

A two-echelon inventory system with supply lead time flexibility

Citation for published version (APA):

Minner, S., Diks, E. B., & Kok, de, A. G. (2003). A two-echelon inventory system with supply lead time flexibility. *IIE Transactions*, 35(2), 117-129. <https://doi.org/10.1080/07408170304383>

DOI:

[10.1080/07408170304383](https://doi.org/10.1080/07408170304383)

Document status and date:

Published: 01/01/2003

Document Version:

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.

A two-echelon inventory system with supply lead time flexibility

S. MINNER¹, E.B. DIKS² and A.G. DE KOK³

¹*Faculty of Economics and Management, Otto-von-Guericke-University of Magdeburg, P.O. Box 4120, 39016 Magdeburg, Germany
E-mail: minner@ww.uni-magdeburg.de*

²*Deloitte Consulting, Orlyplein 50, P.O. Box 58180, 1040 HD Amsterdam, The Netherlands
E-mail: ediks@dc.com*

³*Department of Technology Management, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands
E-mail: a.g.d.kok@tm.tue.nl*

Received November 2000 and accepted January 2002

The impact of manufacturing flexibility on inventory investments in a distribution network consisting of a central depot and a number of local stockpoints is investigated. The lead time of outstanding orders in the pipeline of the central depot can be shortened by the use of flexibility. Stock levels are controlled by a periodic review echelon-order-up-to-policy under service level constraints.

1. Introduction

In this paper we consider the impact of manufacturing flexibility on inventory investment in a distribution network consisting of a central depot and a number of local stockpoints. In practice manufacturing flexibility is exploited by planners that reduce the manufacturing throughput time of a particular production order in case the actual need date is earlier than was initially planned for. Such a reduction in throughput time can be realized by giving this order priority at bottleneck work stations. The rescheduling of orders by giving some orders priority may lead to the delay of other orders unless some excess capacity is available to prevent this happening. The amount of excess capacity needed to maintain due dates depends, among other things, on the frequency of rescheduling. In practice often implicit or explicit information is available about the frequency of rescheduling orders for particular products. Thereby, it is possible to make a trade-off between the frequency of rescheduling orders, which is a measure for manufacturing flexibility, and the capital investment in end products inventory.

To investigate the impact of this type of flexibility we consider a single-product/two-echelon model consisting of a central depot and multiple retailers. The retailers face stochastic demand. The demand for the product at the retailers in subsequent time periods is i.i.d. The lead time of orders from the retailers at the depot is constant, but may be different for different retailers. The lead time of

orders from the depot at the manufacturer of the product is a constant L_0 but may be shortened as explained below. Both the depot and the retailers order according to periodic review echelon-order-up-to-policies. The review period is the same for both depot and retailers.

After ordering of both depot and retailers the situation may occur that the depot has insufficient stock to satisfy the retailer orders. In that case the depot attempts to make outstanding orders available immediately. This speeding up of orders already in the pipeline is exploited until the depot is able to satisfy the retailer orders or further speeding up is impossible. However, how many orders can effectively be made available without any time delay depends on an exogenous stochastic process. The assumption that the extent to which orders can be speeded up is governed by an exogenous stochastic process can be motivated as follows. The opportunity for expediting manufacturing orders depends on the overall workload, the available capacity, and agreements with other customers/depots. In reality a planner at the manufacturer must solve a complex multi-product, multi-period, finite capacity production planning and scheduling problem. The outcome of this planning process is neither observable nor controllable by the individual depot. Therefore, from an individual product's point of view, the outcome of this planning process seems to generate random opportunities for use of manufacturing flexibility. Based on past experience, i.e., historical data, the depot determines the probability that the order to arrive at the

beginning of the next period can be made available instantaneously, the probability that the next two outstanding orders can be delivered instantaneously, etc. It is interesting to note that our description of the stochastic process governing manufacturing flexibility can be used to describe different strategies for allocating flexibility to individual products at the manufacturer.

Before speeding up orders the depot is informed about how many orders in the pipeline can be speeded up. Since orders placed in earlier periods: (i) may already be available; (ii) for these orders materials are replenished earlier at the manufacturer; and (iii) the manufacturing process has progressed, they can be expedited more easily. Therefore, we assume that earlier orders can be speeded up first, that is that the speeding up process of the replenishment pipeline follows a First-In-First-Out (FIFO) priority rule. The availability of speeding up opportunities is exogenous whereas the decision about the use of order expediting is endogenous. Thus it may be that, even after speeding up the maximum number of orders, the depot is still not able to satisfy all retailer orders. The shortage that remains is allocated among all retailers according to some rationing rule (van der Heijden *et al.*, 1997).

The objective of the paper is to get insight into the trade-off between costs of manufacturing flexibility and costs of holding inventory and in particular how manufacturing flexibility should be used for individual products. We consider two problem formulations to investigate this trade-off. The first problem formulation assumes a cost for speeding up an order of a certain age. The second formulation assumes a flexibility budget, i.e., we assume a constraint on the workload associated with speeding up orders. This latter formulation may be more practicable, since it may be hard to allocate costs of manufacturing flexibility to individual orders for individual products. In order to compare the impact of different flexibility policies on inventory investments we determine policies that yield predetermined fill rates at the retailers. The fill rate is defined as the fraction of demand satisfied from stock on hand (the P2 measure in Silver *et al.* (1998)).

To our knowledge no models exist in the OR literature that consider the relation between manufacturing flexibility and capital investment for a periodic demand multi-echelon model. For pure Poisson demand and lot-for-lot installation stock policies (Axsäter, 1993) a similar type of pipeline flexibility is considered by Dada (1992). Furthermore, Moinzadeh and Schmidt (1991), Hausman and Erkip (1994), Verrijdt *et al.* (1998) and Alfredsson and Verrijdt (1999) consider emergency shipment models. As for the model considered by Dada these papers only consider pure Poisson demand. In this paper we consider demand for normal products, which can better be described by a periodic demand model.

The paper is organized as follows. In Section 2, the dynamics of the inventory system are described. The analysis of the model is outlined in Section 3. Section 4

reports on a simulation study to test the quality of the approximations made in the analysis. Managerial insights into supply lead time flexibility are derived from numerical examples in Section 5 and conclusions are drawn in Section 6.

2. The two-echelon inventory model

The two-echelon inventory system consists of a Central Depot (CD) and N retailers as depicted in Fig. 1. The retailers face stochastic customer demand which is independent in time and among retailers. If the external demand at a retailer exceeds the available stock on-hand, the shortage is backordered. In order to avoid large shortages, every retailer has to guarantee a predetermined target service level. The retailers control their inventory according to periodic review order-up-to-policies. These require that, at the beginning of every review period, every retailer i places an order at the CD to raise its inventory position (stock on-hand plus stock in-transit minus backorders) to the order-up-to-level S_i .

The CD controls its inventory according to a periodic review echelon-order-up-to-policy, where the inventory is reviewed immediately after the retailer orders have been received at the CD. The depot's echelon inventory position (stock on-hand plus stock in-transit plus sum of all inventory positions of the retailers) is raised to S_0 by placing an order at the external supplier. The supplier satisfies every order, unless specified otherwise, within a deterministic lead time of $L_0 > 0$ review periods. The CD is allowed to hold stock up to a maximum level of Δ . This parameter is equal to the difference of S_0 and the sum of the retailers order-up-to-levels. For a stockless depot ($\Delta = 0$) all stock arriving at the depot is immediately shipped to the retailers. In the case of infinite depot stock ($\Delta = \infty$), the N stockpoints operate independently.

The CD decides how to allocate the available stock among the retailers. If possible, all retailer inventory

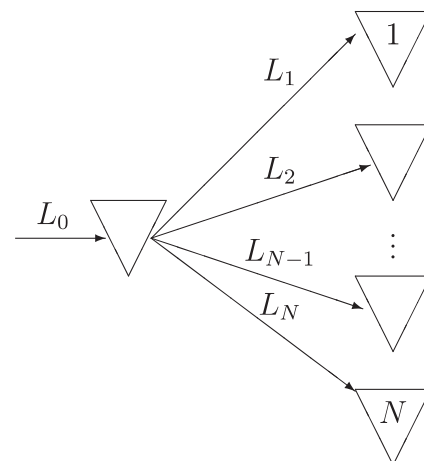


Fig. 1. Two-echelon inventory system.

positions are raised to the desired order-up-to-levels S_i by shipping an order to retailer i , which arrives after L_i review periods. All remaining stock is retained at the CD. In case of stock insufficiency at the CD, the available stock is rationed, i.e., each retailer only receives part of the ordered quantity.

Just before rationing, our model allows for exploiting some flexibility which is the (random) possibility that some orders in the pipeline towards the CD can be made available immediately. If all outstanding orders can be speeded up, i.e., there is the opportunity to reduce all (remaining) lead times to zero we call this “full flexibility”. Specifically, suppose just before rationing the CD has insufficient stock to satisfy all retail order. Then, with probability \mathcal{F}_j an order placed j periods ago (and all outstanding orders placed in earlier periods) can be made available. For this ‘speeding up’ process, we assume a FIFO priority rule without order splitting. Usually it is easier and less expensive to speed up orders placed in earlier periods and it is often impossible to split a lot after the start of processing. Hence, \mathcal{F}_j is monotonously increasing in j . Further, we assume that these probabilities only depend on the age of an order and not on either its size or on the workload given by all orders in the pipeline. This assumption appears to be reasonable if the additional workload for rescheduling orders is dominated by planning activities and setups. The incorporation of such dependencies would require a more detailed modeling of production scheduling and is a matter of further investigations.

If the depot shortage cannot be covered in the case of insufficient flexibility, the shortage is rationed among the retailers as known from the traditional distribution models (van der Heijden *et al.*, 1997). We introduce the following notation. The index i represents a facility. Index 0 corresponds to the CD, whereas an index $1 \leq i \leq N$ corresponds to retailer i . For $i = 0, 1, \dots, N$ we define:

- S_i = Order-up-to-level of facility i .
- L_i = Lead time of facility i .
- $D_{t-s,t}^i$ = Customer demand at retailer i (if $1 \leq i \leq N$), or at all retailers (if $i = 0$), during $[t - s, t)$.
- D_s^i = Customer demand at retailer i (if $1 \leq i \leq N$), or at all retailers (if $i = 0$), during s periods.
- OH_i = Mean stock on-hand at facility i just after an order arrival.
- PS_i = Mean stock in the pipeline to facility i .
- μ_i = Expected single period demand at facility i .
- σ_i = Standard deviation of single period demand at facility i .
- h_i = Inventory holding cost parameter at facility i .

For all integer t we define

- \hat{I}_t^i = Inventory position of retailer i just after the allocation at time t , ($1 \leq i \leq N$).

- X_t = Highest rank of the order in the pipeline towards the CD at time t , which cannot be speeded up.
- Y_t = Number of orders speeded up by the CD at time t .
- θ_t = Number of nonempty orders in the pipeline towards the CD at time t , just before the possible use of flexibility.
- $\hat{\theta}_t$ = Number of nonempty orders in the pipeline towards the CD at time t , just after the possible use of flexibility.

The objective of this paper is to determine the order-up-to-levels $\{S_i\}_{i=0}^N$ and the flexibility function \mathcal{F} , such that the mean holding costs associated with pipeline inventories PS_i and on-hand inventories OH_i , $i = 0, \dots, N$ in the system are minimized under the constraints that the fill rate attained at retailer i is at least the target fill rate β_i^* . The speeding up of orders is incorporated into the optimization problem in two different ways. In Problem 1, we incur costs for rescheduling orders whereas in Problem 2, the workload $E(Y)$ required to reschedule orders is limited by the available capacity Y^* . Let c_r^j denote the cost of speeding up an order of age j , and a_j be the workload of rescheduling an order of age j , $j = 0, \dots, L_0 - 1$. The expected number of orders of age j being rescheduled per period is denoted by p_j . Similar to the assumption concerning the speeding up probabilities \mathcal{F} we assume that the c_r^j and a_j are independent of the associated order sizes. Thus, these problems can be formulated as follows.

• Problem 1

$$\begin{aligned} \min_{\{S_i\}_{i=0}^N, \mathcal{F}} & h_0 \times (PS_0 + OH_0) + \sum_{i=1}^N h_i \times (PS_i + OH_i) \\ & + \sum_{j=0}^{L_0-1} c_r^j \times p_j, \\ \text{subject to} & \beta_i \geq \beta_i^*, \quad i = 1, \dots, N. \end{aligned}$$

• Problem 2

$$\begin{aligned} \min_{\{S_i\}_{i=0}^N, \mathcal{F}} & h_0 \times (PS_0 + OH_0) + \sum_{i=1}^N h_i \times (PS_i + OH_i), \\ \text{subject to} & \beta_i \geq \beta_i^*, \quad i = 1, \dots, N, \\ & \sum_{j=0}^{L_0-1} a_j \times p_j \leq Y^*. \end{aligned}$$

Note that the mean pipeline inventory for any retailer is given by its respective expected demand during the lead time, i.e., $PS_i = L_i \mu_i$, $i = 1, \dots, N$.

3. Analysis

A prerequisite to solve the aforementioned optimization problems is that *given* the control parameters $\{S_i\}_{i=0}^N$ and

$\{\mathcal{F}_j\}_{j=0}^{L_0-1}$ we are able to determine: (i) the fill rate β_i attained at some retailer i ; (ii) the expected number p_j of orders of age j being speeded up per period; and (iii) the mean holding costs in the system. In Sections 3.1–3.3 we subsequently will elaborate on how these aforementioned expressions can be derived.

3.1. Determination of β_i

From a sample path argument it can be shown that the fill rate attained at retailer i equals

$$\beta_i = \lim_{t \rightarrow \infty} \left(1 - \frac{E(D_{t,t+L_i+1}^i - \hat{I}_t^i)^+ - E(D_{t,t+L_i}^i - \hat{I}_t^i)^+}{\mu_i} \right), \quad i = 1, \dots, N. \quad (1)$$

From (1) it is obvious that in order to compute β_i we need an expression for \hat{I}_t^i . To obtain such an expression we closely examine the system dynamics. The inventory position of a retailer i at time t just after rationing depends on whether the CD has sufficient stock available to satisfy all retail orders, and if not, how the available stock is rationed among the retailers. In order to determine whether the CD has sufficient stock we examine which orders in the pipeline are available for speeding up (see Section 3.1.1). Next, we model the *use* of the flexibility process together with the rationing process (see Section 3.1.2).

3.1.1. The availability process

The pipeline of the CD at time t immediately after the regular arrival of an order and the placement of the latest order of size $D_{t-1,t}^0$ is depicted in Fig. 2. Because of the use of the flexibility in previous periods, the pipeline consists of θ_t full and $L_0 - \theta_t$ empty orders.

Let X_t denote the highest rank of the order in the pipeline at time t , which cannot be made available. Then, $X_t = 0$ implies that *all* orders can be speeded up (even the order just placed), whereas $X_t = L_0$ means that no orders can be speeded up. The probability that the order placed j periods ago can be speeded up is given by

$$\mathcal{F}_j = P(X_t \leq j) = \sum_{n=0}^j f_n, \quad j = 0, \dots, L_0 - 1$$

with $f_n = P(X_t = n)$ and $f_{L_0} = 1 - \mathcal{F}_{L_0-1}$.

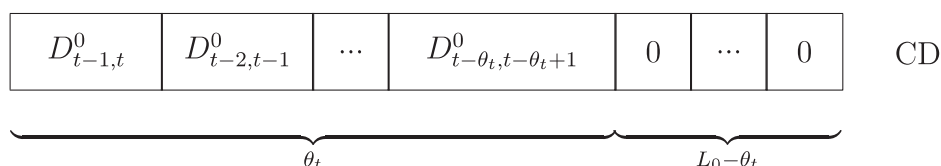


Fig. 2. The pipeline of the central depot.

3.1.2. The ordering process

Let θ_t denote the number of nonempty orders in the pipeline towards the CD, i.e., which have not been pulled to the CD, just before deciding upon the use of the operating flexibility. At least the order placed at the beginning of t is available, thus $1 \leq \theta_t \leq L_0$. At time t , after determining which orders can be speeded up, each retailer places an order at the CD to raise its inventory position to its order-up-to-level. Whether the CD is capable of satisfying all these orders strongly depends on the realization of X_t . We distinguish between two cases.

3.1.2.1. *The case $X_t \geq \theta_t$.* This means that all the orders in the pipeline cannot be speeded up. Then from a sample path argument it follows that the number of products in the pipeline equals $D_{t-\theta_t,t}^0$. Hence, the echelon stock of the CD equals $S_0 - D_{t-\theta_t,t}^0$. When this echelon stock exceeds $\sum_{i=1}^N S_i$ we know that all retail orders can be satisfied.

On the other hand, when $S_0 - D_{t-\theta_t,t}^0$ is less than $\sum_{i=1}^N S_i$ we need to allocate the available echelon stock of the CD appropriately to the retailers. Most models discussed in the literature use linear allocation functions (van der Heijden, 1997), although, an optimal allocation scheme under a cost framework is non-linear (Diks and de Kok, 1998). For an overview and a numerical comparison of different allocation functions we refer to van der Heijden *et al.* (1997). For the remainder of this paper we use the simple variant of the Balanced Stock (BS) rationing policy (van der Heijden, 1997), adapted according to de Kok (1999). In this policy the shortage of the CD (to satisfy all retail orders) is allocated to the retailers by the allocation fractions $\{q_i\}_{i=1}^N$, which are set to

$$q_i = \frac{\sigma_i^2}{2 \sum_{n=1}^N \sigma_n^2} + \frac{\mu_i^2}{2 \sum_{n=1}^N \mu_n^2}, \quad i = 1, \dots, N. \quad (2)$$

Hence,

$$\hat{I}_t^i = S_i - q_i(D_{t-\theta_t,t}^0 - \Delta)^+, \quad i = 1, \dots, N. \quad (3)$$

Note that BS rationing requires the generalized balance assumption (de Kok *et al.*, 1994), which assumes that the inventory position after rationing is always larger than just before rationing.

3.1.2.2. *The case $X_t < \theta_t$.* Here, the CD is able to speed up the $\theta_t - X_t$ consecutive, nonempty orders ‘nearest’ to the CD. If, indeed, the CD decides to speed up all these $\theta_t - X_t$ orders, the number of products left in the pipeline

would be equal to $D_{t-X_i,t}^0$. Then, the echelon stock of the CD would be equal to $S_0 - D_{t-X_i,t}^0$. If this echelon stock would be less than $\sum_{i=1}^N S_i$, we know that all available nonempty orders will be speeded up, and that it is still not sufficient to satisfy all retail orders. On the other hand, if the echelon stock exceeds $\sum_{i=1}^N S_i$, then we know that sufficient products can be made available to satisfy all the retail orders. From this it follows that

$$\hat{I}_t^i = S_i - q_i(D_{t-X_i,t}^0 - \Delta)^+, \quad i = 1, \dots, N. \quad (4)$$

From (3) and (4) it follows that the inventory position of retailer i at time t , just after using the flexibility equals

$$\hat{I}_t^i = S_i - q_i(D_{t-\min(\theta_i, X_i), t}^0 - \Delta)^+, \quad i = 1, \dots, N. \quad (5)$$

To evaluate (1) we need to characterize the stationary behavior of \hat{I}_t^i . Suppose that the stationary behavior of θ_t is given by the distribution $\pi_j = \lim_{t \rightarrow \infty} P(\theta_t = j)$. Then, we are able to evaluate β_i as follows. By first substituting (5) into (1), and next conditioning on X_t and θ_t (for $t \rightarrow \infty$) we obtain

$$\beta_i = 1 - \sum_{j=1}^{L_0} \sum_{l=0}^{L_0} \frac{\pi_j f_l}{\mu_i} \left\{ E(D_{L_i+1}^i + q_i(D_{\min(j,l)}^0 - \Delta)^+ - S_i)^+ - E(D_{L_i}^i + q_i(D_{\min(j,l)}^0 - \Delta)^+ - S_i)^+ \right\}. \quad (6)$$

The fill rate β_i can be computed from (6) by using the following procedure. First, we fit a suitable distribution to the first two moments of $D_{\min(j,l)}^0$. We suggest to use a mixture of two Erlang distributions, since it is closely related to the gamma distribution. Therefore all advantages of using the gamma distribution as an approximation of the true lead time demand distribution (Burgin, 1975) remain valid, but computations are greatly simplified. If we fit a mixed Erlang distribution to the first two moments of a non-negative real random variable X , we mean that X follows an E_{r_1, λ_1} with probability ω_1 , and an E_{r_2, λ_2} with probability ω_2 . In Tijms (1994) several fitting procedures are described to determine the parameters $\lambda_1, \lambda_2, r_1, r_2, \omega_1$, and ω_2 based on the first two moments. Second, after fitting a mixed Erlang distribution to $D_{\min(j,l)}^0$, we compute the first two moments of $(D_{\min(j,l)}^0 - \Delta)^+$. We apply the following formulae to compute the first two moments of a random variable $(X - c)^+$, when X follows a mixed Erlang distribution as described above.

$$E(X - c)^+ = \sum_{i=1}^2 \frac{\omega_i}{\lambda_i} \sum_{n=0}^{r_i-1} (r_i - n) \frac{(\lambda_i c)^n}{n!} e^{-\lambda_i c}, \quad (7)$$

$$E((X - c)^+)^2 = 2 \sum_{i=1}^2 \frac{\omega_i}{\lambda_i^2} \sum_{j=0}^{r_i-1} \sum_{n=0}^j (j+1-n) \frac{(\lambda_i c)^n}{n!} e^{-\lambda_i c}. \quad (8)$$

Third, we fit mixed Erlang distributions to $D_{L_i}^i + q_i(D_{\min(j,l)}^0 - \Delta)^+$ and $D_{L_i+1}^i + q_i(D_{\min(j,l)}^0 - \Delta)^+$, respectively.

Finally, we determine $E(D_{L_i+1}^i + q_i(D_{\min(j,l)}^0 - \Delta)^+ - S_i)^+$ and $E(D_{L_i}^i + q_i(D_{\min(j,l)}^0 - \Delta)^+ - S_i)^+$ by applying (7).

The remainder of this section is devoted to determine the steady-state probabilities $\{\pi_j\}$.

3.1.3. Steady-state probabilities $\{\pi_j\}$

To determine the steady-state distribution of θ_t , where t tends to infinity, we use the following relation

$$\theta_{t+1} = \min(L_0, \theta_t - Y_t + 1). \quad (9)$$

First, we derive an approximation for the distribution of Y_t . Next, we use discrete-time Markov theory to determine $\{\pi_j\}_{j=1}^{L_0}$. The number of orders speeded up by the CD at time t equals

$$Y_t = \begin{cases} \arg \min_{0 \leq i \leq \theta_t - X_t} \{D_{t-\theta_t+i, t}^0 \leq \Delta\} & X_t < \theta_t, \\ 0 & X_t \geq \theta_t. \end{cases} \quad (10)$$

This implies that orders are pulled into the CD (following a FIFO rule) until the remaining pipeline stock is less than or equal to Δ or that all available orders are pulled ($Y_t = \theta_t - X_t$) but the remaining pipeline stock is still larger than Δ . To determine the distribution of Y_t , again, we distinguish between two cases.

3.1.3.1. *The case $X_t \geq \theta_t$.* In this case the CD is not able to speed up any order. Hence, $Y_t = 0$.

3.1.3.2. *The case $X_t < \theta_t$.* From (10) it follows that $P(Y_t = 0) = P(D_{t-\theta_t, t}^0 \leq \Delta)$. For $i = 1, 2, \dots, \theta_t - X_t$ we find

$$\begin{aligned} P(Y_t = i) &= P(X_t \leq \theta_t - i) \times P(D_{t-\theta_t+i-1, t}^0 > \Delta, D_{t-\theta_t+i, t}^0 \leq \Delta) \\ &\quad + P(X_t = \theta_t - i) \times P(D_{t-\theta_t+i, t}^0 > \Delta), \\ &= \mathcal{F}_{\theta_t-i} \times \left(P(D_{t-\theta_t+i, t}^0 \leq \Delta) - P(D_{t-\theta_t+i-1, t}^0 \leq \Delta, D_{t-\theta_t+i, t}^0 \leq \Delta) \right) \\ &\quad + f_{\theta_t-i} \times (1 - P(D_{t-\theta_t+i, t}^0 \leq \Delta)), \\ &= \mathcal{F}_{\theta_t-i} \times \left(F_{D_{t-\theta_t+i, t}^0}^0(\Delta) - F_{D_{t-\theta_t+i-1, t}^0}^0(\Delta) \right) \\ &\quad + f_{\theta_t-i} \times (1 - F_{D_{t-\theta_t+i, t}^0}^0(\Delta)). \end{aligned}$$

So far the derivation has been exact. However, it is rather difficult to compute the distribution of Y_t , since the value of θ_t includes some information about the cumulative orders in the pipeline. If for instance the pull process in a previous period stopped because the depot shortage was covered, then the remaining pipeline stock was less than or equal to Δ whereas the inventory plus the last order pulled exceeded Δ . To obtain an easy and good approximation for Y_t we suggest to approximate Y_t as follows

$$\begin{aligned} P(Y_t = i) &\approx \mathcal{F}_{\theta_t-i} \times \left(F_{D_{\theta_t-i}^0}^0(\Delta) - F_{D_{\theta_t-i+1}^0}^0(\Delta) \right) \\ &\quad + f_{\theta_t-i} \times \left(1 - F_{D_{\theta_t-i}^0}^0(\Delta) \right). \end{aligned}$$

Note that this approximation assumes that the demand during the interval $[t - \theta_t + i, t]$ is *stochastically* identical to the demand during $(\theta_t - i)$ arbitrary periods in time.

From this it can easily be seen that

$$P(Y_t = i) = \begin{cases} (1 - \mathcal{F}_{\theta_t-1}) + \mathcal{F}_{\theta_t-1} \times \alpha_{\theta_t} & i = 0, \\ \mathcal{F}_{\theta_t-i} \times (\alpha_{\theta_t-i} - \alpha_{\theta_t-i+1}) + f_{\theta_t-i} \times (1 - \alpha_{\theta_t-i}) & i = 1, \dots, \theta_t - X_t. \end{cases} \quad (11)$$

$\alpha_j = P(D_j^0 \leq \Delta)$ represents the probability that the depots cumulative demand over j arbitrary periods does not exceed Δ .

Now the process $\{\theta_t\}_t$ can be modeled as a discrete-time Markov chain (see Fig. 3), since from (9) it follows that the realization of θ_{t+1} only depends on θ_t and Y_t (which only depends on θ_t due to the approximation assumption).

The steady-state transition probabilities $\gamma_{i,j} := \lim_{t \rightarrow \infty} P(\theta_{t+1} = j | \theta_t = i)$ are given by

$$\gamma_{i,i+1} = P(Y_t = 0) \approx \alpha_i \times \mathcal{F}_{i-1} + 1 - \mathcal{F}_{i-1}, \quad 1 \leq i \leq L_0 - 1,$$

$$\begin{aligned} \gamma_{i,i+1-j} &= P(Y_t = j), \\ &\approx \mathcal{F}_{i-j} \times (\alpha_{i-j} - \alpha_{i-j+1}) + f_{i-j} \times (1 - \alpha_{i-j}), \\ & \quad 1 \leq i < L_0, \quad 1 \leq j \leq i, \end{aligned}$$

$$\begin{aligned} \gamma_{L_0, L_0+1-j} &= P(Y_t = j), \\ &\approx \mathcal{F}_{L_0-j} \times (\alpha_{L_0-j} - \alpha_{L_0-j+1}) + f_{L_0-j} \times (1 - \alpha_{L_0-j}), \\ & \quad 1 < j \leq L_0, \end{aligned}$$

$$\begin{aligned} \gamma_{L_0, L_0} &= P(Y_t \leq 1) \approx (1 - f_{L_0}) \times \alpha_{L_0-1} + f_{L_0} \\ & \quad + f_{L_0-1} \times (1 - \alpha_{L_0-1}). \end{aligned}$$

Depending on \mathcal{F} and Δ , two special cases can occur. Suppose $\mathcal{F}_i = 0$ for some $0 \leq i < L_0$, then the steady-state probabilities for all states $j \leq i$ are zero. Suppose $\Delta = 0$, $\mathcal{F}_i = 1$ for some $0 \leq i < L_0$ and $\mathcal{F}_{i-1} < 1$ then state i is an absorbing state. For all other cases the steady-state probabilities $\pi_i = \lim_{t \rightarrow \infty} P(\theta_t = i)$ are given by the solution of the following L_0 linear equations:

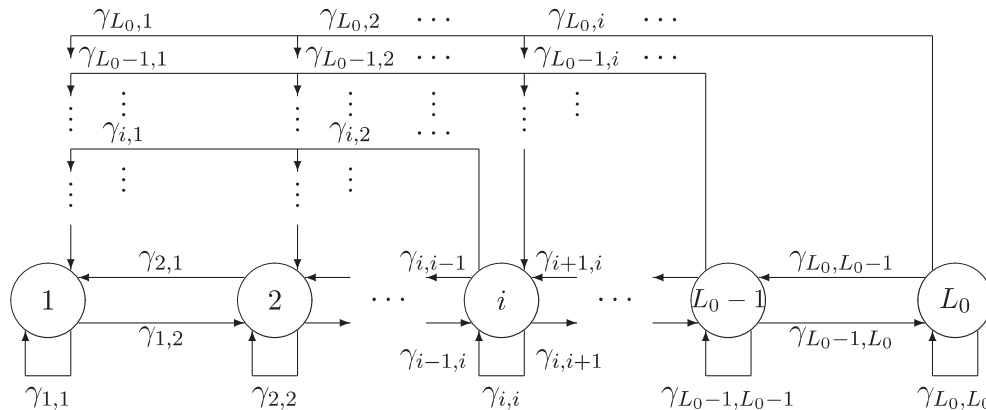


Fig. 3. The discrete-time Markov chain of $\{\theta_t\}_t$.

$$\sum_{n=1}^{L_0} \pi_n = 1, \quad \pi_i = \gamma_{i-1,i} \times \pi_{i-1} + \sum_{j=i}^{L_0} \gamma_{j,i} \times \pi_j, \quad 2 \leq i \leq L_0. \quad (12)$$

These equations can be transformed into $\pi_i = c_i \times \pi_{L_0}$. By defining $c_{L_0} := 1$ we obtain

$$c_{i-1} = \left(c_i - \sum_{j=i}^{L_0} c_j \times \gamma_{j,i} \right) / \gamma_{i-1,i} \quad 2 \leq i \leq L_0, \quad (13)$$

for the coefficients c_i . The π_{L_0} is computed from $\pi_{L_0} \times \sum_{i=1}^{L_0} c_i = 1$.

3.2. Determination of p_j and $E(Y)$

The expected number of orders of age j being speeded up per period is given by the probability that an order of age j is speeded up. In order to be able to speed up the order of age j in period t there have to be at least $j + 1$ full orders in the pipeline to the CD, i.e., $\theta_t \geq j + 1$. Then, the order of age j is speeded up if the number of pulled orders Y_t is larger than or equal to $\theta_t - j$. Therefore,

$$p_j(t) = \sum_{i=j+1}^{L_0} P(\theta_t = i) \times \sum_{k=i-j}^i P(Y_t = k), \quad j = 0, \dots, L_0 - 1. \quad (14)$$

Using the results of Section 3.1, we find

$$p_j = \sum_{i=j+1}^{L_0} \pi_i \times \sum_{k=i-j}^i (\mathcal{F}_{i-k} (\alpha_{i-k} - \alpha_{i-k+1}) + f_{i-k} (1 - \alpha_{i-k})). \quad (15)$$

Then, the expected workload is $E(Y) = \sum_{j=0}^{L_0-1} a_j p_j$.

In the case where the rescheduling cost or the added workload is the same for all outstanding orders we are able to find a simpler expression for the expected number $E(Y)$ of orders being pulled. By definition, the expected

number of orders pulled in period t is given by the expected number of full orders in the depot pipeline at the beginning of period t before exploiting the flexibility minus the expected number of full orders at the end of t after the use of flexibility. Therefore, it is necessary to derive the steady-state probability distribution $\hat{\pi}_i$, $i = 0, \dots, L_0$ of $\hat{\theta}_t$.

From the following theorem it is evident how to compute $E(Y_t)$.

Theorem 1. $E(Y_t) = 1 - \hat{\pi}_{L_0}$.

Proof. By definition it holds that

$$\theta_{t+1} = \min(\hat{\theta}_t, L_0 - 1) + 1. \quad (16)$$

Suppose the pipeline towards the CD at time t , after using flexibility, consists of at least one ‘empty’ order. Then at the beginning of review period $t + 1$, an ‘empty’ order arrives at the CD. Otherwise, a nonempty order arrives at the CD (thus the number of nonempty orders decreases by one). After this arrival a new order is placed in the pipeline (thus the nonempty orders increases by one). From (16) we obtain

$$\pi_i = \begin{cases} \hat{\pi}_{i-1} & 1 \leq i < L_0, \\ \hat{\pi}_{L_0-1} + \hat{\pi}_{L_0} & i = L_0. \end{cases} \quad (17)$$

From the definition of Y_t it follows that

$$E(Y_t) = E(\theta_t) - E(\hat{\theta}_t) = \sum_{i=1}^{L_0} i\pi_i - \sum_{i=0}^{L_0} i\hat{\pi}_i. \quad (18)$$

After substituting (17) in (18), and rewriting the result we obtain $E(Y_t) = 1 - \hat{\pi}_{L_0}$. ■

So to compute $E(Y_t)$ we only need a tractable expression for $\hat{\pi}_{L_0}$. By definition we have

$$\begin{aligned} \hat{\pi}_{L_0} &= \lim_{t \rightarrow \infty} P(\hat{\theta}_t = L_0) = \lim_{t \rightarrow \infty} P(\theta_t = L_0 \wedge \\ Y_t = 0) &= \pi_{L_0}(f_{L_0} + (1 - f_{L_0})\alpha_{L_0}). \end{aligned} \quad (19)$$

3.3. Determination of the objective function

To evaluate the objective function we need to compute OH_i and PS_0 . From a sample path argument it can be shown that the stock on-hand at the CD at time t , just after allocation of the depot stock equals $(\Delta - D_{t-\hat{\theta}_t, t}^0)^+$. Hence,

$$OH_0 = E\left(\Delta - D_{t-\hat{\theta}_t, t}^0\right)^+ \approx \sum_{i=0}^{L_0} \hat{\pi}_i E(\Delta - D_i^0)^+.$$

Note that the distribution $\{\hat{\pi}_i\}$ can easily be obtained from (17) and (19).

The stock on-hand at a retailer i at time $t + L_i$, just after the arrival of an order, equals $E(\hat{I}_t^i - D_{t+L_i}^i)^+$. Hence,

$$OH_i = E(\hat{I}_t^i - D_{t+L_i}^i)^+, \quad i = 1, \dots, N.$$

Substituting (4) into the expression above, and next applying the procedure discussed in Section 3.1.2 yields an approximation of OH_i .

Finally, it can easily be seen that the amount of products in the pipeline towards the CD at time t , just after allocation, equals $D_{t-\hat{\theta}_t, t}^0$. Hence,

$$PS_0 = E\left(D_{t-\hat{\theta}_t, t}^0\right) \approx \sum_{i=0}^{L_0} \hat{\pi}_i E(D_i^0).$$

4. Validation of the algorithm

In this section we test the quality of the approximations derived in Section 3 for all relevant performance characteristics. Note that these approximations are based on the following two assumptions: (i) no imbalance when rationing shortages; and (ii) demand during the interval $(t - \theta_t + i, t]$ is equal in distribution to a sum of $\theta_t - i$ i.i.d. period demands. We set up a simulation experiment consisting of 320 distribution systems each run during 200 000 periods, which ensures stationary behavior. The parameters describing these distribution systems are varied as follows.

- Number of retailers $N \in \{2, 4\}$. The retailers are divided into two groups of identical retailers.
- Depot lead time $L_0 \in \{2, 3\}$. For $L_0 = 2$ we choose $\mathbf{f} \in \{(0, 1, 0), (1, 0, 0)\}$. For $L_0 = 3$, $\mathbf{f} \in \{(1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0)\}$. We set $a_j = 1 \forall j$, therefore we can use the simplified expression for the expected number of rescheduled orders.
- Mean demand $\mu_1 = 10$ for the first and $\mu_2 \in \{10, 20\}$ for the second service group.
- Coefficient of variation ($v = \sigma/\mu$) for the first $v_1 = 0.4$ and the second $v_2 \in \{0.4, 0.8\}$ service group.
- Retailer lead times $L_1 = 1$ and $L_2 \in \{1, 2\}$.
- Service levels $\beta_1^* = 0.9$ and $\beta_2^* \in \{0.9, 0.95\}$.
- Maximum physical depot stock $\Delta \in \{0.5 \times L_0 \times \mu_0, L_0 \times \mu_0\}$.

We compute the average and maximum deviation in service, expected pipeline and on-hand stock at the CD, sum of retailer on-hand stocks and in the expected number of rescheduled orders which are reported in Table 1.

The numbers for service levels and rescheduled orders are absolute errors whereas the deviations in expected

Table 1. Deviations of analytical from simulated performance measures

	β_i	$PS_0 + OH_0$ (%)	$\sum_i OH_i$ (%)	$E(Y)$
Average	0.0006	4.2	0.1	0.0086
Maximum	0.0076	15.6	0.6	0.0718

stock levels are given as relative errors. The errors reported for the sum of depot pipeline and on-hand stock can be reduced to 0.05% on average and 0.4% at maximum when the refinements outlined in the Appendix are applied instead of the ones given in Section 3.3. We conclude that the approximations derived in Section 3 perform well and are suited for use in an optimization scheme. In the next section we present a number of managerial problems related to supply lead time flexibility. The approach developed in this paper enables us to gain insights into these problems. This is of considerable practical interest, since supply flexibility is still a concept, which is not well-understood.

5. Managerial insights into supply lead time flexibility

Now that we have modeled supply lead time flexibility in a realistic way and have shown that the resulting approximations perform well, we return to the objective of our paper. The objective is to evaluate the contribution of supply lead time flexibility to the performance of a supply chain consisting of a depot and multiple retailers. The apparent questions to be answered are:

1. What is the impact of supply lead time flexibility on total operating costs?
2. What is the structure of the optimal rescheduling policy? And, in case the optimal rescheduling policy turns out to be impracticable,
3. Do practicable rescheduling policies exist that are near-optimal?

Questions 2 and 3 are answered in the context of either minimization of total operating costs, including rescheduling costs, subject to fill rate constraints (Problem 1), or minimization of holding costs subject to a rescheduling capacity and fill rate constraints (Problem 2). However, in general we do not know the optimal policy. Instead we determine the optimal policy within a special class of policies that can be modeled by the flexibility function \mathcal{F} . To get insight into the questions posed above we consider the following set of models.

$N = 2$ (identical), $L_0 \in \{1, 2\}$, $L_i = 1$, $h_0 = 1$, $h_i \in \{1, 3\}$, $\beta_i^* = 0.9$, $c_r^0 \in \{10, 20, \dots, 100\}$, $c_r^1 \in \{0.0, 0.25, \dots, 1.0\} \times c_r^0$, $a_0 = 1$, $a_1 \in \{0.0, 0.25, \dots, 1.0\} \times a_0$, $Y^* \in \{0, 0.2, \dots, 1\}$, $\mu_i = 10$, $\sigma_i \in \{4, 8\}$.

Note that for Problem 1 the rescheduling costs c_r^j are relevant, whereas for Problem 2, the a_j and Y^* are relevant. For each case with $L_0 = 1$ we varied Δ between zero and 30. For each case with $L_0 = 2$ we varied Δ between zero and 60. The maximum values of Δ have been selected based on the observation in de Kok *et al.* (1994) that if Δ exceeds 1.5 times the demand during the depot lead time, then an increase in Δ does not have any effect on the retailer stocks. The optimal policies without rescheduling opportunities and different depot lead times are given in

Table 2 for reference purposes. Note that $L = 0$ refers to independent retailers that operate with a lead time of L_i periods.

In the following Tables 3 and 4 we first present some selected optimal policies for $L_0 = 1$ and $L_0 = 2$. ΔC denotes the relative cost reduction compared with the cost in case of no flexibility. Recall that $f_0 = 1$ implies full

Table 2. Optimal policies without rescheduling

L	$h_i = 1$				$h_i = 3$			
	$\sigma_i = 4$		$\sigma_i = 8$		$\sigma_i = 4$		$\sigma_i = 8$	
	Δ^*	C^*	Δ^*	C^*	Δ^*	C^*	Δ^*	C^*
0	0.0	29.1	0.0	48.3	0.0	87.2	0.0	145.0
1	0.0	50.7	0.0	71.6	18.3	111.6	16.9	174.7
2	0.0	72.2	0.0	94.7	40.1	134.6	38.6	200.8

Table 3. Optimal policies for $L_0 = 1$

h_i	σ_i	c_r^0	f_0^*	Δ^*	C^*	$\Delta C(\%)$
1	4	10	1	0.0	39.1	23
1	8	10	1	0.0	58.3	19
1	4	50	0	0.0	50.7	0
1	8	50	0	0.0	71.6	0
3	4	10	1	0.0	97.2	13
3	8	10	1	0.0	155.0	11
3	4	50	0	18.3	111.6	0
3	8	50	0	16.9	174.7	0

Table 4. Optimal policies for $L_0 = 2$

h_i	σ_i	c_r^0	c_r^1	f_0^*	f_1^*	Δ^*	C^*	$\Delta C(\%)$
1	4	10	10	1	0	0.0	39.1	46
1	8	10	10	1	0	0.0	58.3	62
1	4	40	10	0	1	0.0	60.7	16
1	4	40	40	1	0	0.0	69.1	4
1	4	60	15	0	1	0.0	65.7	9
1	4	60	60	0	0	0.0	72.2	0
1	8	40	10	0	1	0.0	81.6	14
1	8	40	40	1	0	0.0	88.3	7
1	8	60	15	0	1	0.0	86.6	9
1	8	60	60	0	0	0.0	94.7	0
3	4	10	10	1	0	0.0	97.2	28
3	8	10	10	1	0	0.0	155.0	23
3	4	40	10	0	1	18.3	121.6	10
3	4	40	40	1	0	0.0	127.2	5
3	4	60	15	0	1	18.3	126.6	6
3	4	60	60	0	0	40.1	134.6	0
3	8	40	10	0	1	16.9	184.3	8
3	8	40	40	1	0	0.0	185.0	8
3	8	60	15	0	1	16.9	189.1	6
3	8	60	60	0	0	38.6	200.8	0

flexibility for all lead times L_0 . For $L_0 = 1$ no flexibility is equivalent to $f_0 = 0$, while for $L_0 = 2$ no flexibility is equivalent to $(f_0, f_1) = (0, 0)$. We deduce the following observations: (i) supply lead time may have a major impact on total costs; (ii) the optimal policy either prescribes a probability of one or zero for being able to pull a certain order if needed. In this respect the optimal policy has a “bang-bang” characteristic; (iii) if the retailers add no value, then the optimal policy is to set Δ equal to zero; and (iv) if the retailers add value, then the optimal value of Δ depends on the size of the rescheduling cost parameters. If all orders can be speeded up at a relatively low cost, there is no need for central inventories. If rescheduling is expensive and therefore completely avoided, Δ is close to the expected demand during the manufacturing lead time. For a sufficiently small cost of rescheduling the next arriving order in the case of $L_0 = 2$ we observe that it is beneficial only to reschedule this order. Then, the optimal policy is to set Δ to the optimal value obtained from a model without rescheduling and a lead time of one period (see Table 2).

For all other values of c_r^j these observations have been confirmed. Of course the observations are not proven formally, yet other numerical experiments with longer lead times confirm our observations. Using the costs of the optimal policies under certain lead times as given in Table 2 together with the rescheduling costs leads us to the definition of critical values for the rescheduling cost parameters, i.e., when to neglect flexibility, when to use full flexibility and when to use partial flexibility (for the

Table 5. Critical values

L_0	h_i	σ_i	$(c_r^0)^F$	$(c_r^1)^F$	$(c_r^0 - c_r^1)^F$
1	1	4	21	–	–
1	1	8	23	–	–
1	3	4	24	–	–
1	3	8	28	–	–
2	1	4	43	21	22
2	1	8	46	23	23
2	3	4	47	23	24
2	3	8	56	26	30

next arriving order only). Let $(c_r^0)^F := C^*(L_0 = 2) - C^*(L_0 = 0)$, $(c_r^1)^F := C^*(L_0 = 2) - C^*(L_0 = 1)$, and $(c_r^0 - c_r^1)^F := C^*(L_0 = 1) - C^*(L_0 = 0)$. For $L_0 = 1$ full flexibility is used if c_r^0 is less than or equal to the critical value $(c_r^0)^F$ and otherwise its use is completely avoided. For $L_0 = 2$, full flexibility is used if $c_r^0 \leq (c_r^0)^F$ and $c_r^0 - c_r^1 \leq (c_r^0 - c_r^1)^F$, no flexibility is used if $c_r^0 > (c_r^0)^F$ and $c_r^1 > (c_r^1)^F$, whereas partial flexibility is used in the remaining cases. The critical values for our base set of models are given in Table 5.

From our analysis we may conclude that computing cost-optimal policies is straightforward. Set a single probability f_j equal to one, where j is chosen from

$$j := \arg \min_{0 \leq L \leq L_0} \{C^*(L) + c_r^L\}. \tag{20}$$

One of the problems one faces when addressing the problem of usage of supply lead time flexibility is the

Table 6. Policies for $L_0 = 1$

h_i	Y^*	σ	Optimal					Bang-bang				
			f_0^*	Δ^*	$E[Y]$	C^*	ΔC (%)	f_0^*	Δ^*	$E[Y]$	C^*	ΔC^p (%)
1	0.2	4	0.20	0.0	0.2	48.8	4	0	0.0	0.0	50.7	4
1	0.2	8	0.20	0.0	0.2	68.6	4	0	0.0	0.0	71.6	4
1	0.4	4	0.40	0.0	0.4	46.3	9	0	0.0	0.0	50.7	10
1	0.4	8	0.40	0.0	0.4	65.0	9	1	20.8	0.4	71.6	10
1	0.6	4	0.60	0.0	0.6	42.8	16	1	18.1	0.6	48.5	13
1	0.6	8	0.60	0.0	0.6	60.6	15	1	15.4	0.6	66.4	10
1	0.8	4	0.80	0.0	0.8	37.3	26	1	15.2	0.8	45.3	21
1	0.8	8	0.80	0.0	0.8	55.2	23	1	10.4	0.8	60.8	10
1	1.0	4	1.00	0.0	1.0	29.1	43	1	0.0	1.0	29.1	0
1	1.0	8	1.00	0.0	1.0	48.3	33	1	0.0	1.0	48.3	0
3	0.2	4	0.34	18.3	0.2	110.4	1	0	18.3	0.0	111.6	1
3	0.2	8	0.37	16.9	0.2	171.3	2	0	16.9	0.0	174.0	2
3	0.4	4	0.61	17.3	0.4	108.7	3	1	20.9	0.4	109.3	1
3	0.4	8	0.74	16.9	0.4	167.7	4	1	20.8	0.4	168.2	0
3	0.6	4	0.80	16.0	0.6	106.5	5	1	18.1	0.6	106.6	0
3	0.6	8	1.00	15.4	0.6	163.1	6	1	15.4	0.6	163.1	0
3	0.8	4	0.80	6.1	0.8	102.6	8	1	15.2	0.8	103.5	1
3	0.8	8	1.00	10.4	0.8	157.5	9	1	10.4	0.8	157.5	0
3	1.0	4	1.00	0.0	1.0	87.2	22	1	0.0	1.0	87.2	0
3	1.0	8	1.00	0.0	1.0	145.0	17	1	0.0	1.0	145.0	0

estimation of c_r^l . Instead we can determine the available rescheduling capacity, e.g., by estimating the workload from rescheduling and estimating the total working hours of planners available for rescheduling. From that one derives the normalized rescheduling capacity. Finally one may divide this number among all products produced so that for each product the rescheduling capacity $Y^* \leq 1$ is determined. This reasoning motivates the formulation of Problem 2. For Problem 2 we follow the same steps as for Problem 1. In Tables 6–8 we present the results for the set of problems defined above.

We can draw the following conclusions: (i) available supply lead time flexibility is completely used in the optimal policy; (ii) bang-bang policies are no longer optimal; (iii) if no value is added at the retailers all stocks are pushed out to the retailers, i.e., $\Delta^* = 0$; (iv) if we have enough

capacity to provide full flexibility then $\Delta^* = 0$; and (v) as flexibility increases the marginal decrease in cost increases, i.e., flexibility pays off if it is high. As the structure of the optimal policies is no longer of the bang-bang type, it seems not easy to implement these policies. Therefore, we investigate the cost penalty ΔC^p of using the optimal bang-bang policy instead of the overall optimal policy. Concerning bang-bang policies we draw the following conclusions: (i) the cost penalty is small in the case where the retailers add value, the cost penalty may be considerable if the retailers add no value; and (ii) in contrast to the optimal policy, the best bang-bang policy does not necessarily use all available flexibility, however in most cases the workload reaches the capacity limit. Further, even if no value is added, inventories are implemented in order to achieve feasibility of a bang-bang policy.

Table 7. Policies for $L_0 = 2$ and $h_i = 1$

Y^*	c_r^l	σ	Optimal						Bang-bang					
			f_0^*	f_1^*	Δ^*	$E[Y]$	C^*	$\Delta C(\%)$	f_0^*	f_1^*	Δ^*	$E[Y]$	C^*	$\Delta C^p(\%)$
0.2	0.25	4	0.00	0.80	0	0.2	58.5	19	0	1	33.2	0.2	63.8	9
0.2	0.25	8	0.00	0.80	0	0.2	78.1	18	0	1	26.3	0.2	80.4	3
0.2	0.50	4	0.00	0.40	0	0.2	67.6	6	0	1	41.5	0.2	72.1	7
0.2	0.50	8	0.00	0.40	0	0.2	87.9	7	0	1	42.0	0.2	93.9	7
0.2	0.75	4	0.00	0.26	0	0.2	69.5	4	0	0	0.0	0.0	72.2	4
0.2	0.75	8	0.05	0.16	0	0.2	90.5	4	0	0	0.0	0.0	94.7	5
0.2	1.00	4	0.10	0.00	0	0.2	70.1	3	0	0	0.0	0.0	72.2	3
0.2	1.00	8	0.10	0.00	0	0.2	91.1	4	0	0	0.0	0.0	94.7	4
0.4	0.25	4	0.20	0.80	0	0.4	48.8	32	0	1	0.0	0.25	50.7	4
0.4	0.25	8	0.20	0.80	0	0.4	68.6	28	0	1	0.0	0.25	71.6	4
0.4	0.50	4	0.00	0.80	0	0.4	58.5	19	0	1	33.2	0.4	63.8	9
0.4	0.50	8	0.00	0.80	0	0.4	78.1	18	0	1	26.3	0.4	80.4	3
0.4	0.75	4	0.00	0.53	0	0.4	65.4	9	0	1	38.8	0.4	69.7	7
0.4	0.75	8	0.04	0.46	0	0.4	85.1	10	0	1	36.6	0.4	89.4	5
0.4	1.00	4	0.22	0.00	0	0.4	67.4	7	0	1	41.5	0.4	72.1	7
0.4	1.00	8	0.22	0.00	0	0.4	86.5	9	0	1	42.0	0.4	93.9	9
0.6	0.25	4	0.46	0.54	0	0.6	45.4	37	1	0	20.0	0.6	50.3	11
0.6	0.25	8	0.46	0.54	0	0.6	63.8	33	1	0	18.6	0.6	70.1	10
0.6	0.50	4	0.20	0.80	0	0.6	48.8	32	0	1	0.0	0.5	50.7	4
0.6	0.50	8	0.20	0.80	0	0.6	68.6	28	0	1	0.0	0.5	71.6	4
0.6	0.75	4	0.00	0.80	0	0.6	58.5	19	0	1	33.2	0.6	63.8	9
0.6	0.75	8	0.36	0.14	0	0.6	77.9	18	0	1	26.3	0.6	80.4	3
0.6	1.00	4	0.37	0.00	0	0.6	63.2	12	0	1	37.5	0.6	68.4	8
0.6	1.00	8	0.37	0.00	0	0.6	80.3	15	0	1	34.1	0.6	87.2	9
0.8	0.25	4	0.73	0.27	0	0.8	39.5	45	1	0	16.2	0.8	46.5	18
0.8	0.25	8	0.73	0.27	0	0.8	57.2	40	1	0	12.1	0.8	62.8	10
0.8	0.50	4	0.60	0.40	0	0.8	42.8	41	1	0	18.1	0.8	48.5	13
0.8	0.50	8	0.60	0.40	0	0.8	60.6	36	1	0	15.1	0.8	66.3	9
0.8	0.75	4	0.20	0.80	0	0.8	48.8	32	0	1	0.0	0.75	50.7	4
0.8	0.75	8	0.62	0.00	0	0.8	67.1	29	0	1	0.0	0.75	71.6	7
0.8	1.00	4	0.55	0.00	0	0.8	55.6	23	0	1	33.2	0.8	63.8	15
0.8	1.00	8	0.55	0.00	0	0.8	71.0	25	0	1	26.3	0.8	80.4	13
1.0	1.00	4	1.00	0.00	0	1.0	29.1	60	1	0	0.0	1.0	29.1	0
1.0	1.00	8	1.00	0.00	0	1.0	48.3	49	1	0	0.0	1.0	48.3	0

Table 8. Policies for $L_0 = 2$ and $h_i = 3$

Y^*	c_r^l	σ	Optimal						Bang-bang					
			f_0^*	f_1^*	Δ^*	$E[Y]$	C^*	$\Delta C(\%)$	f_0^*	f_1^*	Δ^*	$E[Y]$	C^*	$\Delta C^p(\%)$
0.2	0.25	4	0.00	1.00	33.2	0.2	122.0	9	0	1	33.2	0.2	122.0	0
0.2	0.25	8	0.00	1.00	26.3	0.2	179.1	8	0	1	26.3	0.2	179.1	0
0.2	0.50	4	0.00	0.76	39.0	0.2	130.0	3	0	1	41.5	0.2	130.3	0
0.2	0.50	8	0.00	0.95	41.1	0.2	190.9	5	0	1	42.0	0.2	191.0	0
0.2	0.75	4	0.00	0.54	39.6	0.2	131.9	2	0	1	44.6	0.2	133.0	1
0.2	0.75	8	0.00	0.60	40.2	0.2	194.8	3	0	1	48.4	0.2	196.2	1
0.2	1.00	4	0.10	0.31	39.7	0.2	132.7	1	0	0	40.0	0.0	134.6	1
0.2	1.00	8	0.00	0.46	40.5	0.2	196.6	2	0	1	52.5	0.2	199.6	2
0.4	0.25	4	0.32	0.68	17.8	0.4	110.4	18	0	1	18.3	0.25	111.6	1
0.4	0.25	8	0.39	0.61	16.9	0.4	171.7	14	0	1	16.9	0.24	174.7	2
0.4	0.50	4	0.00	1.00	33.2	0.4	122.0	9	0	1	33.2	0.4	122.0	0
0.4	0.50	8	0.00	1.00	26.3	0.4	179.1	11	0	1	26.3	0.4	179.1	0
0.4	0.75	4	0.17	0.77	38.2	0.4	127.8	5	0	1	38.8	0.4	127.8	0
0.4	0.75	8	0.00	1.00	36.6	0.4	186.7	7	0	1	36.6	0.4	186.7	0
0.4	1.00	4	0.14	0.64	39.2	0.4	130.0	3	0	1	41.5	0.4	130.3	0
0.4	1.00	8	0.02	0.90	40.6	0.4	190.9	5	0	1	42.5	0.4	191.0	0
0.6	0.25	4	0.67	0.33	16.8	0.6	108.0	20	1	0	20.0	0.6	108.4	0
0.6	0.25	8	0.83	0.17	16.1	0.6	166.5	17	1	0	18.6	0.6	166.7	0
0.6	0.50	4	0.34	0.66	18.3	0.6	110.4	18	0	1	18.3	0.5	111.6	1
0.6	0.50	8	0.43	0.57	16.9	0.6	171.3	15	1	0	24.0	0.6	172.5	1
0.6	0.75	4	0.00	1.00	33.2	0.6	122.0	9	0	1	33.2	0.6	122.0	0
0.6	0.75	8	0.00	1.00	26.3	0.6	179.1	11	0	1	26.3	0.6	179.1	0
0.6	1.00	4	0.00	0.19	37.5	0.6	126.6	6	0	1	37.5	0.6	126.6	0
0.6	1.00	8	1.00	0.00	35.7	0.6	184.5	8	1	0	35.7	0.6	184.5	0
0.8	0.25	4	0.79	0.21	12.4	0.8	104.5	22	1	0	16.2	0.8	104.7	0
0.8	0.25	8	1.00	0.00	12.1	0.8	159.5	21	1	0	12.1	0.8	159.5	0
0.8	0.50	4	0.82	0.18	16.2	0.8	106.5	21	1	0	18.1	0.8	106.6	0
0.8	0.50	8	1.00	0.00	15.1	0.8	163.0	19	1	0	15.1	0.8	163.0	0
0.8	0.75	4	0.34	0.66	18.2	0.8	110.4	18	0	1	18.3	0.75	111.6	1
0.8	0.75	8	1.00	0.00	21.3	0.8	169.6	16	1	0	21.3	0.8	169.6	0
0.8	1.00	4	0.00	1.00	33.2	0.8	122.0	9	1	0	33.3	0.8	122.0	0
0.8	1.00	8	1.00	0.00	28.4	0.8	177.2	12	1	0	28.4	0.8	177.2	0
1.0	1.00	4	1.00	0.00	0.0	1.0	87.2	35	1	0	0.0	1.0	87.2	0
1.0	1.00	8	1.00	0.00	0.0	1.0	145.0	28	1	0	0.0	1.0	145.0	0

6. Conclusions

In this paper we modeled the presence of manufacturing flexibility at the supplier of a two-stage divergent supply chain. We assumed that complete orders in the pipeline can be speeded up to arrive immediately at the depot according to FIFO. Manufacturing flexibility is modeled by a function $\mathcal{F} = (\mathcal{F}_0, \mathcal{F}_1, \dots, \mathcal{F}_{L_0-1})$, where \mathcal{F}_j equals the probability that an order placed j periods ago can be speeded up. To analyze the system a Markov chain model is solved that yields an approximation of the probability distribution of the number of orders in the pipeline immediately before possible use of flexibility. The analysis yields good approximations for the relevant performance characteristics, such as service levels and costs.

From the analysis of a system with identical retailers we concluded that so-called bang-bang policies are performing well. Bang-bang policies prescribe that certain

orders can be made available with probability one if needed. Which option (i.e., orders) to use depends on the cost of flexibility. Further research will be focused on multi-product models, where flexibility costs or flexibility budgets are determined from models of the manufacturing process itself. This also allows for modeling dependencies between sizes of outstanding orders and rescheduling parameters.

References

Alfredsson, P. and Verrijdt, J.H.C.M. (1999) Modeling emergency supply flexibility in a two-echelon system. *Management Science*, **45**, 1416–1431.
 Axsäter, S. (1993) Continuous review policies for multi-level inventory systems with stochastic demand, in *Logistics of Production and Inventory*, Graves, S., Rinnooy Kan, A. and Zipkin, P. (eds.), Elsevier, Amsterdam, pp. 175–197.

- Burgin, T. (1975) The gamma distribution and inventory control. *Operational Research Quarterly*, **26**, 507–525.
- Dada, M. (1992) A two-echelon inventory system with priority shipments. *Management Science*, **38**, 1140–1153.
- de Kok, A.G. (1999) A final remark on balanced stock rationing. BETA working paper, Eindhoven University of Technology, The Netherlands.
- de Kok, A.G., Lagodimos, A.G. and Seidel, H.P. (1994) Stock allocation in a 2-echelon distribution network under service constraints. Research report TUE/BDK/LBS 94-03, Eindhoven University of Technology, The Netherlands.
- Diks, E.B. and de Kok, A.G. (1998) Optimal control of a divergent multi-echelon inventory system. *European Journal of Operational Research*, **111**, 75–97.
- Hausman, W.H. and Erkip, N.K. (1994) Multi-echelon vs. single-echelon inventory control policies for low-demand items. *Management Science*, **40**, 597–602.
- Moinszadeh, K. and Schmidt, C.P. (1991) An $(S-1, S)$ inventory system with emergency orders. *Operations Research*, **39**, 308–321.
- Silver, E.A., Pyke, D.F. and Peterson, R. (1998) *Decision Systems for Inventory Management and Production Planning*, 3rd edn., Wiley, New York.
- Tijms, H.C. (1994) *Stochastic Models: An Algorithmic Approach*, Wiley, Chichester, UK.
- van der Heijden, M.C. (1997) Supply rationing in multi-echelon systems. *European Journal of Operational Research*, **101**, 532–549.
- van der Heijden, M.C., Diks, E.B. and de Kok, A.G. (1997) Stock allocation in general multi-echelon distribution systems with (R, S) order-up-to-policies. *International Journal of Production Economics*, **49**, 157–174.
- Verrijdt, J.H.C.M., Adan, I. and de Kok, A.G. (1998) A trade-off between emergency repair and inventory investment. *IIE Transactions*, **30**, 119–132.

Appendix

The approximations for depot pipeline and on-hand stocks presented in Section 3.3 can be improved by conditioning on some more information. First, the relation between the distributions of the number of full orders at the beginning of a period $\{\theta_t\}_{(t=1, \dots, L_0)}$ and at the end of a period $\{\hat{\theta}_t\}_{(t=0, \dots, L_0)}$ is exploited. Since $\theta_{t+1} = \min\{L_0, \hat{\theta}_t + 1\}$,

$$P(\hat{\theta} = i) = \begin{cases} P(\theta = i + 1) & \text{if } i < L_0 - 1, \\ P(\theta = L_0 \wedge Y = 0) & \text{if } i = L_0, \\ P(\theta = L_0) - P(\theta = L_0 \wedge Y = 0) & \text{if } i = L_0 - 1, \end{cases}$$

$$= (P(X \leq i) \times (\alpha_i - \alpha_{i+1}) + P(X = i) \times (1 - \alpha_i))$$

$$\times \sum_{\theta=i+1}^{L_0} \pi_\theta$$

$$+ 1_{\{i>0\}} \times \pi_i \times (P(X \geq i) + P(X < i) \times \alpha_i).$$

This formulation includes the following four state transition cases.

1. Pulling is possible ($P(X \leq i) = \mathcal{F}_i$) and it is used ($\alpha_i - \alpha_{i+1}$).
2. Pulling is only possible to some extent ($P(X = i) = f_i$) but more would be necessary to cover the shortage ($1 - \alpha_i$).

3. Pulling is possible ($P(X < i) = \mathcal{F}_{i-1}$) but there is no depot shortage (α_i).
4. No pulling is possible ($P(X \geq i) = 1 - \mathcal{F}_{i-1}$).

Given a number θ of full orders in the pipeline at the beginning of a period the four state transition cases to a result of i full orders at the end of the period provide the following additional information: (i) $D_i^0 \leq \Delta$ and $D_{i+1}^0 > \Delta$; (ii) $D_i^0 > \Delta$; (iii) $D_i^0 \leq \Delta$; and (iv) this case does not provide any additional information.

Replacing the steady-state distribution of $\hat{\theta}$ by the steady-state distribution of θ and the relevant state transition probabilities and by conditioning the expected amount of inventory on the additional information provided in the corresponding case we find

$$OH_0 = \sum_{i=1}^{L_0} ((\mathcal{F}_i \times (\alpha_i - \alpha_{i+1})$$

$$\times E[(\Delta - D_i^0)^+ | D_i^0 \leq \Delta, D_{i+1}^0 > \Delta]$$

$$+ f_i \times (1 - \alpha_i) \times E[(\Delta - D_i^0)^+ | D_i^0 > \Delta]) \times \sum_{\theta=i+1}^{L_0} \pi_\theta$$

$$+ \pi_i \times ((1 - \mathcal{F}_{i-1}) \times E[(\Delta - D_i^0)^+]$$

$$+ \mathcal{F}_{i-1} \times \alpha_i \times E[(\Delta - D_i^0)^+ | D_i^0 \leq \Delta]) + \hat{\pi}_0 \times \Delta.$$

$$PS_0 = \sum_{i=1}^{L_0} ((\mathcal{F}_i \times (\alpha_i - \alpha_{i+1}) \times E[D_i^0 | D_i^0 \leq \Delta, D_{i+1}^0 > \Delta]$$

$$+ f_i \times (1 - \alpha_i) \times E[D_i^0 | D_i^0 > \Delta])$$

$$\times \sum_{\theta=i+1}^{L_0} \pi_\theta + \pi_i \times ((1 - \mathcal{F}_{i-1}) \times E[D_i^0]$$

$$+ \mathcal{F}_{i-1} \times \alpha_i \times E[D_i^0 | D_i^0 \leq \Delta])).$$

Let $\phi_{D_i}(x)$, $F_{D_i}(x)$ denote the i time period demand distribution density and cumulative density function. Using Bayes rule and $P(D_i^0 = x, D_{i+1}^0 > \Delta) = \phi_{D_i}(x) \times (1 - F_{D_i}(\Delta - x))$ we find the following expressions:

$$E[(\Delta - D_i^0)^+ | D_i^0 \leq \Delta, D_{i+1}^0 > \Delta]$$

$$= \int_0^\Delta (\Delta - x)[1 - F_{D_i}(\Delta - x)] dF_{D_i}(x) / (\alpha_i - \alpha_{i+1}),$$

$$= \left(E[(\Delta - D_i^0)^+] - \Delta F_{D_{i+1}}(\Delta) \right. \\ \left. + \int_0^\Delta x F_{D_i}(\Delta - x) dF_{D_i}(x) \right) / (\alpha_i - \alpha_{i+1}).$$

$$E[(\Delta - D_i^0)^+ | D_i^0 > \Delta] = 0,$$

$$E[(\Delta - D_i^0)^+ | D_i^0 \leq \Delta] = E[(\Delta - D_i^0)^+] / \alpha_i.$$

$$E[D_i^0 | D_i^0 \leq \Delta, D_{i+1}^0 > \Delta]$$

$$= \int_0^\Delta x[1 - F_{D_i}(\Delta - x)] dF_{D_i}(x) / (\alpha_i - \alpha_{i+1}),$$

$$= \left(\int_0^\Delta x dF_{D_i}(x) - \int_0^\Delta x F_{D_i}(\Delta - x) dF_{D_i}(x) \right) / (\alpha_i - \alpha_{i+1}).$$

$$E[D_i^0 | D_i^0 > \Delta] = \int_{\Delta}^{\infty} x dF_{D_i}(x) / (1 - \alpha_i),$$

$$E[D_i^0 | D_i^0 \leq \Delta] = \int_0^{\Delta} x dF_{D_i}(x) / \alpha_i.$$

By fitting a mixed Erlang distribution with parameters r , ω , and λ on D_i^0 we find

$$\int_0^{\Delta} x dF_{D_i}(x) = \frac{r}{\lambda} \times ME_{r,\omega,\lambda}(\Delta) - \frac{\omega}{\lambda} \times E_{r,\lambda}(\Delta),$$

$$\int_{\Delta}^{\infty} x dF_{D_i}(x) = E[D_i^0] - \int_0^{\Delta} x dF_{D_i}(x).$$

The integral over the function $x \times \phi_{D_i}(x) \times F_{D_i}(\Delta - x)$ can be analyzed by fitting a $ME_{r,\omega,\lambda}$ -distribution on the single period depot demand. The i -fold convolution ϕ_{D_i} is given by

$$\phi_{D_i}(x) = \sum_{l=0}^i \binom{i}{l} \times \omega^{i-l} \times (1 - \omega)^l \times e_{(r-1)i+l,\lambda}(x).$$

For the resulting product terms we find

$$\int_0^{\Delta} x \times e_{a,\lambda}(x) \times E_{b,\lambda}(\Delta - x) dx$$

$$= \frac{a}{\lambda} \times E_{a+1,\lambda}(\Delta) - \Delta \times e^{-\lambda \times \Delta} \times \frac{(\lambda \Delta)^a}{(a-1)!} \times \sum_{i=0}^{b-1} \frac{(\lambda \Delta)^i}{i!}$$

$$\times \sum_{j=0}^i (-1)^j \times \binom{i}{j} / (j + a + 1),$$

where $e_{a,\lambda}(x)$ denotes the density of an a, λ – Erlang distribution.

Biographies

Stefan Minner is a Research and Teaching Assistant for Business Administration, Production and Logistics at the Faculty of Economics and Management of the Otto-von-Guericke-University Magdeburg. He graduated in Business Administration at the University of Bielefeld in 1994 and received a Ph.D. degree from the University of Magdeburg in 1999. His research interests include operations and supply chain management.

Erik Diks is a manager at Deloitte Consulting within the Dutch Supply Chain Management Department. He received an M.Sc. degree (1993) and a Ph.D. degree (1997) in Mathematics at Eindhoven University of Technology. In the last 4 years he has worked as supply chain consultant on several international projects for large multinationals. His interests are focused on demand and supply planning in the consumer product/food, chemical, telecommunications and automotive sectors.

Ton (A.) G. de Kok is Professor of Production and Operations Management at the Graduate School of Industrial Engineering and Management Science at the Eindhoven University of Technology. He was a Senior Consultant at the Centre for Quantitative Methods, Philips Electronics from 1985 to 1990. From 1990 to 1992 he was Logistics Innovation Manager, Philips Consumer Electronics. He graduated in Mathematics at the Rijksuniversiteit Leiden in 1981 and completed his Ph.D. thesis in 1985 at the Free University in Amsterdam. His current research activities include multi-echelon inventory models, supply chain management, and concurrent engineering.

Contributed by the Inventory Department