Multi-echelon inventory control in divergent systems with shipping frequencies

Citation for published version (APA):

Published: 01/01/1997

Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:
• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.

Download date: 09. Apr. 2021
Multi-echelon inventory control in divergent systems with shipping frequencies
M.C. van der Heijden

BETA-publikatie: P-5
ISSN: 1386-9213
NUGI: 684
Enschede: April 1997
Keywords:
BETA-Research Programme: Operations Management of Supply/Distribution Process
Te publiceren in:
Multi-echelon inventory control in divergent systems with shipping frequencies

Matthieu C. van der Heijden

University of Twente, Faculty of Technology and Management  
Dept. of Management Science and Logistics  
P.O. Box 217, 7500 AE Enschede, The Netherlands  
e-mail: m.c.vanderheijden@sms.utwente.nl

Abstract

We consider inventory control in a two-echelon divergent network, consisting of a central depot and multiple (non-identical) local warehouses. Past research has shown that (R, S) order-up-to control rules for this system type can easily be obtained using a decomposition approach, where the derivation of the order-up-to level S is separated from the derivation of the rationing parameters. We show that a similar approach can be used to solve models including different shipment frequencies at the two levels. For example, a central depot receives replenishment orders every four weeks, while replenishment orders to several local warehouses are shipped weekly. In this way, inventory imbalance can be reduced compared to immediate shipment to the local warehouses after goods receipt by the central depot. A method is presented to determine the control parameters such that target fill rates for the local warehouses are obtained. Extensive experimentation with the model shows that significant stock reduction is only obtained in case of frequent resupply in the downstream part of the network.

Keywords: inventory, distribution, logistics

1. Introduction

A proper choice of inventory management policies in supply chains is essential to provide high customer service levels at minimum costs. One important issue is the allocation of stocks in the supply network. In the context of a divergent (distribution) network, the allocation policy involves two key decisions in particular:
1. balancing of central and local stocks,
2. material rationing to the local stockpoints in the case of shortages at the central depot.

Because these decisions are mutually dependent, optimal inventory control rules are generally not easy to derive. As a consequence, a considerable part of the research on multi-echelon inventory control has focused on models with restrictive assumptions, such as identical retailers, Poisson demand only, zero lead times, no
material rationing and/or clearly suboptimal material rationing rules. These mathematically more tractable models are useful for specific situations and also to get insight the system operation. However, they are not generally applicable for operational control in practice. Besides, it is not clear whether the conclusions drawn from restricted models (e.g. about stock allocation) are also valid for realistic situations involving non-identical retailers, non-Poisson demand, etc.

Therefore we focus on a more general divergent model in this paper, including different demand-, service- and lead time characteristics in different markets. An important addition is the inclusion of different shipment frequencies in the network, e.g. a central depot is replenished monthly, while products are shipped weekly to the local warehouses. In this way, an allocation policy can benefit from both types of risk pooling effects at the central depot (cf. McGavin et al. [1993]):

1. Risk pooling during common lead time, i.e. the allocation of a replenishment order from an outside supplier to local stockpoints is postponed until the order arrives at the central depot. At that time the material is allocated based on the most recent information about local stock status and the material can be transferred to the local stockpoints. In this way, the out of stock risk is pooled during the common lead time between outside supplier and central depot.

2. Risk pooling during the replenishment cycle (= period between the arrival of two consecutive replenishment orders at the central depot). That is, a replenishment order is not completely allocated upon arrival, but a part is kept back to rebalance local stocks at additional shipment opportunities later on in the replenishment cycle. This postponement causes that the out of stock risk is pooled, with the drawback that the central stock is not immediately available to satisfy customer demand.

Note that the main reason to hold central stocks is the second type of risk pooling if the holding costs are the same at all stages in the network. Both types of risk pooling will be included in our model, although the numerical analysis will focus on the second type of risk pooling.

Figure 1. The two-echelon divergent network
We focus on two-echelon systems, in which products are distributed from a supplier to the final customers via one central depot and multiple local stockpoints (see Figure 1). We consider the situation in which the material flow through the network is co-ordinated by a central authority who has full knowledge about the inventory status of the system. Replenishment orders are issued according to a periodic review, order-up-to (R, S) policy, i.e. every R periods the central authority issues a replenishment order that raises the echelon inventory position (= total system stocks plus amount on order minus cumulative backlog at the end stockpoints) to the level S. A method will be derived to calculate the order-up-to level S such that target service levels at the local stockpoints are attained, taking into account the possibilities for risk pooling between replenishments. As service measure, we use the fill rate (= the fraction of demand that immediately can be delivered from stock on hand). In the numerical section we will deal with questions about the effects of the latter type of risk pooling, the timing of intermediate shipments between central depot and local stockpoints, and the balancing of central and local stocks.

The paper is organised as follows. First we discuss related literature (section 2). In section 3 we introduce the mathematical model and the notation. Section 4 deals with the analysis, concluded by a summary of the calculation method. Numerical results are discussed in section 5, including a validation by comparison to discrete event simulation results and an analysis of the effects of stock allocation and shipment frequencies. Finally we end up with our conclusions in section 6.

2. Literature

A large amount of papers on multi-echelon inventory control has appeared in the literature during the last decades, see e.g. Federgruen [1993] for an overview. One of the first papers considering a periodic review model for a divergent network is Eppen and Schrage [1981], who analysed a model with stockless central depot. They present a cost optimal policy based on equal holding- and penalty costs using an allocation rule based on equal stockout probabilities for the local stockpoints. An interesting model is discussed in De Kok [1990], who analyses a divergent model with stockless depot and a (different) constraint for each local stockpoint on the fill rate (the fraction of demand delivered immediately from stock on hand). An improved analysis of the same model is shown in Van der Heijden [1997A], where a decomposition approach is presented to determine the rationing policy and order-up-to level separately. The latter model is extended to general N-echelon divergent systems with stock holding central and intermediate depots in Van der Heijden et al [1997B]. Diks and De Kok [1996] discuss stock optimisation within this model. A discussion on the effects of risk pooling during common lead time is given by Schwarz [1989].

The effect of risk pooling between replenishments has been addressed in several other papers, mainly for periodic review, order-up-to (R, S) policies. For example, Jönsson and Silver [1987] analyse a model with
two shipment opportunities in a replenishment cycle, first at the start of the cycle, followed by an opportunity in one period before the end of the cycle. The stock is entirely allocated at the first opportunity and reallocated amongst the local stockpoints at the second opportunity using mutual transshipments. They show that significant safety stock reductions are possible using redistribution, based on numerical results for the identical retailer case only.

Jackson [1988] also shows significant effects of risk pooling between replenishments in a single-cycle model where the initial stock level is given. In his model, the replenishment cycle is divided in N equal intervals. At the start of each interval, the inventory positions of the local stockpoints are raised to their order-up-to levels as long as sufficient central stock is available. Again, numerical results are limited to identical retailer cases.

McGavin et al [1993] consider allocation policies in a single replenishment cycle with given initial stock as well. They examine various ways to exploit the effect of risk pooling between replenishments. They conclude that a significant part of the risk pooling benefits can be obtained using a simple 50/25 heuristic. That is, the replenishment cycle contains two shipment opportunities, one at the start and one after 50% of the replenishment cycle. For allocation at the second shipment opportunity, 25% of the mean replenishment cycle demand is reserved. Both Jackson and McGavin et al. do not consider the problem of controlling the total system stock by adequate replenishments to the central depot. A similar model with a different cost structure, including the costs of shipments between central depot and local warehouses, is analysed by Güllü and Erkip [1996]. Their model is restricted to two identical local stockpoints however.

A model with multiple shipment opportunities within a replenishment cycle at arbitrary points in time is presented in Graves [1996]. The demand is assumed to follow a Poisson process and the analysis is based on a particular virtual allocation rule. That is, each unit demand causes a stock reservation at each supplying stockpoint upstream in the network. Clearly, this ‘First come, first serve’ rule is not optimal, since it does not account for differences in needs for stocks at downstream stockpoints at allocation times. Also, practical problems may arise when implementing virtual allocation, because on-line stock reservation at upstream stockpoints is necessary for each single demand unit. Although EDI may permit this, such a control policy may cause an undesirable high level of information traffic between sites. Because of this, Graves does not advocate this allocation rule for practical implementation, but he uses it because of its analytical tractability.

Considerably more literature has appeared in the field of multi-echelon inventory control. We refer to Nahmias and Smith [1993], and to Diks et al [1996B] for an overview of service level constrained models. More references about risk pooling and allocation policies can also be found in Schwarz [1989] and McGavin et al [1993].
The contribution of this paper is the following. We include both risk pooling effects of the central depot in our model. Further we allow more than two shipment opportunities within a replenishment cycle, not necessarily at equal intervals. Therefore we are able to study the choice of shipment frequencies and times. The allocation policy uses the most recent information about the system inventory status, avoiding problems with the virtual allocation of Graves [1996] as mentioned above. We do not only consider the allocation decisions given an initial stock level at the start of a replenishment cycle, but include the decision on the order-up-to level \( S \) as well. Finally, we allow that local stockpoints are non-identical, not only in the analysis and but also in the numerical results. In this way, conclusions are better founded. Unfortunately, non-identical local stockpoints (retailers) are ignored in almost all papers when presenting numerical results, while obviously stock imbalance can be more serious then.

3. Model and notation

In this section we describe the model in more detail. At the end we will give the basic assumptions and an overview of the notation used throughout this paper.

3.1. The mathematical model

As mentioned in the introduction, we consider a divergent network consisting of one central depot (having index \( i=0 \)) and \( N \) local stockpoints. The inventories in this network are controlled by an integral periodic review, order-up-to \((R, S_0)\) policy (see Figure 1). That is, every \( R \) time units a central authority issues a replenishment order to an outside supplier that raises the echelon inventory position of the central depot to the level \( S_0 \). The echelon inventory position is defined as the sum of all stocks at the central depot, at the local stockpoints and in the intermediate pipelines minus the sum of the backlogs at the local stockpoints plus the amount on order between production facility and central depot. We assume that the review period \( R \) is given, e.g. as a result of a trade-off between cycle stock and ordering costs or as a given transport frequency allowing to combine the delivery of multiple products from the same supplier.

The replenishment order arrives at the central depot after some deterministic lead time \( L_0 \). Then the central authority has to make the following decisions:
1. How should the replenishment order be allocated to the local stockpoints?
2. Which part of the replenishment order should be retained at the central depot for allocation later on in the replenishment cycle?
3. At which points of time within the replenishment cycle should the remaining central stock be shipped to the local stockpoints?
For the first decision, the initial allocation to the local stockpoints, we use the "balanced stock" rationing rule as introduced by Van der Heijden [1997A]. This rationing rule uses two sets of decision variables, a set of local order-up-to levels $S_i$ (i=1..N) and a set of rationing fractions $p_i$ (i=1..N). If sufficient central stock is available, the inventory position of each local stockpoint is raised to the level $S_i$ and the remaining part is kept at the central depot for subsequent shipments. If the central stock is less than required, say that the difference is an amount $x$, then the local inventory positions are only raised to the level $S_i - p_i * x$.

The second decision is modelled by a decision variable $\Delta_0$, representing the maximal amount of central stock in the central depot immediately after the first allocation to the local stockpoints in a replenishment cycle. Obviously this amount is maximal if no demand has occurred since the replenishment order was issued, so

$$\Delta_0 = S_0 - \sum_{i=1}^{N} S_i$$  \hspace{1cm} (1)

Given $\Delta_0$, the amount of central stock remaining after the first allocation in a replenishment cycle can easily be calculated. As an example, suppose that the first allocation occurs immediately after arrival of the replenishment order at the central depot. Denoting by $D_0[0, L_0]$ the cumulative demand in the system during the common lead time $L_0$, we then have that the amount retained at the central depot equals $\max\{\Delta_0 - D_0[0, L_0], 0\}$. This amount decreases with an increasing total demand during the replenishment order lead time $L_0$. Intuitively it seems reasonable that more stock is transferred if recent demand has been higher.

The third decision implies the choice of shipment opportunities $\tau_m$ within a replenishment cycle for $m=1..n$. Both the number of shipment opportunities $n$ and the timings $\tau_m$ itself are decision variables. In many cases the first shipment opportunity $\tau_1$ will coincide with the arrival of a replenishment order at the central depot. Both former decisions, the central-local stocks trade-off and stock rationing, are made at each time $\tau_m$ using the same rules. That is, the local inventory positions are raised to their order-up-to levels $S_i - p_i x$, where $x$ denotes the total shortage at the central depot.

Summarised, the key decision variables are the following:

- the system order-up-to level $S_0$
- the amount of central stock for risk pooling between replenishments, determined by the parameter $\Delta_0$
- the rationing policy, given by the local order-up-to levels $S_i$ and the rationing fractions $p_i$, and satisfying (1)
- amount and timing of intermediate shipments between central depot and local stockpoints $\tau_m$

### 3.2 Assumptions and notation

We will use the following assumptions in our model:
a) Customer demand occurs at the local stockpoints only.
b) The demand per period is stochastic and stationary in time.
c) The demand is both independent across local stockpoints and across periods in time.
d) All demand that can not be satisfied directly from stock on hand is backlogged.
e) Partial delivery of customer orders is allowed.
f) All lead times are constant.
g) Lot sizing is not used, so any quantity can be ordered and delivered.
h) Shipments from the central depot occur simultaneously to all local stockpoints, i.e. the shipment times $\tau_m$ are the same for all local stockpoints.
i) There are no capacity constraints on production, storage or transport.

Further we give an overview of the key notation used in the sequel:

- $N =$ number of local stockpoints
- $i =$ stockpoint index, $i=0..N$, where $i=0$ denotes the central depot
- $n =$ number of shipments times between central depot and local stockpoints within a replenishment cycle
- $L_i =$ lead time to stockpoint $i$ (and to the central depot for $i=0$)
- $R =$ review period, i.e. the epoch between two consecutive replenishments at the central depot
- $\tau_m =$ time of the $m^{th}$ stock allocation and shipment by the central depot within a replenishment cycle, $m=1..n$. For convenience we measure the $\tau_m$ from the moment that the replenishment order to be allocated is issued, so $L_0 \leq \tau_m < L_0+R$. We define $\tau_0 = 0$ and $\tau_{m+1} = R+\tau_1$.
- $S_i =$ order-up-to level for stockpoint $i$ (and the system order-up-to level for $i=0$)
- $p_i =$ rationing fraction for local stockpoint $i$
- $\Delta_0 =$ maximum amount of stock at the central depot after the first allocation at $\tau_1$, satisfying (1)
- $D_i =$ demand per period at local stockpoint $i$: a random variable with mean $\mu_i$ and variance $\sigma_i^2$
- $D_0 =$ total demand per period: a random variable with mean $\mu_0 = \sum_{i=1}^{N} \mu_i$ and variance $\sigma_0^2 = \sum_{i=1}^{N} \sigma_i^2$
- $D_i[t_1, t_2] =$ demand at stockpoint $i$ in the time interval $[t_1, t_2]$ (and the total demand for $i=0$): a random variable with mean $(t_2 - t_1) * \mu_i$ and variance $(t_2 - t_1) * \sigma_0^2$
- $T_i =$ target fill rate for local stockpoint $i$, i.e. the target fraction of demand at local stockpoint $i$ that should be satisfied directly from stock on hand
- $\psi_i =$ mean physical stock at stockpoint $i$ (and at the central depot for $i=0$)
- $\Omega_i =$ imbalance, caused by local stockpoint $i$ (to be defined in detail in section 4.1)
- $\alpha_m =$ probability that rationing occurs at the $m^{th}$ shipment opportunity $\tau_m$ (see section 4.1 for details)
- $\gamma_m =$ probability that rationing has not occurred yet at the $m^{th}$ shipment opportunity $\tau_m$ (see section 4.2)
- $X^+ =$ max{$X,0$} for any variable $X$
4. Model analysis

In this section we will focus on the determination on the order-up-to level $S_0$ and the allocation rule $\{S_i, p_i\}$, given the decision $\Delta_0$ on the central stock and the timing of shipments $\tau_m$, $m=1..n$. In the numerical section we will deal with the trade-off between central and local stocks using a numerical search over $\Delta_0$ for fixed values of $\tau_m$. Then the values of $\tau_m$ will be varied to get insight in the effects of a varying number and timing of additional shipments.

To analyse the model, we first derive a method to calculate the rationing fractions $p_i$, such that an approximate measure for the expected imbalance is minimised (section 4.1). It will appear that for this we need information about demand- and lead time characteristics, the maximum central stock level $\Delta_0$ and the timing of intermediate shipments $\tau_m$, $m=1..n$. So the rationing fractions do not depend on the target fill rates $\beta_i$. Next we can determine the order-up-to levels $S_i$ ($i=0..N$) such that all target fill rates are attained (section 4.2). Then we give expressions for the mean physical stocks at the various stages in the network (section 4.3). Finally we summarise the method in section 4.4.

4.1. The rationing fractions $p_i$

To discuss the rationing policy, we first have to mention the concept of imbalance. As mentioned in section 3.1, at any shipment opportunity $\tau_m$ the local inventory positions are raised to the order-up-to levels $S_i$, unless insufficient central stock is available. Then the local inventory positions are raised to the levels $S_i-p_ix$, where $x$ denotes the shortage and $p_i$ ($i=1..N$) are rationing fractions to be determined. Such a rationing occurs at most once during a replenishment cycle, because after that there is no central stock left for shipment opportunities later on. However, it is possible that one or more of the local stockpoints have a local inventory position between $S_i-p_ix$ and $S_i$ just before allocation, so these stockpoints should receive a negative amount of material in theory. This situation is called imbalance. Such a situation can not be corrected or at considerable costs only, because it involves transshipments between local stockpoints, possibly via the central depot.

To avoid the latter situation as much as possible, we can choose the rationing policy such that some measure of the imbalance is minimised, see e.g. Zipkin [1984]. A useful rationing policy in this respect is “balance stock” rationing (cf. Van der Heijden [1997A]), that minimises an approximate expression for the mean imbalance. Because this approach seems suitable under several model extensions, we will follow the latter approach (resulting in modified expressions for this particular model).

Without loss of generality, we assume that at some time $t=0$ a replenishment order is issued to an outside supplier, raising the central depot echelon inventory position to the level $S_0$. Hence this order arrives at
time \( t=L_0 \). Now suppose that \( \tau_m (L_0 \leq \tau_m < L_0 + R) \) is the first shipment opportunity at which insufficient central stock is available to raise all local inventory positions to their order-up-to levels \( S_i \). Because of the relation (1), we have that the shortage equals \( x = D_0[0, \tau_m] - \Delta_0 \). Because the local inventory positions could be raised to \( S_i \) at the previous shipment opportunity \( \tau_{m-1} \), the local inventory positions just before allocation equal \( z_i = S_i - D_i[\tau_{m-1}, \tau_m] \). Imbalance is present if the local inventory position just before allocation exceeds the target level after allocation, hence if \( z_i > S_i - p_i x \). We will try to minimise the expected amounts of imbalance

\[
E[\Omega_i] = E[(z_i - S_i + p_i x)^+].
\]

If rationing occurs at \( \tau_m \), these amounts equal

\[
E[\Omega_i | \text{rationing at } \tau_m] = E\left[\left(p_i D_0[0, \tau_m] - p_i \Delta_0 - D_i[\tau_{m-1}, \tau_m]\right)^+ | \text{rationing at } \tau_m\right]
\]

Note that (2) is only valid for \( m \geq 2 \). If \( m=1 \), we assume that the previous shipment occurred at the first shipment opportunity in the previous replenishment cycle, so at \( \tau_1 - R \). The logic behind this is that it is exact if \( \Delta_0 = 0 \), combined with the observation that the probability of rationing at \( \tau_1 \) is only high if \( \Delta_0 \) is relatively small. Based on this assumption, we find that \( z_i = S_i - p_i^*(D_0[-R, \tau_1 - R] - \Delta_0) - D_i[\tau_1 - R, \tau_1] \) for \( m=1 \), and so

\[
E[\Omega_i | \text{rationing at } \tau_1] = E\left[\left(p_i D_0[0, \tau_1] - p_i D_0[-R, \tau_1 - R] - D_1[\tau_1 - R, \tau_1]\right)^+ | \text{rationing at } \tau_1\right]
\]

Rationing occurs at the first shipment opportunity \( \tau_m \) at which the shortage \( x = D_0[0, \tau_m] - \Delta_0 \) is positive. Hence the probability \( \alpha_m \) that rationing occurs at \( \tau_m \) equals

\[
\alpha_m = \Pr\{D_0[0, \tau_{m-1}] \leq \Delta_0 < D_0[0, \tau_m]\}
\]

Further we see from (2)-(3) that the conditional mean imbalance can be written as

\[
E[\Omega_i | \text{rationing at } \tau_m] = E\left[(Y_{im} - a_m)^+\right]
\]

where

\[
\begin{align*}
Y_{im} &= p_i D_0[0, \tau_m] - D_i[\tau_{m-1}, \tau_m] \mid D_0[0, \tau_{m-1}] \leq \Delta_0 < D_0[0, \tau_m], \quad m = 2..n \\
Y_{i1} &= p_i D_0[0, \tau_1] - p_i D_0[-R, \tau_1 - R] - D_1[\tau_1 - R, \tau_1] \mid \Delta_0 < D_0[0, \tau_1] \\
a_m &= p_i \Delta_0, \quad m = 2..n \quad \text{and} \quad a_1 = 0
\end{align*}
\]

By conditioning on the time of rationing, we now find the following expression for the mean imbalance caused by local stockpoint \( i \):

\[
E[\Omega_i] = \sum_{m=1}^{n} \alpha_m E\left[(Y_{im} - a_m)^+\right]
\]
The goal is to minimise $\Sigma E[\Omega_i]$ subject to $\Sigma p_i = 1$, where $\Omega_i$ does not depend on $p_j$ for $j \neq i$. To this end, we approximate the random variables $Y_{im}$ by normally distributed random variables with the same mean and variance, see the appendix. Next we can find near-optimal rationing fractions $p_i$ using a numerical optimisation procedure such as a nested bisection, see Van der Heijden [1997A] for details. Remark that a simple and accurate approximation for the rationing fractions as discussed in Van der Heijden et al [1997B] could not be found for this particular model.

4.2. The order-up-to levels $S_i$

In this section we will show how to determine the order-up to levels $S_i$ given the rationing fractions $p_i$ and the maximum central stock $\Delta_0$. Therefore we examine the behaviour of the inventory level (=physical stock minus backlog) at some local stockpoint $i$ during a replenishment period with length $R$, see Figure 2. This period lasts from the arrival of the first shipment in the replenishment cycle arrives ($t=\tau_1+L_i$) until the arrival of the first shipment in the next replenishment cycle ($t=R+\tau_1+L_i$).

As shown in Figure 2, the time interval $[\tau_1+L_i, R+\tau_1+L_i]$ can be divided in $n$ subcycles, each starting at the time when a shipment from the central depot may arrive, so at $t=\tau_m+L_i$ for $m=1..n$. Note that the $\tau_m$ do not need to be equidistant as is shown in Figure 2. If rationing by the central depot is necessary at shipment opportunity $t=\tau_{m^*}$, the local stockpoint receives its last shipment within the replenishment period at $t=\tau_{m^*}+L_i$.

We see that $m^*=4$ in Figure 2, so no shipment arrives at the start of the 5th subcycle at $t=\tau_5+L_i$.

Extending the well-known solution for the single location (R,S) model, we find the following equation:

$$\sum_{m=1}^{n} \frac{E[\text{shortage just before } (m+1)\text{th shipment arrival} - \text{shortage just after } m\text{th shipment arrival}]}{E[\text{demand during } R]} = 1 - \beta_i$$

(7)
The denominator simply equals \( R \mu_i \). To give an expression for the numerator, we condition on the time of rationing at the central depot as follows:

\[
\sum_{m=1}^{n} \gamma_m A_{im} + \sum_{m=1}^{n} \alpha_m B_{im} = 1 - \beta_i
\]

(8)

where \( \alpha_m = \) probability that rationing by the central depot occurs at the \( m \)th shipment opportunity \( \tau_m \)

\( \gamma_m = \) probability that rationing by the central depot has not yet occurred at \( \tau_m \)

\( A_{im} = \) difference between the expected shortage just before the \( (m+1) \)th shipment arrival and the expected shortage just after the \( m \)th shipment arrival at local stockpoint \( i \), if rationing is not required yet at \( \tau_m \)

\( B_{im} = \) difference between the expected shortage just before the \( (m+1) \)th shipment arrival and the expected shortage just after the \( m \)th shipment arrival at local stockpoint \( i \), if rationing occurs at \( \tau_m \)

An expression for \( \alpha_m \) is given by (4), whilst for \( \gamma_m \) we have that

\[
\gamma_m = 1 - \sum_{k=1}^{m} \alpha_k = \Pr\{D[0, \tau_m] \leq \Delta_0\}
\]

(9)

If rationing is not required at \( \tau_m \), the local inventory position can be raised to the level \( S_i \). Hence the inventory level equals \( S_i - D[\tau_m, \tau_m+L_i] \) just after the \( m \)th shipment arrival and it equals \( S_i - D[\tau_m, \tau_{m+1}+L_i] \) just before the \( (m+1) \)th shipment arrival, where we denote \( \tau_{m+1} = R + \tau_1 \) for convenience. So we find for \( A_{im} \):

\[
A_{im} = E(D[\tau_m, \tau_{m+1}+L_i] - S_i)^+ - E(D[\tau_m, \tau_{m}+L_i] - S_i)^+
\]

(10)

Note that \( A_{im} \) is independent of \( \gamma_m \), because \( \gamma_m \) only depends on the demand in the time interval \([0, \tau_m]\). If rationing is required at \( \tau_m \) however, the local inventory position can only be raised to the level \( S_i - p_i*(D[0, \tau_m] - \Delta_0) \) and the next shipment occurs at time \( R + \tau_1 \). Therefore we find

\[
B_{jm} = E(D[\tau_m, \tau_{m}+L_j] + p_i(D^C[0, \tau_m] - \Delta_0)^+ - S_i)^+
\]

\[
E(D[\tau_m, \tau_{m}+L_j] + p_i(D^C[0, \tau_m] - \Delta_0)^+ - S_i)^+
\]

(11)

where

\[
D^C[0, \tau_m] = D[0, \tau_m] \mid D[0, \tau_{m-1}] \leq \Delta_0 < D[0, \tau_m].
\]

The latter expression accounts for the dependency between \( \alpha_m \) and \( B_{im} \). We can evaluate expression (8) using (4), (9), (10) and (11) if we approximate the various random variables by a convenient probability distribution function using a two-moment fit. The class of phase-type distributions (Erlang mixtures) has proven to be very
suitable to this end, see e.g. Tijms [1994]. Using such an approximation, we can solve equation (8) for $S_i$ using a standard numerical search routine such as bisection. We refer to the appendix for details on the two-moment approximations.

Now we have a method to find the local order-up-to levels $S_i$, given the values of $p_i$ and $\Delta_0$. Subsequently we can obtain the central order-up-to level $S_0$ from equation (1): $S_0 = \Delta_0 + \sum_{i=1}^{N} S_i$

Remark: assumption of no imbalance

The derivation in this section is only valid if target local inventory positions can always be reached when allocating central stock. That is, we assume that the amount of imbalance is negligible, which is plausible because we have chosen the rationing fractions $p_i$ accordingly in the previous section.

4.3. Physical stocks

We will consider the three physical stock locations subsequently, namely the central depot, the local stockpoints and the pipelines. First we consider the central stock, see Figure 3.

![Figure 3. Behaviour of the physical stock at the central depot during a replenishment cycle.](image)

The central physical stock increases at the start of a replenishment cycle and it decreases at each shipment opportunity $\tau_m$. The physical stock is constant between these moments, so we can calculate the mean physical stock at the central depot $\Psi_0$ as a weighted average of the physical stock in each interval. From Figure 3 we see that we have to consider $n+1$ intervals. Denoting by $\Psi_{0,m}$ the mean physical stock in the $m^{th}$ interval and by $r_m$ the length of the $m^{th}$ interval, we have that
\[ \Psi_0 = \frac{1}{R} \sum_{m=1}^{n+1} r_m \Psi_{0,m} \]  

(12)

with \( r_1 = \tau_1 - L_0 \), \( r_{n+1} = R + L_0 \cdot \tau_n \) and \( r_m = \tau_m - \tau_{m-1} \) for \( 2 \leq m \leq n \). It is allowed that \( \tau_1 = L_0 \) (and so \( r_1 = 0 \)). First we determine \( \Psi_{0,m} \) for \( m = 2 \ldots n+1 \). The maximal stock remaining after the \( m \)th shipment equals \( S_0 \cdot \Sigma S_i = \Delta_0 \), which is only present if no demand has occurred during \([0, \tau_m] \), cf. section 3.1. For the actual level, we have to subtract \( D_0[0, \tau_m] \), so the mean physical stock in the interval starting at \( \tau_m \) equals \( \Psi_{0,m+1} = E(\Delta_0 - D_0[0, \tau_m])^+ \). The mean physical stock in the first interval \( \Psi_{0,1} \) equals the mean physical stock in the last interval \( \Psi_{0,n+1} \) plus the mean size of the replenishment order that arrives at \( L_0 \). The latter is obviously equal to the mean replenishment cycle demand \( R\mu_0 \), hence the final result is:

\[ \Psi_0 = \frac{1}{R} \left\{ r_1 \cdot \left( E(\Delta_0 - D_0[0, \tau_n])^+ + R\mu_0 \right) + \sum_{m=2}^{n+1} r_m \cdot E(\Delta_0 - D_0[0, \tau_m])^+ \right\} \]  

(13)

with \( r_1 = \tau_1 - L_0 \), \( r_{n+1} = R + L_0 \cdot \tau_n \) and \( r_m = \tau_m - \tau_{m-1} \) for \( 2 \leq m \leq n \).

For the local stocks we also consider the behaviour of the physical stock during a replenishment cycle, see Figure 2. Analogously to (12), we have

\[ \Psi_i = \frac{1}{R} \sum_{m=1}^{n} (\tau_{m+1} - \tau_m) \cdot \Psi_{i,m} \]  

(14)

where \( \Psi_{i,m} \) denotes the mean physical stock of local stockpoint \( i \) in the \( m \)th subcycle. These values can be approximated as the average of the mean inventory level at the start of a subcycle \( I_{i,m1} \) and at the end of a subcycle \( I_{i,m2} \), unless the probability of stockout at the end of a subcycle is significant. In the latter case, we have to consider two situations:

- the mean inventory level at the end of the subcycle is positive \((I_{i,m2} > 0)\); then \( \Psi_{i,m} = \frac{1}{2} (I_{i,m1} + I_{i,m2}) \)
- the mean inventory level at the end of the subcycle is negative \((I_{i,m2} < 0)\); then the physical stock has decreased to 0 after a fraction \( I_{i,m1}/(I_{i,m1} - I_{i,m2}) \) of the subcycle length, approximately. Only during this fraction of time the mean physical stock equals \( I_{i,m1}/2 \). Hence we find that \( \Psi_{i,m} = \frac{1}{2} I_{i,m1}^2 / (I_{i,m1} - I_{i,m2}) \)

If we split \( I_{i,m2} \) in a positive part \( I_{i,m2}^+ \) and a negative part \( I_{i,m2}^- \), we can combine the two situations above to the approximation

\[ \Psi_{i,m} = \frac{1}{2} I_{i,m2}^+ + \frac{I_{i,m1}^2}{2 \cdot (I_{i,m1} - I_{i,m2})} \]  

(15)

To evaluate the mean inventory levels at the start and at the end of a subcycle, we proceed similarly to the fill rate analysis. Conditioning on the time of rationing at the central depot, we find after some algebra that
Hence the mean physical stock in local stockpoint $i$ can be approximated combining (14)-(18).

Finally, the mean pipeline stock between central depot and local stockpoints is given by Little's formula:

$$\psi_{\text{pipeline}} = \sum_{i=1}^{N} L_i \mu_i$$  (19)

### 4.4. Summary of the analytical method

In the preceding sections we have discussed a method to determine the rationing fractions $p_i$, $i=1..N$ and the order-up-to levels $S_i$, $i=0..N$ required to attain target fill rates $\beta_i$, given the frequency and timing of shipments between central depot and local stockpoints and given the maximal central stock level $\Delta_0$.

Summarised, the proposed method consists of the following steps:

1) Determine the rationing fractions $p_i$ for $i=1..N$ that minimise the (approximate) mean imbalance by minimising $\sum E[Q_i]$ subject to $\sum p_i = 1$ using nested bisection, where $E[Q_i]$ is defined by the combination of (4)-(6), approximating the variables $Y_{m}$ by Erlang mixtures (see appendix).

2) Determine the local order-up-to levels $S_i$, $i=1..N$ from (8) using (4) and (9)-(11) using bisection, again using approximations of random variables by Erlang mixtures (see appendix).

3) Determine the central order-up-to level $S_0$ from (1)

4) Calculate the mean physical stock in the central depot from (13)

5) Calculate the mean physical stock in the local stockpoints from (14)-(18)

6) Calculate the mean pipeline stock from (19)

### 5. Numerical results

The method as presented in the previous section is implemented in a Pascal program. In this section we validate the approximate method by comparison to results from discrete event simulation (section 5.1). Next we analyse the trade-off between central and local stocks for some given shipment schedules by a numerical search over $\Delta_0$ (section 5.2). Finally we examine two-interval policies (two shipment opportunities in a replenishment cycle) by a numerical search over both $\Delta_0$ and $\tau_2$ (section 5.3).
5.1. Validation

To validate our approximate method, we calculate the control policy \( \{p_i, S_i\} \) for a given set of input parameters and next we use simulation to examine whether the target fill rates are actually attained and whether the total stock is approximated accurately. The total stock consists of the stock at all stockpoints plus the pipeline stock in between. For this experiment, we choose the following values for the input parameters:

1) For all cases we assume that the review period equals \( R=5 \) and that all lead times between central depot and local stockpoints equal \( L_i=1 \).

2) The system consists of one central depot and \( N=6 \) local stockpoints, which are divided in two groups with each their own demand and service characteristics. The three local stockpoints within a group are similar with respect to demand characteristics and service requirements.
   - The mean period demand in group 1 equals \( \mu_i=100 \)
   - The mean period demand in group 2 is low (\( \mu_i=100 \)) or high (\( \mu_i=400 \))

3) The target service level equals \( \beta_i=90\% \) or \( \beta_i=99\% \). All four combinations of Group 1 and Group 2 service levels are tested.

4) The demand variation is either relatively low (\( \sigma/\mu_i=0.3 \)) or high (\( \sigma/\mu_i=0.9 \)). All four combinations of Group 1 and Group 2 demand variation levels are tested.

5) The lead time between external supplier and central depot equals either \( L_0=5 \) or \( L_0=15 \).

6) The maximum stock in the central depot is given by \( \Delta_o = c \times L_0 \mu_o \) with \( c=1.0, 1.25 \) or \( 1.5 \).

7) Shipment opportunities occur at each period (\( \tau_m=L_0+(m-1) \) for \( m=1..5 \)) or two times in the replenishment cycle (\( \tau_1=L_0 \) and \( \tau_2=L_0+4 \)). The latter is derived from Jönsson and Silver [1987], who consider redistribution amongst local stockpoints one period before the end of the replenishment cycle.

This means we consider \( 3\times2^2=384 \) cases. In each case, 25,000 periods are simulated to obtain accurate results. The simulations are based compound Poisson demand with Erlang-2 distributed demand per customer. A summary of the results is shown in Table 1. Based on these figures we conclude that the approximate our method is sufficiently accurate. With respect to computational performance, we found that the approximate method required 0.95 seconds CPU time on average using a Pentium 100 Mhz PC.

<table>
<thead>
<tr>
<th>Performance measure</th>
<th>Target fill rate</th>
<th>mean</th>
<th>maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>deviation from target fill rate</td>
<td>( \beta=90 )</td>
<td>0.28</td>
<td>1.94</td>
</tr>
<tr>
<td></td>
<td>( \beta=99 )</td>
<td>0.16</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td>ALL</td>
<td>0.22</td>
<td>1.94</td>
</tr>
<tr>
<td>deviation from total physical stock (%)</td>
<td>( c=1.00 )</td>
<td>0.77 %</td>
<td>2.95 %</td>
</tr>
<tr>
<td></td>
<td>( c=1.25 )</td>
<td>0.65 %</td>
<td>2.38 %</td>
</tr>
<tr>
<td></td>
<td>( c=1.50 )</td>
<td>0.43 %</td>
<td>2.05 %</td>
</tr>
<tr>
<td></td>
<td>ALL</td>
<td>0.62 %</td>
<td>2.95 %</td>
</tr>
</tbody>
</table>

Table 1: Accuracy of the approximation with respect to fill rates and physical stocks.
5.2. Trade-off between central and local stocks

Because the approximate method yields reasonably accurate results, it can be used for a trade-off between central and local stocks. Hence we search for the value of \( \Delta_0 \) that minimise the total system stock required to attain target fill rates \( \beta_i \), given the timing of shipments to the local depots \( \tau_m \), \( m=1..n \). To this end, various numerical procedures are readily available, see e.g. Press et al [1991].

In the numerical experiment, we examined similar cases as for the validation, except for the values of \( \Delta_0 \) which is a decision variable now. In addition, we considered two values for the lead time \( L_i \) from central depot to the local stockpoints, namely \( L_i = 1 \) and \( L_i = 5 \). Further we considered the following shipment schedules between central depot and local stockpoints:

a) \( \tau_m = L_0 + (m-1) \) for \( m=1..5 \), so shipping each period (\( n=5 \))

b) \( \tau_1 = L_0, \tau_2 = L_0 + 2, \tau_3 = L_0 + 4 \) (\( n=3 \))

c) \( \tau_1 = L_0, \tau_2 = L_0 + 3 \) (\( n=2 \))

d) \( \tau_1 = L_0, \tau_2 = L_0 + 4 \) (\( n=2 \))

e) \( \tau_1 = L_0 \) (\( n=1 \))

In total we have \( 5 \times 2^8 = 1280 \) cases. Note that the schedule e) coincides with the two-echelon model in Van der Heijden et al [1997B]. Schedule d) is derived from the policy in Jönsson and Silver [1987], namely a second shipment one period before the end of the replenishment cycle.

First we examine the ratio of central to local stocks, see Table 2. We see that in all cases most stock should be kept at the local warehouses. This is consistent with the known results for related models, see e.g. Graves [1996]. Table 2 shows that even the highest ratio of central to local stocks is less than 1, namely 0.92. Further we see that central stock is not needed if no intermediate shipments are planned. The reason of this outcome is the 'no imbalance' assumption as stated at the end of section 4.2. Simulation results indicate that in that case there is hardly benefit from risk pooling during the replenishment cycle indeed. The more intermediate shipments, the more central stock should be available.

<table>
<thead>
<tr>
<th>schedule</th>
<th>a)</th>
<th>b)</th>
<th>c)</th>
<th>d)</th>
<th>e)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean ratio</td>
<td>0.22</td>
<td>0.13</td>
<td>0.07</td>
<td>0.02</td>
<td>0</td>
</tr>
<tr>
<td>maximal ratio</td>
<td>0.92</td>
<td>0.54</td>
<td>0.29</td>
<td>0.16</td>
<td>0</td>
</tr>
</tbody>
</table>

*Table 2: Overview of ratio central to local stock for each shipment schedule*

The next question is the advantage of additional shipments compared to the situation with a stockless central depot, where all material is shipped immediately after receipt by the central depot to the local stockpoints. We consider the reduction in overall stock (= central stock + local stocks + pipeline stocks in between) for the shipment schedules a)-d) compared to schedule e) with stockless central depot, see Figure 4.
It is clear from Figure 4 that the overall stock reduction is limited. Even for schedule a) we find an average stock reduction of 1.4% and a maximal stock reduction of 6.5%. This seems contradictory with results obtained earlier by Jönsson and Silver [1987] at first sight. However, we can give the following explanation:

- The Jönsson and Silver model is different from our model in the sense that they rebalance local stocks using transshipments, whereas here local stocks are rebalanced using the remaining central stocks.
- We consider reduction in total stock, while Jönsson and Silver measure safety stock reduction only. Because safety stocks are only a small part of total stocks, the reduction seems considerably higher. We found for low service levels ($\beta=0.90$) that the safety stock may even be negative.
- Jönsson and Silver examined different parameter combinations that yield relatively high stock reductions, such as low demand variation ($\sigma/\mu \leq 0.4$) and small lead times ($L \leq 2$) compared to the replenishment cycle length ($4 \leq R \leq 12$). In fact, their analysis requires that $L_0+L_1 \leq R$.

The main issue is whether to measure the effect of additional shipments using total stock reduction or safety stock reduction. On one hand, the decisions involved only influence safety stocks, but on the other hand a company will be focused on total stock costs. Anyway, it may be clear that safety stock reduction isn’t the only important issue. Once proper control rules are determined, it may be considerably more rewarding to reduce cycle stock ($R$) and/or pipeline stock ($L$).

![Figure 4. Total stock reduction for shipment schedules a)-d) compared to a stockless central depot (schedule e)](image)

**5.3. Near-optimal two-interval policies**

A disadvantage of additional shipments between central depot and local stockpoints is clearly the possibility of increased transportation costs. Therefore some authors have considered policies that try to...
achieve the main part of the possible stock reduction by the introduction of one additional shipment only, the so-called two-interval policies. As mentioned before, Jönsson and Silver propose a redistribution opportunity one period before the end of the replenishment cycle. However, then it is not clear how to divide a replenishment cycle in periods: Does a replenishment cycle of 20 days consist of two 10-day periods, four 5-day periods or twenty 1-day periods? A simple policy in this respect is the 50/25 heuristic as introduced by McGavin et al [1993]. That is, 25% of the mean replenishment cycle demand is kept at the central depot for a second shipment after 50% of the replenishment cycle has passed.

In this section, we use our method to examine near-optimal two-interval policies. We assume that two shipments occur in a replenishment cycle, the first immediately after the arrival of a replenishment order at the central depot ($\tau_1 = L_0$) and the second at some other time $\tau_2 (L_0 \leq \tau_2 \leq L_0 + R)$ in the replenishment cycle to be determined. To study this problem, we use a numerical search procedure over the two parameters $\Delta_0$ and $\tau_2$, see Press et al [1991]. To this end, we extended the data set as follows:

- $L_0 = 2, 5$ or $15$
- $L_i = 2, 5, \text{ or } 15$

The demand and service characteristics are the same as in the previous experiments. Also, we take $R=5$ for all cases. Totally, we examine $3^3 \cdot 2^5 = 864$ cases.

We found after numerical optimisation that 482 of the 864 cases (=56%) a second shipment seemed not to be beneficial. For the other cases, we found the policies as depicted in Figure 5. On the horizontal axis, the position of the second shipment within the replenishment cycle is shown ($\tau_2 - L_0$). On the vertical axis, the mean amount of stock reserved for the second shipment is shown as a fraction of the mean replenishment cycle demand, so $E(\Delta_0 - D_0[0,L_0])^+ / R\mu_0$

![Figure 5](image-url)  

*Figure 5. Overview of optimal two-interval policies, in so far it dominates an one-interval policy.*
Figure 5 shows that for the majority of the cases where a two-interval policy is useful, the second allocation should be made after 30%-70% of the replenishment cycle. So the 50%-part of the 50/25 policy of McGavin et al seems reasonable. However, we find that only a small amount of stock should be kept at the central depot for the second shipment. On average, this amount is about 10% and it never exceeds 20% of the mean replenishment cycle demand. This suggests a 50/10 policy rather than a 50/25 policy. However, one should keep in mind that the total stock reduction can be relatively small.

6. Conclusions

In this paper we have presented a computational method to derive the control parameters in a two-echelon distribution system with different shipment frequencies at both levels. Both the accuracy of the method and the computational effectiveness are satisfactory (see section 5.1, Table 1 and explanation):

- the average deviation from target fill rate in an extensive numerical experiment is 0.22 percent points for target fill rates of 90% and 99%
- the mean physical stock in the system is estimated with an average deviation of 0.62%
- the CPU time required is 0.95 sec. on a Pentium-100 PC on average

Further analysis with the model lead to the following conclusions:

1. The total system stock can be reduced by adding shipments opportunities from central depot to local stockpoints within a replenishment cycle. The stock reduction is clearly highest in the case of many shipment opportunities (i.e. many opportunities for rebalancing). However, even if there are five shipment opportunities within a replenishment cycle, the reduction in total system stock is limited to 1.4% on average and 6.5% in the best case for 256 cases tested (section 5.2). Higher percentages, as presented in other papers, are encountered when considering the reduction as percentage safety stock only.

2. In all cases, the mean physical stock at local stockpoints is considerably higher than at the central depot. Even when there are five shipment opportunities within a replenishment cycle, the average ratio of central to local stock is 0.22 only (see Table 2, section 5.2).

3. Within the class of two-interval policies, i.e. two shipments in a replenishment cycle, we found that the optimal timing of the second shipment is varies somewhere between 30% and 70% of the replenishment cycle. The optimal amount of stock to be reserved for the second shipment varies from 0% to 20% of the total replenishment cycle demand (see Figure 6, section 5.3). This suggests a 50/10 heuristic instead of the 50/25 heuristic as proposed by McGavin et al [1993].

The conclusions above are found for the situation in which the total system stock to obtain target fill rates is minimised, i.e. the holding costs at the central depot and the local stockpoints are identical. Of course, these conclusions will change if the holding costs at both levels are different. For example, such a situation is encountered if value is added within the network or if storage costs at local stockpoints and at the central depot
are different. The latter situation occurs frequently in retail logistics, where the local stockpoints are located in expensive urban areas, whilst the central depot is usually located at a cheaper location (cf. Nahmias and Smith [1993]). Note that differentiated holding costs at each stockpoint can easily be taken into account in the numerical optimisation by searching the value of $\Delta_0$ that minimises a weighted sum of average stock levels.

References


Appendix. The mean and variance of the random variables $Y_{im}$

In this appendix we give details on the moments of the random variables $Y_{im}$, which are required to obtain the rationing fractions $p_i$ in a numerical minimisation procedure (see section 4.1). Recall that $Y_{im}$ is defined as

$$Y_{im} = p_i D_0[0, \tau_m] - D_1[\tau_{m-1}, \tau_m] | D_0[0, \tau_{m-1}] \leq \Delta_0 < D_0[0, \tau_m], \quad m = 2..n$$

$$Y_{11} = p_i D_0[0, \tau_1] - p_i D_0[-R, \tau_1 - R] - D_1[\tau_1 - R, \tau_1] | \Delta_0 < D_0[0, \tau_1]$$

For $m \geq 2$ we proceed as follows. We observe that both $D_0[0, \tau_m]$ and $D_1[\tau_{m-1}, \tau_m]$ depend on the condition $D_0[0, \tau_{m-1}] \leq \Delta_0 < D_0[0, \tau_m]$, whereas $D_0[0, \tau_m]$ and $D_1[\tau_{m-1}, \tau_m]$ are also mutually dependent. This causes that the moments of $Y_{im}$ are not easy to derive. Therefore we make the following approximating assumption: $D_1[\tau_{m-1}, \tau_m]$ is approximately independent of $D_0[0, \tau_m]$. This approximation can be justified by the fact that the interval $[0, \tau_m]$ is only a fraction of the interval $[\tau_{m-1}, \tau_m]$, and by the fact that the demand at stockpoint $i$ is only a fraction of the total system demand. Therefore we write

$$E[Y_{im}] = p_i E[D_0[0, \tau_m] | D_0[0, \tau_{m-1}] \leq \Delta_0 < D_0[0, \tau_m]] - (\tau_m - \tau_{m-1}) \mu_i$$

$$\text{Var}[Y_{im}] = p_i^2 \text{Var}[D_0[0, \tau_m] | D_0[0, \tau_{m-1}] \leq \Delta_0 < D_0[0, \tau_m]] + (\tau_m - \tau_{m-1})^2 \sigma_i^2$$

So we need the mean and variance of $D_0[0, \tau_m] | D_0[0, \tau_{m-1}] \leq \Delta_0 < D_0[0, \tau_m]$. We approximate these values as follows:
1. Approximate \( D_0[0, \tau_{m-1}] \) by a probability distribution that can easily be manipulated and that has the same mean \( \tau_{m-1}\mu_0 \) and variance \( \tau_{m-1}\sigma_0^2 \). A good and commonly used candidate is the mixture of an Erlang-(r-1) and an Erlang-r density with the same scale parameter \( \lambda \), defined by

\[
p * f_{r-1, \lambda}(x) + (1-p) * f_{r, \lambda}(x) = p \frac{\lambda^{r-1}x^{r-2}e^{-\lambda x}}{(r-2)!} + (1-p) \frac{\lambda^{r}x^{r-1}e^{-\lambda x}}{(r-1)!}
\]

(22)

with the corresponding probability distribution function

\[
p * \left[1 - F_{r-1, \lambda}(x)\right] + (1-p) * \left[1 - F_{r, \lambda}(x)\right] = \sum_{i=0}^{r-2} \frac{(\lambda x)^i e^{-\lambda x}}{i!} + (1-p) \frac{(\lambda x)^{r-1}e^{-\lambda x}}{(r-1)!}
\]

(23)

We refer to Tijms [1994] for details.

2. Calculate the first two moments of \( D_0[0, \tau_{m-1}] \mid D_0[0, \tau_{m-1}] \leq \Delta_0 \). If the random variable \( X=D_0[0, \tau_{m-1}] \) is mixed Erlang distributed according to (22) and (23), we find

\[
E[X \mid X \leq \Delta_0] = \frac{p(r-1)F_{r, \lambda}(\Delta_0) + (1-p)rF_{r+1, \lambda}(\Delta_0)}{\lambda * \left[pF_{r-1, \lambda}(\Delta_0) + (1-p)F_{r, \lambda}(\Delta_0)\right]}
\]

(24)

\[
E[X^2 \mid X \leq \Delta_0] = \frac{pr(r-1)F_{r+1, \lambda}(\Delta_0) + (1-p)rrF_{r+2, \lambda}(\Delta_0)}{\lambda^2 * \left[pF_{r-1, \lambda}(\Delta_0) + (1-p)F_{r, \lambda}(\Delta_0)\right]}
\]

(25)

3. From this we find:

\[
E[D_0[0, \tau_m] \mid D_0[0, \tau_{m-1}] \leq \Delta_0] = (\tau_m - \tau_{m-1}) * \mu_0 + E[D_0[0, \tau_m] \mid D_0[0, \tau_{m-1}] \leq \Delta_0]
\]

\[
\text{Var}[D_0[0, \tau_m] \mid D_0[0, \tau_{m-1}] \leq \Delta_0] = (\tau_m - \tau_{m-1}) * \sigma_0^2 + \text{Var}[D_0[0, \tau_m] \mid D_0[0, \tau_{m-1}] \leq \Delta_0]
\]

(26)

Now approximate the distribution of \( D_0[0, \tau_m] \mid D_0[0, \tau_{m-1}] \leq \Delta_0 \) by an Erlang-mixture using a two-moment fit.

4. Finally calculate the first two moments of \( D_0[0, \tau_m] \mid D_0[0, \tau_{m-1}] \leq \Delta_0 < D_0[0, \tau_m] \). If the random variable \( X=D_0[0, \tau_m] \mid D_0[0, \tau_{m-1}] \leq \Delta_0 \) is mixed Erlang distributed according to (22) and (23), we find similar expressions as (24) and (25) for \( E[X \mid X > \Delta_0] \) and \( E[X^2 \mid X > \Delta_0] \) where each function \( F_{\alpha}(\Delta_0) \) is replaced by \( 1 - F_{\alpha}(\Delta_0) \).

Hence we find the mean and variance of \( D_0[0, \tau_m] \mid D_0[0, \tau_{m-1}] \leq \Delta_0 < D_0[0, \tau_m] \) using two successive approximations by mixed Erlang distributions.

For \( m=1 \) we derive the mean and variance of \( Y_{i0} \), by partitioning the time interval under consideration, \([-R, \tau_1]\), in four parts. Defining \( T=\min\{R, \tau_1\} \), we have the following parts:
• \([-R, T-R]\) with aggregate demand \(p_D(0)\)
• \([T-R, 0]\) with aggregate demand \(-D_i\)
• \([0, \tau_i-T]\) with aggregate demand \(-p_D(\cdot) + p_D(0) = 0\)
• \([\tau_i-T, \tau_i]\) with aggregate demand \(-D_i + p_D(\cdot) \mid D_0[0, \tau_i] > \Delta_0\)

Using a similar simplifying assumption as for \(m \geq 2\), namely that \(D_i[\tau_1 - R, \tau_i]\) is approximately independent of \(D_0[0, \tau_i]\), it can be derived that

\[
\begin{align*}
E[Y_{11}] &= -R\mu_1 - p_1 T \mu_0 + p_i E\{D_0[\tau_1 - T, \tau_i] \mid D_0[0, \tau_i] > \Delta_0\} \\
\text{Var}[Y_{11}] &= R \sigma_1^2 + p_i^2 T \sigma_0^2 + p_i^2 \text{Var}\{D_0[\tau_1 - T, \tau_i] \mid D_0[0, \tau_i] > \Delta_0\}
\end{align*}
\]

(27)

So we need the mean and variance of \(D_0[\tau_1 - T, \tau_i] \mid D_0[0, \tau_i] > \Delta_0\). We use the following method:

1. Approximate \(\text{Var}\{D_0[0, \tau_i] \mid D_0[0, \tau_i] > \Delta_0\} = s^2\) using an approximation by an Erlang mixture.
2. Estimate \(\text{Var}\{D_0[\tau_1 - T, \tau_i] \mid D_0[0, \tau_i] > \Delta_0\}\) as \(s^2 T / \tau_i\)

To approximate the mean imbalance as given by (6) in section 4.1, we proceed as follows. We approximate the random variables \(Y_{im}, m=1..n\) by normally distributed variables with the same mean \(\mu_{Yim}\) and standard deviation \(\sigma_{Yim}\). Then \(E[(Y_{im} - a_m)^+]\) is given by the loss function:

\[
E\left[(Y_{im} - a_m)^+\right] = \sigma_{Yim} \phi\left(\frac{\mu_{Yim} - a_m}{\sigma_{Yim}}\right) + (\mu_{Yim} - a_m) \Phi\left(\frac{\mu_{Yim} - a_m}{\sigma_{Yim}}\right)
\]

(28)

where \(\phi(.)\) and \(\Phi(.)\) denote the standard normal density and distribution function respectively. Next we use a numerical procedure to minimise \(\sum_{i=1}^{N} E[\Omega_i] = \sum_{i=1}^{N} \sum_{m=1}^{n} \alpha_m E\left[(Y_{im} - a_m)^+\right]\). We refer to Van der Heijden [1997A] for details.

Remark:
As mentioned in section 4.2 (equations (8), (10) and (11)), we need the variables \(A_{jm}\) and \(B_{jm}\) to calculate the order-up-to levels \(S_i\). Because the random variables have the same structure as discussed above (in particular \(D_0[0, \tau_m] = D_0[0, \tau_m] \mid D_0[0, \tau_{m-1}] \leq \Delta_0 < D_0[0, \tau_m]\)), the above mentioned procedure can be used to approximate \(A_{jm}\) and \(B_{jm}\) as well.