriate for each item at each level (echelon), cannot hold in our system.

The leadtime demand method is similar to the classical method in single-echelon inventory systems. That is, the inventory position after ordering minus the demand during the leadtime is the inventory level (positive for on-hand inventory and negative for backorders). Given the inventory position after ordering, as long as the demand during the leadtime can be characterized, the inventory level is known and hence the cost incurred can be computed. In a continuous-review multi-echelon study, Simon [31] used a binomial disaggregation to characterize the steady state of the inventory level for Poisson process demand. Shanker [32] extended this work to a similar system with compound Poisson process demand. Graves [33] proposed a two-moment approximation for the steady-state distribution and showed that this approximation has a better performance than METRIC in a numerical study.

The flow-unit method keeps track of each supply unit as it moves through the system. Our work also use this method. Specifically, it characterizes the time between the placement of an order and the occurrence of its assigned demand unit. If this time is longer than the leadtime, backlogging cost

is incurred; otherwise, holding cost is incurred. Axsäter [34] first used this method to evaluate the cost of a continuous-review inventory system with one-for-one replenishment policy and independent Poisson demand. Later on, Axsäter [35] extended this work to systems with batch ordering. Many extensions have been studied, for example, the installation or echelon stock (R, Q) policies with Poisson or compound Poisson demand processes, as proposed in [36–38].

Our work also shares a common theme with the cost-accounting work. Levi et al. [39] proposed a new marginal cost accounting scheme for a single-echelon inventory system. We both focus on costs incurred by the units that ordered in the same period. However, in their scheme, they focused on the backlogging cost incurred by these units for a single period due to non-stationary inventory parameters (see [40] for capacitated models and [41] for serial systems). However, we focus on the duration of a state (in stock or not) of one unit in the inventory system. In addition, we consider a pure distribution network with a MOQ requirement. Therefore, our work is fundamentally different from theirs.

2.3. MOQ Requirement

In the area of MOQ, Fisher and Raman [42] first considered a two-period model with a MOQ requirement in each period. They worked out the optimal order quantity by using stochastic programming methods. Zhao and Katehakis [1] introduced the concept of M-increasing function and first partially characterized the optimal policy for multi-period inventory systems with MOQ. For the uncharacterized part, the authors gave easily computable upper bounds and asymptotic lower bounds for these intervals. However, for the characterized part, the optimal policy is complicatedly structured and difficult to implement in practice. Zhou et al. [43] proposed a two-parameter heuristic policy for a single-echelon inventory system with a MOQ requirement and demonstrated that the performance of this policy is close to the optimal policy, except for a few cases when the coefficient of the demand distribution is very small. For the same model, Kiesmüller et al. [44] proposed a modified base-stock policy, which is a special case of the policy proposed by Zhou et al. [43]. In the numerical study, the authors showed that this policy has a good performance in most cases. This modified base-stock policy has only one parameter and is easy to implement. Zhu et al. [45] studied a single-item periodic-review stochastic inventory system with both minimum order quantity (MOQ) and batch ordering requirements.

The most closely related papers to our work are those in [17,18]. Axsäter [17] studied a OWMR model with Poisson demand where each facility applies a periodic-reviewed order-up-to-S policy, and then Forsberg [18] extended this work to a OWMR model with compound Poisson demand. Both groups used the flow-unit method to study a OWMR model where the warehouse is assumed to adopt a virtual allocation policy. The major difference between our paper and the above two is that we assume the warehouse faces a MOQ requirement and thus adopts a different inventory policy (different from their order-up-to-S policies). Note that the MOQ requirement brings new challenge to keep track of a single unit, whereas in [17,18], the warehouse does not need to consider MOQ and each ordered unit from the retailers will trigger a virtual replenishment of the warehouse.

Finally, in addition to the above three streams of literature, our work is also related to papers on the uncertainty management in supply chains and on the use of classic models. For example, Fera et al. [46] defined a system dynamics model to assess competitiveness coming from the positioning of the order in different supply chain locations. Miranda et al. [47] showed the improvements of an alternative approach for the calculation of the multi-item economic order quantity in presence of space constraints. Fera et al. [48] presented an algorithm that helps the aggregator of energy loads in carrying out its daily decisions, using an economic objective function.

3. Model Description

We consider one warehouse (Installation 0) and N retailers (Installations 1, 2, \dots , N), as shown in Figure 1. Retailers face independent Poisson demand processes. Unfulfilled demands at each

retailer are backlogged. Retailers replenish their stocks from the warehouse, and in turn the warehouse replenishes from an external supplier. The lead time is assumed to be L_i periods between the warehouse and retailer i , and L_0 periods between the external supplier and the warehouse. Orders that cannot be immediately satisfied at the warehouse are also backlogged. There are linear holding costs at both the warehouse and all retailers, while the backlogging costs are only incurred at retailers. We also assume the warehouse and retailers select units to satisfy the demands based on a first-ordered-first-consumed rule, and hence they select the oldest units in stock to satisfy the most recent demands. For example, the units ordered by the warehouse in period n can be used to satisfy retailers' demand only when all the units ordered by the warehouse in period $n - 1$ are used out.

Notably, there is a MOQ requirement for the warehouse. It restricts that the warehouse, if it orders, must order at least M units, where M is a specified minimum order quantity. No such enforcement is imposed on retailers, and each retailer is allowed to order any quantity. The supply chain is operated under installation-stock based control. We assume that each retailer i adopts a base-stock policy with the order-up-to level S_i . The warehouse, due to the MOQ requirement, adopts a refined base-stock policy with reorder point S_0 (which is discussed below). Since each retailer adopts a base-stock policy, we assume that each demand unit at the retailer will trigger a virtual order to the warehouse and the warehouse allocates its inventory based on the virtual allocation policy.

Our objective is to minimize the long-run average system cost, including the ordering, holding, and backlogging cost, over an infinite horizon. Note that we set the ordering cost to be zero because linear ordering cost can be ignored under the average cost criterion (see, e.g., [49]). In addition, note that M , S_0 , and S_i ($i \in N$) are all integers.

The sequence of the events operate as follows. (1) The warehouse orders and then the periodic delivery from the external supplier arrives. (2) Retailers place orders to request replenishment from the warehouse. (3) The periodic delivery from the warehouse arrives at each retailer. (4) The stochastic demand takes place at all retailers. Inventory and backorders are assessed and then costs are charged. The primary notations used in the paper are summarized in Table 1.

Table 1. Primary notations.

Symbol	Description
M	minimum order quantity (an integer)
λ_i	demand rate at installation $i = 0, 1, \dots, N$ and $\lambda_0 = \sum_{i=1}^N \lambda_i$
h_i	holding cost per unit per period at installation $i = 0, 1, \dots, N$
p_i	backlogging cost per unit per period at installation $i = 1, \dots, N$
S_i	the policy parameter (an integer) for retailer $i = 1, \dots, N$
S_0	the policy parameter (an integer) for the warehouse
x^+	$\max(0, x)$, the maximum between zero and x
$Z^{[a,b]}$	A set of integers which are not less than a and not greater than b

4. A Refined Base-Stock Policy for MOQ

In this section, we focus on a single-echelon stochastic inventory system with a MOQ requirement. Kiesmüller et al. [44] proposed a generalized one-parameter base-stock policy (referred as S policy) that considers the MOQ requirement. Their policy works as follows: no order is placed when the inventory position is not less than the level S ; otherwise, an order is placed to raise the inventory to S . However, if this order is smaller than the MOQ (i.e., M), the order quantity is increased to M . Let

X_n be the inventory position before ordering in period n , and then the order quantity in period $n + 1$, q_{n+1} , can be described as

$$q_{n+1} = \begin{cases} 0, & \text{if } X_n \geq S, \\ M, & \text{if } S - M < X_n < S, \\ S - X_n, & \text{if } X_n \leq S - M. \end{cases}$$

We can see that the principle of the above S policy is to raise the inventory position to the level S as long as the inventory position is below S . If the difference between the current level and the aimed level (S) is less than M , then the order quantity is raised to M due to the MOQ requirement. Therefore, this policy is a generalized base-stock policy with the base-stock level S . In a numerical test, Kiesmüller et al. [44] showed this policy has near-optimal performance in most cases.

However, there is one implicit drawback of the S policy. Suppose that the unit holding cost is quite large, then the firm has no incentive to hold much inventory, and S can be negative. Assume $-M + 1 \leq S < 0$ (we show that the optimal S cannot be less than $-M + 1$ below), in the case that the inventory position before ordering is less than $S - M$, to raise the inventory position to S , the order quantity should be $S - X_n$ according to the S policy. However, if S is negative, even if the inventory position is raised to S , it will still incur backlogging cost. Obviously, compared to this, raising the inventory position to zero is a better alternative. The same problem exists when the inventory position is less than S and higher than $-M$. The firm has no incentive to keep the backlogging state in such cases and should try to raise the inventory position to zero at least.

Therefore, to tackle the above drawback, we propose a refined base-stock policy, which works as follows: if the inventory level is less than the level S , an order is placed to raise the inventory position to the maximum between S and zero (if this order is smaller than M , the order quantity is increased to M); otherwise, no order is placed. Formally, the order quantity q_n is defined by

$$q_n = \begin{cases} 0, & \text{if } X_n \geq S, \\ M, & \text{if } \max\{-M, S - M\} \leq X_n < S, \\ -X_n, & \text{if } S - M \leq X_n < -M, \\ \max\{S, 0\} - X_n, & \text{if } X_n < S - M. \end{cases}$$

In our policy, if the inventory level is less than S , we place an order to keep the inventory position non-negative. However, in the S policy, it is possible that the inventory is still in the backlogging state after ordering. This is the major difference between our policy and the S policy.

Analysis Based on Markov Chain

Particularly, our policy is reduced to the S policy if $S \geq 0$. In such case, we can use the same method to analyze the system as Kiesmüller et al. [44]. Considering this, in the remainder of this section, we deal with the case where $-M + 1 \leq S < 0$.

Given $-M + 1 \leq S < 0$ and inventory position X_n , the order quantity q_n can be rewritten as (see Figure 2 for illustration)

$$q_n = \begin{cases} 0, & \text{if } X_n \geq S, \\ M, & \text{if } -M \leq X_n < S, \\ -X_n, & \text{if } X_n < -M. \end{cases}$$

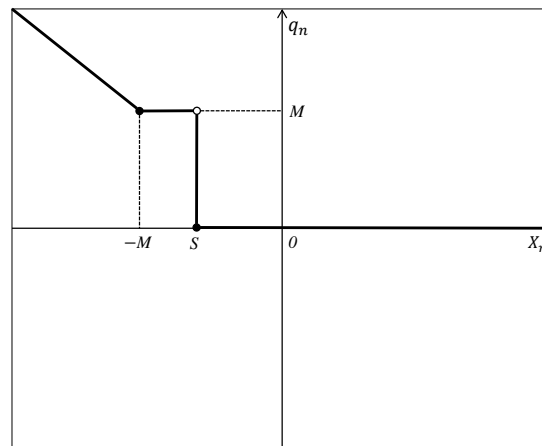


Figure 2. Illustration of order quantity q_n .

Let Y_n be the inventory position after ordering in period n . The inventory balance equation follows from (see Figure 3 for illustration)

$$Y_{n+1} = \begin{cases} Y_n - D_n, & \text{if } Y_n - D_n \geq S, \\ Y_n - D_n + M, & \text{if } -M \leq Y_n - D_n < S, \\ 0, & \text{if } Y_n - D_n < -M, \end{cases}$$

where D_n is the demand in period n .

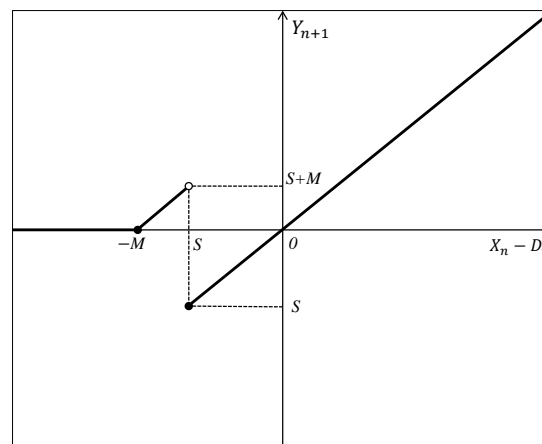


Figure 3. Illustration of inventory balance equation.

It is clear that Y_n is a Markov chain with a finite state space $Z^{[S, S+M-1]}$, i.e., $\{S, S+1, \dots, S+M-1\}$. Then, the transition probabilities $f_{a,b} = \text{Prob}(Y_{n+1} = b | Y_n = a)$ can be derived as (see Figure 4 for illustration)

$$f_{a,b} = \begin{cases} \text{Prob}(D_n = a + M - b) + \text{Prob}(D_n = (a - b)^+), & \text{if } 0 < b \leq S + M - 1, \\ \text{Prob}(D_n = (a - b)^+) + \text{Prob}(D_n \geq a + M), & \text{if } b = 0, \\ \text{Prob}(D_n = (a - b)^+), & \text{if } S \leq b \leq -1, \end{cases}$$

where $(a - b)^+$ equals to $a - b$ if $a \geq b$, and zero otherwise.

Second, assume $S^* < y^* - M + 1$. Since $C(y)$ is non-increasing when $y \leq y^*$, we get $C(S^* - 1 + i) \geq C(S^* + i)$, $\forall i \in \mathbb{Z}^{[1, M]}$. It implies

$$L(S^*) = \sum_{i=1}^M \alpha_i C(S^* - 1 + i) \geq \sum_{i=1}^M \alpha_i C(S^* + i) = L(S^* + 1).$$

This also contradicts the definition of S^* . Therefore, we arrive at

$$y^* - M + 1 \leq S^* \leq y^*.$$

Moreover, $y^* \leq S^* + M - 1$ can be derived from $y^* - M + 1 \leq S^*$. \square

Proposition 1 helps us to search the optimal level S^* . The intuition of this proposition is to keep y^* in the set of integers $\mathbb{Z}^{[S, S+M-1]}$, i.e., $\{S, S+1, \dots, S+M-1\}$. Recall that $\mathbb{Z}^{[S, S+M-1]}$ is the state space for the inventory position after ordering. If one S cannot guarantee that y^* falls in this set, then this S cannot be the optimal one. Proposition 1 provides a lower bound $y^* - M + 1$ and an upper bound y^* for S^* . Obviously, $S^* \geq -M + 1$, which validates our previous statement that $S \geq -M + 1$.

To conclude this section, we propose a refined base-stock policy for the stochastic inventory system with a MOQ requirement. In the remainder of this paper, we assume the warehouse in our OWMR model adopts this policy to control its inventory with $S_0 = S$.

5. A Position-Based Cost-Accounting Scheme

We introduce a new cost-accounting scheme and use this scheme to evaluate the cost of our model in this section. We first focus on the inventory position of the warehouse. Note that the warehouse faces a superposition of the order processes of all retailers. Since each retailer applies a base-stock policy, each demand unit at retailers will trigger a virtual order to the warehouse. Due to the superposition property of Poisson process, the warehouse faces a Poisson demand process with arrival rate $\lambda_0 = \sum_{i=1}^N \lambda_i$. Then, we can characterize the *inventory position* of the warehouse as in the single-echelon system studied in Section 4. We next suppose that, in the steady state, the distribution of the warehouse inventory position \tilde{a} is known (see Equation (1)). With a little abuse of notation, we use $\text{Prob}(Y = y) = \alpha_{y-S+1}$ to denote the steady-state probability that the warehouse inventory position is y .

We compute the costs associated with the units ordered by the warehouse in the current period (in which the ordering happens), that is: (1) if no order is placed at the warehouse, then no cost is incurred in that period; and (2) if some units are ordered by the warehouse in one period, we compute all the costs associated with these units. Note that, in the latter case, the costs associated with these units may incur in other periods, e.g., holding costs incurred in the future periods; however, under our framework, we assign all these costs to the current period.

In each period, if the warehouse places an order, each ordered unit will have a corresponding position at the warehouse for that period. For each position at the warehouse, there will be a corresponding expected system cost C , which is the total cost associated with the unit at this position, including holding cost at the warehouse, holding and backlogging cost at retailers. As long as one unit is ordered by the warehouse, it will have a position at the warehouse and its expected overall cost through the system is determined. In other words, for a unit with position P_0 , its overall cost depends on the position P_0 . Therefore, our cost-accounting scheme is a *position-based* framework.

The long-run average cost $L(S_0, S_1, \dots, S_N)$ can be characterized as

$$L(S_0, S_1, \dots, S_N) = \sum_{y=S_0}^{S_0+M-1} \sum_{q \geq M} \text{Prob}(Y = y, Q = q) \sum_{k=0}^{q-1} C(y - k, S_1, \dots, S_N) \quad (2)$$

where $C(y - k, S_1, \dots, S_N)$ is the expected cost incurred by the unit with position $P_0 = y - k$ when it is ordered by the warehouse. Since $y \in Z^{[S_0, S_0+M-1]}$, $q \geq M$, and $k \in Z^{[0, q-1]}$, both $P_0 > 0$ and $P_0 \leq 0$ are possible.

Remark 1. The position of a unit (before it is used to satisfy retailer demands) at the warehouse will decrease as the number of ordered units increases. In our position-based scheme, we only focus on the position of a unit in the period when this unit is ordered by the warehouse. Therefore, hereinafter, we use “a unit with position P_0 ” to denote the unit whose position is P_0 in the specific period when it is order by the warehouse.

Remark 2. In Equation (2), $\text{Prob}(Y = y, Q \geq M)$ is the probability that inventory position after placing an order equals to y . This probability is less than the steady state probability $\text{Prob}(Y = y)$ discussed earlier, and we can easily verify it from $\text{Prob}(Y = y) = \text{Prob}(Y = y, Q \geq M) + \text{Prob}(Y = y, Q = 0)$. We do not consider the case $Q = 0$, because no cost is incurred if no ordering.

In the sequel, we first discuss how to compute the cost function $C(P_0, S_1, \dots, S_N)$, and then derive the inventory position probability $\text{Prob}(Y = y, Q = q)$.

5.1. Computation of Cost Function

For an arbitrary unit (hereinafter referred as unit w) with position P_0 when it is ordered by the warehouse, the cost function $C(P_0, S_1, \dots, S_N)$ is formulated as

$$C(P_0, S_1, \dots, S_N) = V(P_0) + \sum_{i=1}^N \frac{\lambda_i}{\lambda_0} \Pi_i(P_0, S_i),$$

where $V(P_0)$ is the expected warehouse holding cost incurred by unit w , and $\Pi_i(P_0, S_i)$ is the expected backlogging and holding cost incurred by unit w at retailer i . Because the warehouse faces a Poisson process demand and adopts the virtual allocation policy, this unit is assigned to retailer i with probability λ_i/λ_0 .

Figure 5 shows an example of our position-based cost function, where there are a demand unit of position $P_0 = 6$ at the warehouse, and two retailers, one with order-up-to level $S_1 = 3$ and the other with $S_2 = 2$. Therefore, the expected holding cost at the warehouse is $V(6)$, and the expected backlogging and holding costs at Retailers 1 and 2 are $\Pi_1(6, 3)$ and $\Pi_2(6, 2)$, respectively. The total cost is $C(6, 3, 2) = V(6) + \lambda_1 \Pi_1(6, 3)/\lambda_0 + \lambda_2 \Pi_2(6, 2)/\lambda_0$.

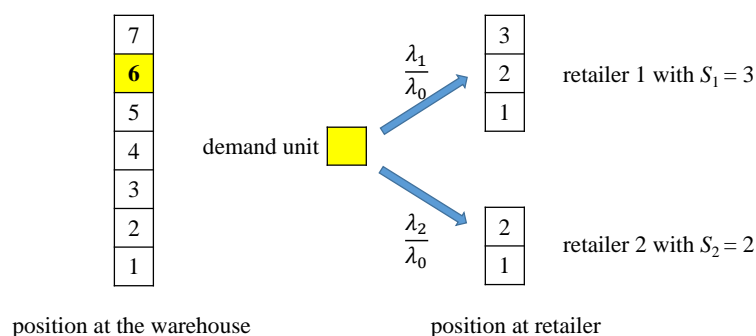


Figure 5. An example of position-based cost function.

We extend the notations to be used in this section in Table 2.

Table 2. Extended notations.

Symbol	Description
T_0	the time difference between the period when unit w is ordered by the warehouse and the period when unit w is ordered by a retailer
T_i	the time difference between the period when unit w is ordered by retailer i and the period when unit w is ordered by a customer
P_i	the position of unit w at retailer i when it is ordered by retailer i
Δ	the number of delayed periods at the warehouse due to the possible stock out
$D_i^{T_i}$	the number of units demanded at retailer i over T_i periods
$D_0^{T_0}$	the total number of units demanded at warehouse over T_0 periods.

5.1.1. Computation of $V(P_0)$.

Suppose that unit w is ordered by warehouse in period n . Since the lead time is L_0 periods, this unit arrives at the warehouse in period $n + L_0$. Recall that T_0 is the time difference between the period when unit w is ordered by the warehouse and the period when unit w is ordered by a retailer. If $T_0 \geq L_0$, the unit first arrives at the warehouse, and then is requested by a retailer after some periods. Therefore, the number of periods with holding cost is $L_0 - T_0$. If $T_0 < L_0$, the unit has not reached the warehouse when it is ordered by a retailer, and there will be no holding cost.

Consequently, for a single unit with position P_0 , the expected holding cost incurred at the warehouse, $V(P_0)$, can be expressed as

$$V(P_0) = E[h_0(T_0 - L_0)^+ | P_0] = \sum_{t_0 \geq L_0} h_0(t_0 - L_0) \cdot \text{Prob}(T_0 = t_0 | P_0) \quad (3)$$

where $\text{Prob}(T_0 = t_0 | P_0)$ gives the distribution of T_0 that is dependent on P_0 .

In the following, our effort is devoted to determine the distribution of T_0 . The position P_0 is determined at the moment unit w is ordered by the warehouse, therefore, unit w will be used to satisfy the P_0 th warehouse demand after it is ordered by the warehouse. As mentioned earlier, both $P_0 > 0$ and $P_0 \leq 0$ are possible.

For a given $P_0 = p_0 > 0$, the distribution of T_0 can be characterized by

$$\text{Prob}(T_0 = t_0 | P_0 = p_0 > 0) = \begin{cases} \text{Prob}(D_0^{t_0+1} \geq p_0, D_0^{t_0} < p_0), & \text{if } t_0 \geq 1, \\ \text{Prob}(D_0^1 \geq p_0), & \text{if } t_0 = 0, \end{cases} \quad (4)$$

where $D_0^{t_0}$ represents the total demand at the warehouse over t_0 periods,

$$\begin{aligned} \text{Prob}(D_0^{t_0+1} \geq p_0, D_0^{t_0} < p_0) &= \text{Prob}(D_0^{t_0+1} \geq p_0) - \text{Prob}(D_0^{t_0+1} \geq p_0, D_0^{t_0} \geq p_0) \\ &= \text{Prob}(D_0^{t_0+1} \geq p_0) - \text{Prob}(D_0^{t_0} \geq p_0) \\ &= \sum_{k=0}^{p_0-1} \frac{\lambda_0^k}{k!} e^{-\lambda_0 t_0} [(t_0)^k - e^{-\lambda_0} (t_0 + 1)^k], \end{aligned}$$

and

$$\text{Prob}(D_0^1 \geq p_0) = 1 - \sum_{k=0}^{p_0-1} \frac{\lambda_0^k}{k!} e^{-\lambda_0}.$$

We next study the distribution of T_0 if $P_0 \leq 0$, followed by two cases. The first case is $S_0 > 0$ and $P_0 \leq 0$, which means unit w is used to satisfy the retailer's order in the last period. This implies that unit w is requested by one retailer in the last period, and thus $T_0 = -1$.

The second case is $S_0 \leq 0$ and $P_0 \leq 0$. Since $P_0 \leq 0$, unit w is used to satisfy the retailer's order earlier, implying that unit w is requested by one retailer in one of the earlier periods, and thus T_0 can take the values from $\{-1, -2, \dots\}$. Specifically, if $T_0 = -1$, unit w is requested by one retailer one period before it is ordered by the warehouse. Hence, in the period when unit w is requested by one retailer, the inventory position of the warehouse at the end of that period is less than S_0 , and this will trigger one order including unit w from the warehouse to the supplier. Therefore, for any unit w with a position $P_0 < S_0$, it must have $T_0 = -1$, i.e.,

$$\text{Prob}(T_0 = -1 | P_0 = p_0 < S_0) = 1, \quad (5)$$

which means this unit is used to satisfy the demand in the last period.

We are left to analyze the position $P_0 = p_0 \in Z^{[S_0, 0]}$. The probability of $T_0 = -1$ is

$$\text{Prob}(T_0 = -1 | P_0 = p_0) = \sum_{i=p_0}^{S_0+M-1} \text{Prob}(D_0^1 > i - S_0) \cdot \text{Prob}(y_{t-1} = i), \text{ where } p \in Z^{[S_0, 0]}. \quad (6)$$

If $T_0 = -2$, unit w is requested two periods earlier before it is ordered by the warehouse. To compute the probability of $T_0 = -2$, we need to condition on the inventory position one period earlier. Suppose unit w is ordered by the warehouse in period n , $T_0 = -2$ implies it is requested by one retailer in period $n - 2$. This means that the inventory position at the end of period $n - 2$ is not less than S_0 , while the inventory position at the end of period $n - 1$ must be less than S_0 . Following the same logic, we can get the probability of $T_0 = -n$ ($n \geq 2$) as follows

$$\begin{aligned} & \text{Prob}(T_0 = -n | P_0 = p_0) \\ &= \sum_{i=p_0}^{S_0+M-1} \sum_{k=S_0}^{p_0} \text{Prob}(D_0^1 = i - k) \cdot \text{Prob}(D_0^{n-1} > k - S_0, D_0^{n-2} \leq k - S_0) \cdot \text{Prob}(y_{t-n} = i), \end{aligned} \quad (7)$$

where $p_0 \in Z^{[S_0, 0]}$.

To conclude, for a single unit with position P_0 , based on the values of P_0 , the detailed probabilities $\text{Prob}(T_0 = t_0 | P_0 = p_0)$ are given in Equations (4)–(7), and the computation in Equation (3) of $V(P_0)$ is completed.

5.1.2. Computation of $\Pi_i(P_0, S_i)$.

We are left to compute the cost for a single unit incurred at retailer i , $\Pi_i(P_0, S_i)$. Let $\Delta = (L_0 - T_0)^+$ denote the possible delay at the warehouse when unit w is ordered by a retailer. Note that $\Delta = \delta$ means the unit will be shipped to the retailer after δ periods it is ordered by the retailer. If $L_i + \Delta < T_i$, the unit has already been at retailer i for $T_i - L_i - \Delta$ periods when it is ordered by customer demand. In this case, it will incur a holding cost $h_i * (T_i - L_i - \Delta)$. If $L_i + \Delta \geq T_i$, the unit is still on the way to retailer i and will arrive after $L_i + \Delta - T_i$ periods, then it will incur backlogging cost $b_i * (L_i + \Delta - T_i)$. Consequently, the cost incurred by this unit at retailer i , $\Pi_i(P_0, S_i)$, is computed by

$$\begin{aligned} \Pi_i(P_0, S_i) &= E[h_i(T_i - L_i - \Delta)^+ + p_i(L_i + \Delta - T_i)^+ | P_0, S_i] \\ &= \sum_{t_i} \sum_{\delta} \{h_i(t_i - L_i - \delta)^+ + p_i(L_i + \delta - t_i)^+\} \cdot \text{Prob}(\Delta = \delta | P_0) \cdot \text{Prob}(T_i = t_i | P_0, S_i) \end{aligned} \quad (8)$$

where $\text{Prob}(\Delta = \delta | P_0)$ gives the distribution of Δ that is dependent on P_0 , and $\text{Prob}(T_i = t_i | P_0, S_i)$ gives the distribution of T_i that is dependent on both P_0 and S_i .

We first derive the probability $\text{Prob}(\Delta = \delta | P_0)$. Since $\Delta = (L_0 - T_0)^+$, we get

(1) if $P_0 > 0$,

$$Prob(\Delta = \delta | P_0) = \begin{cases} Prob(T_0 = L_0 - \delta | P_0), & \text{if } 1 \leq \delta \leq L_0, \\ \sum_{t_0 \geq L_0} Prob(T_0 = t_0 | P_0), & \text{if } \delta = 0, \end{cases}$$

(2) if $P_0 \leq 0$ and $P_0 < S_0$,

$$Prob(\Delta = \delta | P_0) = 1, \text{ if } \delta = L_0 + 1,$$

(3) if $S_0 \leq P_0 \leq 0$,

$$Prob(\Delta = \delta | P_0) = Prob(T_0 = L_0 - \delta | P_0), \text{ if } \delta \geq L_0 + 1.$$

Next, we work on the probability $Prob(T_i = t_i | P_0, S_i)$. At the retailer level, after placing an order to the warehouse, the inventory position of retailer i reaches its base stock S_i , as do the other retailers (according to the base-stock policy). Note that the inventory position of retailer i here includes the units that may have not reached the warehouse but have already been assigned to retailer i .

At the warehouse level, for a unit w with position P_0 , suppose it is ordered by the warehouse in period $n + t_0$, and is ordered by retailer i in period n . That is, the time difference between the two periods is t_0 . The total demand over these t_0 periods is $D_0^{t_0}$. Hence, in period $n + t_0$, there are also $D_0^{t_0} - P_0$ units ordered by the warehouse after unit w . For any unit among them, if it is assigned to retailer i , it will have a higher position at retailer i than unit w . This is because unit w is virtually ordered earlier than these units by retailer i . Let $R_0 = D_0^{t_0} - P_0$, and we decompose these R_0 units and explore how many will “be above” (have a higher position at retailer i if is assigned to retailer i) unit w . Let $\{R_i | i \in N\}$ denote the numbers of such units. Obviously, $\sum_{i=1}^N R_i = R_0$. The unit w has a position of $P_i = S_i - R_i$ at retailer i in period $n + t_0$, and thus it will be used to satisfy the P_i th demand starting from period $n + t_0$. At the warehouse level, when a new demand unit arrives, it is previously ordered from retailer i with a probability λ_i / λ_0 . Consequently, given that R_0 takes the value of r_0 ($r_0 \geq 0$), the conditional probability that r_i units (out of these r_0 units) are ordered from retailer i , can be determined as

$$Prob(R_i = r_i | R_0 = r_0) = C_{r_0}^{r_i} \left(\frac{\lambda_i}{\lambda_0} \right)^{r_i} \left(1 - \frac{\lambda_i}{\lambda_0} \right)^{r_0 - r_i}, \quad r_i \leq r_0. \quad (9)$$

Remark 3. In [31], a similar binomial distribution is proposed to disaggregate the backorders. In our method, we disaggregate all the demand units after unit w by Equation (9); however, Simon [31] disaggregated only the possible backorders of one period.

Now, for the unit w with position $P_0 = p_0 > 0$ at the warehouse, we have

$$\begin{aligned} Prob(R_0 = r_0, T_0 = t_0 | P_0 = p_0) &= Prob(D_0^{t_0} < p_0, D_0^{t_0+1} = p_0 + r_0) \\ &= \sum_{l=r_0+1}^{p_0+r_0} Prob(D_0^{t_0} = p_0 + r_0 - l) \cdot Prob(D_0^1 = l), \end{aligned} \quad (10)$$

where $D_0^{t_0}$ denotes the demand over t_0 periods (i.e., from period n to period $n + t_0 - 1$) and D_0^1 denotes the demand in period $n + t_0$. Then, we can compute the joint distribution of R_i and T_0 by using Equations (9) and (10) as follows

$$Prob(R_i = r_i, T_0 = t_0 | P_0 = p_0) = \sum_{r_0 \geq r_i} Prob(R_i = r_i | R_0 = r_0) \cdot Prob(R_0 = r_0, T_0 = t_0 | P_0 = p_0). \quad (11)$$

Since the position at retailer i is $P_i = S_i - R_i$, from Equation (11), we can know the joint distribution of P_i and T_0 as

$$Prob(P_i = S_i - r_i, T_0 = t_0 | P_0 = p_0) = Prob(R_i = r_i, T_0 = t_0 | P_0 = p_0).$$

The distribution for P_i then is defined by

$$\text{Prob}(P_i = S_i - r_i | P_0 = p_0) = \sum t_0 \cdot \text{Prob}(R_i = r_i, T_0 = t_0 | P_0 = p_0) \cdot \text{Prob}(T_0 = t_0 | P_0 = p_0). \quad (12)$$

Similar to the process at the warehouse level, if $P_i = p_i > 0$, we can derive the following distribution of T_i

$$\text{Prob}(T_i = t_i | P_i = p_i > 0) = \begin{cases} \text{Prob}(D_i^{t_i+1} \geq p_i, D_i^{t_i} < p_i), & \text{if } t_i \geq 1, \\ \text{Prob}(D_i^1 \geq p_i), & \text{if } t_i = 0, \end{cases} \quad (13)$$

where

$$\text{Prob}(D_i^{t_i+1} \geq p_i, D_i^{t_i} < p_i) = \sum_{k=0}^{p_i-1} \frac{\lambda_i^k}{k!} e^{-\lambda_i t_i} [(t_i)^k - e^{-\lambda_i} (t_i + 1)^k]$$

and

$$\text{Prob}(D_i^1 \geq p_i) = 1 - \sum_{k=0}^{p_i-1} \frac{\lambda_i^k}{k!} e^{-\lambda_i}$$

If $P_i \leq 0$, the unit is ordered in the last period; therefore, $T_i = -1$ and we have

$$\text{Prob}(T_i = -1 | P_i = p_i \leq 0) = 1. \quad (14)$$

This is similar to the case $P_0 \leq 0$ and $S_0 \geq 0$ where $T_0 = -1$ at the warehouse.

Remark 4. Equation (13) implies that the distribution of T_i depends on P_i , whose distribution depends on P_0 from Equation (12). This means that the distribution of T_i also depends on P_0 .

To conclude, the probability $\text{Prob}(T_i = t_i | P_0, S_i)$ can be computed based on Equations (12)–(14). Up to now, given the position at the warehouse P_0 , we can compute the cost associated with unit w by Equations (3) and (8). We next provide the derivation of the distribution of the inventory position P_0 .

5.2. Computation of Inventory Position Probability

We extend the notations to be used in this section in Table 3. Please note that Y_n represents the inventory positions at the warehouse, which is updated immediately after ordering.

Table 3. Extended notations II.

Symbol	Description
X_n	the inventory position of the warehouse in period n before ordering to the supplier
Y_n	the inventory position of the warehouse in period n after ordering to the supplier before receiving retailer orders
Q_n	the quantity ordered by the warehouse to the supplier in period n
D_n	the demand arrived at the warehouse in period n

In period n , given an inventory position y_n after an ordering $Q_n \geq M$, we need to compute the inventory position probability $\text{Prob}(Y_n = y_n, Q_n \geq M)$. We choose to conditioning on the inventory position in period $n - 1$, i.e., Y_{n-1} . For any $y_i \in Z^{[S_0, S_0+M-1]}$, we have

$$\text{Prob}(Y_n = y_n, Q_n \geq M) = \sum_{y_i=S_0}^{S_0+M-1} \text{Prob}(Y_n = y_n, Q_n \geq M | Y_{n-1} = y_i) \cdot \text{Prob}(Y_{n-1} = y_i). \quad (15)$$

For the conditional probability, depending on the value of S_0 , we have two cases:

(1) if $S_0 \geq 0$,

$$\text{Prob}(Y_n = y_n, Q_n \geq M | Y_{n-1} = y_i) = \begin{cases} \text{Prob}(D_{n-1} \geq y_i - (S_0 - M)), & \text{if } y_n = S_0, \\ \text{Prob}(D_{n-1} = y_i - (y_n - M)), & \text{if } y_n \in Z^{[S_0+1, S_0+M-1]}, \end{cases} \quad (16)$$

(2) if $-M + 1 \leq S_0 < 0$,

$$\text{Prob}(Y_n = y_n, Q_n \geq M | Y_{n-1} = y_i) = \begin{cases} 0, & \text{if } y_n \in Z^{[S_0, -1]}, \\ \text{Prob}(D_{n-1} \geq y_i + M), & \text{if } y_n = 0, \\ \text{Prob}(D_{n-1} = y_i - (y_n - M)), & \text{if } y_n \in Z^{[1, S_0+M-1]}. \end{cases} \quad (17)$$

Figures 6 and 7 plot the relationships between Y_n and $Y_{n-1} - D_{n-1}$ with $S_0 \geq 0$ and $-M + 1 \leq S_0 < 0$, respectively. In both figures, the red bold lines represent the Y_n after an ordering $Q_n \geq M$, and the black bold lines represent a zero ordering, i.e., $Q_n = 0$.

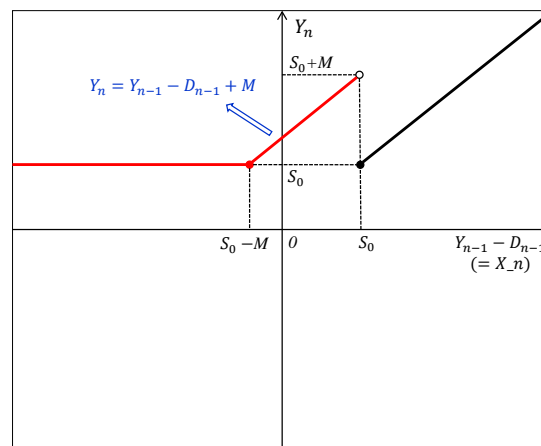


Figure 6. Inventory position with $S_0 \geq 0$.

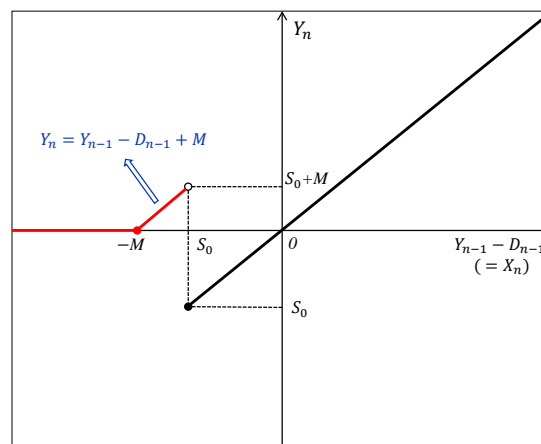


Figure 7. Inventory position with $-M + 1 \leq S_0 < 0$.

Next, we discuss how to derive the distribution of the position P_0 of an arbitrary unit w at the warehouse, since it is a prerequisite in the computation of cost function $C(P_0, S_1, \dots, S_N)$. Similar to the above, we have two cases to follow: the first case is $S_0 \geq 0$, and the second case is $-M + 1 \leq S_0 < 0$.

Case 1: $S_0 \geq 0$. Based on Equations (15) and (16), Y_n takes values from $\{S_0, S_0 + 1, \dots, S_0 + M - 1\}$. If $Y_n \geq S_0 + 1$, then $Q_n = M$ and $X_n = Y_n - M$ (see Figure 6). This implies that, in period n , the warehouse orders M units in period n . The unit w is uniformly distributed in these units, and therefore

$$P_0 \sim \text{Uniform}[Y_n - M + 1, Y_n],$$

where $\text{Uniform}[a, b]$ represents a Uniform distribution in the interval $[a, b]$, and $Y_n - M + 1 = X_n + 1$, i.e., the starting position.

If $Y_n = S_0$, then Q_n can be larger than M . Unit w is still uniformly distributed in these Q_n units. To obtain the distribution of P_0 , we must know the distribution of X_n . We next characterize the distribution of X_n for $Y_n = S_0$. Since $Y_n = S_0$, we have $X_n \leq S_0 - M$ (see Figure 6), and the distribution of X_n is defined by

$$\text{Prob}(X_n = x_n, Q_n \geq M) = \sum_{y_i=S_0}^{S_0+M-1} \text{Prob}(Y_{n-1} = y_i) \cdot \text{Prob}(D_{n-1} = y_i - x_n), \quad x_n \leq S_0 - M.$$

The order quantity is $Q_n = S_0 - X_n$ in period n . The unit w is uniformly distributed in these Q_n units, hence we have

$$P_0 \sim \text{Uniform}[X_n + 1, S_0]$$

We should note that the distribution of P_0 depends only on P_0 and is independent of S_i .

Case 2: $-M + 1 \leq S_0 < 0$. Based on Equations (15) and (17), Y_n takes values from $\{0, 1, \dots, S_0 + M - 1\}$. If $Y_n \geq 1$, then $Q_n = M$ and $X_n = Y_n - M$ (see Figure 7). Similar to the previous analysis, we have

$$P_0 \sim \text{Uniform}[Y_n - M + 1, Y_n].$$

If $Y_n = 0$, then Q_n can be larger than M . We have $X_n \leq -M$ (see Figure 7), and the distribution of X_n is defined by

$$\text{Prob}(X_n = x_n, Q_n \geq M) = \sum_{y_i=S_0}^{S_0+M-1} \text{Prob}(Y_{n-1} = y_i) \cdot \text{Prob}(D_{n-1} = y_i - x_n), \quad x_n \leq -M.$$

Similar to the previous analysis, the distribution of P_0 is

$$P_0 \sim \text{Uniform}[X_n + 1, S_0].$$

Remark 5. Our position-based cost-accounting scheme uses flow-unit method to evaluate the cost of the whole system. Levi et al. [39] used a similar idea of holding cost accounting; however, due to their non-stationary parameter setting, they evaluated the backlogging cost for a single period, not for a single unit in our scheme. Axsäter [34] used a flow-unit to evaluate the cost of a continuous-review model without the MOQ requirement. A distinct feature of our work is to assign the cost of a single unit to its position at the warehouse, and we characterize the distribution of this position.

6. Optimization

We have shown how to evaluate the cost of our model with given inventory parameters $\{S_0, S_1, \dots, S_N\}$. In this section, we discuss how to optimize our model and find the optimal parameter $\{S_0^*, S_1^*, \dots, S_N^*\}$ that minimize the long-run average cost.

We first note that $V(P_0)$ depends only on P_0 , which in turn depends only on S_0 , and $\Pi_i(P_0, S_i)$ is dependent on $S_j (j \neq i)$. This implies that, for a given S_0 , we can optimize each retailer i separately with respect to S_i . Let $S_i^*(S_0)$ denote the optimal value of S_i for a given S_0 .

Proposition 2. For a given S_0 , $L(S_0, S_1, \dots, S_N)$ is convex in each S_i .

Proof of Proposition 2. $L(S_0, S_1, \dots, S_N)$ denotes the long-run average system cost, which consists of the cost incurred at the warehouse and the costs incurred at all retailers. For a given S_0 , the cost incurred at the warehouse is independent of S_i . We next only consider the cost incurred at retailer i . If this cost is convex in S_i , then we can have the total cost L is convex in S_i , recalling that we can optimize each retailer i separately with respect to S_i . For a given S_0 , the distribution of backlogging quantity can be determined, and we can decompose these backlogging units using the same way as Equation (9). Given the distribution of backlogging quantity for retailer i , we can know the distribution of delay time at the warehouse caused by these backlogging units. With a little abuse of notation, we use Δ_i to denote the delay time at the warehouse for retailer i . Now, for a given S_0 , we know the retailer's leadtime L_i and the distribution of its delay time Δ_i . To optimize retailer i , Δ_i can be seen as an additional leadtime, and then the problem is reduced to a single-echelon retailer i with a leadtime $L_i + \Delta_i$. For each possible value of Δ_i , the cost incurred at retailer i is a convex function of S_i , which can be easily obtained from the typical Newsvendor problem with a base-stock policy. Taking the expectation with respect to Δ_i , we can conclude that the expected cost is convex in S_i and so is the total cost L . \square

Proposition 2 implies that when other parameters are fixed, for each specific retailer, there exists a unique order up to level such that the total inventory cost $L(S_0, S_1, \dots, S_N)$ is minimized. This result is consistent with that in continuous-review model without MOQ requirement (see, e.g., [34]). Using Proposition 2, we can find the corresponding $S_i^*(S_0)$ for each S_0 . We next try to find some bounds for S_0 and S_i .

Proposition 3 (Lower bound). $S_0^* \geq -M + 1$ and $S_i^* \geq 0$ for each $i = 1, 2, \dots, N$.

Proof of Proposition 3. Since $S_0^* \geq -M + 1$ is straightforward by Proposition 1, we only need to prove $S_i^* \geq 0$ for each $i = 1, 2, \dots, N$. We prove this by contradiction. Suppose there exist an optimal $S_i < 0$ for an arbitrary retailer $i = \{1, 2, \dots, N\}$. Then, compare this system to another system with the same parameters except for $S_i = 0$. We can easily find that the cost incurred at the warehouse and retailer $1, 2, \dots, i-1, i+1, \dots, N$ are all the same in the two different systems. The only difference is the cost incurred at retailer i . For retailer i , $S_i < 0$ means it always has backlogging units in every period and the expected cost is $\mathbb{E}[b_i(S_i - D)]$. Note that there is no holding cost in each period if $S_i < 0$. Similarly, if $S_i = 0$, the expected cost is $\mathbb{E}[b_i(0 - D)]$ and there is also no holding cost. Obviously, the system with $S_i = 0$ has a better perform than the system with $S_i < 0$. This contradicts the optimality of $S_i < 0$. Therefore, we have $S_i^* \geq 0$. \square

Intuitively, the installations have no incentive to keep the inventory position always below zero. More specifically, it can never be optimal for a warehouse to keep the inventory too low ($S_0^* < -M + 1$) or for a retailer to hold negative inventory ($S_i^* < 0$). Proposition 3 provides a lower bound for S_0 and S_i . A loose upper bound for S_0^* is \hat{S}_0 such that $\text{Prob}(T_0 < L_0) < \epsilon$, where $\text{Prob}(T_0 < L_0)$ is the probability that T_0 takes a value smaller than L_0 for $S_0 = \hat{S}_0$ and ϵ is a tiny value that is close to zero. In others words, if $S_0 = \hat{S}_0$, the possibility that there is a delay at the warehouse is small enough. This implies that backlogging at the warehouse seldom occurs.

Now, we have the lower bound and upper bound for S_0 , and for each possible integer in the interval, we can search the corresponding optimal $S_i^*(S_0)$. Therefore, we can find the globally optimal parameters $S_0^*, S_1^*, \dots, S_N^*$.

We next work to derive tight bounds for the inventory parameters. First, we have the following lemma.

Lemma 1. $S_i^*(S_0) \geq S_i^*(S_0 + 1), \forall i = 1, \dots, N$.

Proof of Lemma 1. For a given S_0 , the problem of determining the inventory parameter S_i at retailer i is reduced to the classical Newsvendor problem whose leadtime is stochastic due to the possible delay

at the warehouse. Furthermore, if we consider the distribution of the random delay encountered at the warehouse, obviously, the cumulative distribution function for S_0 is stochastically larger than that for $S_0 + 1$. Therefore, the optimal inventory order-up-to level for S_0 is not less than that for $S_0 + 1$. \square

Lemma 1 implies that when the warehouse holds more inventory, the optimal order-up-to level of a retailer tends to decrease. The intuition is that, when the warehouse holds more inventory, the inventory shortage at the warehouse is less likely to happen, or, in other words, it is more likely for the warehouse to satisfy a retailer's order, and thus the retailers tends to hold less inventory.

Define

$$S_i^l = F_{1i}^{-1}\left(\frac{b_i}{h_i + b_i}\right), \quad \forall i = 1, \dots, N, \quad (18)$$

where F_{1i}^{-1} is the inverse cumulative demand distribution functions of retailer i over $L_i + 1$ periods. Obviously, S_i^l is the optimal inventory parameters for retailer i in a single-echelon model with leadtime L_i periods. In other words, S_i^l is the optimal S_i when there is no delay at the warehouse. Then, we have the following proposition.

Proposition 4 (Lower bound). S_i^l is the lower bound for S_i^* , i.e., $S_i^l \leq S_i^*$, where S_i^l is given in Equation (18).

Proof of Proposition 4. Recall that $S_i^*(S_0) \geq S_i^*(S_0 + 1)$. By induction, we have

$$S_i^*(S_0) \geq S_i^*(S_0 + 1) \geq S_i^*(S_0 + 2) \geq \dots \geq S_i^*(+\infty),$$

where $S_i^*(+\infty)$ is the optimal S_i when S_0 goes to infinity. When S_0 is sufficiently large, we know that there is no delay at the warehouse, and the optimal S_i is S_i^l by definition, i.e. $S_i^*(+\infty) = S_i^l$. \square

The intuition of Proposition 4 is straightforward. If the warehouse has enough inventory, retailers have no incentive to keep much inventory; on the contrary, if there is a long delay at the warehouse, retailers have much incentive to keep more inventory to avoid possible shortages. The biggest difference of retailer i in our model from that in the single-echelon model is that retailer faces possible delay at the warehouse in case of shortages. This delay can be as long as L_0 periods, and as short as 0 period. Then, retailer i can be treated as a single-echelon retailer with leadtime $L_i + \Delta$ periods ($0 \leq \Delta \leq L_0$). Therefore, S_i^l is a lower bound for S_i^* .

Remark 6. In [34], there is also an upper bound for S_i , however, their upper bound does not hold in our problem. This is because, in [34], the maximum possible delay at the warehouse is L_0 periods, while in our problem the maximum delay can be longer than L_0 periods due to the existence of MOQ constraint.

We next provide a more tight upper bound for S_0 .

Proposition 5 (Upper bound). S_0^u is the upper bound for S_0^* , i.e., $S_0^* \leq S_0^u$, where S_0^u is obtained by minimizing $L(S_0, S_1^l, \dots, S_N^l)$ over the set $\Psi_u = \{S_0 \leq \hat{S}_0 | S_i^*(S_0) = S_i^l, \forall i = 1, \dots, N\}$.

Proof of Proposition 5. We know that $S_i^*(S_0^u) = S_i^l$ and $S_i^*(S_0^l) = S_i^u$. For $S_0 > S_0^u$, the corresponding optimal S_i is $S_i^*(S_0) = S_i^l$ due to Lemma 1 and Proposition 4. By the definition of S_0^u , we can have $L(S_0, S_1^l, \dots, S_N^l) \geq L(S_0^u, S_1^l, \dots, S_N^l)$ for such S_0 . Therefore, the optimal S_0 must satisfy $S_0^* \leq S_0^u$. \square

Recall that $S_0^* \geq -M + 1$, thus S_0^* must satisfy

$$-M + 1 \leq S_0^* \leq S_0^u,$$

Moreover, we know that, for each given S_0 , $L(S_0, S_1, \dots, S_N)$ is convex in each S_i , and $S_i^* \geq S_i^l$. Clearly, to obtain the optimal inventory parameters, we only need to enumerate $S_0 \in \mathbb{Z}^{[-M+1, S_0^u]}$, and the corresponding $S_i^*(S_0)$.

7. Conclusions

A two-echelon inventory system with a MOQ requirement is analytically studied in this paper, where there are one warehouse and multiple retailers. In each period, retailers replenish their stocks from the warehouse, and the warehouse in turn replenishes from an external supplier, but the warehouse must order either none or a quantity no less than MOQ. The MOQ requirement is quite common in China and other low cost manufacturing countries, where low profit margins force manufacturers to pursue large production quantities to break even and achieve economies of scale. For the retailers, we assume they adopt the base stock policy. However, for the warehouse, due to the MOQ requirement, we propose a new heuristic ordering policy, called refined base stock policy, which particularly considers the case of negative base stock S .

Our objective is to minimize the long-run average system cost over an infinite horizon. To this end, we present a position-based cost-accounting scheme to exactly evaluate the system cost. Specifically, once the warehouse places an order, each ordered unit will have a corresponding position at the warehouse; and for each of such positions, we associate it with a corresponding expected system cost, including holding cost at the warehouse, and holding and backlogging cost at retailers. Built upon this position-based scheme, we derive lower bound and upper bound for the inventory parameters, which facilitate the search of optimal policy. In addition, by optimizing the system, we can also derive some managerial insights. For example, when the warehouse holds more inventory, the optimal order-up-to level of a retailer tends to decrease.

Author Contributions: H.S., T.T. and H.Z. developed the model, conducted the analysis, and wrote the paper together.

Funding: This research was supported by the MOE (Ministry of Education in China) Project of Humanities and Social Sciences (Grant No. 17YJC630120), the National Natural Science Foundation of China (Grant Nos. 71801231, 71902018, and 71601033), and the Fundamental Research Funds for the Central Universities (Grant No. 17wkpy18).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhao, Y.; Katehakis, M.N. On the structure of optimal ordering policies for stochastic inventory systems with minimum order quantity. *Probab. Eng. Inf. Sci.* **2006**, *20*, 257–270. [[CrossRef](#)]
2. Graves, S.C. A multiechelon inventory model with fixed replenishment intervals. *Manag. Sci.* **1996**, *42*, 1–18. [[CrossRef](#)]
3. Simchi-Levi, D.; Zhao, Y. Performance evaluation of stochastic multi-echelon inventory systems: A survey. *Adv. Oper. Res.* **2011**, *2012*, 126254. [[CrossRef](#)]
4. Clark, A.J.; Scarf, H. Optimal policies for a multi-echelon inventory problem. *Manag. Sci.* **1960**, *6*, 475–490. [[CrossRef](#)]
5. Federgruen, A.; Zipkin, P. Computational issues in an infinite-horizon, multiechelon inventory model. *Oper. Res.* **1984**, *32*, 818–836. [[CrossRef](#)]
6. Chen, F. Optimal policies for multi-echelon inventory problems with batch ordering. *Oper. Res.* **2000**, *48*, 376–389. [[CrossRef](#)]
7. Shang, K.H.; Zhou, S.X. Optimal and heuristic echelon (r, nQ, T) policies in serial inventory systems with fixed costs. *Oper. Res.* **2010**, *58*, 414–427. [[CrossRef](#)]
8. Eppen, G.; Schrage, L. Centralized ordering policies in a multi-warehouse system with lead times and random demand. *Multi-Level Prod. Control. Syst. Theory Pract.* **1981**, *16*, 51–67.
9. Federgruen, A.; Zipkin, P. Allocation policies and cost approximations for multilocation inventory systems. *Naval Res. Logist. Q.* **1984**, *31*, 97–129. [[CrossRef](#)]
10. Federgruen, A.; Zipkin, P. Approximations of dynamic, multilocation production and inventory problems. *Manag. Sci.* **1984**, *30*, 69–84. [[CrossRef](#)]
11. Van Houtum, G.; Inderfurth, K.; Zijm, W.H. Materials coordination in stochastic multi-echelon systems. *Eur. J. Oper. Res.* **1996**, *95*, 1–23. [[CrossRef](#)]

12. Verrijdt, J.; De Kok, A. Distribution planning for a divergent depotless two-echelon network under service constraints. *Eur. J. Oper. Res.* **1996**, *89*, 341–354. [\[CrossRef\]](#)
13. Van der Heijden, M.; Diks, E.; De Kok, A. Stock allocation in general multi-echelon distribution systems with (R, S) order-up-to-policies. *Int. J. Prod. Econ.* **1997**, *49*, 157–174. [\[CrossRef\]](#)
14. Diks, E.; De Kok, A. Optimal control of a divergent multi-echelon inventory system. *Eur. J. Oper. Res.* **1998**, *111*, 75–97. [\[CrossRef\]](#)
15. Doğru, M.K.; De Kok, A.; Van Houtum, G. A numerical study on the effect of the balance assumption in one-warehouse multi-retailer inventory systems. *Flex. Serv. Manuf. J.* **2009**, *21*, 114–147. [\[CrossRef\]](#)
16. Marklund, J.; Rosling, K. Lower bounds and heuristics for supply chain stock allocation. *Oper. Res.* **2012**, *60*, 92–105. [\[CrossRef\]](#)
17. Axsäter, S. Optimization of order-up-to-s policies in two-echelon inventory systems with periodic review. *Nav. Res. Logist.* **1993**, *40*, 245–253. [\[CrossRef\]](#)
18. Forsberg, R. Optimization of order-up-to-S policies for two-level inventory systems with compound Poisson demand. *Eur. J. Oper. Res.* **1995**, *81*, 143–153. [\[CrossRef\]](#)
19. Axsäter, S.; Marklund, J.; Silver, E.A. Heuristic methods for centralized control of one-warehouse, N-retailer inventory systems. *Manuf. Serv. Oper. Manag.* **2002**, *4*, 75–97. [\[CrossRef\]](#)
20. Jackson, P.L. Stock allocation in a two-echelon distribution system or “what to do until your ship comes in”. *Manag. Sci.* **1988**, *34*, 880–895. [\[CrossRef\]](#)
21. van der Heijden, M.C. Multi-echelon inventory control in divergent systems with shipping frequencies. *Eur. J. Oper. Res.* **1999**, *116*, 331–351. [\[CrossRef\]](#)
22. Erkip, N. Approximate Policies In Multi-Echelon Inventory Systems. Ph.D. Thesis, Stanford University, Stanford, CA, USA, 1984.
23. Jackson, P.L.; Muckstadt, J.A. Risk pooling in a two-period, two-echelon inventory stocking and allocation problem. *Nav. Res. Logist.* **1989**, *36*, 1–26. [\[CrossRef\]](#)
24. McGavin, E.J.; Schwarz, L.B.; Ward, J.E. Two-interval inventory-allocation policies in a one-warehouse N-identical-retailer distribution system. *Manag. Sci.* **1993**, *39*, 1092–1107. [\[CrossRef\]](#)
25. Sherbrooke, C.C. METRIC: A multi-echelon technique for recoverable item control. *Oper. Res.* **1968**, *16*, 122–141. [\[CrossRef\]](#)
26. Sherbrooke, C.C. VARI-METRIC: Improved approximations for multi-indenture, multi-echelon availability models. *Oper. Res.* **1986**, *34*, 311–319. [\[CrossRef\]](#)
27. Costantino, F.; Di Gravio, G.; Tronci, M. Multi-echelon, multi-indenture spare parts inventory control subject to system availability and budget constraints. *Reliab. Eng. Syst. Saf.* **2013**, *119*, 95–101. [\[CrossRef\]](#)
28. Patriarca, R.; Costantino, F.; Di Gravio, G. Inventory model for a multi-echelon system with unidirectional lateral transshipment. *Expert Syst. Appl.* **2016**, *65*, 372–382. [\[CrossRef\]](#)
29. Patriarca, R.; Costantino, F.; Di Gravio, G.; Tronci, M. Inventory optimization for a customer airline in a Performance Based Contract. *J. Air Transp. Manag.* **2016**, *57*, 206–216. [\[CrossRef\]](#)
30. Costantino, F.; Di Gravio, G.; Patriarca, R.; Petrella, L. Spare parts management for irregular demand items. *Omega* **2018**, *81*, 57–66. [\[CrossRef\]](#)
31. Simon, R.M. Stationary properties of a two-echelon inventory model for low demand items. *Oper. Res.* **1971**, *19*, 761–773. [\[CrossRef\]](#)
32. Shanker, K. Exact analysis of a two-echelon inventory system for recoverable items under batch inspection policy. *Nav. Res. Logist. Q.* **1981**, *28*, 579–601. [\[CrossRef\]](#)
33. Graves, S.C. A multi-echelon inventory model for a repairable item with one-for-one replenishment. *Manag. Sci.* **1985**, *31*, 1247–1256. [\[CrossRef\]](#)
34. Axsäter, S. Simple solution procedures for a class of two-echelon inventory problems. *Oper. Res.* **1990**, *38*, 64–69. [\[CrossRef\]](#)
35. Axsäter, S. Exact and approximate evaluation of batch-ordering policies for two-level inventory systems. *Oper. Res.* **1993**, *41*, 777–785. [\[CrossRef\]](#)
36. Axsäter, S. Simple evaluation of echelon stock (R, Q) policies for two-level inventory systems. *IIE Trans.* **1997**, *29*, 661–669. [\[CrossRef\]](#)
37. Axsäter, S. Exact analysis of continuous review (R, Q) policies in two-echelon inventory systems with compound Poisson demand. *Oper. Res.* **2000**, *48*, 686–696. [\[CrossRef\]](#)

38. Forsberg, R. Exact evaluation of (R, Q)-policies for two-level inventory systems with Poisson demand. *Eur. J. Oper. Res.* **1997**, *96*, 130–138. [[CrossRef](#)]
39. Levi, R.; Pál, M.; Roundy, R.O.; Shmoys, D.B. Approximation algorithms for stochastic inventory control models. *Math. Oper. Res.* **2007**, *32*, 284–302. [[CrossRef](#)]
40. Levi, R.; Roundy, R.O.; Shmoys, D.B.; Truong, V.A. Approximation algorithms for capacitated stochastic inventory control models. *Oper. Res.* **2008**, *56*, 1184–1199. [[CrossRef](#)]
41. Levi, R.; Roundy, R.O.; Truong, V.A.; Wang, X.S. Provably near-optimal balancing policies for multi-echelon stochastic inventory control models. *Math. Oper. Res.* **2016**, *42*, 256–276. [[CrossRef](#)]
42. Fisher, M.; Raman, A. Reducing the cost of demand uncertainty through accurate response to early sales. *Oper. Res.* **1996**, *44*, 87–99. [[CrossRef](#)]
43. Zhou, B.; Zhao, Y.; Katehakis, M.N. Effective control policies for stochastic inventory systems with a minimum order quantity and linear costs. *Int. J. Prod. Econ.* **2007**, *106*, 523–531. [[CrossRef](#)]
44. Kiesmüller, G.P.; De Kok, A.; Dabia, S. Single item inventory control under periodic review and a minimum order quantity. *Int. J. Prod. Econ.* **2011**, *133*, 280–285. [[CrossRef](#)]
45. Zhu, H.; Liu, X.; Chen, Y.F. Effective inventory control policies with a minimum order quantity and batch ordering. *Int. J. Prod. Econ.* **2015**, *168*, 21–30. [[CrossRef](#)]
46. Fera, M.; Fruggiero, F.; Lambiase, A.; Macchiaroli, R.; Miranda, S. The role of uncertainty in supply chains under dynamic modeling. *Int. J. Ind. Eng. Comput.* **2017**, *8*, 119–140. [[CrossRef](#)]
47. Miranda, S.; Fera, M.; Iannone, R.; Riemma, S. A multi-item constrained EOQ calculation algorithm with exit condition: A comparative analysis. *IFAC-PapersOnLine* **2015**, *48*, 1314–1319. [[CrossRef](#)]
48. Fera, M.; Macchiaroli, R.; Iannone, R.; Miranda, S.; Riemma, S. Economic evaluation model for the energy Demand Response. *Energy* **2016**, *112*, 457–468. [[CrossRef](#)]
49. Zheng, Y.S.; Federgruen, A. Finding optimal (s, S) policies is about as simple as evaluating a single policy. *Oper. Res.* **1991**, *39*, 654–665. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).