Stock allocation in a two-echelon distribution network under service constraints

Citation for published version (APA):
Stock Allocation in a Two-echelon Distribution Network under Service Constraints

A.G. de Kok*
A.G. Lagodimos**
H.P. Seidel***

Research Report TUE/BDK/LBS/94-03
April 1994

* Graduate School of Industrial Engineering and Management Science
  Eindhoven University of Technology
  P.O. Box 513, Paviljoen F16
  NL-5600 MB Eindhoven
  The Netherlands

** Vector Consultants
  175 El. Venizelou Ave.
  171 23 N. Smirni
  Greece

*** KPMG Consultants
  Beemdstraat 1
  5653 MA Eindhoven
  The Netherlands

This paper should not be quoted or referred to without the prior written permission of the author
Abstract

Consideration is given to a two-echelon distribution network focusing on the determination of cost optimal stocknorms under service level constraints. The network we study allows holding stock at both echelons and implements periodic review order-up-to-S policies at both echelons. In the event of material shortages at the depot material rationing is needed. The policy we use allows for each stockpoint to attain a desired service level, which need not be the same across all end stockpoints. We develop distribution-free expressions describing the system dynamics and models for the total system cost and the service constraints. We propose heuristic algorithms for service level computations and use an empirical optimisation procedure to determine the optimal stocknorms under different system settings. Many theoretical results are validated through simulation, which demonstrates the adequacy of the approximations involved in the analysis.
1. Introduction

The problem of establishing appropriate stocknorms for the control of multi-echelon production and distribution networks is well recognised. Aiming at assisting practitioners with this task, many research papers have appeared to date, dealing with various facets of the problem under different operating environments. Comprehensive reviews of this particularly active area of research have often appeared (see, for example, Clark [1], Nahmias [2] and Inderfurth[3]). Leaving aside the particularities of the networks studied, two variations of the problem have received most attention: namely the determination of cost-optimal [5]-[12] and service-related [13]-[19] stocknorms.

One variation of the problem that has received considerably less attention concerns the determination of cost-optimal stocknorms under service level constraints [20]-[21]. Bearing in mind the difficulty in specifying stockout costs in practice (see, for example, Schneider [22] and Tijms and Groenevelt [23]), this approach offers a good practical compromise between the two previous extremes. Following this approach, the aim of this paper is the determination of optimal stocknorms for a two-echelon distribution network under individual fill rate constraints at each end stockpoint.

The distribution network we study consists of one central depot supplying N local warehouses (end stockpoints) where the external demand is realised. The ordering policies used at all stockpoints are echelon-based periodic review order-up-to \( S \) policies. It is well established now that these policies correspond in practice to those implemented by base stock control and, under certain conditions, by material requirements planning systems (see Lagodimos [24] and Axsater and Rosling [25]).

In contrast to many previous studies that considered the depot as an administrative unit that cannot hold inventory, all stockpoints are allowed to hold inventory. At the start of each period, local warehouses release orders aimed at restoring their inventory position up to \( S \). If the depot inventory is sufficient, all orders are completely satisfied. Otherwise, the available depot inventory is rationed among requisitioning local warehouses according to some predetermined rationing policy. Several rationing policies have been studied in the past [6], [8], [18], [26], [27]. Details of the policy we consider are given later. It is important to stress, that, in contrast to other rationing policies, this policy allows for attaining different service targets at end stockpoints.

The use of material rationing for allocating scarce depot inventory is not a universally accepted mechanism for regulating the system operation. Another mechanism, which has mainly been studied for networks using continuous review ordering policies, is to delay end stockpoint orders until material becomes available at the depot [14], [16]. In fact, all other papers that (to our knowledge) deal with the problem we study here for networks using periodic review policies and allowing stock holding throughout have adopted this mechanism. Specifically, Rosenbaum [20] assumes partial end stockpoint replenishment orders fulfilment and delay of their unsatisfied portion, while Schneider et al. [21] assume that the replenishment order is delayed when it cannot be satisfied in full.

The remainder of this paper is organised as follows. Section 2 gives the details of the system operation and presents the corresponding dynamic equations. Section 3 deals with the problem of modelling and evaluating the fill rate at individual end stockpoints and presents heuristic algorithms for performing the evaluation. In section 4 we present simulation results showing the accuracy of the approximations derived in section 3. Section 5 presents and solves the cost optimisation problem under fill rate constraints. Finally, section 6 discusses the findings and gives the conclusions of this research together with recommendations for further work.

2. System operation

In this section we give the operating assumptions of the system we study and discuss the details of the material rationing policy we consider. We also develop equations describing the system dynamics which form the basis for establishing cost and service models.
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>(d_i)</td>
<td>demand at end stockpoint (i) in an arbitrary period</td>
</tr>
<tr>
<td>(\mu_i)</td>
<td>(E(d_i))</td>
</tr>
<tr>
<td>(\sigma_i)</td>
<td>(\sigma(d_i))</td>
</tr>
<tr>
<td>(D)</td>
<td>aggregated demand at all stockpoints in an arbitrary period</td>
</tr>
<tr>
<td>(d_{i,t+k})</td>
<td>demand at end stockpoint (i) in ((t,t+k])</td>
</tr>
<tr>
<td>(D_{t,t+k})</td>
<td>aggregate system demand in ((t,t+k])</td>
</tr>
<tr>
<td>(L)</td>
<td>depot lead time</td>
</tr>
<tr>
<td>(l_i)</td>
<td>lead time at end stockpoint (i)</td>
</tr>
<tr>
<td>(I^i_t)</td>
<td>inventory position at end stockpoint (i) after ordering at time (t)</td>
</tr>
<tr>
<td>(I_t)</td>
<td>aggregated upper echelon inventory after ordering at time (t); (I_t = \sum_{k=1}^{N} I^k_t)</td>
</tr>
<tr>
<td>(SS^i)</td>
<td>planned safety stock at end stockpoint (i)</td>
</tr>
<tr>
<td>(SS^0)</td>
<td>planned safety stock at the depot</td>
</tr>
<tr>
<td>(SS)</td>
<td>system wide safety stock, (SS = SS^0 + \sum_{k=1}^{N} SS^k)</td>
</tr>
<tr>
<td>(R)</td>
<td>mean aggregate demand over depot lead time; (R = L \sum_{k=1}^{N} \mu_k)</td>
</tr>
<tr>
<td>(U^i_t)</td>
<td>projected net inventory of end stockpoint (i) at the end of period (t+l_t+1)</td>
</tr>
<tr>
<td>(U_t)</td>
<td>system wide projected net inventory; (\sum_{k=1}^{N} U^k_t)</td>
</tr>
<tr>
<td>(p_i)</td>
<td>allocation fraction for end stockpoint (i)</td>
</tr>
<tr>
<td>(q^i_t)</td>
<td>quantity allocated to end stockpoint (i) at the beginning of period (t)</td>
</tr>
<tr>
<td>(S^i)</td>
<td>order-up-to level at end stockpoint (i)</td>
</tr>
<tr>
<td>(S)</td>
<td>aggregate upper echelon order-up-to level; (S = \sum_{k=1}^{N} S^i)</td>
</tr>
<tr>
<td>(W)</td>
<td>echelon order-up-to level at the depot</td>
</tr>
<tr>
<td>(\Delta)</td>
<td>(W \cdot S)</td>
</tr>
<tr>
<td>(\beta^*_i)</td>
<td>target fill rate for end stockpoint (i)</td>
</tr>
</tbody>
</table>

Table 1. List of notation used
2.1 Operating assumptions

The notation we use is shown in Table 1. Most of our analysis is based on the following assumptions:

1. The demand is imposed at the local warehouses and all unsatisfied demand is backordered.
2. No transhipments between local warehouses are allowed.

Some additional assumptions concerning the distribution of the imposed demand will be introduced when needed. As stated in the Introduction, our analysis considers networks using order-up-to $S$ policies throughout. Note that all the results also apply to systems using nested $(S,T)$ policies at both echelons, provided that lead times are integer multiples of the review period $T$.

2.2 Material rationing

The rationing policy we consider will be referred to as Appropriate Share (AS) rationing. It can be viewed as an adaptation of an allocation policy introduced by de Kok [18] for two-echelon depotless (i.e., where the depot cannot hold inventory) networks, to the systems we study here. The purpose of AS rationing is to ensure that a prespecified independent service target can be attained at end stockpoint.

In order to define AS rationing rigorously we introduce a quantity $U_t^i$, which will be referred to as the projected net inventory of end stockpoint $i$ at the end of period $t+1$:

$$U_t^i = I_t^i - (I_{t+1}^i)\mu_i.$$

$U_t^i$ represents our best estimate for the net inventory magnitude at the end of period $t+1$ as known at the beginning of period $t$. Note that, in cases where stockpoint $i$ reaches its order-up-to level (i.e., $I_t^i = S$), by the standard definition, $U_t^i$ becomes identical to the planned safety stock $SS_i$ at this stockpoint. In the following we also use the systemwide projected net inventory $U_t$, which is the sum of the projected net inventory of all end stockpoints.

AS rationing: ration the available depot inventory so that the projected net inventory $U_t^i$ at each end stockpoint $i$ over the systemwide projected net inventory $U_t$ equals a prespecified allocation fraction $p_i$:

$$\frac{U_t^i}{U_t} = p_i. \quad (1)$$

Clearly, we need that $\sum_{j=1}^{N} p_j = 1$. In the sequel, we refer to the positive material quantity $q_t^i$, allocated by the AS rationing at stockpoint $i$ and satisfying its corresponding equation for $p_i$, as its appropriate share. We can easily show that:

$$q_t^i = I_t^i - (I_{t+1}^i - d_{t+1}^i).$$

The magnitude of each $p_i$ will not be known in advance. In fact, for the general definition of AS rationing given, we have $2N+1$ decision variables which should be determined to obtain the desired system performance; namely, $W_t,\{S_i\}$ and $\{p_j\}$. In the following we concentrate on a restricted version of AS rationing, referred to as Consistent Appropriate Share (CAS) rationing, which drastically reduces the decision variables involved.
CAS rationing: ration the depot inventory according to AS rationing by choosing the allocation fractions \( p_i \), such that

\[
p_i = \frac{S^i - (I_i + 1)\mu_i}{\sum_{j=1}^N S^j - (I_j + 1)\mu_j}
\]

The rational of CAS rationing is that, at any time it attempts to keep the ratio of the projected net inventory at any end stockpoint over the system projected net inventory constant. This happens both when there is sufficient depot material to satisfy all upper echelon orders and when material rationing is necessary. With CAS rationing the variables which need to be determined are reduced to \( N+2 \); namely, \( W, S \) and \( \{p_j\} \).

Some general comments on AS rationing are necessary. Since it uses information on the inventory content of all end stockpoints, AS rationing is a push rationing policy (see classification by Silver and Peterson [28]). Interestingly, AS rationing (specifically CAS rationing) constitutes a generalisation of the well known Fair Share (FS) rationing policy. This policy was introduced for a depotless network by Eppen and Schrage [6] and later extended for the networks we study here (see van Donselaar and Wijngaard [15] and Lagodimos [34]). In fact, for normally distributed demands and when we aim to achieve identical stockout probabilities at all end stockpoints (as assumed by FS rationing), then CAS rationing reduces to FS rationing. We can easily show that, in this case, the allocation fraction for each end stockpoint \( i \), becomes:

\[
p_i = \frac{\sigma_i}{\sum_{j=1}^N \sigma_j}
\]

as required by the FS rationing policy. We stress that AS rationing is more general since it holds for any demand distribution and can be used for any service definition desired (not just the probability of stockout).

### 2.3 System dynamics

A prerequisite for developing multi-echelon cost and service models is to characterise the stochastic behaviour of the inventory content at each end stockpoint over time. We start by stating (without proof) an exact result applicable to any rationing policy.

**Proposition** (Lagodimos [30]): The aggregate inventory position of the upper echelon, immediately after ordering at period \( t \), is given by:

\[
I_t = \min (W - D_{t-L,t}, S)
\]

Therefore, the aggregate upper echelon inventory is fully determined by the aggregate system inventory (\( L \) time periods earlier) and the total system demand in \([t-L,t)\). In order, however, to determine the inventory content of individual stockpoints the effects of the rationing policy need to be considered. Similarly to de Kok [18], we introduce the following assumption:
Generalised Balanced Inventories (GBI) assumption:

In the event of material rationing, each end stockpoint always obtains its appropriate share \( q_i^j \), i.e. 
\[ P\{q_i^j \geq 0\} = 1, \; i=1, \ldots, N \]

In the same way that AS rationing generalises FS rationing, the GBI assumption generalises the Balanced Inventories (BI) assumption used in conjunction with FS rationing ([16], [26], [30], [31]). A discussion of the relationship between these assumptions and a first investigation of the severity of the GBI assumption for depotless two-echelon networks under AS rationing is given by Verrijdt and de Kok [33].

The GBI assumption allows us to model the inventory content of each end stockpoint as a function of the aggregate upper echelon inventory only. Using (1), (2) and (3), after some straightforward algebra, we obtain for \( I_i^j \):

\[
I_i^j = p_i \left[ S - (D_i - L_i - (W - S))^+ \right] + (l_i + 1) \mu_i, \tag{4}
\]

where \( a^+ = \max(0, a) \). Observe that the above is a function of aggregate variables only. Another useful expression for \( I_i^j \) is obtained by introducing the quantity

\[ \Delta := W - S, \]

which represents the difference between the echelon order-up-to level at the depot and the aggregate order-up-to level at the upper echelon. Substituting \( \Delta \) in (4), we can re-write \( I_i^j \) as:

\[
I_i^j = c_i - p_i (D_i - L_i - \Delta)^+, \tag{5}
\]

where \( c_i = p_i (S - \sum_{j=1}^{N} (l_j + 1) \mu_j) + (l_i + 1) \mu_i \).

There are two limiting cases of particular interest. Firstly, when \( W \) becomes very large relatively to \( S \) (i.e. \( \Delta \to \infty \)), the two-echelon network is decomposed in \( N \) independent single echelon systems. Secondly, when \( W = S \) (i.e. \( \Delta = 0 \)), the depot never holds any inventory and the system becomes identical to the two-echelon depotless system studied by de Kok [18] (this follows directly from (3)).

3. Service determination

We now present expressions for the fill rate at individual end stockpoints and deal with the evaluation of stocknorms for attaining desired fill rate levels at end stockpoints. The heuristic algorithms we propose to determine these stocknorms are validated through simulation. While our analysis is valid for any demand distribution, many results use the assumption that end stockpoints demand over their respective lead time follow a Mixed Erlang (ME) distribution. The reasons for choosing this distribution is that it can be easily fitted to any mean and variance combination of random variables (see Tijms [34]) and that it leads to computationally tractable results.
3.1 General service models

The fill rate ($\beta$) is defined as the fraction of the demand satisfied directly from stock and is one of the most popular measures of customer service used in practice. Since it effectively measures the amount that a system is out of stock, it is often used as a surrogate for stockout costs per unit of stockout. In general (see Schneider [22]):

$$\beta = \frac{E[(d_t^i - l_t^i)\cdot]}{\mu}$$

For the network under CAS rationing, substitution of (5) above, yields a distribution-free expression for the fill rate at any end stockpoint $i$:

$$\beta_i = 1 - \frac{E[p_i(D_{t-L,t} - \Delta)^+] - E[d_{t+i-1}^i - c_i^i - E[p_i(D_{t-L,t} - \Delta)^+] - E[d_{t+i-1}^i - c_i^i] \mu_i}$$

(6)

where $c_i$ is defined in (5). Observe that each term in the numerator of (6) has the form $E[(X - \Delta)^+]$.

If $X$ and $Y$ are ME distributed, these terms can be routinely evaluated. We now give the result for the case where $X$ and $Y$ follow a pure Erlang distributions. For ME distributions we only have to evaluate weighted combinations of the expression corresponding to the pure Erlang case. Let $X$ and $Y$ follow an $E_{k,\mu}$ and an $E_{l,\nu}$ distribution respectively. We can show that

$$E[(X - \Delta)^+ Y - c]^+] = (1 - \sum_{n=0}^{k-1} \frac{e^{-\mu \Delta} \left(\frac{\mu}{\nu}\right)^i}{i!} \sum_{j=0}^{L-1} \frac{(l-j)f}{j!} e^{-\nu c \left(\frac{\nu}{\nu}\right)^j}$$

$$+ \sum_{l=0}^{k-1} \frac{(l-i)}{i!} e^{-\mu (\Delta + c) \left(\frac{\Delta + \mu}{\nu}\right)^i}$$

$$- \sum_{l=0}^{k-1} \frac{i}{l!} e^{-\mu (\Delta + c) \left(\frac{\Delta + \mu}{\nu}\right)^i}$$

$$+ \sum_{j=0}^{l-1} \sum_{m=0}^{n+1} \frac{(l-j) e^{-\nu (\Delta + c)} (-1)^n \mu^{km-n} \left[\frac{\nu (\Delta + c)}{(\mu - \nu)^{n+k-1}} \left(\frac{\nu (\Delta + c)}{(\mu - \nu)^{n+k-1}} \right)^i \right] n!}{j! (j-n)! \mu^{k+n-1} \nu^{k+n-1} \left(\frac{\nu (\Delta + c)}{(\mu - \nu)^{n+k-1}} \right)^i}$$

which can be evaluated using numerical methods.

3.2 Heuristic algorithms

On the basis of the previous analysis, the problem of determining stocknorms which ensure individual fill rate targets at all end stockpoints, corresponds to the solution of the following system of $N+1$ non-linear equations with $N+2$ unknowns, $\Delta$ and $\{p_i\}$:
where \( f(S+\Delta S,p_j) \) is the right hand side of (6) and \( \beta^*_i \) is the prespecified target fill rate with \( 0 \leq \beta^*_i \leq 1 \). Clearly, the above system has an infinite number of solutions. However, given a value of \( \Delta \), the system has a unique solution.

We now present three heuristic algorithms which provide accurate and fast solutions to this problem. These constitute adaptations of previous algorithms developed for depotless two-echelon networks to the present system. All heuristics make use of the solution for the problem when \( \Delta \) tends to infinity. In this case the problem decomposes to the solution of \( N \) independent equations (with one unknown \( S^i \)) of the form:

\[
\beta^*_i = 1 - \frac{E[(d_{i,t+1}^i + S^i)^+] - E[(d_{i,t+1}^i - S^i)^+]}{\mu_i}.
\]

**Heuristic 1**

This is an adaptation of the heuristic introduced by de Kok [18], and later refined by Verrijdt and de Kok [33], for a depotless two-echelon network.

**Step 1** Choose \( S \) as \( S = \sum_{i=1}^{N} S^i \), where all \( S^i \) are calculated using (8).

**Step 2** Calculate \( P_i \) for the current value of \( S \) from \( f(S+\Delta S,p_j) = \beta^*_i \)

(this can be done using bisection, since \( f(S+\Delta, S, p_j) \) is an increasing function of \( p_j \)).

**Step 3** If \( \sum_{j=1}^{N} p_j = 1 \) (close enough), then Stop; Else

if \( \sum_{j=1}^{N} P_j > 1 \), then increase \( S \)

if \( \sum_{j=1}^{N} P_j < 1 \), then decrease \( S \)

Stop Use \( S \) as an approximation for the solution of (7)

The above algorithm is a nested bisection scheme. The computational burden is related to step 2, where we have to solve \( N \) equations through bisection and iterate several times, depending on the stopping criterion in step 3.
Heuristic 2

This heuristic algorithm was also proposed in de Kok [18] and is based on the exogenous determination of \{p_j\}, obtained by assuming that the system consists of \(N\) independent single echelon networks. These \{p_j\} values are then used throughout the calculations. The justification for using these values stems from the fact that they are exact for \(\Delta = \infty\). However, simulation results have demonstrated that \{p_j\} are quite insensitive to both the value of \(\Delta\) and \(L\).

Step 1 Calculate \(p_i\) using (8) and then (2).

Step 2 Calculate \(S^i\) from \(f(S + \Delta, S, p_i) = \beta_i^*\).

Step 3 Calculate \(S\) from \(S = \frac{1}{N} \sum_{j=1}^{N} S^j\).

We should note that Heuristic 2 involves only \(2N\) one-dimensional bisections, associated with step 2.

Heuristic 3

This heuristic is otherwise identical to Heuristic 2, except that it assumes that the variables in the nominator of (6) are ME or gamma distributed. Observe that \(p_i(D_{t-L,t} - \Delta)^+\) in (6) has a probability mass in zero and so may not be ME distributed. However, if we convolute this variable with \(d_{t,t+1}^i\) or \(d_{t,t+1}^i\) (to obtain the numerator of (6)), the probability mass in zero disperses along the positive axis. Thus, we still assume that \(p_i(D_{t-L,t} - \Delta)^+ \pm d_{t,t+1}^i\) and \(p_i(D_{t-L,t} - \Delta)^+ \pm d_{t,t+1}^i\) are ME distributed.

This assumption allows us to use a fast inversion algorithm developed by Verrijdt and de Kok [33], outlined in the Appendix. We only need to compute the first two moments of \(p_i(D_{t-L,t} - \Delta)^+ \pm d_{t,t+1}^i\) and \(p_i(D_{t-L,t} - \Delta)^+ \pm d_{t,t+1}^i\), a routine matter, if we assume that \(D_{t-L,t}\) is ME distributed (see Appendix). Thus we obtain heuristic 3:

Step 1 Calculate \(p_i\) from (8) and (2).

Step 2 Calculate \(S^i\) from \(f(S + \Delta, S, p_i) = \beta_i^*\), using the inversion algorithm in Verrijdt and de Kok [33].

Step 3 Calculate \(S\) from \(S = \frac{1}{N} \sum_{j=1}^{N} S^j\).

This heuristic is much faster than Heuristic 2, especially for situations where the ME distribution have very large scale parameters (e.g. large end stockpoint lead times). Simulations have shown that it yields almost identical results as Heuristic 2.
4. Simulation results

We now test the validity of the proposed service models. Since the differences between the results we found for different heuristics are negligible (in all cases tested), we only report simulations obtained using Heuristic 3. For more extensive simulation experiments we refer to Seidel [35]. We used the following input data:

* \( N = 3 \)
* \( \beta_1^* = 0.95, \beta_2^* = 0.95, \beta_3^* = 0.90 \)
* 5 demand sets
  * \( D = 1: \) \((\mu_i, \sigma_i) = (100, 25)\) \(i = 1, 2, 3\)
  * \( D = 2: \) \((\mu_i, \sigma_i) = (100, 75)\) \(i = 1, 2, 3\)
  * \( D = 3: \) \((\mu_i, \sigma_i) = (100, 200)\) \(i = 1, 2, 3\)
  * \( D = 4: \) \((\mu_1, \sigma_1) = (100, 25)\)
    \((\mu_2, \sigma_2) = (100, 50)\)
    \((\mu_3, \sigma_3) = (100, 75)\)
  * \( D = 5: \) \((\mu_1, \sigma_1) = (100, 100)\)
    \((\mu_2, \sigma_2) = (100, 150)\)
    \((\mu_3, \sigma_3) = (100, 200)\)
* Lead times \([L, l_1, l_2, l_3] \)
  * \( l_1 = 1: \) \([L, l_1, l_2, l_3] = [5, 1, 2, 3]\)
  * \( l_1 = 2: \) \([L, l_1, l_2, l_3] = [4, 2, 3, 4]\)
  * \( l_1 = 3: \) \([L, l_1, l_2, l_3] = [3, 3, 4, 5]\)
* Squared coefficient of variation of the end stockpoint lead times \([vc_1, vc_2, vc_3]\), varied at 3 levels:
  * \(vc = 1: \) \([0.0, 0.0, 0.0]\) (deterministic lead times)
  * \(vc = 2: \) \([0.2, 0.2, 0.2]\) (stochastic lead times)
  * \(vc = 3: \) \([0.0, 0.2, 0.4]\) (stochastic lead times)

For all input combinations, we determined the order-up-to levels, for seven \( \Delta \) values, defined as multiples of \( R = L \sum_{j=1}^{N} \mu_j \), as follows:

\[ \Delta_1 = 0, \Delta_2 = 0.4R, \Delta_3 = 0.8R, \Delta_4 = R, \Delta_5 = 1.2R, \Delta_6 = 1.4R, \Delta_7 = 4R \]

For each case, we simulated the system over 50,000 periods (for the deterministic lead times) and over 100,000 periods (in case of stochastic lead times). In all cases the end stockpoints period demand was drawn from the appropriate ME distribution.
<table>
<thead>
<tr>
<th>Demand D=1</th>
<th>$\Delta_1$</th>
<th>$\Delta_2$</th>
<th>$\Delta_3$</th>
<th>$\Delta_4$</th>
<th>$\Delta_5$</th>
<th>$\Delta_6$</th>
<th>$\Delta_7$</th>
</tr>
</thead>
<tbody>
<tr>
<td>vc=1:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wareh 1:</td>
<td>0.946</td>
<td>0.946</td>
<td>0.947</td>
<td>0.947</td>
<td>0.950</td>
<td>0.950</td>
<td>0.950</td>
</tr>
<tr>
<td>wareh 2:</td>
<td>0.938</td>
<td>0.939</td>
<td>0.941</td>
<td>0.941</td>
<td>0.946</td>
<td>0.947</td>
<td>0.947</td>
</tr>
<tr>
<td>wareh 3:</td>
<td>0.905</td>
<td>0.906</td>
<td>0.908</td>
<td>0.900</td>
<td>0.895</td>
<td>0.895</td>
<td>0.895</td>
</tr>
<tr>
<td>vc=2:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wareh 1:</td>
<td>0.945</td>
<td>0.946</td>
<td>0.946</td>
<td>0.945</td>
<td>0.945</td>
<td>0.945</td>
<td>0.945</td>
</tr>
<tr>
<td>wareh 2:</td>
<td>0.943</td>
<td>0.944</td>
<td>0.944</td>
<td>0.944</td>
<td>0.944</td>
<td>0.944</td>
<td>0.941</td>
</tr>
<tr>
<td>wareh 3:</td>
<td>0.893</td>
<td>0.894</td>
<td>0.894</td>
<td>0.892</td>
<td>0.891</td>
<td>0.891</td>
<td>0.891</td>
</tr>
<tr>
<td>vc=3:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wareh 1:</td>
<td>0.951</td>
<td>0.952</td>
<td>0.951</td>
<td>0.950</td>
<td>0.949</td>
<td>0.949</td>
<td>0.949</td>
</tr>
<tr>
<td>wareh 2:</td>
<td>0.945</td>
<td>0.946</td>
<td>0.945</td>
<td>0.944</td>
<td>0.944</td>
<td>0.944</td>
<td>0.944</td>
</tr>
<tr>
<td>wareh 3:</td>
<td>0.894</td>
<td>0.895</td>
<td>0.894</td>
<td>0.901</td>
<td>0.890</td>
<td>0.890</td>
<td>0.890</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Demand D=5</th>
<th>$\Delta_1$</th>
<th>$\Delta_2$</th>
<th>$\Delta_3$</th>
<th>$\Delta_4$</th>
<th>$\Delta_5$</th>
<th>$\Delta_6$</th>
<th>$\Delta_7$</th>
</tr>
</thead>
<tbody>
<tr>
<td>vc=1:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wareh 1:</td>
<td>0.920</td>
<td>0.926</td>
<td>0.936</td>
<td>0.941</td>
<td>0.945</td>
<td>0.945</td>
<td>0.947</td>
</tr>
<tr>
<td>wareh 2:</td>
<td>0.942</td>
<td>0.947</td>
<td>0.949</td>
<td>0.950</td>
<td>0.950</td>
<td>0.949</td>
<td>0.951</td>
</tr>
<tr>
<td>wareh 3:</td>
<td>0.892</td>
<td>0.898</td>
<td>0.898</td>
<td>0.897</td>
<td>0.896</td>
<td>0.896</td>
<td>0.893</td>
</tr>
<tr>
<td>vc=2:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wareh 1:</td>
<td>0.914</td>
<td>0.919</td>
<td>0.929</td>
<td>0.934</td>
<td>0.939</td>
<td>0.942</td>
<td>0.947</td>
</tr>
<tr>
<td>wareh 2:</td>
<td>0.943</td>
<td>0.946</td>
<td>0.948</td>
<td>0.949</td>
<td>0.949</td>
<td>0.949</td>
<td>0.951</td>
</tr>
<tr>
<td>wareh 3:</td>
<td>0.892</td>
<td>0.897</td>
<td>0.896</td>
<td>0.894</td>
<td>0.893</td>
<td>0.893</td>
<td>0.894</td>
</tr>
<tr>
<td>vc=3:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wareh 1:</td>
<td>0.913</td>
<td>0.918</td>
<td>0.929</td>
<td>0.936</td>
<td>0.941</td>
<td>0.945</td>
<td>0.951</td>
</tr>
<tr>
<td>wareh 2:</td>
<td>0.942</td>
<td>0.945</td>
<td>0.948</td>
<td>0.948</td>
<td>0.949</td>
<td>0.949</td>
<td>0.951</td>
</tr>
<tr>
<td>wareh 3:</td>
<td>0.893</td>
<td>0.897</td>
<td>0.895</td>
<td>0.894</td>
<td>0.893</td>
<td>0.892</td>
<td>0.893</td>
</tr>
</tbody>
</table>

Table 2: Simulation results for some input combinations (It=1).

Table 2 gives an example of the simulation results obtained (for demand sets D=1 and 5). Given the approximations involved in the models, the results are generally satisfactory, providing good agreement between theoretical and simulated values. One issue, however, deserves some further comments. There are cases where the simulated fill rate approaches its target values very closely (observe demand set D=1). But, there are also cases, where the observed discrepancies between the realised and the target fill rate values increase (observe some cases for demand set D=5).

<table>
<thead>
<tr>
<th>Demand D=5</th>
<th>$\Delta_0$</th>
<th>$\Delta_1$</th>
<th>$\Delta_2$</th>
<th>$\Delta_3$</th>
<th>$\Delta_4$</th>
<th>$\Delta_5$</th>
<th>$\Delta_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>It=1:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vc=1:</td>
<td>55%</td>
<td>55%</td>
<td>40%</td>
<td>27%</td>
<td>16%</td>
<td>10%</td>
<td>0%</td>
</tr>
<tr>
<td>vc=2:</td>
<td>56%</td>
<td>55%</td>
<td>39%</td>
<td>27%</td>
<td>17%</td>
<td>9%</td>
<td>0%</td>
</tr>
<tr>
<td>vc=3:</td>
<td>56%</td>
<td>55%</td>
<td>39%</td>
<td>27%</td>
<td>17%</td>
<td>10%</td>
<td>0%</td>
</tr>
</tbody>
</table>

Table 3: Percentage of imbalance in some simulation experiments (It=1).
Table 3 gives the percentage of occasions in which the GBI assumption is not satisfied, for demand set $D = 5$. Note that the GBI assumption is not valid if one or more of the $q_i^*$ resulting from the rationing decision are negative (see Zipkin [31] and van Donselaar [32] for a more restricted definition of imbalance). Associating the discrepancies in Table 2 with the results in Table 3 we may conclude that the satisfaction or not of the GBI assumption is the principle cause for the observed discrepancies. Observe, however, that the proposed models are fairly robust to the presence of imbalances. In fact, the experiments demonstrate that an imbalance of less than circa 30% does not seriously affect the accuracy of the proposed fill rate models. In line with [30]-[33], theoretical work is under way in order to model the probability of imbalances (i.e. $P(\exists i: q_i^* < 0)$) for the generalised balance definition introduced in this paper.

5. Optimal stock allocation

In this section we develop an optimization model for the determination of cost-optimal stocknorms under fill rate constraints and use an empirical technique to investigate the optimal stock allocation within the system.

We start by presenting the cost model. Since the system uses order-up-to $S$ policies throughout and stockout costs are replaced by fill rate constraints, the model only incorporates holding costs which may be different at each echelon. Let us define

$$h_0 := \text{holding cost at the central depot per unit inventory on hand at the start of review period}$$

$$h_i := \text{holding cost at end stockpoint i per unit inventory on hand at the start of a review period}$$

Note that the inventory on hand at the central depot remains constant during a review period, while the inventory on hand at end stockpoints decreases during a review period. Since we only have information on the demand per review period we have chosen the convention of incurring holding costs for each item on hand at the start of a review period. Let $X_t$ and $X_t^i$ be the inventory on hand at the depot and end stockpoint $i$ immediately after an allocation and a receipt of an order, respectively. Dropping the time indices, we have to solve the following optimisation problem:

$$\min_{\Delta} h_0 E[X] + \sum_{j=1}^{N} h_j E[X_j]$$  \hspace{1cm} (9)

s.t. $f(S+\Delta, S, p_i) = \beta_i^* \hspace{1cm} \forall i$

We need expressions for $E[X] = E[X_1]$ and $E[X_1^i] = E[X_1^i]$ which may be obtained directly from the system dynamics. Omitting trivial derivation details, it follows from (3) and (5) that:

$$E[X_1^i] = E[(c_i - p_i(D_t-L_t-\Delta)^+ - D_t^i)^+]$$

$$E[X_t] = (\Delta - D_t-L_t)^+$$
Since we can apply the heuristics given in section 3 for each \( \Delta \) to satisfy the service constraints, in order to solve (9), we only have to perform a one-dimensional search on \( \Delta^* \), to obtain the globally minimum cost. Examining the properties of the objective function, we found that the problem does not have a convex cost function. However, since the proposed heuristics are quite fast, we can apply any minimization procedure for functions with one argument to identify all local minima.

Using such a search procedure, we usually found only two local minima in the cases we examined: for \( \Delta=0 \) and for some \( \Delta \) greater than 0. Typically we found that \( \Delta^* < 2R \). This may be explained from the fact that for \( \Delta > 2R \), the average stock on hand at the end stockpoints remains constant, while the depot stock increases linearly with \( \Delta \). Table 4 presents numerical results for the demand sets \( D=1, 3 \) and 5 (with \( L=1 \)). The bold figures denote the minimal cost found for given values of \( h_0 \) and \( h_i \); where \( h_i = 1 \) (for \( i = 1, 2, 3 \)). Note that the optimal value of \( \Delta \) is close to \( R \) and that the overall holding cost function is very flat for \( 0 \leq \Delta \leq R \). For further details of the analysis and more numerical results we refer to Seidel [35].

<table>
<thead>
<tr>
<th>( \Delta )</th>
<th>( D = 1 )</th>
<th>( D = 3 )</th>
<th>( D = 5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( h_0/h_i )</td>
<td>( h_0/h_i )</td>
<td>( h_0/h_i )</td>
<td>( h_0/h_i )</td>
</tr>
<tr>
<td>1</td>
<td>0.75</td>
<td>0.5</td>
<td>0.25</td>
</tr>
<tr>
<td>0</td>
<td>429</td>
<td>429</td>
<td>429</td>
</tr>
<tr>
<td>300</td>
<td>430</td>
<td>430</td>
<td>430</td>
</tr>
<tr>
<td>600</td>
<td>430</td>
<td>430</td>
<td>430</td>
</tr>
<tr>
<td>900</td>
<td>432</td>
<td>432</td>
<td>432</td>
</tr>
<tr>
<td>1200</td>
<td>434</td>
<td>434</td>
<td>434</td>
</tr>
<tr>
<td>1500</td>
<td>445</td>
<td>435</td>
<td>426</td>
</tr>
<tr>
<td>1800</td>
<td>688</td>
<td>613</td>
<td>538</td>
</tr>
<tr>
<td>2100</td>
<td>988</td>
<td>838</td>
<td>688</td>
</tr>
<tr>
<td>2400</td>
<td>1288</td>
<td>1063</td>
<td>838</td>
</tr>
<tr>
<td>2700</td>
<td>1588</td>
<td>1288</td>
<td>988</td>
</tr>
<tr>
<td>3000</td>
<td>1888</td>
<td>1513</td>
<td>1138</td>
</tr>
</tbody>
</table>

Table 4. Total holding cost as a function of \( \Delta \) for particular data sets.

Another interesting problem, closely related to the above, is the allocation of safety stocks implied by the optimal solution. Using the standard definition of safety stocks in multi-echelon inventory systems, for any value of \( W, S \) and \( \{p_i\} \), we can determine the corresponding safety stocks at the depot and the upper echelon by solving the equations (see Lagodimos [30]):

\[
SS^i = S^i - (l_r+1)\mu_i
\]
Table 5 gives the values of $SS$, $SS^0$, and $\sum_{j=1}^{N} SS^j$ corresponding to the minimal costs policies in Table 4. Observe that, in most cases, all the system positive safety stock is allocated at the upper echelon, while the optimal policy calls for the positioning of negative safety stock at the depot. This is in line with the results by Badinelli and Schwarz [17] and Lagodimos and Anderson [19] who have dealt with similar problems under different system settings.

<table>
<thead>
<tr>
<th>$h_0/h_1$</th>
<th>D=1</th>
<th>D=3</th>
<th>D=5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$SS^0$</td>
<td>$\sum_{j=1}^{N} SS^j$</td>
<td>SS</td>
</tr>
<tr>
<td>1.00</td>
<td>-1500</td>
<td>1629</td>
<td>129</td>
</tr>
<tr>
<td>0.75</td>
<td>-1500</td>
<td>1629</td>
<td>129</td>
</tr>
<tr>
<td>0.50</td>
<td>0</td>
<td>145</td>
<td>129</td>
</tr>
<tr>
<td>0.25</td>
<td>0</td>
<td>145</td>
<td>129</td>
</tr>
</tbody>
</table>

Table 5. Safety stock allocation for cost-optimal strategies.

6. Discussion and conclusions

In this paper we have derived models for the fill rates at end stockpoints in a two-echelon distribution network, using periodic review order-up-to policies throughout. All stockpoints of the network are allowed to hold inventory. The models were developed for a new rationing policy, AS rationing, which constitutes a direct generalisation of the popular FS rationing policy. They all make use of the GBI assumption, which implies that inventory imbalances never occur. In fact, the GBI assumption generalises the well-known inventory balance assumption associated with FS rationing.

In order to compute fill rates we proposed three heuristic algorithm. These algorithms are adaptations of previous heuristics, developed for depotless two-echelon networks, for the system we studied. Although some of the results of the paper assume that lead time demand is ME or gamma distributed, the analysis presented can be applied to other demand distributions as well. The assumption of the ME and gamma distributed demand, however, is essential for the application of heuristics. Computer simulation reveals that the proposed fill rate models provide excellent results when the GBI assumption is satisfied. In fact, the GBI assumption proved to be fairly robust, allowing to obtain good agreement even in cases where the percentage of imbalance reached circa 70%. Theoretical work is under way for developing analytical models of the probability that the GBI assumption is satisfied for a given network (see [6], [30]-[33] for similar investigations).

Dealing with the cost-optimal stock allocation under fill rate constraints, our analysis demonstrated that the policy of keeping $\Delta = 0$ provides near-optimal cost values even if $h_0/h_1$ is small (0.5 say). Apparently, an increase of the central depot stock yields such a small reduction of end stockpoint stocks that one should be careful in holding much central stock. The same was found to hold true.
when dealing with the associated problem of safety stock positioning. Our results to this problem support the conjecture that, under the satisfaction of the GBI assumption, most available safety stock should be allocated at the upper echelon. In fact, such a conjecture was reached by other authors for similar problems under FS rationing ([12], [19]).

There are several research possibilities stemming from the results presented here. Most important is the extension of our analysis to arbitrary N-echelon divergent networks. A possibly fruitful approach for achieving this was initially proposed in Seidel [35] and was later applied to N-echelon serial systems by van Houtum and Zijm [12]. Key idea is to approximate the convolution of a distribution with probability mass in zero and a distribution with a proper density by an ME distribution, along the lines sketched w.r.t. heuristic 3. Work along these lines is under development and will be soon reported.
7. References


[17] Badinelli, R.D., Schwarz, 1988, "Backorders optimization in a one-warehouse N-identical retailers..."


17
Research, under review.


Verrijdt and de Kok [33] proposed a fast inversion algorithm to solve the equations in Step 2 of Heuristic 2 based on the assumption of gamma distributed random variables. This inversion scheme is based on the following idea. Let X and X+Y be gamma distributed. Define the function \( \tilde{\beta}(x) \) by

\[
\tilde{\beta}(x) := 1 - \frac{E[(X+Y-x)^+]-E[(X-x)^+]}{E[Y]}
\]

Note that \( \tilde{\beta}(x) \) is strictly increasing in x and \( \tilde{\beta}(0)=0 \) and \( \tilde{\beta}(\infty)=1 \). Therefore, \( \tilde{\beta}(x) \) may be regarded as a legitimate probability distribution function of some random variable \( X_\beta \); that is:

\[
\tilde{\beta}(x) = P\{X_\beta \leq x\}
\]

If we assume that \( X_\beta \) is gamma distributed then we can solve the equation \( \beta(x)=\tilde{\beta}^* \) by inversion of the gamma distribution, i.e.

\[
x = \Gamma^{-1}_{\alpha,\mu}(\tilde{\beta}^*)
\]

The corresponding parameters of the gamma distribution \( \Gamma_{\alpha,\mu} \) may be evaluated from the first two moments of \( X_\beta \).

\[
\alpha = \frac{E^2[X_\beta]}{\sigma^2(X_\beta)} \quad , \quad \mu = \frac{\alpha}{E[X_\beta]}
\]

\[
E[X_\beta] = \frac{E[(X+Y)^2] - E[X^2]}{2E(Y)}
\]

\[
E[X_\beta^2] = \frac{E[(X+Y)^3] - E[X^3]}{3E[Y]}
\]

Under the assumption of X and Y being gamma distributed it is easy to compute \( E[X_\beta] \) and \( E[X_\beta^2] \).

There exist many algorithms to compute \( \Gamma^{-1}_{\alpha,\mu}(.) \).