

**Near cost-optimal inventory control
policies for divergent networks
under fill rate constraints
Dr.ir. M.C. van der Heijden**

~~PR-15~~ WP-10

BETA-publicatie	:	PR-15
ISSN	:	1386-9213; PR-15
NUGI	:	684
Enschede	:	Oktober 1997
Keywords	:	inventory, distribution, logistics, optimisation
BETA-Research Programme	:	Operations Management of Supply/Distribution Processes
Te publiceren in	:	International Journal of Production Economics

Near cost-optimal inventory control policies for divergent networks under fill rate constraints

Matthieu 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 deal with the optimisation of stock levels in general divergent networks under a periodic review, order-up-to (R, S) policy. The goal is to attain target fill rates, while the total holding costs in the entire network are minimised. To this end, we first present a method for the fast calculation of the control parameters, given central and intermediate stock levels. Next we develop an approximate procedure to determine stock levels sequentially. Extensive numerical experimentation shows that this procedure yields satisfactory results. It also shows that significant stocks at intermediate stockpoints are only useful if unit holding costs in these stockpoints are considerably less than in the end stockpoints that deliver directly to the final customers.

Keywords: inventory, distribution, logistics, optimisation

1. Introduction and literature

The determination of optimal inventory levels in a network of stockpoints is a classical problem. Especially in the last decades this subject has gained interest as a consequence of the increasing attention for integral supply chain management. This is supported by the fast development of information technology, which facilitates the collection and exchange of the information. Appropriate planning and control tools are required to utilise the technological possibilities. These developments have stimulated many research activities. Since the pioneering work of Clark and Scarf [1960], many researchers have developed algorithms to analyse various series, assembly and distribution systems.

Generally, the effective and efficient control of product flows through a logistic network should be based on cost considerations (holding and ordering costs) on one hand and customer service considerations on the other. Well-known appropriate performance indicators for customer service include the non-stockout probability ($\hat{=}$ probability that the inventory just before the arrival of a replenishment order is nonnegative) and the fill rate ($\hat{=}$ the fraction of demand satisfied from stock on hand immediately). The latter customer service measure is widely used in practise. However, combining cost and fill rate targets in one model is

generally not easy. Most research proceeds from one of the two viewpoints, either minimising costs (see Federgruen [1993] and Van Houtum et al [1996] for an overview) or attaining target service levels (see Diks et al [1996A] for an overview).

When minimising costs, customer service is usually included by imposing penalty costs on shortages. The basic work in this field is from Clark and Scarf [1960] for a series system with order-up-to policy. Model extensions include the analysis of two-echelon divergent systems by Eppen and Schrage [1981] and the analysis of assembly systems by Langenhoff and Zijm [1990]. In addition, a lot of work has been done on two-echelon divergent system with lot sizing, see e.g. Svoronos and Zipkin [1988], Axsäter [1995] and Chen and Zheng [1997]. Recently, Diks and De Kok [1996C] developed a method to find cost-optimal policies for general N-echelon divergent systems with periodic review and an order-up-to policy. All these models use penalty costs to include customer service considerations. This may cause problems for implementation in practice, because penalty costs might be difficult to establish. In these situations, the problems can be solved if a direct relation between cost parameters and customer service can be given. With respect to the stockout probability, such a relation exists, cf. Diks and De Kok [1996B] for divergent networks with a periodic review, order-up-to policy. However, an explicit relation between penalty costs and the commonly used fill rate is not available to our knowledge.

When focussing on target service levels, most papers deal with performance evaluation (stock levels, service level) for a given set of control parameters, see e.g. Deuermeier and Schwarz [1981], de Kok [1990B] and Van der Heijden [1992]. Next, the problem is how to determine the control parameters such, that the target service levels to the final customers are attained. Usually, many sets of control parameters yield the same target service levels. One possible procedure is the calculation of the control parameters given the stock levels at intermediate nodes (i.e. nodes that do *not* deliver directly to the final customers), see e.g. Van der Heijden et al [1996A]. Next, one can evaluate various scenarios for the intermediate stock levels to find a cost-effective solution. An explicit optimisation is not carried out. A comparable approach is to determine the control parameters given service levels at intermediate nodes, see e.g. Van der Heijden [1992]. Once more, the next step is scenario evaluation to find a cost-effective set of control parameters, and explicit optimisation is not reported. Only in a few cases stock levels are explicitly optimised under service level constraints, see e.g. Rosenbaum [1981] and Schneider et al [1995] for approximate analyses of a two-echelon systems under installation stock policies.

Therefore, we deal with explicit stock optimisation under fill rate constraints in this paper. We consider echelon stock, periodic review, order-up-to control policies in general N-echelon divergent systems. That is, each stockpoint in the network has a unique supplier but it may deliver material to multiple other stockpoints. We proceed from the analysis by Van der Heijden et al [1997], where the control parameters are determined such that target fill rates are attained, given intermediate stock levels. So in fact we use a fill rate oriented

approach. The stock levels required at the end stockpoints are a result of these calculations, so that the total holding costs can be evaluated. Next some numerical search procedure is required to find a cost optimal solution. In principle, this seems straightforward in the case of two-echelon networks, because the cost function to be optimised has only one parameter (the central stock level). Therefore, a brute force approach is probably adequate then. However, the number of parameters in the cost function may increase considerably when the number of echelons increases. Then a brute force numerical search may require too much computation time, because the evaluation of a single scenario requires some significant effort. Half a second CPU-time for a single function evaluation may seem only little time, but it becomes cumbersome if hundreds or even thousands of function evaluations are required. Besides, it is not assured that the procedure will end in a global optimum.

Therefore, we shall start with a fast approximate method to calculate the control parameters. Numerical experiments will show that we can speed up a single evaluation of the cost function by a factor more than 10 compared to the approach by Van der Heijden et al [1997], without significant loss of accuracy. Next, we develop an approximate optimisation procedure, based on an experimental analysis of the cost function structure. It will appear that we can find near cost-optimal control policies for large divergent networks within a reasonable amount of time. The contribution of this paper consists of two parts:

- We describe a fast method to calculate control parameters such that target fill rates are achieved, given the stock levels at central and intermediate nodes.
- We give an optimisation heuristic that finds near cost-optimal control policies under fill rate constraints for large divergent networks within reasonable amount of time.

The paper is organised as follows. In section 2, we give a detailed model description and we introduce the basic notation to be used throughout the paper. Section 3 deals with a computationally efficient method to establish the control parameters (scenario analysis). We derive an optimisation procedure for two-echelon systems based on this method in section 4. Direct extension to general N-echelon divergent systems is not possible, but we are able to develop an optimisation heuristic considering the structure of the cost function (section 5). All methods are extensively tested in several numerical experiments. Finally, we end up with a discussion of the results and some directions for further research in section 6.

2. Model and notation

2.1. Model description

An example of a N-echelon divergent system is shown in Figure 1. Stockpoint 0 is supplied by some external source (e.g. a production facility). The stockpoints that deliver directly to the final customers are called *end stockpoints*. All other stockpoints are called *intermediate stockpoints*. The direction towards the end

stockpoints is called the *downstream* direction, whereas the *upstream* direction leads to stockpoint 0. To satisfy (stochastic) demand by the final customers, the products are supplied via the network.

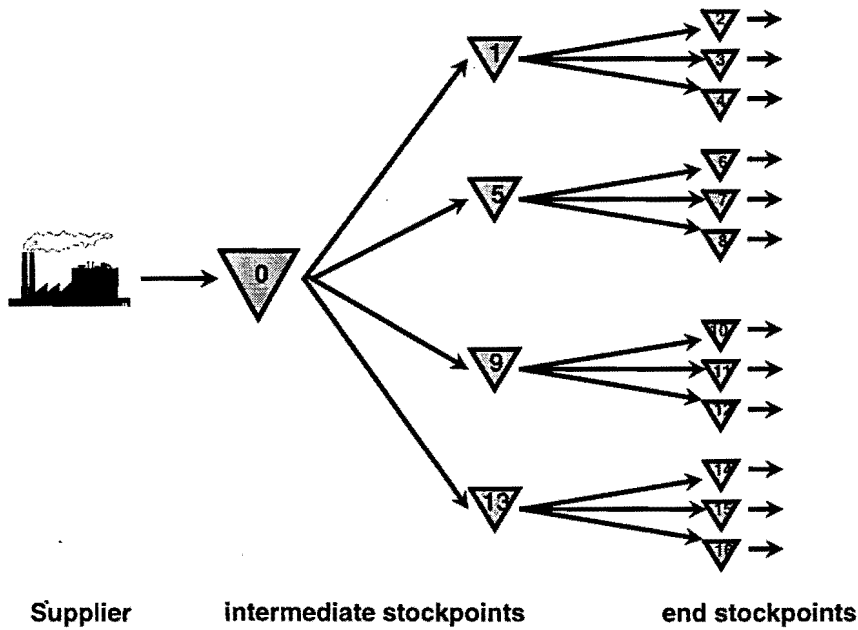


Figure 1. Example of a general N-echelon divergent network

The inventories in the network are controlled using echelon stock, periodic review, order-up-to (R, S_i) policies. That is, each stockpoint i issues periodically (with equal time interval R) a replenishment order at its supplier. The order size is chosen such that the echelon stock, defined as the physical stock plus all downstream stocks minus all downstream backlogs plus the amount on order at the supplier, is raised to the order-up-to level S_i . The lead time of a replenishment order to stockpoint i is deterministic and equal to L_i , independent of the actual order size.

The performance of a control policy is measured by holding costs on one hand and fill rates on the other. The holding costs incurred at stockpoint i equal h_i per product unit per time unit. Pipeline stocks can be ignored in principle, because these are independent of the control policy. However, they are relevant when comparing the *relative* cost difference between alternative solutions. Fixed ordering costs are ignored, since a replenishment order is shipped to all stockpoints each R periods. The choice of the replenishment interval R is a separate decision, e.g. based on a trade-off between order costs and cycle stock costs. Here we focus on the distribution of safety stocks in the network.

An important aspect of this problem is material rationing in the case of shortages. If a supplier has insufficient inventory to deliver all replenishment orders to its successors, all order quantities should be adjusted according to some rationing rule. Various approaches to ration available stocks have been suggested in the literature, see Van der Heijden et al [1997A] for an overview and a numerical comparison. The latter paper concludes that the so-called Balanced Stock rationing rule is the best available if it comes to attaining

target service levels. This rationing rule tries to minimise *imbalance*, the phenomenon that a rationing rule causes negative order quantities. Because imbalance is generally ignored when calculating order-up-to levels (it is difficult to incorporate), the actual fill rates attained using Balance Stock rationing are close to target. Another advantage is that the rationing parameters can be determined based on demand- and lead time characteristics only, independent of target fill rates and holding costs. In addition, one simple variant leads to straightforward calculations. We refer to Van der Heijden et al [1997A] for more details on imbalance and rationing policies.

Therefore, we shall proceed from the simple variant of Balanced Stock rationing in the sequel. That is, when the shortage at a supplying stockpoint k equals x , the echelon stocks of its successors $i=1..N$ are not raised to the order-up-to levels S_i but to the levels $S_i - p_i x$ where p_i are rationing fractions such that $\sum_{i=1}^N p_i = 1$, calculated as

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

where σ_i^2 denotes the variance of the echelon demand per period at successor i (=all demand at or downstream from stockpoint i). Note that the rationing fractions p_i as defined by (1) sum up to 1 indeed.

The key question remaining is: which order-up-to levels S_i should be used at each stockpoint to obtain target fill rates per end stockpoint at minimum total holding costs? Our analysis of this problem will be based on the following assumptions:

- a) Customer demand occurs at the end stockpoints only.
- b) The demand per period is stochastic and stationary in time.
- c) The demand is both independent across end 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) There are no capacity constraints on production, storage or transport.

2.2. Basic notation

Before proceeding, we give the basic notation to be used in the sequel. The examples refer to Figure 1.

Network configuration:

- i = stockpoint index, $i \geq 0$, where $i=0$ denotes the most upstream stockpoint without loss of generality
- $pre(i)$ = supplier of stockpoint i (predecessor), e.g. $pre(11)=9$
- $succ(i)$ = set of stockpoints that are directly supplied by stockpoint i (successors), e.g. $succ(0) = \{1, 5, 9, 13\}$

ech(i) = set of stockpoints in the echelon at stockpoint i, e.g. ech(5) = {5, 6, 7, 8} and ech(0) = {0, 1, ..., 16}

up(i) = set of all stockpoints on the path to the most upstream stockpoint, e.g. up(10)={0, 9}

N_i = number of successors of stockpoint i, e.g. $N_0 = 4$ and $N_9 = 3$.

Model input

R = length of the replenishment cycle

L_i = (deterministic) lead times to stockpoint i from its supplier

h_i = holding costs per product unit and per time unit at stockpoint i and in the pipeline to its successors

β_i = target fill rate at end stockpoint i

μ_i, σ_i = mean and standard deviation of the period demand at end stockpoint i

Model output

S_i = order-up-to level at stockpoint i

p_i = rationing fraction of stockpoint i, to be calculated from (1)

ψ_i = mean inventory at stockpoint i, just before the arrival of a replenishment order

$\bar{\Psi}_i$ = time average inventory at stockpoint i

$\tilde{\Psi}_i$ = time average inventory in the pipeline to stockpoint i

Supplementary notation

Z^+ = $\max\{Z, 0\}$ for any variable Z

$\Delta_i = S_i - \sum_{j \in \text{succ}(i)} S_j$

$D_{i,t}$ = echelon demand at stockpoint i in a period with length t (= local demand if i is an end stockpoint)

$X_0 = D_{0,L_0}$, the total system demand during the upstream lead time L_0

$Y_i = (X_i - \Delta_i)^+$ for $i \geq 0$

$X_i = D_{i,L_i} + p_i Y_{\text{pre}(i)}$ for $i \geq 1$

$F_i(x)$ = probability distribution function of the random variable X_i , $i \geq 0$

3. Fast calculation of the control parameters

In this section, we deal with the efficient calculation of the order-up-to levels S_i , if all values of Δ_i are given.

We call this a Δ -scenario. In the subsequent two sections we discuss numerical search methods to find cost optimal values of all Δ_i . With respect to Δ -scenarios, we first summarise the results obtained by Van der Heijden et al [1997A] in section 3.1 (for details we refer to the original paper). Next, we present a method to gain computational efficiency with limited loss of accuracy in section 3.2. The results of a numerical test are discussed in section 3.3.

3.1. Results available

Given all values of Δ_i , the order-up-to levels S_i of all end stockpoints i can be derived by solving the equation

$$\frac{E[(X_i + D_{i,R} - S_i)^+] - E[(X_i - S_i)^+]}{R\mu_i} = 1 - \beta_i \quad (2)$$

The denominator of the left-hand side equals the mean demand during a replenishment cycle (=period between the arrival of two consecutive replenishment orders). The numerator represents the difference between the expected shortage at the end and at the start of the replenishment cycle. Note that the probability distribution functions of Y_i and X_i can be determined recursively in principle. It is convenient to use successive two-moment approximations for these distributions (e.g. by Gamma distributions). Next, the definition of Δ_i gives the order-up-to levels at all intermediate stockpoints:

$$S_i = \Delta_i + \sum_{j \in \text{succ}(i)} S_j \quad (3)$$

The average stock levels of end nodes can be calculated as

$$\Psi_i = E[(S_i - D_{i,R} - X_i)^+] \quad (4)$$

$$\bar{\Psi}_i \approx \frac{1}{6} E[(S_i - X_i)^+] + \frac{4}{6} E[(S_i - D_{i,R/2} - X_i)^+] + \frac{1}{6} E[(S_i - D_{i,R} - X_i)^+] \quad (5)$$

For intermediate nodes we have

$$\Psi_i = \bar{\Psi}_i = E[(\Delta_i - X_i)^+] = \Delta_i - E[X_i] + E[Y_i] \quad (6)$$

Finally the average pipeline stock to node i can be calculated based on Little's formula as

$$\tilde{\Psi}_i = E[D_{i,L_i}] \quad (7)$$

Note that we can assume $\Delta_i \geq 0$ without loss of generality, because for each $\Delta_i < 0$ we find that $\Psi_i = 0$.

So the steps required evaluating a Δ -scenario are:

1. Choose some values for Δ_i for all intermediate nodes. From (6), we see that this is equivalent to choosing intermediate stock levels.
2. Calculate all rationing fractions in the network from (1)
3. Calculate all order-up-to levels at end nodes by solving (2) numerically (e.g. using bisection)
4. Calculate all order-up-to levels at intermediate nodes from (3)
5. Calculate all stock levels from (4)-(7)

It is important to point out that this simple method is possible because of the Balanced Stock rationing rule, since all rationing fractions p_i can be calculated separately from the order-up-to levels S_i .

3.2. Fast approximate method

The main bottleneck in the computational efficiency is solving equation (2) numerically for each end node. For example, the system as depicted in Figure 1 has 12 end nodes, hence a single scenario evaluation requires 12 calls to a numerical search procedure. Although very efficient procedures are available, this may become a problem if hundreds or thousands of scenarios should be evaluated in an optimisation procedure.

However, we can avoid the numerical search procedure using the approach of De Kok [1989] for standard single node inventory systems. Extension to our multiechelon (R, S) system is straightforward. The idea is the following. We see from equation (2) that the fill rate β_i is a function of the order-up-to level S_i , so we can write $\beta_i(S_i)$. If we draw a graph of this function on the interval $S_i \in [0, \infty)$, we see a remarkable similarity to a gamma distribution. Therefore we approximate the quasi-probability distribution function $\beta_i(S_i)$ by a gamma distribution having the same first two moments m_{i1} and m_{i2} , respectively. Expressions for m_{i1} and m_{i2} can easily be derived from equation (2):

$$m_{i1} = E[X_i] + \frac{\sigma_i^2}{2\mu_i} + \frac{1}{2} R\mu_i \quad (8)$$

$$\begin{aligned} m_{i2} &= E[X_i^2] + \frac{1}{R\mu_i} E[X_i] E[D_{i,R}^2] + \frac{1}{3R\mu_i} E[D_{i,R}^3] \\ &\approx E[X_i^2] + E[X_i] \left\{ \frac{\sigma_i^2}{\mu_i} + R\mu_i \right\} + \frac{(R\mu_i^2 + \sigma_i^2)(R\mu_i^2 + 2\sigma_i^2)}{3R\mu_i^2} \end{aligned} \quad (9)$$

For the latter approximation arises from the assumption that the demand during a replenishment cycle $D_{i,R}$ is gamma distributed. This assumption is actually not necessary, but it is convenient as we avoid using the third moment of $D_{i,R}$.

Next De Kok [1989] applies the procedure of Van der Veen [1981] for approximate inversion of the gamma distribution. This procedure is based on interpolation between the percentiles of the normal and the exponential distribution using the coefficient of variation as weight. The final result is

$$S_i \approx m_{i1} + k_{i0} \sqrt{m_{i2} - m_{i1}^2} + (k_{i1} - k_{i0}) \left(\frac{m_{i2}}{m_{i1}} - m_{i1} \right) \quad (10)$$

where $k_{i1} = -1 - \ln(1 - \beta_i)$, $k_{i0} = \Phi^{-1}(\beta_i)$ and m_{i1} and m_{i2} are given by (8) and (9). Note that Abramowitz and Stegun [1972] give a simple and excellent approximation for the inverse standard normal distribution function $\Phi^{-1}(\cdot)$.

Hence, the order-up-to-level S_i required meeting the fill rate β_i can directly be approximated from (10). The only modifications of De Kok's procedure for single stockpoint policies are the expressions for the first two moments (8) and (9). More details on this approach can also be found in Van der Heijden [1992], who applies it to evaluate scenario's in general N-echelon divergent networks with (R, S) *installation* stock policies.

3.3. Method accuracy and computational efficiency

We compared the accuracy of both methods from the previous subsections by comparison to simulation results. We used the same experimental design as in Van der Heijden et al [1997A]. That is, we analysed 384 cases for 2-echelon systems and 3132 cases for 3-echelon systems (see Appendix for parameter settings and the original paper for details). Table 1 shows statistics on approximation accuracy and computational

efficiency. The columns “numerical inversion” and “approximate inversion” refer to the method from section 3.1 and section 3.2 respectively. Note that all CPU-times mentioned in this paper are measured on a Pentium-100Mhz PC.

Performance measure	Target fill rate	2-echelon systems		3-echelon systems	
		numerical inversion	approximate inversion	numerical inversion	approximate inversion
mean deviation from target fill rate (percent points)	$\beta=90$	0.38	0.54	0.37	0.49
	$\beta=99$	0.14	0.27	0.21	0.23
	ALL	0.26	0.40	0.29	0.36
max. deviation from target fill rate (percent points)	$\beta=90$	2.97	2.43	2.55	2.25
	$\beta=99$	0.75	0.68	1.13	0.65
	ALL	2.97	2.43	2.55	2.25
mean CPU time (x 1/100 sec.)	ALL	12.2	1.0	71.2	3.4

Table 1: Approximation accuracy and computational efficiency.

We see that the approximation accuracy of both methods is good. The approximate inversion method is somewhat less accurate on average, but the worst case behaviour is even slightly better. This suggests that multiple approximation errors may compensate. The difference in computational efficiency is however very large (some factor 10-20 on average). The results suggest that the accuracy difference between two methods decreases with system size, while the efficiency difference increases with system size. These observations are in favour of the fast approximate method from section 3.2 when it comes to large-scale optimisation.

4. Optimisation procedure for two-echelon networks

In this section, we discuss a method to find cost-optimal order-up-to levels in two-echelon systems. We start with an exploratory analysis of the cost function. The approximate optimisation method is presented in section 4.2. The results of a numerical experiment to test the method accuracy are shown in section 4.3.

4.1. Cost function

Obviously, the cost function to be minimised equals the sum of the holding costs. When we consider time average stock levels, which seems most plausible, we have $\bar{Z}(\Delta_0) = \sum_{i \in \text{ech}(0)} h_i \left\{ \bar{\Psi}_i(\Delta_0) + \sum_{j \in \text{succ}(i)} \tilde{\Psi}_j \right\}$, where

$\bar{\Psi}_i(\Delta_0)$ are given by (5) for end nodes and by (6) for intermediate nodes and $\tilde{\Psi}_i$ is given by (7). Here the values of $S_i(\Delta_0)$ are given by (10), so that the target fill rates are attained. Note that we do not include the pipeline stock to the most upstream stockpoint $i=0$, because this is an external supply process. Further, we charge the pipeline stock against the holding costs of the supplier. The latter is not relevant for the optimum location, only for the magnitude of the optimum cost level.

In the literature about cost oriented models (cf. Van Houtum et al. [1997]), one tends to include holding costs on the stock levels at the end of a replenishment cycle only, so $Z(\Delta_0) = \sum_{i \in \text{ech}(0)} h_i \Psi_i(\Delta_0)$. We consider both cost functions $\bar{Z}(\Delta_0)$ and $Z(\Delta_0)$. However, note that we actually know these cost functions *by approximation only*, since the order-up-to levels (10) and the end stocks (5) are approximations. This causes additional complexity in the optimisation procedure, as we shall see.

Equation (6) shows that the magnitude of any Δ_i will usually be around $E[X_i]$ if some central stock is present. Therefore it is convenient to express Δ_i as $\Delta_i = a_i E[X_i]$, where $a_i \geq 0$ because $\Delta_i \geq 0$. We immediately see that the amount of intermediate stock is relatively small if a_i is considerably less than 1, while there is usually significant central stock if a_i is close to or greater than 1. So we rewrite the function to be optimised as

$$\begin{aligned} \bar{Z}(a_0) &= \sum_{i \in \text{ech}(0)} h_i \left\{ \bar{\Psi}_i(a_0) + \sum_{j \in \text{succ}(i)} \tilde{\Psi}_j \right\} \\ \text{and } Z(a_0) &= \sum_{i \in \text{ech}(0)} h_i \Psi_i(a_0) \\ \text{with } a_0 &= \Delta_0 / E[X_0] \end{aligned} \tag{11}$$

To get some insight in the behaviour of the cost functions, we examined a number of instances. As a typical example, Figure 2 shows the cost functions for a two-echelon model with one central depot and two local stockpoints. In addition, the average actual fill rate as obtained from discrete event simulation is shown. In this example, both the review period and all lead times equal one period, so $R=1$ and $L_i=1$ ($i=0..2$). Both local stockpoints face period demand with mean $\mu_i=100$ and standard deviation $\sigma_i=40$ ($i=1,2$). The target fill rates are $\beta_i=0.95$ ($i=1,2$). The local holding costs equal $h_i=1$ ($i=1,2$) and the central holding costs equal $h_0=0.25$.

Firstly, we see from Figure 2 that the location of the minima is about the same for both cost functions $\bar{Z}(a_0)$ and $Z(a_0)$. This appears to be a common phenomenon, so it doesn't really seem to matter which is minimised. This observation supports the cost-oriented approach to focus on end of period stocks. It facilitates optimisation as well, because $Z(a_0)$ is generally easier to handle than $\bar{Z}(a_0)$.

Secondly, Figure 2 shows that both costs functions do *not* need to have a single minimum. This is a consequence of the fact that $\bar{Z}(a_0)$ and $Z(a_0)$ are approximate cost functions only. There is a local minimum somewhere around $a_0=1$ and a local minimum at $a_0=0$ (no central stock). Further analysis shows that this is caused by the fact that the approximation error is *not* uniform over all values of a_0 . Considering the actual fill rates obtained from simulation, it appears that the actual fill rate at $a_0=0$ is slightly lower than around $a_0=1$. There are two possible causes for this behaviour. Firstly, an approximation error when calculating the order-

up-to levels S_i is made by ignoring inventory imbalance, being the amount of material that can not material that can not be allocated using the rationing rule. Inventory rationing occurs frequently when only little central stock is available, so if a_0 is small. Therefore, the actual fill rate will be somewhat below target for small values of a_0 . In such a situation, the true cost optimal solution is the local minimum around $a_0=1$. Secondly, some numerical errors in the approximation may depend on a_0 . An example is the sequential approximation of Y_i and X_i by gamma distributions.

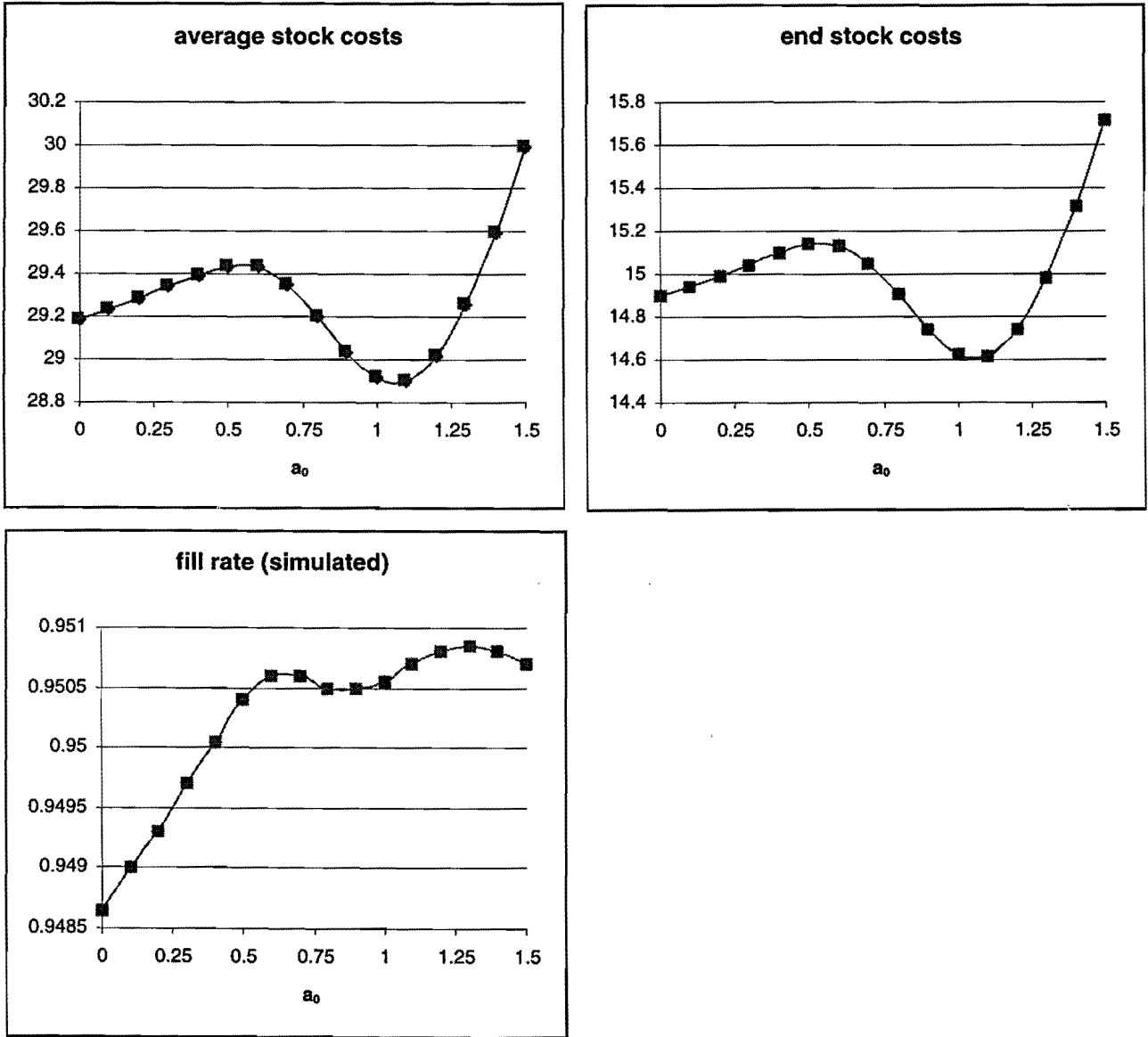


Figure 2. Relation between a_0 and respectively average stock costs $\bar{Z}(a_0)$, end stock costs $Z(a_0)$ and average simulated fill rate β_i .

The presence of multiple local minima complicates the optimisation procedure. This is particularly true when optimising larger networks, as the number of parameters a_i in the cost function equals the number of intermediate stockpoints. It may occur that the true global minimum is around $a_i=1$, while the global minimum in the *approximate* cost function is at $a_i=0$ due to ignorance of imbalance and/or non-uniform approximation errors. Fortunately, we know that the approximation accuracy is rather high for this method

(see section 3). If we accept the global minimum of the approximate cost function, we may decide to choose for a stockless central depot while some central stock may be advantageous, and very probably not the other way round. However, the cost difference will be small and then it may be preferable to avoid central stock anyway as a stockless central depot has generally lower fixed costs and variable handling costs than a stock holding depot. Therefore, we shall focus on the global minimum of the approximate cost function.

To find the global minimum of the cost function, we proceed as follows. First we search for the local minimum around $a_i=1$. Next we calculate the cost function in $a_i=0$. The global minimum is the best of these two options. In a general N-echelon network, we have to do this for each parameter a_i during each step of the optimisation procedure. Although the global convergence of this procedure could not be proved, numerical experiments showed that in this way the global minimum is found in all cases of a test for two- end three-echelon models (a computationally very intensive grid search was used as a benchmark).

To find the local minimum around $a_0=1$, we have to define a limited search area. Equation (6) shows that the central stock level Ψ_0 is almost zero if $\Pr\{X_0 < \Delta_0\} = \Pr\{X_0 < a_0 E[X_0]\}$ is almost zero. In all those cases we have virtually a stockless central depot. The probability $\Pr\{X_0 < a_0 E[X_0]\}$ is small if $a_0 E[X_0] < E[X_0] - k\sigma[X_0]$, where $k=2.33$ for example. This value of k arises from the 0.99 percentile of the normal distribution. It seems reasonable that the distribution of X_0 is close to normal, since it represents an aggregate demand process (over all end stockpoints and over the lead time L_0). So we may define a search area in terms of $a_0 \geq 1 - k \cdot cv[X_0]$, where the coefficient of variation $cv[X_0]$ is defined as the quotient of the standard deviation and the mean. A robust way to find the local minimum is to start with some value for k (e.g. $k=0.25$) and next gradually increasing the value for k until we have found a local minimum usually we find the local minimum for some value of $k < 2.33$.

So the conclusions from this section are:

1. We will focus on minimising $Z(\cdot)$, which is no restriction because the optimal value of a_0 is about the same for both cost functions $Z(\cdot)$ and $\bar{Z}(\cdot)$.
2. The approximate and true cost functions may have different global optima, but the cost differences are probably small. As a consequence however, we may decide to omit central stock whereas some positive central stock is actually slightly better from the perspective of total holding costs.
3. We assume that the global optimum of the approximate cost function is the minimum of the local optimum around $a_0=1$ and the local optimum at $a_0=0$ for all i .
4. To find the local optimum around $a_0=1$, we search in a region $a_0 \geq 1 - k \cdot cv[X_0]$ where k is gradually increasing from 0.25 to (say) 2.33. If the cost function appears to be minimal at the lower bound for a_0 , we may conclude that a stockless depot is cost optimal. We should start our search in a small area around a_0 and increase it stepwise to avoid convergence to the wrong local minimum.

4.2. Optimisation

To optimise two-echelon systems, we proceed as follows. First we give an explicit expression for the cost function $Z(\cdot)$ such that the fill rate constraints are satisfied. Next, we determine $Z'(a_0)$, the first derivative with respect to a_0 . It is most convenient to express this function as $Z'(a_0) = \frac{dZ(\Delta_0)}{d\Delta_0} \frac{d\Delta_0}{da_0} = Z'(\Delta_0)E[X_0]$, because the derivative of the cost function with respect to Δ_0 , $Z'(\Delta_0)$, results in a somewhat less complicated expression. Finally, we find the optimal value of a_0 by solving the equation $Z'(\Delta_0) = 0$ using a standard numerical search routine, yielding an optimal value $a_0^* = \Delta_0^* / E[X_0]$. An explicit expression for the optimal a_0^* could not be derived.

To derive an explicit expression for $Z(\Delta_0)$, we note that expression (4) for the expected stocks at the end of a replenishment cycle at end nodes can be written as

$$\Psi_i(\Delta_0) = S_i(\Delta_0) - E[X_i] - R\mu_i + E[\text{shortage end replenishment cycle}]$$

If S_i is chosen such that the target fill rate is attained, we have that the last term equals $(1 - \beta_i)R\mu_i$, so

$$\Psi_i(\Delta_0) = S_i(\Delta_0) - E[X_i] - R\mu_i\beta_i \quad (12)$$

From (6), (11) and (12) we find

$$Z(\Delta_0) = h_0 \{ \Delta - E[X_0] + E[Y_0(\Delta_0)] \} + \sum_{i=1}^{N_0} h_i \{ S_i(\Delta_0) - E[X_i(\Delta_0)] - R\mu_i\beta_i \} \quad (13)$$

Next, we can find the derivative $Z'(\Delta_0)$:

$$Z'(\Delta_0) = h_0 F_0(\Delta_0) + \sum_{i=1}^{N_0} h_i \{ S'_i(\Delta_0) - p_i [1 - F_0(\Delta_0)] \} \quad (14)$$

where $F_0(\cdot)$ denotes the probability distribution function of X_0 and

$$\begin{aligned} S'_i(\Delta_0) = m'_{i1}(\Delta_0) & \left\{ 1 - \frac{m_{i1}(\Delta_0)k_{i0}}{\sqrt{m_{i2}(\Delta_0) - m_{i1}^2(\Delta_0)}} + (k_{i0} - k_{i1}) \left(1 + \frac{m_{i2}(\Delta_0)}{m_{i1}^2(\Delta_0)} \right) \right\} \\ & + m'_{i2}(\Delta_0) \left\{ \frac{k_{i0}}{2\sqrt{m_{i2}(\Delta_0) - m_{i1}^2(\Delta_0)}} + \frac{(k_{i1} - k_{i0})}{m_{i1}(\Delta_0)} \right\} \end{aligned} \quad (15)$$

Here $m_{i1}(\Delta_0)$ and $m_{i2}(\Delta_0)$ are given by (8) and (9), where X_i is a function of Δ_0 . After some algebra, we obtain the following derivatives to Δ_0 :

$$m'_{i1}(\Delta_0) = -p_i [1 - F_0(\Delta_0)] \quad (16)$$

$$m'_{i2}(\Delta_0) = -2p_i^2 E[Y_0(\Delta_0)] - p_i [1 - F_0(\Delta_0)] \left\{ \frac{\sigma_i^2}{\mu_i} + (2L_i + R)\mu_i \right\} \quad (17)$$

Note that $F_0(\cdot)$ can be evaluated using a gamma approximation for X_0 by fitting the first two moments. The optimal value Δ_0^* and hence the optimal value $a_0^* = \Delta_0^* / E[X_0]$ can be obtained by solving $Z'(\Delta_0) = 0$ numerically using (14)-(17).

4.3. Numerical results

We tested the numerical search procedure on 512 cases (see Appendix). The CPU-time required is 0.08 seconds per case on average. Further, we examined the fraction of intermediate stock in the optimal solution, defined as $\frac{\bar{\Psi}_0}{\bar{\Psi}_0 + \sum_{i \in \text{succ}(0)} \bar{\Psi}_i}$. Table 2 shows the overall results. We see that the fraction of intermediate stock in

the optimal solution is usually quite low, even if the holding cost difference between intermediate and local stocks are high. So most stock should be kept local, even if the holding costs are much higher.

h_0	Average fraction intermediate stock	Max. fraction intermediate stock
0.25	0.07	0.37
0.50	0.06	0.23
0.75	0.03	0.14
1.00	0.01	0.09

Table 2. Average and maximum fraction of central stocks (in all cases we have that $h_i=1, i \geq 1$)

Note that this exact optimisation procedure was validated with a grid search as a benchmark. The grid search evaluated all possible combinations of a_0 with $0 \leq a_0 \leq 1.5$ with step 0.05. The grid search did not yield better results than the optimisation procedure, thereby founding the assumption that we only have local optima around $a_0=1$ and at $a_0=0$.

5. Optimisation procedure for N-echelon networks

Similar to the previous section, we start with an exploratory analysis of the cost function for three-echelon models in section 5.1. Section 5.2. deals with the approximate optimisation method. Finally, we show numerical results in section 5.3 (three-echelon models) and section 5.4 (four-echelon models).

5.1. Cost function

The cost function $Z(\cdot)$ has as many parameters as the number of intermediate stockpoints in the network. For example, the cost function for the network as depicted in Figure 1 can be written as $Z(a_0, a_1, a_5, a_9, a_{13})$ or $Z(\Delta_0, \Delta_1, \Delta_5, \Delta_9, \Delta_{13})$. Of course, it is difficult to display such a function graphically. Therefore we choose a three-echelon network in which the characteristics of all two-echelon subnetworks are identical with respect to demand characteristics, lead times and target fill rate. As a consequence, a cost function with two parameters remains. We consider a system that distributes products via one upstream node, two successors and four end stockpoints to the customers. All four end stockpoints face period demand with mean 10 and

standard deviation 4. The target fill-rates equal 0.95. Both the review period R and all lead times are equal to 1 period. The holding costs equal $h_0=0.25$ at the most upstream stockpoint, $h_1=h_2=0.5$ at the two successors and $h_3=h_4=h_5=h_6=1$ at the four end stockpoints. Figure 3 shows the cost functions $\bar{Z}(a_0, a_1)$ and $Z(a_0, a_1)$ as well as the corresponding average actual (simulated) fill rate.

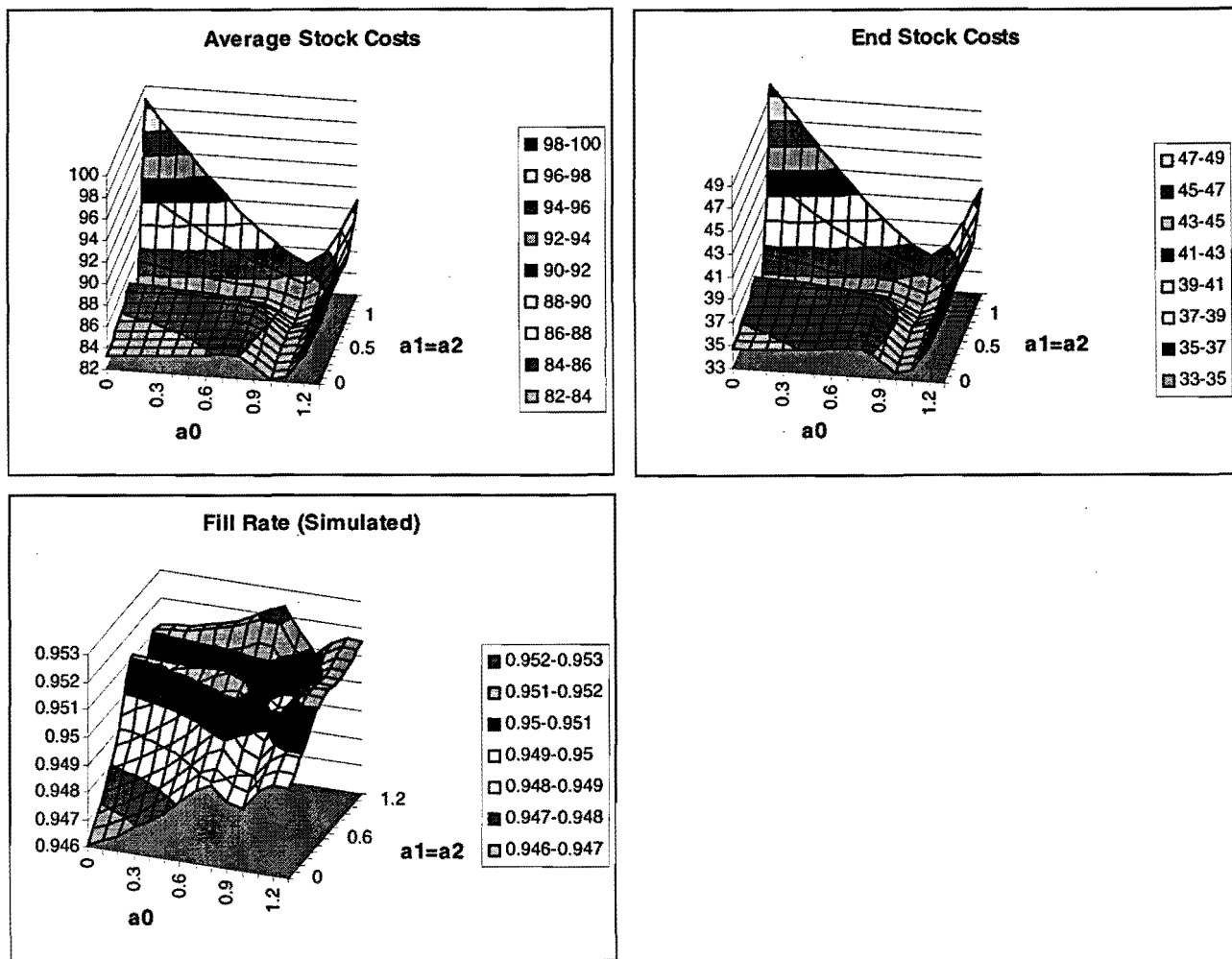


Figure 3. Relation between a_0 , $a_1 (= a_2)$ and respectively average stock costs $\bar{Z}(a_0)$, end stock costs $Z(a_0)$ and average simulated fill rate β_i .

Figure 3 shows just one example, but other cases lead to similar observations. We find the following:

1. The location of the optima are approximately equal for both cost functions $\bar{Z}(\cdot)$ and $Z(\cdot)$. Hence we may choose to minimise the most simple function, being $Z(\cdot)$
2. We have multiple local optima, at $a_i=0$ and around $a_i=1$. The actual fill rate is lower in the neighbourhood of $a_i=0$. Therefore, the model shows less stock requirements in these regions, resulting in multiple local minima. Once more, we search the local optimum around $a_i=1$, compare this to the case $a_i=0$ and choose the best option.
3. It seems that the cost minimum over a_0 for given a_1 is almost independent of a_1 and the other way round. An optimisation heuristic may exploit this phenomenon by optimising $Z(a_0, a_1, \dots)$ separately with

respect to each parameter a_i , given the other parameters a_j , $i \neq j$. Note that this observation does *not* hold for the parameters Δ_i ! For example, if a_0 changes while a_1 is constant, the definitions of Y_i and X_i show that Y_0 changes. So $E[X_1]$ changes, so $\Delta_1 = a_1 E[X_1]$ changes although a_1 is constant. Hence, the parameterisation in a_i instead of Δ_i is crucial.

5.2. Approximate optimisation method

To explain the method, we first define the level of a stockpoint in the divergent network. Each end stockpoint has level 0 and for all intermediate stockpoints, we define the level as one plus the highest level of all successors. Referring to Figure 1, we have the following levels:

- the stockpoints 2-4, 6-8, 10-12 and 14-16 have level 0
- the stockpoints 1, 5, 9 and 13 have level 1
- stockpoint 0 has level 2.

When optimising general N-echelon divergent networks, of course we would like to proceed similarly to the approach for two-echelon models: Determine the derivatives of $Z(\cdot)$ with respect to all a_i , set these to zero and solve this set of equations for all a_i . Unfortunately, this is not so easy. The reason for this is that it is cumbersome to find explicit expressions for the derivatives of all stock levels Ψ_i to all parameters a_j corresponding to stockpoints with level 2 or more (for example the derivative of Ψ_1 to a_0). To see what we can do, we observe the following:

1. The parameter a_i only influences downstream stocks, i.e. the values of Ψ_j for $j \in \text{ech}(i)$. As an example from Figure 1, we have that a_1 , a_5 , a_9 and a_{13} do not affect Ψ_0 . Hence we do not need to re-evaluate the whole network when changing the value of a_i , but only the subnetwork defined by $\text{ech}(i)$.
2. We can define a *partial cost function* $Z_i(a_i)$ for each stockpoint i with level 1 (i.e. for each two-echelon subnetwork covering end stockpoints). Of course, these partial cost functions depend on all a_j for $j \in \text{up}(i)$. However, for any given set of a_j ($j \in \text{up}(i)$) the derivatives of this partial cost function to a_i , $Z'_i(a_i) = Z'_i(\Delta_i)E[X_i]$, can still be evaluated according to (14)-(17).

The second observation implies that we can easily optimise all two-echelon subsystems in the downstream part of the network, *given* all upstream stocks (determined by all parameters a_j of the corresponding stockpoints). Consequently, the optimisation of a three-echelon network as depicted in Figure 1 is straightforward. We can use a standard numerical minimisation routine over a_0 , where we start in a small area around $a_0=1$ and increase this area stepwise. At each iteration, we find the optimal values of a_1 , a_5 , a_9 and a_{13} using some bisection-type routine. Hence we optimise four partial cost functions at each iteration. This is possible, because only very little computation time required for each partial optimisation (see section 4.3).

Although the optimisation of larger divergent systems is possible as well, the number of parameters in the numerical minimisation procedure increases quickly. Therefore, an approximate heuristic is required.

Referring to the analysis of the cost function, we propose the following heuristic:

1. Determine first estimates for a_i one by one, assuming that $a_j=0 \forall j \neq i$.
2. Adjust the values of a_i , given all other a_j as determined in step 1. Start with all stockpoints having level 1. These adjustments can be made using a simple bisection of the derivative function. Continue in the upstream direction (i.e. with increasing level) until a_0 has been corrected.

Each parameter search is done in a region around $a_i=1$ ($a_i \geq 1-k*cv[X_i]$). After each single parameter search we evaluate whether $a_i=0$ is a better option. If so, we move to $a_i=0$. Step 2 can be extended by repeating the sequential calculation of a_i , working alternately in the upstream direction (increasing level) and downstream direction (decreasing level).

So the issue remaining is how to find an optimal value for some single parameter a_i , given all other parameters a_j . For all stockpoints with level 1 this can be done using some bisection-type procedure, and for the other intermediate stockpoints (level ≥ 2) this can be done using a standard numerical search in a single variable around $a_i=1$. It is worthwhile to mention that some computational efficiency can be gained in step 1. If $a_j=0 \forall j \neq i$, we have that $\Psi_j=0$ for all *intermediate* stockpoints $j \neq i$. Hence we should only be concerned with Ψ_i and with Ψ_k for all *end* stockpoints $k \in \text{ech}(i)$. Some algebra shows that in that case

$$\Psi'_i(\Delta_i) = F_i(\Delta_i) \quad (18)$$

$$\Psi'_k(\Delta_i) = \frac{dS_k(\Delta_i)}{d\Delta_i} - q_{ki}[1 - F_i(\Delta_i)] \quad (19)$$

where q_{ki} denote the product of all rationing fractions on the path between node i and node k . For example, in Figure 1 we have that $q_{0,12} = p_9 p_{12}$ en $q_{9,12} = p_{12}$. The derivative of $S_k(\Delta_i)$ is obtained similarly to (15), where the functions $m_{i1}(\Delta_0)$ and $m_{i2}(\Delta_0)$ should be replaced by $m_{ki}(\Delta_i)$ and $m_{k2}(\Delta_i)$ respectively. The derivatives (16) and (17) are replaced by

$$m'_{ki}(\Delta_i) = -q_{ki}[1 - F_i(\Delta_i)] \quad (20)$$

$$m'_{k2}(\Delta_i) = -2q_{ki}^2 E[Y_i(\Delta_i)] - p_i [1 - F_i(\Delta_i)] \left\{ \frac{\sigma_k^2}{\mu_k} + R\mu_k \right\} - 2[1 - F_i(\Delta_i)] \sum_{j \in \text{path}(i,k)} q_{jk} E[D_{j,L_j}] \quad (21)$$

where $\text{path}(i, k)$ denotes the set of all stockpoints on the path from node i to node k , excluding i . For example, in Figure 1 we have that $\text{path}(0,12) = \{9, 12\}$ and $\text{path}(9, 12) = \{12\}$. The equations (18)-(21) together give an explicit expression for the derivative of the approximate cost function. Hence we can use bisection once more to find the best a_i one by one, given that that $a_j=0 \forall j \neq i$.

5.3. Numerical results for three-echelon networks

We tested the numerical search procedure on 780 cases. In each three-echelon network, the products are distributed from one upstream stockpoint ($i=0$) via two successors ($i=1, 2$) to 4 or 12 end stockpoints. We refer to the Appendix for details on the parameter settings. In principle, we have to optimise over three parameters, a_0 , a_1 and a_2 . We chose the same demand, lead time and service characteristics for both ech(1) and ech (2), so that $a_1=a_2$ holds. This was done to keep the computational effort for the grid search (to check the global optimality) within limits. We compared the some variants of the heuristic to the exact optimisation procedure, namely step 1 followed by 0, 1, 2 or 3 “loops” as described by step 2. Multiple loops are carried out in alternating order (upstream – downstream – upstream- etc.).

To examine whether the exact procedure did not converge to a local minimum, it was validated with a grid search as a benchmark. The grid search evaluated all possible combinations of a_i with $0 \leq a_i \leq 1.5$ with step 0.05 (i.e. $16^2=256$ scenarios). The grid search did not yield better results than the optimisation procedure as described in 5.2 for all cases, once more confirming that we only have local minima around $a_i=1$ and at $a_i=0$. Note that the numerical optimisation procedure required 11.9 sec. per case on average.

	heuristic			
	0 loops	1 loop: upstream	2 loops: up-down	3 loops: up-down-up
Average deviation (%)	0.356 %	0.067 %	0.050 %	0.048 %
Maximum deviation (%)	4.41 %	1.69 %	1.66 %	1.66 %
Fraction cases within 0.1% of optimum	74.0%	88.1%	91.5%	91.8%
Fraction cases within 1% of optimum	85.3%	98.3%	98.6%	98.7%
Average CPU-time (sec.)	0.18 sec.	0.72 sec.	0.87 sec.	1.40 sec.

Table 3. Computational efficiency and accuracy of (approximate) optimisation methods

Table 3 shows that the heuristic performs well for most cases. In a few cases, we end up in a local minimum. The cost differences are small only in these cases. Additional loops cannot contribute to finding the global minimum then. A single loop seems sufficient and the CPU-time is small. It is important to point out that Table 3 only shows the deviation from optimum with respect to the approximate cost function $Z(\cdot)$. More interesting is the cost function $\bar{Z}(\cdot)$, because it includes cycle stocks and pipeline stocks. For this cost function, the relative deviation is less than shown in Table 3. For example, the relative deviation of the worst case with respect to $\bar{Z}(\cdot)$ decreases from 1.69% to 1.08% when using one loop.

h_0	$h_1=h_2$	Average fraction	Max. fraction
0.25	0.25	0.12	0.38
	0.50	0.07	0.34
	1.00	0.05	0.27
0.50	0.50	0.04	0.34
	1.00	0.01	0.21

Table 4. Average and maximum fraction of intermediate stocks (in all cases we have that $h_i=1, i \geq 1$)

Further, we examined the fraction of intermediate stocks in the optimal solution. Table 4 shows the overall results. We see that the fraction of intermediate stock in the optimal solution is still quite low, even in the case of large holding cost differences.

5.4. Numerical results for four-echelon models

We studied 412 cases of four-echelon networks (see Appendix for parameter settings). In order to be able to find the true global minimum, we take the same parameter settings for each two-echelon subnetwork.

Therefore, we have to optimise over three parameters a_i only. The near global minimum was found using a grid search, because it is difficult to avoid local optima in a brute force numerical search procedure in more than one parameter. Consequently, the optimisation heuristics may be slightly better than the results of the grid search (not more than 0.7%). In those cases, we set the deviation from optimum to 0.

	heuristic			
	0 loops	1 loop: upstream	2 loops: up-down	3 loops: up-down-up
Average deviation (%)	0.120%	0.025%	0.018%	0.014%
Maximum deviation (%)	3.39%	0.88%	0.82%	0.80%
Fraction cases within 0.1% of optimum	89.7%	94.7%	96.2%	96.6%
Fraction cases within 1% of optimum	93.8%	100%	100%	100%
Average CPU-time (sec.)	1.0 sec.	11.5 sec.	18.5 sec.	28.4 sec.

Table 5. Computational efficiency and accuracy of (approximate) optimisation methods

We can draw similar conclusions as from the three-echelon systems. The inclusion of one correction loop is justified, but a second loop has little additional value. Although the computational effort increases, it is still within reasonable limits (which is surely not true for the grid search). One might end up in a local minimum occasionally, but then the cost difference with the global minimum is only small. This cost difference is even less when considering $\bar{Z}(\cdot)$ instead of $Z(\cdot)$. For example, the relative deviation of the worst case with respect to $\bar{Z}(\cdot)$ decreases from 0.88% to 0.26% when using one loop. The conclusions with respect to the fraction of intermediate stocks in the system remain valid as well (or get even stronger), see Table 6.

h_i (level 2)	h_k (level 1)	Average fraction	Max. fraction
0.25	0.50	0.02	0.16
0.25	1.00	0.01	0.13
0.50	0.50	0.02	0.16
0.50	1.00	0.00	0.04

Table 6. Average and maximum fraction of intermediate stocks ($h_i=1$ for all end nodes i and $h_0=0.25$)

6. Discussion

Large scale cost optimisation of divergent networks under fill rate constraints is difficult for two reasons:

- the cost function is only known by approximation and this approximate function may have multiple local minima.
- the number of parameters of the (approximate) cost function increases quickly with the size of the network.

Therefore, straightforward numerical optimisation is cumbersome. However, an optimisation heuristic consisting of a sequence of one-parameter optimisations yields accurate results in reasonable CPU-time. In some cases, only local (non-global) optimum is found, but then the cost difference with the true global minimum is small.

It is important to point out that we based the optimisation on the Balanced Stock rationing rule. The advantage of this rule is that the deviation of actual fill rate from target is generally very small. The disadvantage is that this rationing rule does not account for differences in holding costs between successors of one stockpoint. If these differences are large, use of Balanced Stock rationing might be inappropriate. It is likely that better policies from a cost perspective may exist then. However, it is also possible that such a cost-optimal allocation increases imbalance, so that target service levels are not attained. Then we may have to choose between a policy that guarantees target service levels with low, but not minimal costs and a policy that guarantees low costs with reasonable service levels, but lower than target. For the moment we prefer a policy of the first type, because often service considerations dominate cost considerations. In addition, it can be expected that in practise the most significant differences in holding costs occur between different levels in the divergent network rather than between stockpoints at the same level. Of course, further research on this topic is useful (comparison of a cost-oriented and service-oriented approach). Next to this, a number of practically relevant aspects should be included in the model, such as lot sizing and stochastic lead times. This will require a lot of effort, since these aspects strongly increase the model complexity.

References

1. Abramowitz, M. and I.A. Stegun (1972), *Handbook of mathematical functions*, Dover Publications Inc, New York.
2. Axsäter, S. (1995), "Approximate evaluation of batch-ordering policies for a one-warehouse, N non-identical retailer system under compound Poisson demand", *Naval Research Logistics* 42, 807-819.
3. Chen, F. and Y-S. Zheng (1997), "One-warehouse multiretailer systems with centralized stock information", *Operations Research* 45 no. 2, 275-287.
4. Clark, A.J. and H. Scarf (1960), "Optimal policies for a multi-echelon inventory problem", *Management Science* 6, 475-490.
5. Diks, E.B., A.G. de Kok and A.G. Lagodimos (1996A), "Multi-echelon systems: a service measure perspective", *European Journal of Operational Research* 95 no. 2, 241-263.
6. Diks, E.B., and A.G. de Kok (1996B), "Cost effective control policies for multi-echelon distribution systems", Memorandum COSOR 96-35, Eindhoven University of Technology.

7. Diks, E.B., and A.G. de Kok (1996C), "Optimal control of a divergent N-echelon inventory system", memorandum COSOR 96-09, Eindhoven University of Technology.
8. Deuermeyer, B.L. and L.B. Schwarz (1981), "A model for the analysis of system service level in warehouse-retailer distribution systems: the identical retailer case", in: *Multi-level Production - Inventory Control Systems: Theory and Practice*, L.B. Schwarz (ed.), North-Holland, Amsterdam.
9. Eppen, G. and L. Schrage. (1981), "Centralized ordering policies in a multi-warehouse system with lead times and random demand", in: *Multi-level Production - Inventory Control Systems: Theory and Practice*, L.B. Schwarz (ed.), North-Holland, Amsterdam.
10. Federgruen, A. (1993), "Centralized planning models for multi-echelon inventory systems under uncertainty", in: *Logistics of production and inventory*, S.C. Graves, A.H.G. Rinnooy Kan and P.H. Zipkin (ed.), North-Holland, Amsterdam.
11. Hadley, G., and T.M. Whitin (1963), *Analysis of inventory systems*, Prentice-Hall Inc., Englewood Cliffs, New Jersey.
12. Heijden, M.C. van der (1992), "Analysing divergent logistic networks with local (R, S) inventory control", *International Journal of Production Economics* 27, 187-219.
13. Heijden, M.C. van der, E.B. Diks and A.G. de Kok (1997A), "Stock allocation in general multi-echelon distribution systems with (R, S) order-up-to policies", *International Journal of Production Economics* 49, 157-174.
14. Heijden, M.C. van der, and A.G. de Kok (1997B), "Estimating stock levels in periodic review inventory systems", working paper University of Twente, Faculty of Technology and Management (submitted for publication).
15. Houtum, G.J. van, K. Inderfurth and W.H.M. Zijm (1996), "Materials coordination in stochastic multi-echelon systems", *European Journal of Operational Research* 95 no. 1, 1-23.
16. Kok, A.G. de (1990A), "Basics of inventory management models", FEW, 520-525, Tilburg University.
17. Kok, A.G. de (1990B), "Hierarchical production planning for consumer goods", *European Journal of Operational Research* 45, 55-69.
18. Langenhoff, L.J.G., and W.H.M. Zijm (1990), "An analytical theory of multi-echelon production/distribution systems", *Statistica Neerlandica* 44, 149-174/
19. Rosenbaum, B.A. [1981], "Inventory placement in a two-echelon inventory system: an application", in: *Multi-level Production - Inventory Control Systems: Theory and Practice*, L.B. Schwarz (ed.), North-Holland, Amsterdam.
20. Schneider, H., D.B. Rinks and P. Kelle (1995), "Power approximations for a two-echelon inventory system using service levels", *Production and Operations Management* 4, 381-400.
21. Svoronos, A. and P. Zipkin (1988), "Estimating the performance of multi-level inventory systems", *Operations Research* 36, 57-72.
22. Veen, B. van der (1981), "Safety stocks – an example of theory and practice in O.R.", *European Journal of Operational Research* 6, 367-371.

Appendix. Parameter values in the numerical experiment

In this appendix, we only state the parameter values in the numerical experiments to validate the accuracy of the approximation. In all cases we choose the review period equal to $R=1$. For more details on the choice of the parameter combinations, we refer to Van der Heijden et al [1997A]. With respect to the holding costs h_i , we note that we only considered cases with non-decreasing unit holding costs when moving in the downstream direction. If the unit holding costs decrease, there is no reason to keep intermediate stocks in that particular part of the divergent network.

A.1. Approximation accuracy for two-echelon systems in section 3.

parameter	description	values in experiment
N_0	number of end stockpoints	2, 6
$\mu_i, i \geq 1$	the mean demand per period at end stockpoint i	10, 30
$cv_i = \sigma_i / \mu_i, i \geq 1$	coefficient of variation of demand per period at end stockpoint i	0.4, 0.8
β_i	Target fill rate at end stockpoint i	90%, 99%
L_0	Lead time from external supplier to the most upstream stockpoint	1, 3
L_i	Lead time to each end stockpoint i	1
a_0	constant, describing the level of stock at stockpoint 0: $\Delta_0 = a_0 E[X_0]$	0, 0.8, 1.2

A.2. Approximation accuracy for three-echelon systems in section 3.

parameter	description	values in experiment
N_0	number of successors of stockpoint 0	2, 4
$N_j, j \in \text{succ}(0)$	number of end stockpoints for each successor j of stockpoint 0	2, 6
$\mu_i, i \geq 1$	the mean demand per period at end stockpoint i	10, 30
$cv_i = \sigma_i / \mu_i, i \geq 1$	coefficient of variation of demand per period at end stockpoint i	0.4, 0.8
β_i	Target fill rate at end stockpoint i	90%, 99%
L_0	Lead time from external supplier to the most upstream stockpoint	1, 3
L_j	Lead time from stockpoint 0 to its successors j	1, 2
L_i	Lead time to each end stockpoint i	1
a_0	constant, describing the level of stock at stockpoint 0: $\Delta_0 = a_0 E[X_0]$	0, 1.2
a_j	constant, describing the level of stock at each successor j of stockpoint 0: $\Delta_j = a_j E[X_j]$	0, 1.2

A.3. Optimisation accuracy for two-echelon systems in section 4.

parameter	description	values in experiment
N_0	number of end stockpoints	2, 6
$\mu_i, i \geq 1$	the mean demand per period at end stockpoint i	10, 30
$cv_i = \sigma_i / \mu_i, i \geq 1$	coefficient of variation of demand per period at end stockpoint i	0.4, 0.8
β_i	Target fill rate at end stockpoint i	90%, 99%
h_0	holding costs per unit product per unit time at stockpoint 0	0.25, 0.5, 0.75, 1.0
h_i	holding costs per unit product per unit time at end stockpoint i	1.0
β_i	Target fill rate at end stockpoint i	90%, 99%
L_0	Lead time from external supplier to the most upstream stockpoint	1, 3
L_i	Lead time to each end stockpoint i	1
a_0	constant, describing the level of stock at stockpoint 0: $\Delta_0 = a_0 E[X_0]$	0, 0.8, 1.2

A.4. Optimisation accuracy for three-echelon systems in section 4.

parameter	description	values in experiment
N_0	number of successors (level 1) of stockpoint 0 (level 2)	2
$N_j, j \in \text{succ}(0)$	number of end stockpoints (level 0) for each successor j (level 1) of stockpoint 0	2, 6
$\mu_i, i \geq 1$	the mean demand per period at end stockpoint i	10, 30
$cv_i = \sigma_i / \mu_i, i \geq 1$	coefficient of variation of demand per period at end stockpoint i	0.4, 0.8
β_i	Target fill rate at end stockpoint i	90%, 99%
L_0	Lead time from external supplier to the most upstream stockpoint	1, 3
L_j	Lead time from stockpoint 0 to its successors j	1, 2
L_i	Lead time to each end stockpoint i	1
h_0	holding costs per unit product per unit time at stockpoint 0	0.25, 0.5
$h_j, j \in \text{succ}(0)$	holding costs per unit product per unit time at each successor j of stockpoint 0	0.25, 0.5, 1.0
$h_i, i \in \text{succ}(j)$	holding costs per unit product per unit time at end stockpoint i	1.0

A.5. Optimisation accuracy for four-echelon systems in section 4.

parameter	description	values in experiment
N_0	number of successors (level 2) of stockpoint 0 (level 3)	2
$N_j, j \in \text{succ}(0)$	number of stockpoints (level 1) per stockpoint at level 2	2, 4
$N_k, k \in \text{succ}(j)$	number of end stockpoints (level 0) per stockpoint at level 1	2, 6
$\mu_i, i \geq 1$	the mean demand per period at end stockpoint i	10, 30
$cv_i = \sigma_i / \mu_i, i \geq 1$	coefficient of variation of demand per period at end stockpoint i	0.4, 0.8
β_i	Target fill rate at end stockpoint i	90%, 99%
L_i	all lead times (equal throughout the network)	1
h_0	holding costs per unit product per unit time at stockpoint 0 (level 3)	0.25
$h_j, j \in \text{succ}(0)$	holding costs per unit product per unit time at stockpoint j with level 2	0.25, 0.5
$h_k, k \in \text{succ}(j)$	holding costs per unit product per unit time at stockpoint k with level 1	0.5, 1.0
$h_i, i \in \text{succ}(k)$	holding costs per unit product per unit time at end stockpoint i (level 0)	1.0

Note that we only considered short lead times ($L_i=1$), because cycle stock and pipeline stock are dominant above safety stocks in the case of long, deterministic lead times in large divergent systems. The largest network considered consists of 59 stockpoints, of which 11 intermediate nodes (1 at level 3, 2 at level 2 and 8 at level 1). Hence, the cost function to be optimised has 11 parameters a_i in these cases.