Controlling Divergent Multi-Echelon Systems

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Technische Universiteit Eindhoven, op gezag van
de Rector Magnificus, prof.dr. M. Rem, voor
een commissie aangewezen door het College van
Dekanen in het openbaar te verdedigen op
woensdag 17 september 1997 om 16.00 uur

door

Erik Bas Diks

geboren te Veldhoven
Dit proefschrift is goedgekeurd door de promotoren:

prof. dr. A.G. de Kok
en
prof. dr. J. Wessels

Diks, Erik Bas
Controlling Divergent Multi-Echelon Systems / Erik Bas Diks. -
Eindhoven: Eindhoven University of Technology
Thesis Technische Universiteit Eindhoven. -
With index, ref. - With summary in Dutch
ISBN 90-9010994-0
Subject headings: inventory control, distribution control.

Drukt Print Partners Ipskamp, Enschede

©1997 by E.B. Diks, Eindhoven, The Netherlands

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior permission of the author.
Preface

This monograph reports on the research I have done from 1 September, 1993, till 1 July, 1997, at the Eindhoven University of Technology. It is concerned with the analysis of divergent multi-echelon inventory systems. Several algorithms have been developed to determine appropriate or even cost-optimal control policies. In the development of these algorithms special attention is paid to the applicability to practice.

Although, only my name appears on the cover of this monograph I am indebted to many people who made this work possible. Therefore, I would like to express my gratitude towards them.

First of all, I would like to thank Ton de Kok. His excellent supervision and inexhaustible stream of suggestions and ideas contributed much to the contents of this monograph. Also I highly appreciated his enthusiastic way of supporting me. He always managed to find some time to hear me, and comment on my research activities.

Secondly, I would like to thank Jaap Wessels, Henk Zijm, and Frank van der Deyn Schouten for their comments and suggestions in the final stage of the writing process.

A special word of thanks goes to Matthieu van der Heijden. During the last one and a half year I have worked intensively with him. Therefore, most of the work presented in this monograph is the result of our collaboration. I really enjoyed working with him, which was not only pleasant but also led to several publications in international journals.

I would like to thank Stefan Minner, which was a guest at our department for two months. I enjoyed working with him on our paper on distribution flexibility.

Also I would like to thank all the members of the department of Operations Research. Especially, my office mates, which provided a pleasant atmosphere.

Finally, I would like to thank my family and my friends for their continuing support and interest in my work.

Eindhoven, July 1997
Erik B. Diks
Contents

1 Introduction
  1.1 Objective and motivation ............................................. 1
  1.2 The model under consideration ........................................ 4
    1.2.1 The control of a single location stockpoint ...................... 4
    1.2.2 Interaction between stockpoints ................................. 5
    1.2.3 Objectives ................................................................ 6
    1.2.4 Literature .................................................................. 6
  1.3 General divergent multi-echelon models: an overview ............. 8
    1.3.1 Demand process ...................................................... 9
    1.3.2 Lead time .................................................................. 9
    1.3.3 Ordering policy ....................................................... 9
    1.3.4 Control concept ...................................................... 10
  1.4 Outline of the monograph .............................................. 13

2 The basic model
  2.1 Introduction .................................................................. 17
  2.2 Notation ......................................................................... 17
    2.2.1 System layout .......................................................... 18
    2.2.2 Analysis ..................................................................... 18
  2.3 Model description .......................................................... 20
  2.4 System dynamics ........................................................... 21
    2.4.1 Service levels ........................................................... 23
    2.5 Inbalance ....................................................................... 28

3 Service level constraints: fixed lead time case ................. 35
  3.1 Introduction .................................................................... 35
  3.2 Controlling a 2-echelon system ....................................... 36
    3.2.1 Consistent Appropriate Share rationing ......................... 36
    3.2.2 Balanced Stock rationing .......................................... 39
    3.2.3 Adapted Consistent Appropriate Share rationing .............. 40
  3.3 Controlling an N-echelon system ..................................... 41
    3.3.1 Consistent Appropriate Share rationing ......................... 41
    3.3.2 Balanced Stock rationing .......................................... 45
  3.4 Numerical experiment for 2-echelon models .................... 45
    3.4.1 Experimental design for 2-echelon models ..................... 44
    3.4.2 Results for 2-echelon models ................................... 45
  3.5 Numerical experiment for 3-echelon models .................... 48
4 Service level constraints: random lead times
4.1 Introduction ................................................................. 55
4.2 Modeling random lead times ........................................... 56
4.3 Analysis of 2-echelon systems without central stock ................. 57
  4.3.1 Determination of the order-up-to-levels ......................... 58
  4.3.2 Calculation of the demand characteristics ....................... 59
  4.3.3 Calculation of the allocation fractions ......................... 62
  4.3.4 Physical stock .................................................. 63
4.4 Analysis of 2-echelon systems with central stock .................... 63
  4.4.1 Determination of the order-up-to-levels ......................... 64
  4.4.2 Calculation of the demand characteristics ....................... 64
  4.4.3 Calculation of the allocation fractions ......................... 65
  4.4.4 Physical stock .................................................. 65
4.5 Analysis of general N-echelon systems ................................ 67
  4.5.1 Determination of the order-up-to-levels ......................... 67
  4.5.2 Calculation of the demand characteristics ....................... 67
  4.5.3 Calculation of the allocation fractions ......................... 69
  4.5.4 Physical stock .................................................. 69
4.6 Summary of the algorithm ............................................. 71
4.7 Model to generate lead times without crossing ...................... 72
  4.7.1 Single location system .......................................... 72
  4.7.2 Analysis ....................................................... 74
  4.7.3 Divergent multi-echelon systems ................................ 74
4.8 Numerical validation .................................................. 76
  4.8.1 Experimental design for 2-echelon models ....................... 77
  4.8.2 Results for 2-echelon models ................................... 78
  4.8.3 Experimental design for 3-echelon models ....................... 79
  4.8.4 Results for 3-echelon models ................................... 79
  4.8.5 Computational effort ........................................... 80
4.9 Sensitivity analysis .................................................. 80
4.10 Conclusions ......................................................... 82

5 Cost minimization: optimal allocation functions ........................ 83
5.1 Introduction ............................................................. 83
5.2 Model description ..................................................... 85
5.3 Analysis cost function ............................................... 85
  5.3.1 Cost function ................................................... 86
  5.3.2 Properties of the allocation function ......................... 89
  5.3.3 Properties of the cost function ................................ 92
5.4 Optimal control policy ................................................ 96
  5.4.1 Optimal allocation function ................................... 96
Contents

5.4.2 Optimal order-up-to-level ........................................ 100
5.5 Optimality of decomposition approach .............................. 100
5.6 Extensions .................................................................. 102
  5.6.1 Arbitrary fixed lead times ....................................... 103
  5.6.2 Different cost structure ......................................... 104
5.7 Conclusions .................................................................. 105

6 Cost minimization: linear allocation functions ..................... 107
6.1 Introduction .................................................................. 107
6.2 Analysis ....................................................................... 108
  6.2.1 Specific class of linear allocation functions .................. 108
  6.2.2 Analysis cost function ........................................... 110
6.3 Algorithm ..................................................................... 112
6.4 Numerical results ......................................................... 115
6.5 Extension of the decomposition algorithm ......................... 117
6.6 Algorithm ..................................................................... 119
6.7 Numerical results ......................................................... 120
6.8 Error made by linearizing allocation functions ................. 122
6.9 Conclusions .................................................................. 124

7 Lateral transshipments ...................................................... 127
7.1 Introduction .................................................................. 127
7.2 Model description ......................................................... 128
7.3 System dynamics ........................................................ 129
  7.3.1 Rationing policy at the CD ..................................... 130
  7.3.2 Rebalancing policy at the retailers .............................. 131
7.4 Determination of the control parameters ......................... 132
  7.4.1 Determination of the control parameters of the rationing policy ........................................... 132
  7.4.2 Determination of the control parameters of the rebalancing policy ........................................ 134
  7.4.3 Determination of $\Delta_1$ ........................................ 135
7.5 Extension to divergent $N$-echelon system ......................... 135
7.6 Numerical results ......................................................... 138
7.7 Extension of the model .................................................. 140
7.8 Conclusions and further research .................................... 141

8 Conclusions and suggestions for future research .................. 143
8.1 Cost minimization given service level constraints ............... 145
8.2 Other possible model extensions ..................................... 146

Bibliography ..................................................................... 147

Samenvatting (Summary in Dutch) ....................................... 153

Curriculum Vitae ................................................................ 157
1

Introduction

1.1 Objective and motivation

In this monograph, we study single-item divergent multi-echelon inventory systems. A divergent multi-echelon system is a network of stockpoints in which every stockpoint has a unique supplier but may have several successors. As an illustration of the applicability of multi-echelon systems we give two examples. In these examples some interesting questions arise, which are examined in this monograph. The first example is a distribution system of a single item. The second example is a hierarchical planning approach to allocate the available products of a product group to the multiple product types within the group.

In the first example, we consider the distribution system of a single item as depicted in Figure 1.1, where a triangle represents a stockpoint. A production facility in New York supplies a large European central depot in Brussels. This depot supplies national distribution centers in Germany and the Netherlands. Finally, those national distribution centers supply regional depots from where market demand is satisfied. The customer demand is satisfied at the regional depots by the stock on hand.

![Diagram of distribution network](image)

Figure 1.1. A distribution network as an example of a divergent multi-echelon system.
In case of insufficient stock the unsatisfied demand is backlogged. The central depot in Brussels issues a replenishment order periodically. The order size is chosen such that the stock level is raised to some order-up-to-level. After some lead time (production time plus distribution time) the replenishment orders arrive at the central depot. Some part of the order may be kept at the central depot as backup inventory. The remaining part is allocated among the national depots. The allocated amounts are dispatched immediately to the national depots, where they arrive after some distribution time. In this monograph we provide insight in the following questions:

- How much safety stock to attain in this distribution system?
- Where to locate this safety stock in the system?
- How to deal with shortages in the central depot and in the national depots?

To answer these questions we need to know the objective of the distribution system. Two different objectives are addressed in this monograph. First, the system is controlled such that some cost function is minimized, for example, the expected systemwide holding and penalty costs per period. Second, the system is controlled such that prespecified customer service levels are attained. The latter objective has gained in popularity, since nowadays for many companies high service levels are a competitive edge. For both objectives it is eminent to understand the material flow in the system in order to answer the questions raised above. Therefore an important part of this monograph is devoted to the analysis of the system dynamics of divergent multi-echelon systems.

In the second example, we consider a single depot which faces demand for a single product group; see Figure 1.2. This product group consists of a number of product types. Demand is satisfied immediately from stock on hand and backlogged otherwise. Periodically, an orientation on the production volume at product group level is passed to the production facility. This orientation is chosen such that the total amount of products is raised to some order-up-to-level. The time between orientation and the delivery from production consists of: (1) an orientation lead time, during which the total volume of the product group is fixed, but changes in the product mix are still possible, and (2) a frozen lead time, during which both the total volume and the product mix are fixed.

![Figure 1.2. A hierarchical planning approach as an example of a divergent multi-echelon system.](image)
1.1. Objective and motivation

In this hierarchical planning approach there are several questions of interest:

- How many items for the product group as a whole should be ordered periodically?
- How to allocate the available products within the product group to the different product types?

In this monograph we address these questions by analyzing a model of divergent multi-echelon systems.

Both the distribution system and the hierarchical planning approach can be modeled as a divergent multi-echelon system. In this monograph we are concerned with determining how to control the material in this system such that one of the aforementioned objectives is satisfied. Quite a lot of research has been done in this area. We distinguish between two types of models. First, there are models describing explicitly how each control parameter affects the material flow in the system. This makes it possible to optimize the system, i.e., determine all the control parameters in the system such that some objective (e.g., target customer service level) is satisfied. Unfortunately, due to the restrictive assumptions made in these models, the applicability to real-world supply chains is usually limited. Second, there are the more general applicable models which are used in practice. These do not explicitly model how the control parameters affect the material flow in the system, and therefore are hard to optimize.

Let us elaborate on the disadvantages of both types of models. As already mentioned, the applicability of the first type of model in practice is limited. There are several reasons for this. First, only small systems consisting of one central depot supplying a number of end-stockpoints can be studied. In practice, however, we need to be able to analyze larger systems (e.g., the distribution system depicted in Figure 1.1). Second, rather restrictive assumptions with respect to the end-stockpoints are made. For instance, it is assumed that all end-stockpoints are identical, or at least that the lot sizes at each end-stockpoint are identical. Both these assumptions are invalid in practice, since the EOQ formula (cf. Harris [1913]) tells us that lot sizes depend on the demand at each end-stockpoint and the cost structure of each end-stockpoint, which are quite different in practice.

The second type of model consists of Distribution Resource Planning (DRP) models, which have been implemented in many companies as the front-end of their integrated logistics control systems. DRP (cf. Martin [1990]), is the equivalent MRP, Manufacturing Resource Planning (cf. Vollmann, Berry & Whybark [1992]), for the distribution chain. The planning logic of DRP consolidates demand forecasts at different stockpoints into time-phased dependent demand at intermediate stockpoints and ultimately into time-phased demand at the manufacturing location. This top-down logic does not explicitly take into account possible (future) shortages at stockpoints. To circumvent this problem so-called re-scheduling messages are generated to inform a planner that the logic identified a shortage at a certain stockpoint and the planner is supposed to solve this shortage. However, if the planner solves this problem, inevitably his solution impacts a number of decisions already made by the DRP system at downstream stockpoints and most likely also at upstream stockpoints. Hence the planner is forced to overrule the decisions of the planning system, since these decisions are not consistent. This manual replanning process can be quite time-consuming and intricate. Due to this intangible replanning process it is very hard to determine how to control the system such that operational objectives regarding customer service at end-stockpoints are satisfied. Therefore optimization of a DRP system is a time-consuming and intractable process.

The objective of this monograph is to develop algorithms that optimize the material flow in the system, i.e., that determine the control parameters such that either each end-stockpoint attains its target service level (service measure objective), or the expected total costs to operate the system are minimized (cost measure objective). Hence in this monograph we consider optimization models instead
of DRP models. In the development of the algorithms to solve these models we focus on numerical tractability and applicability, in order to provide a tool to control the material flow in divergent multi-
echelon systems (e.g., production/distribution systems). Our model differs from the models developed so far by its large applicability and its optimization potential. The strength of our model is that it can handle

- A general divergent network.
  We consider a divergent $N$-echelon system (instead of a 2-echelon system). The $N$ denotes the highest number of stockpoints on a path between the unique most upstream stockpoint to an end-stockpoint.
- Heterogeneous end-stockpoints.
  We do not require that all end-stockpoints are identical (or the lot sizes of end-stockpoints are identical).
- Optimization.
  Unlike DRP we can optimize the system.

Finally, we give an outline of the remainder of this chapter. The subject of this monograph is the control of divergent multi-echelon systems. In Section 1.2 we introduce the model which is analyzed throughout this monograph. A general overview of the most important contributions to the control of divergent multi-echelon systems is given in Section 1.3. The outline of this monograph is given in Section 1.4.

1.2 The model under consideration

In this monograph, we consider a single-item divergent $N$-echelon system. In such a system we distinguish between two different types of performance measures: internal and external measures. The latter are related to the service provided to customers at the most downstream stockpoint (e.g., the customer service level and customer waiting times). The former is related to internal customer service and relevant costs (e.g., set-up, holding and transportation costs). We emphasize that, in some sense, internal service is irrelevant as long as external service is according to prespecified targets at minimum internal costs. A prerequisite to determine these performance measures is to have models which analyze the system dynamics (i.e., operational characteristics) of divergent multi-echelon systems. When such analysis yields insight in the evolution of material stocks of a divergent multi-echelon system over time we are able to determine the service and costs measures. In Section 1.2.1 we present the major elements that constitute a single location stockpoint in our model. An extensive discussion of our model assumptions is given in Chapter 2. Next, in Section 1.2.2, we address the interaction between these single location stockpoints. In Section 1.2.4, we give an overview of the literature related to the analysis of the model formulated in Sections 1.2.1 and 1.2.2. From this we identify some interesting research directions, which has been carried out in some parts of this monograph.

1.2.1 The control of a single location stockpoint

Which control policy is the most appropriate one to control a stockpoint in a multi-echelon system depends on: the demand process, the lead time process, the ordering process, and the control concept. Let us elaborate on these elements.

**Demand process**

The nature of a product largely determines the behavior of the customer demand process. In this monograph we concentrate on consumables, which usually have high demands (and low costs). The demand
1.2. The model under consideration

process for these products may be described by the mean and variance of the demand over a certain period. We model the demand process by fitting an appropriate distribution to this mean and variance.

Lead time

The lead time is defined as the time between placing an order and its arrival. In this monograph we investigate both the fixed and random lead time case. In a distribution environment the lead time can usually be regarded as fixed. However, in a manufacturing environment the lead time usually is liable to fluctuations.

Ordering policy

The most upstream stockpoint which is supplied by the external supplier (e.g., the central depot in Brussels; see Figure 1.1) inspects its stock level every \( R \) periods (also called a review period). After each inspection an order is placed such that the stock level returns to a target order-up-to-level. All the other stockpoints in the system also adopt an order-up-to-policy. They place an order immediately after an order arrival at its supplier. Note that for the fixed lead time case this means that every stockpoint inspects the stock level every \( R \) periods. For the random lead time case the review period is random (except for the most upstream stockpoint) due to the random interarrival times of orders.

Control concept

Every stockpoint in the system is controlled by a so-called echelon stock policy. Clark & Scarf (1960) introduced the concept of echelon stock. The echelon stock of a stockpoint equals all stock at this stockpoint plus in transit to or on hand at any downstream stockpoints minus the backorders at its downstream end-stockpoints. The echelon inventory position of a stockpoint is defined as the sum of its echelon stock and the material in transit to this stockpoint.

1.2.2 Interaction between stockpoints

We assume that only the end-stockpoints face customer demand. This customer demand is satisfied from the stock on hand and backlogged otherwise. Every \( R \) periods the most upstream stockpoint issues a replenishment order at the external supplier to raise the echelon inventory position to its order-up-to-level. This order arrives after a certain lead time. Then, every successor places an order such that its echelon inventory position is raised to its order-up-to-level. There are two possibilities:

1. The physical stock is sufficient to raise the echelon inventory position of all successors to their order-up-to-levels. Then the required amounts are sent to the successors and excess stock is kept at the most upstream stockpoint to be allocated at the next occasion.

2. The physical stock is not sufficient to reach the order-up-to-levels. Then material rationing is required to allocate the available physical stock over its successors appropriately. For this purpose we introduce allocation functions.

A similar procedure is applied at the other intermediate stockpoints, i.e., when a replenishment order arrives at an intermediate stockpoint this order is allocated. Again we have the two possibilities as stated above.

We like to pay special attention to the notion of imbalance, which is specific for divergent systems controlled by echelon stock policies. Eppen & Schrage (1981) define imbalance as the phenomenon that the allocation policy is not able to allocate the available products such that each end-stockpoint faces an equal stockout probability. Verrijdt & De Kok (1996) generalize the definition by defining imbalance as the phenomenon that the allocation policy of an intermediate stockpoint allocates a negative
quantity to at least one of its successors. Since the analysis of divergent systems under echelon stock policies requires that no imbalance occurs, we make the balance assumption: the allocation policy always allocates nonnegative stock quantities.

1.2.3 Objectives
In this monograph, we develop several algorithms to determine the control parameters of the divergent multi-echelon system as formulated in Sections 1.2.1 and 1.2.2. In this development we distinguish between two objectives: the service measure objective and the cost measure objective. Let us briefly elaborate on these objectives.

The service measure objective
This objective requires a target service level at each end-stockpoint. In this monograph we consider three different service measures (cf. Lagodimos [1992]):

- The non-stockout probability (α): the probability that the net stock (the stock on hand minus backorders) is nonnegative at the end of an arbitrary replenishment cycle.
- The fill rate (β): the fraction of the period demand that is satisfied directly from the stock on hand.
- The modified fill rate (γ): one minus the ratio of the average shortage immediately before arrival of a replenishment order and the average demand during an arbitrary replenishment cycle.

The cost measure objective
This objective minimizes the expected total costs per period. Note that this objective requires a cost framework. In this monograph we only consider holding costs at each stockpoint and penalty costs at the end-stockpoints.

For both objectives several algorithms are developed. Depending on the objective each algorithm has its disadvantages. For example, a prerequisite for applying an algorithm considering the former objective is that the mean amount of stock at each intermediate stockpoint has to be known. But also an algorithm considering the latter objective has its disadvantages. Let us address the two most important disadvantages. First, the penalty costs at each end-stockpoint has to be known. In practice, however, one usually is not able to assign these costs, since often the primary consideration of the penalty costs is the loss of customer goodwill (Schneider [1981]). To deal with this, one usually determines the penalty costs indirectly by a certain service measure. Even in situations where penalty costs can be determined (e.g., in the case of contractual arrangements regarding shortage penalties), service measures are needed in order to have direct information on physical performance of the supply chain. Second, the applicability in practice is limited due to the rather restrictive assumptions that are required (i.e., identical holding and penalty costs at the end-stockpoints, and identical lead times at the end-stockpoints). In this monograph we extend some of the existing algorithms such that these can be applied to a larger class of divergent multi-echelon systems.

1.2.4 Literature
In this section, we discuss the literature on divergent multi-echelon systems, which control their inventory by periodic review echelon order-up-to-policies. In this discussion, we distinguish between the two objectives of Section 1.2.3.
1.2. The model under consideration

Service measure objective
An important contribution in this field is the paper of De Kok [1990], which constitutes the basis for many papers. De Kok analyzed how to control a divergent 2-echelon system with a stackless depot such that target fill rates are attained. De Kok, Lagodimos & Seidel [1994] generalized the results to a 2-echelon system where the depot is allowed to hold stock. They also introduce the Consistent Appropriate Share (CAS) rationing policy, which is a generalization of the Fair Share (FS) rationing policy of Eppen & Schrage [1981]. Verrijdt & De Kok [1995] extended the results to arbitrary divergent N-echelon systems without intermediate stocks. Numerical experiments reveal that the CAS rationing policy performs well for systems where the end-stockpoints have similar service level requirements. However, when the service level requirements differ significantly the performance is rather poor. This is caused by the severe violation of the balance assumption. For this reason Verrijdt & De Kok [1996] developed a modification of the CAS rationing policy to cope with significantly differing target service levels at the end-stockpoints. Recently, Van der Heijden [1997a] developed the Balanced Stock (BS) rationing policy for arbitrary divergent N-echelon systems with stockless intermediate stockpoints.

In Chapter 3, we extend the results of Verrijdt & De Kok [1995] and Van der Heijden [1997b] to arbitrary divergent N-echelon systems in which each stockpoint is allowed to hold stock. Furthermore, we detected an inconsistency in the logic of the CAS rationing rule. For that purpose we introduce an adaptation of the CAS rationing rule.

Cost measure objective
Most literature on multi-echelon systems focuses on cost-optimization issues. Usually the cost framework just consists of holding costs at each stockpoint and penalty costs at the end-stockpoints. The holding costs are incurred prior to every order arrival at a stockpoint, and are proportional to its stock on hand plus all products in transit to its successors. It usually is assumed that the holding cost rate is non-decreasing on every path from the upstream stockpoint to an end-stockpoint. This assumption is standard; in most applications, holding cost rates consist primarily of the cost of capital, and are thus increasing with the cumulative value added (cf. Federgruen [1993]). The penalty costs are incurred prior to every order arrival at an end-stockpoint, and are proportional to the number of backorders.

The start of research on multi-echelon inventory models is generally attributed to Clark & Scarf [1960], who studied an N-echelon serial system (i.e., a system with N stockpoints in series) operating under periodic review ordering policies. They proved that the cost-optimal policy for the N-echelon serial system with discounted penalty costs and holding costs, is characterized by N so-called echelon order-up-to-levels. These echelon order-up-to-levels can be computed by decomposing the problem (exactly) into N separate single location problems, which can be specified and solved recursively. The model and solution were originally suggested by Clark [1958]. In order to prove the optimality of these policies we need that only at the most upstream stockpoint set-up costs can occur.

The work of Clark & Scarf mainly concentrates on determining the order-up-to-level at each stockpoint. Eppen & Schrage [1981] focus on determining appropriate allocation functions. They examine a divergent 2-echelon system in which a central depot supplies multiple end-stockpoints. Some restrictive assumptions are made. First, the depot cannot hold any stock. It acts as a transshipment center, not as a stocking point. This assumption is appropriate, for example, when the depot serves merely as a coordinator, allocating incoming products immediately. Note that the example in Section 1.1 concerning the hierarchical production planning approach can also be modeled as a divergent 2-echelon system with a stockless depot. Second, the holding and penalty costs at the end-stockpoints are identical. Third, the demand period at each end-stockpoint is normally distributed. Fourth, the lead
time of the end-stockpoints are identical. Finally, each time an allocation is made, the depot receives enough material from the supplier to be able to allocate the material to each end-stockpoint so that an equal fractile point is achieved on an appropriately chosen demand function. In other words, after ordering each end-stockpoint faces an equal stockout probability. This assumption is referred to as the allocation assumption, and is a special case of the balance assumption as introduced in Section 1.2.2. Simulation studies (cf. Van Doornelaar & Wijngaard [1987]) indicate that this assumption is not a serious restriction. The resulting allocation rule is referred to as Fair Share (FS) rationing. Under these assumptions Eppen & Schrage derive an optimal order-up-to-policy at the depot, assuming no set-up costs. In case of fixed set-up costs, an approximately optimal policy is derived.

Federgruen & Zipkin [1984] extend the model and the results of Eppen & Schrage. Holding and penalty costs do not have to be identical across the end-stockpoints and period demands at the end-stockpoints do not have to be normally distributed, but may belong to a larger class of demand distributions (e.g., Erlang and gamma distribution). Furthermore, set-up costs are incurred by orders placed by the depot. In their paper the allocation of material at the depot is determined by frequently solving the myopic allocation problem. This problem minimizes the expected costs in the period that the allocation actually takes effect, ignoring costs in all subsequent periods. Another important distinction is that Eppen & Schrage requires that the ordering policy is an order-up-to-policy. Given this restriction, they show how to compute the optimal order-up-to-level. In the paper of Federgruen & Zipkin [1984], however, the ordering policy is determined by an (approximate) dynamic program. It turns out that the myopic allocations are optimal given the allocation assumption. For an excellent overview concerning dynamic programming techniques to analyze multi-echelon systems we refer to the survey paper by Federgruen [1993].

Langenhoff & Zijm [1990] and Zijm & Van Houtum [1994] extended the Eppen & Schrage model by allowing the demand at each end-stockpoint per period to be arbitrarily distributed. From an average cost analysis they concluded that decomposition results for a divergent 2-echelon system are exact given the balance assumption (although, both papers do not provide any proof for this). For an extensive analysis (including the proof) we refer to Van Houtum & Zijm [1991b]. For a detailed overview of these average cost analyses we refer to the survey paper of Van Houtum, Inderfurth & Zijm [1996]. From these aforementioned papers it was felt that the existing results could be generalized to divergent \( N \)-echelon systems with less restrictive assumptions, and that from an extensive cost analysis more insight could be obtained in the structural properties of the optimal policy. Therefore in Chapter 5 we prove that, indeed, these restrictive assumptions can be relaxed, and that decomposition is exact for a divergent \( N \)-echelon system given the balance assumption. Decomposition reduces the complex control problem to (1) determining the optimal order-up-to-level at each stockpoint, and (2) determining the optimal allocation functions at each intermediate stockpoint.

1.3 General divergent multi-echelon models: an overview

So far we concentrated on divergent multi-echelon system controlled by period review echelon order-up-to policies. In Section 1.2.4, we gave an overview of the literature concerning the control of that system. Besides this, a lot of literature has appeared on other models of divergent multi-echelon systems. In this section, we present an overview of this literature by again distinguishing between the major elements that constitute a single location stockpoint and affects it control. In Section 1.3.1, we address how to model the customer demand process. In Section 1.3.2, we discuss whether lead times can be regarded fixed or random. In Section 1.3.3, we present some practically useful continuous and periodic review ordering policies. The ordering policies either follow the already mentioned echelon
1.3. General divergent multi-echelon models: an overview

The stock concept, or the installation stock concept, which bases its decisions on more local information. In Section 1.3.4, we present the operating details of both control concepts, and demonstrate the differences.

1.3.1 Demand process

The nature of a product largely determines the behavior of the customer demand process. We distinguish between products with low demands and high demands.

An important class of products which usually have low demands are the repairables (which usually have high cost). In the literature the demand process for these products is usually assumed to be a compound Poisson process, or simply a Poisson process. One of the first papers which models the system interactions of repairables in divergent multi-echelon systems is the METRIC model of Sherbrooke [1968]. METRIC stands for 'Multi-Echelon Technique for Recoverable Item Control'. This model constitutes the basis for many papers on divergent multi-echelon systems. Several extensions of METRIC have been developed over the years. For extensive reviews we refer to Nahmias [1981], Mahini & Gelders [1990], Cho & Pardar [1991] and Axssiter [1993].

Another important class of products are the consumables (which usually have low cost). For these products the demand process may be described by the mean and variance of the demand over a certain period as well as a compound Poisson process. The former approach models the demand process by fitting an appropriate distribution to its mean and variance. In this monograph we use this approach instead of the (compound) Poisson assumption.

1.3.2 Lead time

As already mentioned, we define the lead time as the time between placing an order and its arrival. This lead time can have several interpretations. For instance, in the distribution network as depicted in Figure 1.1 the lead time from New York to Brussels consists of a production lead time plus a distribution lead time. This production lead time is the time required to produce the products. It is usually liable to fluctuations due to material unavailability (e.g., in order to assemble we need all required sub-assemblies), or due to a shortage of capacity (e.g., a machine is needed to produce the product). The distribution lead time only consists of handling & transportation time and therefore is not liable to fluctuations in most cases. So, the lead time in a production/assembly environment usually is regarded as random, unlike the lead time in a distribution environment. Both the fixed and random lead time cases are investigated in this monograph. For the random lead time case we require that subsequent orders do not cross in time.

1.3.3 Ordering policy

In principle there are two major ordering disciplines: continuous review policies and periodic review policies. In case of continuous review, the stock level is monitored after each customer demand and, immediately after the stock level drops below a reorder point, an order can be placed to replenish the stock. In case of periodic review, the stock level is inspected periodically, such that orders are generated at review moments only. In this case the demand per review period may have an arbitrary distribution.

In practice there is need for useful replenishment policies that are easy to implement. Examples of such policies are the order point, order-up-to-policy and the order point, order-quantity policy. The order point, order-up-to-policy is characterized by the control parameters \(s, S\) (if \(s\) equals \(S\), then the \(s\) is suppressed). This means that at any time when the stock level equals or drops below \(s\), an order is placed immediately. The quantity of this order is such that the stock level returns to a target order-
up-to-level $S$. This policy is optimal for a single location inventory system with constant set-up costs, linear holding and penalty costs, fixed replenishment lead time and which allows for backordering of unsatisfied demand. The proof has been given by Scarf [1960] for a finite horizon, and by Lichhart [1963a,b] for an infinite horizon. Under some assumptions (among other things replenishment orders do not cross in time) Kaplan [1970] extends this result by proving that the $(s, S)$ policy is also optimal for the random lead time case. The periodic review analogue of the $(s, S)$ policy is denoted by $(R, r, S)$. This means that every $R$ periods the inventory is inspected, and orders are generated at those review moments only. The order-point, order-quantity policy is characterized by $(s, nQ)$. This means that when the stock level $x$ falls to or below reorder-point $s$ an order of $nQ$ products is placed where $Q$ is the order-quantity (also called lot size) and $n$ is the minimum integer with $x + nQ > s$. The periodic review analogue of the $(s, nQ)$ policy is denoted by $(R, r, nQ)$. Note that in case the demand process is a discrete process without batching the $(s, nQ)$ policy reduces to the better known $(s, Q)$ policy.

Which control policy is the most appropriate one to control the supply chain depends on several factors. First, the set-up costs play an important role. When these costs are considerable there is an incentive to minimize the number of orders placed. In a $(s, S)$ policy (or $(R, r, S)$ policy) this means that $S - s$ should be large, while in a $(s, nQ)$ policy (or $(R, r, nQ)$ policy) this means that $Q$ should be large. The appropriate lot size $S - s$ (or $Q$) depends on the tradeoff between the set-up costs and the holding costs. For instance, for a repairable lot size usually equals one, i.e., $s = S - 1$ (or $Q = 1$). While for a consumable lot size usually exceeds one. Second, the flexibility of monitoring the stock plays an important role. This flexibility determines whether we can consider the inventory system under continuous review or periodic review. Although a decrease of the length of the review period can be observed, we feel that periodic review reflects practice in most (large) companies, in particular the consumer market. Finally, whether to use an order-up-to-policy or an order-point policy depends heavily on the flexibility of the supplier. Some suppliers require that the amount ordered is a multiple of the lot size $Q$. Quite some literature is devoted to determining appropriate lot sizes. The lot-sizing flexibility depends on the order discipline. For a stockpoint controlled by a periodic review policy typically holds: the shorter the review period, the less likely the supplier is able to satisfy the demand in any quantity (cf. Van Houtum, Inderfurth & Zijm [1996]).

1.3.4 Control concept

The control of multi-echelon systems is often completely decentralized in the sense that ordering decisions at a stockpoint are solely based on its inventory position, which is defined as the sum of all planned orders at this stockpoint plus its physical stock minus its backorders. This concept of controlling the inventory of a stockpoint is referred to as the installation stock concept. An obvious advantage of an installation stock policy is that it does not require any information of the inventory situation at other stockpoints. However, due to this lack of information about the entire system the cost effectiveness of these policies is limited. For instance, excessive demand may not be identified at upstream stockpoints due to the delay in information through resulting replenishment orders upstream. One way of taking such information into account is to control the inventory based on the echelon stock. Since in echelon stock policies the order decisions are based on the complete knowledge of how much stock is downstream, we need information about the product flow through the system. Due to the developments in the area of information technology this is no longer a problem.

An important difference between installation stock and echelon stock policies is pointed out by Chen & Zheng [1994]. In the former policies the inventory position of a stockpoint includes all outstanding orders, i.e., in transit to this stockpoint or backordered at the supplier, while for echelon stock
policies the echelon inventory position of a stockpoint only includes the in transit orders to this stockpoint. Consequently, a stockpoint using an installation stock policy can always raise its inventory position to a desired level. If some part of the order cannot be delivered immediately it is backordered at its supplier. So we may model every stockpoint as a single location inventory system with a random lead time, i.e., the transportation time plus an additional waiting time. In echelon stock policies this lead time exactly equals the transportation time. However, in these policies it is more difficult to determine the echelon inventory position, since the stockpoint cannot be regarded as a single location inventory system.

Axsiöter & Rosling [1993] compared installation stock policies and echelon stock policies. They proved (under some initial conditions on the stock levels and unit demand) that, when every stockpoint in a multi-echelon system is controlled by a continuous time order-up-to-policy, an installation stock policy can always be replaced by an equivalent echelon stock policy, and vice versa. When every stockpoint in a serial system is controlled by a continuous time $(s, nQ)$ policy an installation stock policy can always be replaced by an equivalent echelon stock policy, but not vice versa. This result also hold for $(R, s, nQ)$ policies. Furthermore, Axsiöter & Rosling [1993] showed that an echelon stock policy that cannot be duplicated by an installation stock policy may yield lower (holding) costs. So, in serial systems echelon stock policies dominate installation stock policies. Whether this also holds for divergent systems is unclear up to now.

As pointed out by Diks, De Kok & Lagodimos [1996], the complexity of the analysis for the installation stock and echelon stock concept manifests at different aspects. For installation stock policies the major difficulties are the determination of the demand process at upstream stages and the determination of the delay time characteristics. For echelon stock policies the major problems are the imbalance and the analysis of divergent systems with lot sizing. In the remainder of this section we review some important papers using the installation stock concept and the echelon stock concept, respectively.

Installation stock policies

Deuermeyer & Schwarz [1981] analyze installation stock $(s, nQ)$ policies in a divergent 2-echelon system. A METRIC based approach is adopted to analyze the system, i.e., the whole system is decomposed into several single location systems. Furthermore, several assumptions are made:

1. Fixed lead times.
2. In case of shortages at the depot a retail order is delayed until it can be satisfied completely (no-splitting assumption).
3. Identical lot sizes for all end-stockpoints.
4. Poisson demand at the end-stockpoints.

The same model was examined by Lee & Moinzadeh [1987] and Svoronos & Zipkin [1988]. Svoronos & Zipkin propose several refinements of the technique of Deuermeyer & Schwarz, and performed a numerical study which indicates that due to these refinements their approach is more accurate than Deuermeyer & Schwarz and Lee & Moinzadeh. Furthermore, they claim that assumption 1 can be relaxed.

Assumption 2 usually is violated, since it is quite common to split lots in case of material shortages. Lot splitting also impacts the cash flow of the supplying companies and their capital tied up in inventory. Indeed, if lot-splitting is applied, remnant stock does not occur in case of shortages and customers can be invoiced immediately for the products shipped. Moreover, customers may be able to start operations before the complete order arrives. Diks, De Kok & Lagodimos [1996] pointed out
that more research is required on this topic.

In practice assumption 3 is invalid. E.g., suppose that the end-stockpoints are in fact wholesalers, power retailers and other stockpoints of the company that owns the depot. Then the EOQ formula (Harris [1913]) tells us that lot sizes depend on the demand at each end-stockpoint and its cost structure. These are in practice quite different. To our knowledge few results are available dealing with non identical lot sizes. Rosenbaum [1981a,b] developed a heuristic model for a divergent 2-echelon system with non identical end-stockpoints. However, in our opinion the applicability is limited due to the various assumptions made. Extension to larger systems with three or more echelons is not very tractable.

Axssater, Forsberg & Zhang [1994] relax assumption 3 by assuming that one of the end-stockpoints has the smallest order quantity and all other quantities in the system are integer multiples of this smallest order quantity. Furthermore, assumption 4 is relaxed by considering different compound Poisson processes with discrete batch sizes. They replace the compound Poisson process at every end-stockpoint by an 'equivalent' Poisson process such that the ratio between the mean and standard deviation is the same as for the real distribution. Next, the algorithm of Axssater [1990a] provides the order-quantities in the Poisson demand model. These are used to compute the order quantities in the original model (compound Poisson demand). The algorithm of Axssater is exact for two end-stockpoints but approximative for more than two end-stockpoints. So far the extension to systems with three or more echelons is not available. Principally the approach of Axssater, Forsberg & Zhang is developed for systems where every stockpoint is controlled by a continuous time order-up-to policy, since then the order-up-to-levels in the Poisson model can be determined optimally by applying the algorithm of Axssater [1990b]. Then, the extension to systems with three or more echelons is straightforward.

Van der Heijden [1992] analyzes (R, S) policies in divergent N-echelon systems. An important feature of the model is that it incorporates lot sizing by allowing for a different review period at each end-stockpoint. Furthermore, the model of Van der Heijden is less restrictive than the models of Deuermeyer & Schwarz [1981], Lee & Moizadze [1987], and Svoronos & Zipkin [1988], since it considers random lead times and compound Poisson processes at the end-stockpoints. A similar system is considered by Schneider, Rinks & Kelle [1995]. They do not allow for different review periods at the end-stockpoint, but consider (s, S) policies to incorporate lot sizing. As in Deuermeyer & Schwarz [1981] the analysis is also based on the aforementioned METRIC base approach.

Echelon stock policies

The echelon stock concept is not applied as widely as the installation stock concept. Furthermore, most papers using echelon stock policies only consider a serial system. Clark & Scarf [1960] proved the optimality of the periodic review echelon stock policies for serial systems with set-up costs only at the highest echelon. When each stockpoint in the serial system faces set-up costs, Clark & Scarf [1962] suggested (s, S) policies. Chen & Zheng [1994] study (s, n, Q) policies in serial systems, and developed a recursive procedure to derive the exact steady-state echelon inventory positions, which can be used to evaluate the long-run average holding and penalty costs, as well as service measures. Their results apply to both continuous review systems with compound Poisson demand and periodic review systems with independent, identically distributed demands. For the special case of continuous review systems and Poisson demand De Bolt & Graves [1985] derive approximately optimal (s, Q) policies. The results of Clark & Scarf [1960] are generalized to assembly systems by Rosling [1989], Langenhourf & Zijm [1990], and Vischiers [1996].

In this monograph, we study echelon stock (R, S) policies in divergent echelon systems. These
policies are investigated in many papers (see Section 1.2.4). All these papers do not incorporate lot-sizing. Besides the paper of Van Donselaar [1990] to the best of our knowledge few results are available for these systems. Van Donselaar [1990] analyzes (s, nQ) policies in a divergent 2-echelon system. Being a heuristic approach, and the large number of assumptions made in this paper (all end-stockpoints are identical with respect to lead time, lot size and demand characteristics) the applicability in practice is doubtful, as well as the possibility to extend it to larger N-echelon systems.

1.4 Outline of the monograph

This monograph is concerned with the development of control policies for single-item divergent multi-echelon systems. In Chapter 2, we present a model of a divergent N-echelon system under echelon stock control. In this model, every stockpoint in the system is controlled by a periodic review echelon order-up-to policy, i.e., every review period each stockpoint places an order at its predecessor to raise the echelon inventory position to a fixed order-up-to-level. In case the predecessor has insufficient stock available this stock is rationed among all successors by some rationing policy. This model constitutes the basis for the remaining chapters of this monograph. Besides this model we also introduce some notation which is used throughout this monograph. To determine service and cost measures in the system we need to characterize the stochastic behavior of the inventory at each stockpoint over time. By applying the sample path technique we obtain a formula for the echelon inventory position of each stockpoint. The distribution function of this inventory position is given by an incomplete convolution of continuous distribution functions. We use a recursive technique developed by De Kok to approximate this function. The technique involves fitting mixtures of Erlang distributions to the first two moments of nonnegative random variables.

In Chapter 3, we present several algorithms to determine all the control parameters in the system such that the customer service level constraints are satisfied. The algorithms are based on the work of Verrijdt & De Kok [1995] and Van der Heijden [1997b]. For divergent N-echelon systems where only the end-stockpoints are allowed to hold stock, Verrijdt & De Kok [1995] and Van der Heijden [1997b] introduced CAS rationing and BS rationing, respectively. We extend their results to divergent N-echelon systems where all stockpoints are allowed to hold stock. Moreover, we performed an extensive numerical experiment to get insight in the performance and differences of all the variants of CAS and BS rationing. The numerical experiment consists of an analytic evaluation and simulation of many instances. It provides good suggestions to both practitioners and to academics for direct application and further research. The main conclusion of this experiment is that BS rationing performs better than CAS rationing.

A shortcoming of the model addressed in Chapter 3 is that all the lead times are fixed. However, in real-world supply chains the lead times are subject to randomness. Lead times to downstream stockpoints usually are transportation times and thus usually reliable, but lead times to upstream stockpoints usually represent manufacturing times. These manufacturing times are random due to, e.g., capacity constraints, machine-breakdowns, and batching decisions. So far no analysis is available for multi-echelon models with echelon stock policies and random lead times with the exception of the paper of Zipkin [1991], who deals with continuous review (S - 1, S) policies. Therefore, in Chapter 4, we extend the model presented in Chapter 3 by incorporating random lead times. Special attention is given to the impact of the randomness of the lead times on the amount of safety stock needed in the system to satisfy fill rate constraints. We developed an algorithm to determine the control parameters such that target fill rates at end-stockpoints are attained. Due to the superior performance of BS rationing, we restrict to this rationing policy. Again we apply the sample path approach to determine the system
dynamics. However, this approach is only valid when the orders do not cross in time. To test the accuracy of the developed algorithm we perform an extensive numerical experiment. In this experiment we make use of a novel technique for generating a random lead time process without crossing. This technique models a random lead time by the sojourn time of customer in a $G1/G/1$-queue plus a fixed time.

An important drawback of the algorithms developed in Chapters 3 and 4 is the assumption that the mean amounts of stock held at the intermediate stockpoints have to be known in order to apply the algorithms. By varying the mean amount of stock held at each intermediate stockpoint and using standard Newtonian search methods one can find the cost-optimal mean stock levels subject to the service level constraints. However, the model with a cost measure objective does not require such a Newtonian search method, since from the analysis it follows how much stock to hold at intermediate stockpoints. Furthermore, it was felt that more research should be done to extend the existing results on these models to generally applicable divergent $N$-echelon systems, and that from analyzing these models thoroughly we could obtain insight in the structural properties of the cost-optimal policy. Van Houtum & Zijm [1991b] proved exact decomposition results for a divergent 2-echelon system given the balance assumption. They imposed the same stringent assumptions as Eppen & Schrage [1981]. In Chapter 5 we prove that these stringent assumptions can be relaxed, and that decomposition is exact for a divergent $N$-echelon system given the balance assumption. Decomposition reduces the complex control problem to (1) determining the optimal order-up-to-level at each stockpoint, and (2) determining the optimal allocation functions at each intermediate stockpoint. We used the result of Gong, De Kok & Ding [1994] to prove that the first problem coincides with the classical newsboy problem. The second problem is solved by applying the Lagrangian multiplier technique. From this we obtain insight in the structure of the optimal allocation functions. An algorithm is developed to actually determine these functions. Based on several properties of the optimal allocation functions we classify these functions into four classes.

It is rather cumbersome and time-consuming to determine the optimal allocation functions. Hence, there is a need for a more practically useful approach. In Chapter 6, we restrict to a specific class of linear allocation functions. Under this additional restriction we develop an easy and fast algorithm to determine the parameters of a control policy which may not be cost-optimal, but yields approximately optimal costs.

A possible way to decrease the holding costs needed to operate the system, but still guarantee target service levels at the end-stockpoints, is by allowing lateral transshipments. In the situation where some end-stockpoints have excess inventory while others face shortages, lateral transshipments have gained in popularity as the appropriate recourse action to avoid shortages. However, by allowing lateral transshipments extra transportation (and handling) costs are involved. In Chapter 7, we analyze divergent multi-echelon systems in which lateral transshipments are allowed. By comparing the model with transshipments to a similar model without transshipment, we determine when it is profitable to allow for lateral transshipments.

Finally, in Chapter 8, a summary of the main results is given, and some topics for future results are proposed.

Most of the results in this monograph has already appeared or will soon appear in the literature. In this chapter we used Diks, De Kok & Lagodimos [1996] to give an overview of the literature. In Chapter 3, we compare various allocation functions and accompanying numerical methods by an extensive numerical study. These results are based on Van der Heijden, Diks & De Kok [1997]. In Chapter 4, we extend the model by allowing random lead times. The modeling of random lead times without
I.4. Outline of the monograph

crossing is based on Diks & Van der Heijden [1997], and the control of the divergent multi-echelon inventory system is based on Van der Heijden, Diks & De Kok [1996]. In Chapter 5, we prove that the decomposition is exact and derive the characteristics of the cost-optimal control policy. These results are based on Diks & De Kok [1998b]. In Chapter 6, we use these characteristics to develop practically useful allocation functions. These results are based on Diks & De Kok [1998a] and Diks & De Kok [1996a]. The transshipment model discussed in Chapter 7 is based on Diks & De Kok [1996a] and Diks & De Kok [1997].
Introduction
2

The basic model

2.1 Introduction

In this monograph, we analyze divergent multi-echelon systems in which inventory is controlled by echelon stock policies. To explain the echelon stock concept properly we introduce the echelon stock and the echelon inventory position of a stockpoint in Definition 2.1.

Definition 2.1. The echelon stock of a stockpoint equals the sum of its physical stock plus the amount in transit to or on hand at its downstream stockpoints minus possible backorders at its end-stockpoints. The echelon inventory position of a stockpoint equals its echelon stock plus all material in transit to that stockpoint.

In this chapter, we present a model for divergent $N$-echelon systems under echelon stock control. Every stockpoint in the system is controlled by an echelon order-up-to policy. This means that each time a stockpoint inspects its inventory it places an order at its predecessor so as to raise its echelon inventory position to a fixed order-up-to-level. In case the predecessor has insufficient stock available this stock is rationed among all successors by some allocation policy. This model constitutes the basis for the remaining chapters of this monograph. In Chapters 3 and 4, we develop algorithms to determine the order-up-to-level and the allocation functions of each stockpoint such that customer service level requirements are satisfied. In Chapters 5 and 6, we develop algorithms to determine these control parameters such that the expected total costs per period are minimized. In Chapter 7, we discuss a somewhat different model, which is introduced in detail only then. At the beginning of each of these chapters we briefly address the minor differences with the basic model (if there are any).

The chapter is organized as follows. In Section 2.2, we introduce some notation which is used throughout this monograph. In Section 2.3, we present the basic model and its assumptions. Next, in Section 2.4, we analyze the system dynamics of this model assuming all the control parameters are known. This is done by applying a sample path approach, i.e., by following the material flow through the system over time. From the system dynamics we determine several service measures. Finally, in Section 2.5, we address the important balance assumption, which is required to apply echelon stock policies.
2.2 Notation

For our convenience, we introduce some notation, which is used throughout this monograph. In Section 2.2.1, we introduce the notation with respect to the system layout. In Section 2.2.2, we introduce some notation in order to analyze the system dynamics of a divergent $N$-echelon system properly.

2.2.1 System layout

In this section, we introduce the notation with respect to the system layout of a divergent $N$-echelon system. The notation is illustrated by considering the divergent 4-echelon inventory system as depicted in Figure 2.1.

![Diagram of a divergent 4-echelon inventory system]

**Figure 2.1. Schematic representation of a divergent 4-echelon inventory system.**

The notation below is explained by a description and an example in brackets referring to the situation of Figure 2.1:

- **ech(i)**: All stockpoints that constitute the echelon of stockpoint $i$ (e.g., $ech(5) = \{5, 8, 9\}$).
- **pre(i)**: Preceding stockpoint of stockpoint $i$ (e.g., $pre(8) = 5$).
- $U_i$: All stockpoints on path from supplier to stockpoint $i$ (e.g., $U_1 = \emptyset$ and $U_6 = \{1, 3\}$).
- $V_i$: All stockpoints which are supplied by stockpoint $i$ (e.g., $V_1 = \{2, 3, 4\}$).
- $W_i$: All stockpoints with low level code $i$ (e.g., $W_3 = \{3, 4\}$).
- $E$: All end-stockpoints (e.g., $E = \{2, 6, 8, 9, 10\}$).
- $E_i$: All end-stockpoints in $ech(i)$ (e.g., $E_3 = \{6, 8, 9\}$).
- $M$: All intermediate stockpoints (e.g., $M = \{1, 3, 4, 5, 7\}$).
- $N$: Number of stages in inventory system (e.g., $N = 4$).

A low level code (LLC) is assigned to every stockpoint. By definition the low level code of an end-stockpoint $i$ equals 1, i.e., $LLC(i) = 1$. For an intermediate stockpoint $i$ we have $LLC(i) = 1 + \max_{j \in V_i} LLC(j)$. For example, we have that $LLC(5) = 2$.

2.2.2 Analysis

In this section, we introduce some notation in order to analyze the system dynamics of a divergent $N$-echelon system properly.
2.2. Notation

Lead times and demand

\( R \)  
Duration of a review period (in periods).

\( \mu_i \)  
Mean of one-period demand at stockpoint \( i \).

\( \sigma_i \)  
Standard deviation of one-period demand at stockpoint \( i \).

\( D_{E_i}^t \)  
The customer demand at the stockpoints in \( E_i \) during \( t \) periods.

\( D_{E_i,t_1,t_2} \)  
The demand at the stockpoints in \( E_i \) during \( [t_1,t_2] \).

\( L_i \)  
Lead time of stockpoint \( i \) (in periods).

Stockpoint status

For all order arrival times \( t \geq 0 \):

\( l_i^t \)  
The echelon inventory position of stockpoint \( i \) at time \( t \) (just before allocation).

\( l_i^{t^*} \)  
The echelon inventory position of stockpoint \( i \) at time \( t^* \) (just after allocation).

\( j_i^{t^*} \)  
The echelon stock of stockpoint \( i \) at time \( t^* \) (just before order arrival).

\( j_i^t \)  
The echelon stock of stockpoint \( i \) at time \( t \) (just after order arrival).

\( q_i(x) \)  
The number of products allocated to stockpoint \( i \) at time \( t \).

\( z_i(x) \)  
The echelon inventory position of stockpoint \( i \) just after allocation if the echelon stock of its supplier just before allocation equals \( x \).

Control parameters

\( q_i \)  
Allocation fraction from stockpoint \(pre(i)\) to stockpoint \( i \).

\( S_i \)  
Order-up-to-level of stockpoint \( i \) with respect to the echelon inventory position.

\( \Delta_i \)  
Maximum physical stock at intermediate stockpoint \( i \), i.e., \( \Delta_i = S_i - \sum_{k \in E_i} S_k \) for \( i \in M \).

Cost measures

\( h_i \)  
Value added per product at stockpoint \( i \).

\( p_i \)  
Penalty costs incurred per backlogged product at end-stockpoint \( i \).

Service measures

\( \alpha_i \)  
Probability that the echelon stock of stockpoint \( i \) is nonnegative at the end of an arbitrary replenishment cycle (* denotes the target value).

\( \beta_i \)  
Fraction of the demand satisfied immediately from the stock on hand at stockpoint \( i \).

\( \gamma_i \)  
One minus the ratio of the average shortage at end-stockpoint \( i \) before arrival of a replenishment order and the average demand at end-stockpoint \( i \) during an arbitrary replenishment cycle.

Mathematical notation

\( E[X] \)  
The mean of random variable \( X \).

\( var[X] \)  
The variance of random variable \( X \).

\( c(X) \)  
The coefficient of variation of random variable \( X \).

\( \Phi, \phi \)  
The cdf/pdf of the standard normal distribution.

\( \equiv \)  
Identically distributed.

\( |A| \)  
The number of elements in set \( A \).

\( i \in A \setminus B \)  
\( i \in A \) and \( i \notin B \).

\( \lfloor x \rfloor \)  
The largest integer smaller than or equal to \( x \).

\( x^+ \)  
\( \max(0, x) \) for any expression \( x \).
2.3 Model description

Consider a single-item divergent multi-echelon inventory system (see e.g., Figure 2.1). This system consists of a unique most upstream stockpoint (denoted by index 1) supplying a number of successors, which on their turn may supply several stockpoints. If a stockpoint supplies products to at least one stockpoint we refer to this stockpoint as an intermediate stockpoint, otherwise we refer to it as an end-stockpoint. Stockpoint 1 places orders at an external supplier by a periodic review policy. That is, it issues replenishment orders at the external supplier periodically, say every $R$ periods. We assume that this external supplier is able to guarantee that every replenishment order arrives after a fixed lead time $L_1$. Due to this assumption we do not have to take into account the order dependencies between all the stockpoints supplied by this external supplier. Immediately upon the arrival of a shipment at stockpoint 1 its successors inspect their inventory, and place a replenishment order. The order placed by a successor $j \in V_j$ is such that its echelon inventory position is raised to its order-up-to-level $S_j$.

Now there are two possibilities:

- The physical stock at stockpoint 1 is sufficient to raise the echelon inventory position of each successor $j$ to its order-up-to-level $S_j$. Then the required amounts are sent to the successors and excess stock is kept at stockpoint 1 to be allocated at the next occasion.

- The physical stock is not sufficient to reach the successors’ order-up-to-levels $S_j$. Then material rationing is required to allocate the available stock to its successors appropriately. For this purpose we introduce an allocation function.

An identical allocation procedure is applied to the other intermediate stockpoints. That is, upon the arrival of a shipment at an intermediate stockpoint $i$ each successor $j \in V_i$ inspects its inventory, and places a replenishment order to raise its echelon inventory position to its order-up-to-level $S_j$. Again, we distinguish between the two possibilities as given above.

We assume that the model above satisfies the following assumptions:

1. Customer demand only occurs at end-stockpoints.
2. The one-period demand is random and stationary in time.
3. The demand is both independent across end-stockpoints and across periods in time.
4. All customer demand that cannot be satisfied immediately from the stock on hand is backordered.
5. Partial delivery of customer orders is allowed.
6. No lot-sizing restrictions on replenishment orders.
7. There are no capacity constraints on production, storage or transport.

Let us briefly discuss these assumptions. Assumption 1 can be made without loss of generality, since we can add an end-stockpoint with lead time zero for each intermediate stockpoint facing customer demand. In assumption 2, we assumed the stationarity to simplify the analysis. In Section 2.4, we show that the results can be extended to non-stationary demand. With respect to assumption 3, we note that the extension to demand that is correlated across end-stockpoints is straightforward. If the one-period demands of all end-stockpoints are independent then $E[D_i]$ and $\sigma^2[D_i]$ are simply calculated from $E[D_i] = l \sum_{j \in E} \mu_j$ and $\sigma^2[D_i] = l \sum_{j \in E} \sigma_j^2$. These expressions can easily be modified to include correlations between end-stockpoints. First, the expressions for $E[D_i]$ remain the same. Second, defining $\rho_{jk}$ as the correlation between the one-period demand of two stockpoints $j$ and $k$, we have the following modified expression for the variance: $\sigma^2[D_i] = l \sum_{j \in E} \sum_{k \in E} \rho_{jk} \sigma_j \sigma_k$ (where
2.4 System dynamics

\( \rho_{ij} := 1 \) by definition. Then the analysis in the sequel still applies. However, the introduction of correlations between demand in subsequent periods is not straightforward (cf. Indekeu [1995] and Lagodimos, De Kok & Verrijdt [1995]). Furthermore, throughout this monograph we assume that all customer demand which cannot be satisfied immediately from stock on hand is backordered (assumption 4). An example of a class of models which does not require assumption 4 are the lost-sales models. The analysis of models in which the decision to backorder is based on some criterion, is often cumbersome. Assumption 5 is quite common in case of material shortages. Due to this assumption customers may be able to start operations before the complete order arrives. Assumption 6 is required to keep the analysis tractable. However, some papers do consider lot-sizing restrictions in multi-echelon systems under echelon stock control. One is the paper of De Bodt & Graves [1985], which studies echelon stock \((n, Q)\) policies in serial systems. To the best of our knowledge, they are the first to study simple and easily implementable echelon stock policies in multi-echelon systems with set-up costs at each stockpoint. They provide approximate performance measures under a nestedness assumption: whenever a stockpoint receives a shipment, a batch must be sent down immediately down to its successors. Another is the paper of Chen & Zheng [1994], which studied echelon stock \((s, n Q)\) and \((R, s, n Q)\) policies in serial systems. They provide an exact algorithm to compute the steady-state echelon inventory positions of each stockpoint, which can be used to evaluate the long-run average holding and penalty costs as well as other performance measures. Few of these papers consider divergent multi-echelon systems. To the best of our knowledge there is only the paper of Van Donselaar [1990], which analyzes echelon stock \((s, n Q)\)-policies in a divergent 2-echelon system. Due to its heuristic approach and all the assumptions made in the paper the applicability in practice is doubtful, as well as the possibility to extend it to larger \(N\)-echelon systems. So, this typically is a topic for further research. Finally, in assumption 7 we assume that there are no capacity constraints. Van Houtum, Indekeu & Zijm [1996] discussed this assumption. They stated that the most direct way to model restricted capacity is by assuming that some stockpoints have a fixed capacity determined by the maximum number of products that can be ordered each period. An alternative way is by replacing a stockpoint by a queuing system (cf. Lee & Zipkin [1994] and Veatch & Wein [1994]). Federguen & Zipkin [1986] proved that an order-up-to-policy is cost optimal for a single location stockpoint with capacity constraints. However, the extension which incorporates capacity restrictions in serial systems is not straightforward (cf. Speck and Wal [1991a,b]). Glasserman & Tayur [1996] developed a simple approximation for serial systems with limited capacity. For further support on this subject we refer to the related papers of Glasserman & Tayur [1994], and Glasserman [1997].

2.4 System dynamics

A prerequisite for determining service and cost measures in a multi-echelon system is to describe the stochastic process of the inventory at each stockpoint over time. In this section the system dynamics of the system as described in Section 2.3 is determined by applying a sample path approach, i.e., by following the material flow through the system over time. From the system dynamics we derive tractable expressions for several service measures (Section 2.4.1).

Consider the most upstream stockpoint denoted by index 1. At the beginning of period \( t - L_1 \) it raises the echelon inventory position to \( S_1 \). Since the lead time equals \( L_1 \), this order arrives at the beginning of period \( t \). So the echelon stock of stockpoint 1 just after the arrival of this order equals

\[ \hat{I}_1^t = S_1 - D_{1, L_1, t}^1. \] (2.1)

If \( \hat{I}_j^t \geq \sum_{\text{in } V_1} S_t \) then every stockpoint \( j \in V_1 \) is able to raise its echelon inventory position to its
order-up-to-level. Thus,
\[ D^{l}_{0,i,t} \leq \Delta_i \quad \Rightarrow \quad I^l_i := S_i, \quad j \in V_i, \] (2.2)

However, if \( \hat{I}^l_j < \sum_{i \in V_i} S_i \), then the complete echelon stock of stockpoint 1 is allocated to its successors \( j \in V_i \) by using some allocation functions \( z_j \). Thus,
\[ D^{l}_{j-1,t} > \Delta_i \quad \Rightarrow \quad I^l_j := z_j \left( S_i - D^{l}_{j-1,t} \right) \], \quad j \in V_i. \] (2.3)

By definition the echelon stock of stockpoint \( j \) at time \( t \) (just before allocation) equals \( I^l_j \). Hence, the amount of products actually allocated to stockpoint \( j \) at time \( t \) equals
\[ q^l_j := I^l_j - \hat{I}^l_j, \quad j \in V_i. \] (2.4)

Unfortunately, for some successor \( j \) it can happen that this amount \( q^l_j \) is negative. To explain this, we note that the echelon inventory position of stockpoint \( j \) at time \( t \) just before allocation equals
\[ I^l_j = I^l_{j,R} - D^{l}_{j-1,R}, \quad j \in V_i. \] (2.5)

The above equality immediately results from a simple path argument. Substituting (2.5) in (2.4) yields
\[ q^l_j = I^l_{j,R} - \hat{I}^l_{j,R} + D^{l}_{j-1,R}, \quad j \in V_i. \] (2.6)

The sizes of \( \hat{I}^l_{j,R} \) and \( I^l_j \) depend on the echelon stock of stockpoint 1 at time \( t - R \) and \( t \), respectively, and the allocation function \( z_j \) (see equations (2.2) and (2.3)). It may well be that \( \hat{I}^l_{j,R} \) is larger than \( I^l_j \), e.g., when the demand \( D^{l}_{j-1,R} \) is large (resulting in little echelon stock to allocate to its successors at time \( t \)), whereas \( D^{l}_{j-1,R} \) is small. In that case \( q^l_j \) is negative. When for at least one successor \( j \) it happens that \( q^l_j < 0 \) then there is imbalance at stockpoint 1. In the analysis in this chapter we assume that there does not exist any imbalance. This assumption is referred to as the balance assumption. Since we know that this assumption can be violated, it is important to know the impact of the assumption on the validity of the results obtained by the analysis. In the numerical study in Section 3.4–3.5 we investigate this impact. How to measure imbalance and how to cope with imbalance in real-world situations are topics that are addressed in Section 2.5.

In practice there is a need for simple allocation functions, that are easy to implement. Therefore in a major part of this monograph we consider a linear allocation function \( z_j(x) \), which is defined by
\[ z_j(x) := S_i - q^l_j \left( \sum_{i \in V_i} S_i - x \right), \quad j \in V_i. \] (2.7)

By definition all the \( x \) products are allocated. Hence, \( \sum_{i \in V_i} z_j(x) = x \), which implies that \( \sum_{i \in V_i} q^l_j = 1 \). The \( q^l_j \) are referred to as the allocation fractions of stockpoint \( i \). We assume that all allocation fractions are positive. The allocation function given by (2.7) is used in several papers (e.g., De Kok, Lagodimos \\& Seidel [1994] and Van der Heijden [1997b]). In Chapter 5, we do not assume that this allocation function. Then we derive the cost-optimal allocation function. Since we use (2.7) in the Chapters 3, 4, 6 and 7 we restrict to this linear allocation function in the remainder of this section. From (2.2), (2.3) and (2.7) it follows that
\[ \hat{I}^l_j = S_i - q^l_j \left( D^{l}_{j-1,R} - \Delta_i \right), \quad j \in V_i. \] (2.8)

Next, we consider an arbitrary successor of stockpoint 1, say \( j \). According to (2.8) this stockpoint places an order at stockpoint 1 and after possible rationing the echelon inventory position equals the order-up-to-level \( S_i \) minus a fraction \( q^l_j \) of the amount of products short at stockpoint 1. This (partial)
2.4. System dynamics

Order arrives at stockpoint \( j \) at the beginning of period \( t + L_j \). Hence, the echelon stock of stockpoint \( j \) at the beginning of period \( t + L_j \) equals

\[
I^j_{t+L_j} = \tilde{I}^j_t - D^j_{t+L_j}.
\]

If \( I^j_{t+L_j} \geq \sum_{a \in V_j} S_a \), then every stockpoint \( k \in V_j \) is able to raise its echelon inventory position to its order-up-to-level. Thus,

\[
I^j_t - \sum_{a \in V_j} S_a \geq D^j_{t+L_j} \implies I^j_{t+L_j} := S_k, \quad k \in V_j.
\]  

(2.9)

However, if \( I^j_{t+L_j} < \sum_{a \in V_j} S_a \), then the echelon stock of stockpoint \( j \) is allocated to its successors \( k \in V_j \) by using the allocation functions \( \tau_{k|a \in V_j} \). Thus,

\[
I^j_t - \sum_{a \in V_j} S_a < D^j_{t+L_j} \implies I^j_{t+L_j} := z_k \left( I^j_t - D^j_{t+L_j} \right), \quad k \in V_j.
\]  

(2.10)

Using the allocation rule as defined in (2.7), and (2.9)-(2.10) yields

\[
I^j_{t+L_j} = S_k - q_k \left( D^j_{t+L_j} - \left( I^j_t - \sum_{a \in V_j} S_a \right)^+ \right), \quad k \in V_j.
\]  

(2.11)

Substitution of (2.8) in (2.11) yields

\[
I^j_{t+L_j} = S_k - q_k \left( D^j_{t+L_j} - \Delta_j + q_k \left( D^j_{t+L_j} - \Delta_k \right)^+ \right)^+, \quad k \in V_j.
\]  

(2.12)

With equations (2.8) and (2.12) we are able to describe the stochastic process of the inventory at stockpoint \( j \) and \( k \), respectively, over time. Note that we do not require stationary demand, however, let us for sake of clarity restrict to stationary demand (see assumption 2 in Section 2.3). Then, by defining \( X_i := D^i_{t+L_i} - \Delta_k \), we are able to simplify (2.8) and (2.12)

\[
\tilde{I}^j_t \overset{d}{=} S_j - q_j X^+_j, \quad j \in V_1,
\]
\[
I^k_t \overset{d}{=} S_k - q_k \left( X_k + q_k X^+_k \right)^+, \quad k \in V_j.
\]

Using similar arguments as above it is possible to derive an expression for the echelon inventory position of any stockpoint. Suppose that a stockpoint \( i \) is supplied by \( i_{-1} \) and \( i_{1} \) for \( n = 1, \ldots, (r-2) \), with \( i_1 \) denoting the most upstream stockpoint. Then, it can be shown that

\[
\tilde{I}^i_t \overset{d}{=} S_i - q_i \left( X_i + q_i X^+_i \right)^+, \quad \ldots, + q_n \left( X_n + q_n X^+_n \right)^+\right)^+.
\]  

(2.13)

2.4.1 Service levels

In order to determine good (or even optimal) control parameters we need to determine the cost of holding inventory, the set-up costs and the penalty costs. The penalty costs are often ascribed to the possible loss of customer goodwill. In practice one usually is not able to assign these costs. Hence, they are determined indirectly by a certain service measure. So these service measures are used as operational surrogates. Even in situations where penalty costs can be determined (e.g., in the case of contractual arrangements regarding shortage penalties), service measures are needed in order to have direct information on operational performance of the supply chain (Schneider [1981]). In this monograph we consider the major three service measures, as introduced in Chapter 1.2.4. In this section we show how to determine these measures given all the control parameters. We distinguish between internal service and external customer service.
The non-stockout probability ($\alpha$).

The internal non-stockout probability is defined as the probability that an intermediate stockpoint $i$ is able to raise the echelon inventory positions of every successor to its order-up-to-level $S_i$. Hence,

$$\alpha_i := P_r \left( \bar{L}_j - D_{i, r + L_i} \geq \sum_{j \in V_i} S_j \right), \quad i \in M.$$  

The non-stockout probability of an end-stockpoint $i$ is defined as the probability that all its customer demand per replenishment cycle can be satisfied immediately from the stock on hand. Hence,

$$\alpha_i := P_r (\bar{L}_j - D_{i, r + L_i} \geq 0), \quad i \in E. \quad (2.14)$$

The fill rate ($\beta$).

This service measure is widely used in practice (cf. De Kok [1990] and Lagodimos [1992]). The internal fill rate at an intermediate stockpoint is defined as the fraction of the demand immediately satisfied from the stock on hand. Hence,

$$\beta_i := \frac{E \left[ \sum_{j \in V_i} q_i^j \right]}{E \left[ \sum_{j \in V_i} (S_j - \bar{L}_j) \right]}, \quad i \in M. \quad (2.15)$$

The definition of (2.15) can be simplified by two observations. First, the numerator of (2.15) equals the expected total amount of products allocated to the successors of stockpoint $i$ at time $t$. Using (2.6) we obtain

$$E \left[ \sum_{j \in V_i} q_i^j \right] = E \left[ \sum_{j \in V_i} (\bar{L}_j - (\bar{L}_j - D_{i, r, x, j}^j)) \right] = \sum_{j \in V_i} P \mu_j. \quad (2.16)$$

Second, the denominator of (2.15) equals the expected total number of products which is needed to raise the echelon inventory position of every successor to its order-up-to-level. Using (2.4) and (2.16) we obtain

$$E \left[ \sum_{j \in V_i} (S_j - \bar{L}_j) \right] = E \left[ \sum_{j \in V_i} (S_j - (\bar{L}_j - q_i^j)) \right] = \sum_{j \in V_i} (P \mu_j + E [S_j - \bar{L}_j]). \quad (2.17)$$

Substitution of (2.16) and (2.17) in (2.15) yields

$$\beta_i = \frac{\sum_{j \in V_i} P \mu_j}{\sum_{j \in V_i} (P \mu_j + E [S_j - \bar{L}_j])}, \quad i \in M.$$  

Unlike the $\alpha$ service level definition, the fill rate at an intermediate stockpoint does not give any information about how each of its successors experiences this fill rate. This can be determined by computing $\beta_i^j$, which denotes the fill rate issued at stockpoint $i$ to a particular successor $j \in V_i$. It is easy to see that

$$\beta_i^j = \frac{P \mu_j}{P \mu_j + E [S_j - \bar{L}_j]}, \quad i \in M \quad \text{and} \quad j \in V_i.$$  

Much more important is the (external) fill rate of an end-stockpoint $i$ (cf. Hadley & Whitin [1963]). This can be computed by one minus the ratio of the expected amount backordered at end-stockpoint $i$ during an arbitrary replenishment cycle and the expected demand at end-stockpoint $i$ during this cycle. The amount backordered during a replenishment cycle equals the amount of outstanding backorders
2.4. System dynamics

at the end of this cycle minus the amount at the begin of this cycle. Using a sample path argument it follows that for the replenishment cycle \( t + L_i + L_i + R \) of end-stockpoint \( i \) we have

\[
\text{Number of backorders (begin cycle)} = (D_{i,n+L_i} - \bar{L}_i)^+.
\]

\[
\text{Number of backorders (end cycle)} = (D_{i,n+L_i+R} - \bar{L}_i)^+.
\]

Since the mean demand at end-stockpoint \( i \) during the aforementioned replenishment cycle equals \( R\mu_i \), the fill rate equals

\[
\beta_i = 1 - \frac{E[(D_{i,n+L_i+R} - \bar{L}_i)^+] - E[(D_{i,n+L_i} - \bar{L}_i)^+]}{R\mu_i}, \quad i \in E.
\]

Under stationary demand the fill rate of an end-stockpoint \( i \) is given by

\[
\beta_i = 1 - \frac{E[(D_{i,n+L_i+R} - \bar{L}_i)^+] - E[(D_{i,n} - \bar{L}_i)^+]}{R\mu_i}, \quad i \in E. \tag{2.18}
\]

The modified fill rate \( \gamma_i \)

Besides the amount of products which need to be backordered at the end of a replenishment cycle, the \( \gamma \) service level also takes into account how long a customer has to wait until delivery of a backordered product. Unlike installation stock policies, the demand at intermediate stockpoints which cannot be satisfied immediately is lost. Hence, the \( \gamma \) service level is only introduced for end-stockpoints. It equals

\[
\gamma_i := 1 - \frac{E[(D_{i,n} - \bar{L}_i)^+]}{R\mu_i}, \quad i \in E. \tag{2.19}
\]

Note that \( \beta_i \) and \( \gamma_i \) are similar service level definitions. Therefore the latter is referred to as the modified fill rate. Specifically, if end-stockpoint \( i \) requires a high target service level, then usually the amount of backorders at the beginning of a replenishment cycle is small. Hence, \( E[(D_{i,n} - \bar{L}_i)^+] \) is small, and thus \( \beta_i \) is approximately equal to \( \gamma_i \).

When all the control parameters are given we are able to describe the stochastic process of the stock level in every stockpoint (see equation (2.13)). From this, mathematical expressions can be derived which enable us to compute all the aforementioned service measures. Unfortunately, it is quite cumbersome to evaluate these expressions exactly for multi-echelon systems with more than two stages. Therefore we propose to use the procedure as introduced by Seidel & De Kok [1990] to approximate the service measures. We explain the procedure by considering the non-stockout probability of an end-stockpoint in a divergent 3-echelon system. The generalization to larger systems and other service level measures is straightforward. Suppose the most upstream stockpoint \( 1 \) supplies a stockpoint \( j \), which supplies the considered end-stockpoint \( k \). From (2.12) and (2.14) it can be shown that

\[
a_k = \Pr \left( D_{i_n} + q_k (D_{i_n} + q_k (D_{i_n} - \Delta_i)^+ - \Delta_i)^+ + S_i \leq 0 \right).
\]

To approximate \( a_k \) we proceed as follows.

First, we fit a suitable distribution to the first two moments of \( D_{i_n} \). 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 the computations are greatly simplified. If we fit a mixed Erlang distributions to the first two moments of a nonnegative real random variable \( X \), we mean that \( X \) follows an \( E_{n,k} \),
with probability $\theta_1$, and an $E_{\nu,\lambda_3}$ with probability $\theta_2$. Hence the probability density function $f_X$ is defined by

$$f_X(x) := \theta_1 \lambda_1^{n-1} \left(\frac{x}{\nu} - 1\right)! e^{-\lambda_1 x} + \theta_2 \lambda_2^{n-1} \left(\frac{x}{\nu} - 1\right)! e^{-\lambda_2 x}.$$  

By using the fitting procedure of Tijms [1994] we can easily obtain the parameters $\lambda_1$, $\lambda_2$, $r_1$, $r_2$, $\theta_1$ and $\theta_2$ by matching the first two moments. If $c[X] < 1$ then we approximate the distribution of $X$ by the mixed Erlang distribution with parameters

$$r_1 := \lfloor 1/c^2[X] \rfloor, \quad r_2 := r_1 + 1, \quad \lambda_1 := \frac{r_2 - \theta}{\mu X}, \quad \lambda_2 := \lambda_1, \quad \theta_1 := \frac{1}{1 + c^2[X]} \left( r_2 c^2[X] - \sqrt{r_2(1 + c^2[X]) - r_2^2 c^4[X]} \right), \quad \theta_2 := 1 - \theta_1.$$  

If $c[X] \geq 1$ then we approximate the distribution of $X$ by the so-called Coxian-2 distribution with the gamma-normalization. Then the parameters of the mixed Erlang distribution are

$$r_1 := 1, \quad r_2 := 1, \quad \lambda_1 := \frac{2 \lambda_1}{E[X]} \left( 1 + \frac{c^2[X] - \frac{1}{2}}{c^2[X] + 1} \right), \quad \lambda_2 := \frac{4 E[X]}{E[X]} - \lambda_2, \quad \theta_1 := \frac{\lambda_1 E[X] - 1}{\lambda_2 - \lambda_1}, \quad \theta_2 := 1 - \theta_1.$$  

Second, after fitting a mixed Erlang distribution to the first two moments of $D_{\tau_m}^+$, we compute the mean and variance of $(D_{\tau_m}^+ - \Delta t)^{+}$. We apply the following formulae to compute the first two moments of a random variable $(X - c)^+$ when $X$ follows an $E_{\nu,\lambda_3}$ distribution with probability $\theta_1$, and an $E_{\nu,\lambda_2}$ distribution with probability $\theta_2 := 1 - \theta_1$. The first moment is given by

$$E[X - c]^+ = \int_c^{\infty} (x - c) dF_X(x) = \int_0^{\infty} (1 - F_X(x + c)) \, dx$$

$$= \int_c^{\infty} (1 - F_X(x)) \, dx = \int_0^{\infty} \sum_{j=0}^{\infty} \theta_j \frac{(-\lambda_1 x)^j}{j!} e^{-\lambda_1 x} \, dx$$

$$= \sum_{j=1}^{\infty} \frac{\theta_j (-1)^{j-1}}{\lambda_1^j j!} \int_c^{\infty} \lambda_1^{j-1} x^j e^{-\lambda_1 x} \, dx = \sum_{j=1}^{\infty} \frac{\theta_j (-1)^{j-1}}{\lambda_1^j j!} \sum_{n=0}^{j} \frac{(\lambda_1 x)^n}{n!} e^{-\lambda_1 x}$$

$$= \sum_{j=1}^{\infty} \frac{\theta_j (-1)^{j-1}}{\lambda_1^j} \sum_{n=0}^{j} (r_1 - n) \frac{(\lambda_1 x)^n}{n!} e^{-\lambda_1 x}. \quad (2.20a)$$
2.4. System dynamics

The second moment is given by

\[
E[(X - c)^2] = \int_{0}^{\infty} (x - c)^2 \, dF_X(x) = 2 \int_{0}^{\infty} x(1 - F_X(x + c)) \, dx
\]

\[
= 2 \int_{0}^{\infty} x \sum_{i=0}^{\infty} \frac{\lambda_i}{i!} e^{-\lambda_i} \sum_{j=0}^{i} \frac{1}{j!} \sum_{n=0}^{j} \left( \frac{i}{n} \right) e^{-n} x^n e^{-\lambda_i} \, dx
\]

\[
= 2 \sum_{i=1}^{\infty} \frac{\lambda_i}{i!} e^{-\lambda_i} \sum_{n=0}^{i} \left( \frac{i}{n} \right) e^{-n} (n + 1)! \frac{1}{\lambda_i n + 2} \sum_{j=0}^{i} \frac{1}{j!} \sum_{n=0}^{j} \left( \frac{i}{n} \right) e^{-n} x^n e^{-\lambda_i} \, dx
\]

\[
= 2 \sum_{i=1}^{\infty} \frac{\lambda_i}{i!} e^{-\lambda_i} \sum_{n=0}^{i} \left( \frac{i}{n} \right) e^{-n} (n + 1)! \frac{1}{\lambda_i n + 2} \sum_{j=0}^{i} \frac{1}{j!} \sum_{n=0}^{j} \left( \frac{i}{n} \right) e^{-n} x^n e^{-\lambda_i}.
\]  

(2.20b)

Third, we fit a mixed Erlang distribution to the first two moments of \(D_{x+1}^z + q_t(D_{x+1}^z - \Delta_1)^+\). Again, applying (2.20) yields the first two moments of \((D_{x+1}^z + q_t(D_{x+1}^z - \Delta_1)^+ - \Delta_2)^+\). Finally, we fit a distribution to the first two moments of \(D_{x+1}^z + q_t(D_{x+1}^z - \Delta_1)^+ - \Delta_2)^+\). Now \(\alpha_t\) is approximated by the probability that this random variable \(D_{x+1}^z + q_t(D_{x+1}^z - \Delta_1)^+ - \Delta_2)^+\) is less than \(S_t\).

We formalized the above procedure into the algorithm as depicted in Figure 2.2. Like in the derivation of (2.13) we consider a path from stockpoint \(i_t\) to \(i_{t+1}\), where \(i_{t+1}\) is supplied by \(i_t\) for \(n = 1, 2, \ldots, r - 1\). The basis of the algorithm is given by the following recursive relation

\[
\Gamma_{u} := (D_{x+1}^z + q_t\Gamma_{x+1}^- - \Delta_1)^+, \quad n = 1, 2, \ldots, r - 1,
\]

1. \(E[\Gamma_{u}^-] = 0, \text{var}[\Gamma_{u}^-] = 0 \) and \(n := 1\).
2. Fit a mixed Erlang distribution \(G_{x+1}^-\) to the first two moments of \(D_{x+1}^z + q_t\Gamma_{x+1}^-\), which are computed from

\[
E[D_{x+1}^z + q_t\Gamma_{x+1}^-] = L_{x+1} \mu_{x+1} + q_t E[\Gamma_{x+1}^-],
\]

\[
\text{var}[D_{x+1}^z + q_t\Gamma_{x+1}^-] = L_{x+1} \sigma_{x+1}^2 + q_t^2 \text{var}[\Gamma_{x+1}^-].
\]

3. By applying (2.20) we determine the first two moments of \(\Gamma_{x+1}^-\).
4. If \(n < r - 1\) then \(n := n + 1\) and return to step 2.
5. To determine \(\alpha_t\) we fit a mixed Erlang distribution \(G_t\) to the first two moments of \(D_{x+1}^z + q_t\Gamma_{x+1}^- + \Delta_1\). Then,

\[
\alpha_t = G_t(S_t).
\]

To compute other service measures we use that \(D_{x+1}^z \leq S_t - q_t\Gamma_{x+1}^-\).

---

**Figure 2.2.** Algorithm to compute service measures.
Now the algorithm just consists of fitting a mixed Erlang distribution to the first two moments of the random variable in the \( \mathcal{C} \) expression, beginning from \( n = 1 \) to \( n = r - 1 \). An exact algorithm to evaluate the distribution of an expression very similar to \( P_s \) of (2.13) is presented recently by Van Houtum & Zijm [1997]. Instead of a mixture of two Erlang distributions, they fit an Erlang distribution to the first two moments of each random variable. They concluded that the algorithm of Figure 2.2 is very accurate, and that only little additional accuracy is obtained by an exact evaluation of (2.13).

2.5 Imbalance

In Section 2.4, we introduced imbalance as the phenomenon where the rationing policy of an intermediate stockpoint allocates a negative quantity to at least one of its successors. In practice this would imply that products which were already allocated (and in the pipeline) have to be reallocated, by shipping the products from the stockpoints with negative allocation quantities, to those with positive allocation quantities. Imbalance does not result in any problem if, after each rationing decision, there would be a sufficiently large "demandless" period (e.g., week-end), which allows for the reallocation. However, usually such a reallocation period does not exist, or is undesirable due to the additional transportation costs involved.

In this monograph, we analyze divergent multi-echelon systems which are controlled by echelon stock policies. This analysis requires that there does not occur any imbalance. This assumption is referred to as the balance assumption, and coincides with the assumption as introduced by Verrijdt & De Kok [1996]. It is a generalization of the allocation assumption of Eppen & Schrage [1981] (which assumes that the allocation quantities are always sufficient to ensure equal stockout probabilities for all end-stockpoints). Unfortunately, as already mentioned in Section 2.4, generally by using an echelon stock policy some imbalance will occur. Thus, in reality (and in a simulation model) the balance assumption is violated. This violation is irrelevant as long as the results obtained by the analysis are still applicable, i.e., external service is according to prespecified targets. To identify the systems for which the imbalance has a considerable impact on the service levels attained, there is a need for an analytical measure to quantify the imbalance. In this monograph we address two definitions. The first measure, introduced by Seidel & De Kok [1990], and later analyzed by Verrijdt & De Kok [1996], defines the imbalance at an intermediate stockpoint \( i \) by the probability that the rationing policy allocates a negative quantity to one of its successors. Hence,

\[
\Omega_i^P(t) := \Pr(\exists \ j \in V_i : q_i^j < 0), \quad i \in M.
\]

The second measure, introduced by Van der Heijden [1997b], not only takes into account whether there is a successor with a negative allocation quantity, but it also incorporates how large the negative allocation quantities are. Van der Heijden defines the imbalance at an intermediate stockpoint \( i \) by the modulus of the sum of all negative allocation quantities. Hence,

\[
\Omega_i^N(t) := \sum_{j \in V_i} \Omega_j^P(t), \quad \Omega_i^N(t) := E[|q_i^j|], \quad i \in M.
\]

Other definitions are given by Zipkin [1984], Van Donselaar [1990] and Van Donselaar & Wijngaard [1993].

To illustrate how the imbalance can affect the service levels attained at the end-stockpoints, we consider a numerical example. In this example, we compare the target fill rates at the end-stockpoints with the actual fill rates, which are computed by discrete event simulation. Before addressing the numerical example, we explain how the simulation model deals with imbalance. Suppose at time \( t \) imbalance occurs at stockpoint \( i \). This means that the rationing policy allocates a negative amount to at least one
2.5. Imbalance

of its successors. Let us denote the set of successors for which \( q^j \leq 0 \) by \( V^i(t) \). Then we suggest to allocate nothing to every successor \( j \in V^i(t) \), and only part of the 'original' order to a successor \( j \notin V^i(t) \), namely

\[
\text{Amount allocated to stockpoint } j \notin V^i(t) := \left( \frac{\sum_{t \in V^i} q^i / \sum_{j \notin V^i(t)} q^j}{q^j} \right) q^j.
\]

One might develop more sophisticated procedures to deal with imbalance, e.g., for the 2-echelon system Langenhoff & Zijm [1990] presented an optimization problem to determine the allocation quantities minimizing the expected total holding and penalty costs per period, and Diks [1995] determines the allocation quantities by taking the allocation fractions into account. However, since the procedure as described above is easily implemented, very fast and yields good results we apply this procedure in all the simulation studies performed in the remainder of this monograph.

The numerical example is based on the divergent 3-echelon system as introduced by Bertrand & De Kok [1995]. In this system the most upstream stockpoint supplies 6 stockpoints, and each of these 6 stockpoints supplies a group of 4 stockpoints (see Figure 2.3). The review period \( R \) equals 1. The lead times of stockpoint 1 is equal to 5, the lead times of the stockpoints 2 to 7 all are equal to 2, and the lead time within a group is equal to 1. The demand characteristics of the stockpoints within one group are identical. These characteristics are depicted in Figure 2.3, where \( \mu \) denotes the mean, and \( c^2 \) denotes the squared coefficient of variation of the one-period demand at an end-stockpoint in each group (G2 to G7). To control this inventory system we use a Consistent Appropriate Share (CAS) rationing policy, which is explained thoroughly in Chapter 3. The control parameters (i.e., the allocation fractions and the order-up-to-levels) are computed such that every end-stockpoint attains a target fill rate \( \beta^* \). Note that \( \beta^* \) is identical for all end-stockpoints, and takes on the following three values: 0.9, 0.95 and 0.99.

As will turn out in Chapter 3, we require that \( \Delta_1 \) is known before we can compute the aforementioned control parameters. Hence, we assume that \( \Delta_3 = \Delta_4 = \Delta_5 = \Delta_7 = 0 \), and \( \Delta_2 = \Delta_6 \). The values of \( \Delta_1 \) and \( \Delta_2 \) are varied.

![Figure 2.3. Schematic representation of a divergent 3-echelon system.](image)

Table 2.1 depicts the amount of imbalance at each intermediate stockpoint and the fill rate at each end-stockpoint obtained by simulation (using the control parameters from the analysis). When \( \Delta_1 = \)

<table>
<thead>
<tr>
<th>G</th>
<th>( \mu )</th>
<th>( c^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>83</td>
<td>1.42</td>
</tr>
<tr>
<td>3</td>
<td>125</td>
<td>0.74</td>
</tr>
<tr>
<td>4</td>
<td>83</td>
<td>1.05</td>
</tr>
<tr>
<td>5</td>
<td>83</td>
<td>0.65</td>
</tr>
<tr>
<td>6</td>
<td>156</td>
<td>0.74</td>
</tr>
<tr>
<td>7</td>
<td>156</td>
<td>0.92</td>
</tr>
</tbody>
</table>
$\Delta_2 = 0$ every intermediate stockpoint in the system does not hold any stock. This results in a considerable amount of imbalance: $\Omega_{i}^{K} = 0.55$, $\Omega_{i}^{V} = 0.42$ and $\Omega_{i}^{K} = 0.35$. Due to this imbalance the service level experienced by a customer differs significantly from the target fill rate ($\tilde{\gamma}$ denotes the estimated fill rate at an end-stockpoint in group $i$). A well-known remedy to diminish the imbalance is to keep some stock back in the chain. In the model this corresponds with increasing one or more of the $\Delta_i$. Only raising $\Delta_2$ and $\Delta_1$ hardly has any effect. However, if at the same time $\Delta_i$ is increased the gap between the attained fill rates and the target fill rates decreases considerably (see $\tilde{\gamma}_j$ and $\tilde{\gamma}_i$). $\Delta_1$ is varied as 0, 14000 and 16000, which results in a mean stock of approximately 0.2 and 0.8 times the mean one-period demands. Analogously $\Delta_2$ is varied as 0 and 750, which results in a mean stock of approximately 0 and 0.4 times the mean one-period demands. From the results in Table 2.1 we conclude that the control parameters obtained by the analysis can be applied when the imbalance is sufficiently low.

Empirically Seidel & De Kok [1990] concluded that when $\Omega_{i}^{V}$ exceeds 0.3 the imbalance will have severe consequences for the attained customer service levels. Therefore as a rule of thumb they suggest to retain stock at stockpoint $i$, when $\Omega_{i}^{V}$ exceeds 0.3. Specifically, they increase $\Delta_i$ until $\Omega_{i}^{V}$ drops below 0.3. It is rather time-consuming to determine $\Delta_i$ such that $\Omega_{i}^{V}$ drops below 0.3, since we need to analyze and simulate the model probably several times. Therefore there is a need for an analytical expression for $\Omega_{i}^{V}$. This enables to determine this $\Delta_i$, which very likely results in a system for which the difference between the target fill rate and the actual fill rate is acceptable. It is extremely difficult to derive an analytically tractable expression for this measure $\Omega_{i}^{V}(\cdot)$. Therefore Verrijdt & De Kok [1996] decided to introduce the surrogate measure $\Omega_{i}^{V}(\cdot)$, which denotes the probability that stockpoint $i$ allocates a negative amount to its successor $j$ at time $t$. Thus,

$$\Omega_{i}^{V}(\cdot) := \Pr(q_i^j < 0), \quad j \in V_i.$$

In order to approximate $\Omega_{i}^{V}(\cdot)$ they assume that the balance assumption was not violated at time $t - R$.

Then, according to (2.6) we have that

$$q_i^j = l_i^j - l_{i-1}^j + D_i^{j-1}, \quad j \in V_i.$$

When restricting to a divergent 2-echelon system consisting of a central depot (denoted by index 1) supplying a number of end-stockpoints, substitution of (2.8) in the formula above yields

$$q_i^j = q_i (D_i^{j-L_i-R_i, L_i} - \Delta_i)^+ + D_i^{j-L_i, L_i}, \quad j \in V_i.$$

In the expression above the random variables are dependent. To approximate $\Omega_{i}^{V}(\cdot)$ Verrijdt & De Kok assumed that the depot cannot hold any stock (so $\Delta_1 = 0$). Furthermore, they distinguished between $R \leq L_1$ and $R > L_1$. It can be shown that $q_i^j = X - Y$, with

$$X := \begin{cases} \begin{array}{ll} q_i D_i^{j-L_i, L_i} & R \leq L_1, \\ q_i D_i^{j-L_i-R_i, L_i} + (1-q_i) D_i^{j-L_i, L_i} & R > L_1, \end{array} \end{cases}$$

and

$$Y := \begin{cases} \begin{array}{ll} q_i \sum_{j \in \alpha} D_i^{j-L_i, L_i} & R \leq L_1, \\ q_i \sum_{j \in \alpha} D_i^{j-L_i-R_i, L_i} & R > L_1, \end{array} \end{cases}$$

Note that $X$ and $Y$ are mutually independent, nonnegative random variables. By fitting appropriate distribution functions to the first two moments of $X$ and $Y$, we approximate $\Omega_{i}^{V}(\cdot)$ by computing $\Pr(X < Y)$. 


<table>
<thead>
<tr>
<th>$\beta^*$</th>
<th>$\Delta_1$</th>
<th>$\Delta_2$</th>
<th>$\hat{\Omega}_i^{x,k}$</th>
<th>$\hat{\Omega}_i^{y,k}$</th>
<th>$\hat{\Omega}_i^{x,k}$</th>
<th>$\hat{\Omega}_i^{x,k}$</th>
<th>$\hat{\Omega}_i^{x,k}$</th>
<th>$\hat{\Omega}_i^{x,k}$</th>
<th>$\hat{\Omega}_i^{x,k}$</th>
<th>$\hat{\beta}_i$</th>
<th>$\hat{\beta}_2$</th>
<th>$\hat{\beta}_3$</th>
<th>$\hat{\beta}_4$</th>
<th>$\hat{\beta}_5$</th>
<th>$\hat{\beta}_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.90</td>
<td>0</td>
<td>0</td>
<td>0.09</td>
<td>0.55</td>
<td>0.23</td>
<td>0.42</td>
<td>0.16</td>
<td>0.22</td>
<td>0.34</td>
<td>0.868</td>
<td>0.888</td>
<td>0.878</td>
<td>0.890</td>
<td>0.886</td>
<td>0.877</td>
</tr>
<tr>
<td>0.90</td>
<td>14000</td>
<td>0</td>
<td>0.03</td>
<td>0.54</td>
<td>0.22</td>
<td>0.42</td>
<td>0.16</td>
<td>0.22</td>
<td>0.35</td>
<td>0.884</td>
<td>0.892</td>
<td>0.887</td>
<td>0.893</td>
<td>0.890</td>
<td>0.888</td>
</tr>
<tr>
<td>0.90</td>
<td>16000</td>
<td>0</td>
<td>0.00</td>
<td>0.54</td>
<td>0.22</td>
<td>0.41</td>
<td>0.15</td>
<td>0.23</td>
<td>0.35</td>
<td>0.884</td>
<td>0.892</td>
<td>0.887</td>
<td>0.893</td>
<td>0.890</td>
<td>0.887</td>
</tr>
<tr>
<td>0.90</td>
<td>0</td>
<td>750</td>
<td>0.09</td>
<td>0.55</td>
<td>0.23</td>
<td>0.42</td>
<td>0.16</td>
<td>0.22</td>
<td>0.34</td>
<td>0.868</td>
<td>0.888</td>
<td>0.875</td>
<td>0.890</td>
<td>0.886</td>
<td>0.882</td>
</tr>
<tr>
<td>0.90</td>
<td>14000</td>
<td>750</td>
<td>0.03</td>
<td>0.27</td>
<td>0.22</td>
<td>0.19</td>
<td>0.16</td>
<td>0.22</td>
<td>0.35</td>
<td>0.896</td>
<td>0.892</td>
<td>0.895</td>
<td>0.893</td>
<td>0.889</td>
<td>0.886</td>
</tr>
<tr>
<td>0.90</td>
<td>16000</td>
<td>750</td>
<td>0.00</td>
<td>0.22</td>
<td>0.22</td>
<td>0.15</td>
<td>0.15</td>
<td>0.22</td>
<td>0.35</td>
<td>0.896</td>
<td>0.892</td>
<td>0.894</td>
<td>0.893</td>
<td>0.889</td>
<td>0.887</td>
</tr>
<tr>
<td>0.95</td>
<td>0</td>
<td>0</td>
<td>0.08</td>
<td>0.55</td>
<td>0.23</td>
<td>0.42</td>
<td>0.17</td>
<td>0.22</td>
<td>0.34</td>
<td>0.908</td>
<td>0.936</td>
<td>0.921</td>
<td>0.937</td>
<td>0.934</td>
<td>0.926</td>
</tr>
<tr>
<td>0.95</td>
<td>14000</td>
<td>0</td>
<td>0.03</td>
<td>0.54</td>
<td>0.23</td>
<td>0.42</td>
<td>0.16</td>
<td>0.22</td>
<td>0.34</td>
<td>0.933</td>
<td>0.942</td>
<td>0.936</td>
<td>0.942</td>
<td>0.940</td>
<td>0.938</td>
</tr>
<tr>
<td>0.95</td>
<td>16000</td>
<td>0</td>
<td>0.00</td>
<td>0.54</td>
<td>0.22</td>
<td>0.41</td>
<td>0.15</td>
<td>0.23</td>
<td>0.35</td>
<td>0.933</td>
<td>0.941</td>
<td>0.936</td>
<td>0.942</td>
<td>0.940</td>
<td>0.938</td>
</tr>
<tr>
<td>0.95</td>
<td>0</td>
<td>750</td>
<td>0.09</td>
<td>0.55</td>
<td>0.23</td>
<td>0.42</td>
<td>0.16</td>
<td>0.22</td>
<td>0.34</td>
<td>0.917</td>
<td>0.935</td>
<td>0.921</td>
<td>0.936</td>
<td>0.933</td>
<td>0.930</td>
</tr>
<tr>
<td>0.95</td>
<td>14000</td>
<td>750</td>
<td>0.03</td>
<td>0.27</td>
<td>0.22</td>
<td>0.18</td>
<td>0.16</td>
<td>0.22</td>
<td>0.35</td>
<td>0.947</td>
<td>0.942</td>
<td>0.946</td>
<td>0.943</td>
<td>0.940</td>
<td>0.938</td>
</tr>
<tr>
<td>0.95</td>
<td>16000</td>
<td>750</td>
<td>0.00</td>
<td>0.22</td>
<td>0.22</td>
<td>0.15</td>
<td>0.15</td>
<td>0.22</td>
<td>0.35</td>
<td>0.947</td>
<td>0.941</td>
<td>0.946</td>
<td>0.943</td>
<td>0.940</td>
<td>0.938</td>
</tr>
<tr>
<td>0.99</td>
<td>0</td>
<td>0</td>
<td>0.08</td>
<td>0.55</td>
<td>0.23</td>
<td>0.42</td>
<td>0.17</td>
<td>0.22</td>
<td>0.34</td>
<td>0.963</td>
<td>0.980</td>
<td>0.970</td>
<td>0.980</td>
<td>0.979</td>
<td>0.974</td>
</tr>
<tr>
<td>0.99</td>
<td>14000</td>
<td>0</td>
<td>0.03</td>
<td>0.54</td>
<td>0.23</td>
<td>0.42</td>
<td>0.16</td>
<td>0.22</td>
<td>0.35</td>
<td>0.981</td>
<td>0.985</td>
<td>0.982</td>
<td>0.986</td>
<td>0.985</td>
<td>0.984</td>
</tr>
<tr>
<td>0.99</td>
<td>16000</td>
<td>0</td>
<td>0.00</td>
<td>0.54</td>
<td>0.22</td>
<td>0.41</td>
<td>0.15</td>
<td>0.23</td>
<td>0.35</td>
<td>0.981</td>
<td>0.985</td>
<td>0.982</td>
<td>0.986</td>
<td>0.985</td>
<td>0.984</td>
</tr>
<tr>
<td>0.99</td>
<td>0</td>
<td>750</td>
<td>0.08</td>
<td>0.55</td>
<td>0.23</td>
<td>0.42</td>
<td>0.17</td>
<td>0.22</td>
<td>0.34</td>
<td>0.964</td>
<td>0.981</td>
<td>0.974</td>
<td>0.982</td>
<td>0.980</td>
<td>0.971</td>
</tr>
<tr>
<td>0.99</td>
<td>14000</td>
<td>750</td>
<td>0.03</td>
<td>0.27</td>
<td>0.22</td>
<td>0.18</td>
<td>0.16</td>
<td>0.22</td>
<td>0.35</td>
<td>0.989</td>
<td>0.986</td>
<td>0.989</td>
<td>0.986</td>
<td>0.985</td>
<td>0.984</td>
</tr>
<tr>
<td>0.99</td>
<td>16000</td>
<td>750</td>
<td>0.00</td>
<td>0.22</td>
<td>0.22</td>
<td>0.15</td>
<td>0.15</td>
<td>0.22</td>
<td>0.35</td>
<td>0.989</td>
<td>0.985</td>
<td>0.989</td>
<td>0.986</td>
<td>0.985</td>
<td>0.984</td>
</tr>
</tbody>
</table>

Table 2.1. The simulation results of the inventory system as depicted in Figure 2.3, where $\beta^*$ denotes the target fill rate, $\Delta_1$ and $\Delta_2$ are control parameters, $\hat{\Omega}_i^{x,k}$ denotes the estimated imbalance at stockpoint $i$ and $\hat{\beta}_i$ denotes the estimated fill rate of an end-stockpoint in group $i$. 

---

31
Diks [1995] extended the results of Verrijdt & De Kok by deriving a tractable approximation for \( \Omega_{ik}^X(t) \) when also the depot is allowed to hold stock \((\Delta_1 \geq 0)\), but \( R \leq L_1 \). After considerable algebra it follows that

\[
\Omega_{ik}^X(t) \approx \int_0^\infty \int_{cz+\Lambda_1}^\infty (1 - F_Y(cz + \Lambda_1 - w)) F_X(w - cz) dF_W(w) dF_Z(z),
\]

with \( X := D_{R-R_i-i-1}^j, Y := D_{R,R_i,i-1}^j, Z := D_{R_i,R_i,i-1}^j, W := \sum_{s \neq j} D_{R_i-R_i,i}^j \), and \( c := \frac{R_i - R}{W} \).

Note that \( X, Y, Z \) and \( W \) are independent nonnegative random variables. Again by fitting distribution functions to the first two moments of these random variables we are able to evaluate \( \Omega_{ik}^X(t) \). For some special cases (e.g., all random variables are exponentially distributed) it is possible to derive an analytically tractable expression. Otherwise we suggest to approximate \( \Omega_{ik}^X(t) \) by using numerical integration. Diks [1995] approximated the aforementioned random variables by the mixed Erlang distributions as presented in Section 2.4.1. He also derived an approximation for \( \Omega_{ik}^X(t) \). This approximation is simply obtained by neglecting the dependencies between the amount of products allocated to every successor, but taking into account that all allocation quantities \( q^j_i \) cannot be negative at the same time. Hence,

\[
\Omega_{ik}^X(t) \approx 1 - \prod_{s \in V_i} (1 - \Omega_{ik}^X(t)) + (-1)^{|W|} \prod_{s \in V_i} \Omega_{ik}^X(t). \tag{2.22}
\]

Note that approximation (2.22) is exact, when the depot only has two successors. By substituting the approximation of \( \Omega_{ik}^X(t) \) in (2.22) we obtain an approximation for \( \Omega_{ik}^X(t) \). Diks [1995] tested the performance of the approximations (2.21) and (2.22) by comparing them with the imbalance observed in a discrete-event simulation model. He constructed an experimental design consisting of 288 instances of a 2-echelon system, where a central depot (denoted by the index 0) supplies 6 end-stockpoints. Besides varying the demand characteristics, the service level requirements, the mean amount of stock held at the depot, also the lead time of the depot \((L_1 \in \{2, 9\})\), and the lead time of each end-stockpoint \((L_j \in \{1, 3\})\) is varied. Figure 2.4 (a) depicts \( \Omega_{ik}^X(t) \) obtained by simulation as a function of approximation (2.21) (we suppress \( t \) since \( \Omega_{ik}^X(t) \) is identically distributed for all \( t \)). From this figure we conclude that the approximation tends to underestimate \( \Omega_{ik}^X \). This can be explained as follows. In the simulation model some of the imbalance at time \( t \) is caused by imbalance at time \( t - R \), whereas the approximation of (2.21) assumes that no imbalance occurs at time \( t - R \). Figure 2.4 (b) depicts \( \Omega_{ik}^X \) obtained by simulation as a function of approximation (2.22). From this figure we conclude that approximation (2.22) is a rough approximation, which overestimates \( \Omega_{ik}^X \). This overestimation immediately results from (2.22) and the fact that we underestimate \( \Omega_{ik}^X \).

Due to new developments in the field of rationing policies (see Chapter 3) the need for more accurate approximations has diminished. Van der Heijden [1997b] developed the Balanced Stock rationing policy (referred to as BS rationing) which allocates the material such that the imbalance hardly affects the results obtained by the analysis. The BS rationing policy tries to set the allocation fractions \( \{q_i\}_i \in V_i \) such that \( \Omega_{ik}^X(t) \) is minimized. For that purpose we need a tractable expression for \( \Omega_{ik}^X(t) \). Like Verrijdt & De Kok, Van der Heijden derived such an expression by restricting to a divergent 2-echelon system with a stockless central depot (denoted by index 1), and assuming that end-stockpoint \( j \) does not face any imbalance at time \( t - R \). Hence,

\[
q_i^j = q_i^j \left( D^j_{i-1, R_i - i} - D^j_{R_i - 1, i} \right) + D^j_{i, R_i - i}, \quad j \in V_1. \tag{2.23}
\]
2.5. Imbalance

![Graphs showing performance of the approximations for Ω^k_j and Ω^k_j.](image)

**Figure 2.4.** Performance of the approximations for Ω^k_j and Ω^k_j.

From this we derive an approximation for Ω^H_j(t) (see Theorem 2.1).

**Theorem 2.1.**

\[
\Omega^H_j(t) \approx \sigma_{G_j} \Phi \left( \frac{\mu_{G_j}}{\sigma_{G_j}} \right) + \mu_{G_j} \Phi \left( \frac{\mu_{G_j}}{\sigma_{G_j}} \right), \quad j \in V_1, \tag{2.24}
\]

with

\[
\mu_{G_j} := -R\mu_j, \quad \sigma_{G_j}^2 := 2q_j^2 I_1 \sigma_j^2 + (R - 2q_j I_1) \sigma_j^2 \quad \text{and} \quad I_1 := \min(R, L_1).
\]

**Proof.** Since \( q_j^2 \) can take on values on the entire interval \((-\infty, \infty)\) we approximate \( Y := -q_j^2 \) by a normal distribution with mean \( \mu_{G_j} \) and variance \( \sigma_{G_j}^2 \). From equation (2.23) we can simply derive this mean and variance straightforwardly. Furthermore, we have that \( \Omega^H_j(t) = E[Y^+]. \) Hence,

\[
\Omega^H_j(t) \approx \int_{-\infty}^{\infty} y dF_Y(y) \quad \text{with} \quad Y \sim N(\mu_{G_j}, \sigma_{G_j}^2)
\]

\[
= \int_{-\mu_{G_j}/\sigma_{G_j}}^{\infty} (\mu_{G_j} + x \sigma_{G_j}) d\Phi(x)
\]

\[
= \mu_{G_j} \Phi \left( \frac{\mu_{G_j}}{\sigma_{G_j}} \right) + \sigma_{G_j} \left[ \int_{-\mu_{G_j}/\sigma_{G_j}}^{\infty} xe^{-x^2/2} dx \right]
\]

\[
= \sigma_{G_j} \Phi \left( \frac{\mu_{G_j}}{\sigma_{G_j}} \right) + \mu_{G_j} \Phi \left( \frac{\mu_{G_j}}{\sigma_{G_j}} \right).
\]

\[\square\]
The basic model

A numerical study of Van der Heijden shows that the approximation above is a very rough one. Presumably because the right side tail behavior of \( q' \) influences \( \Omega(t) \) considerably. The simple normal approximation is tuned on the mean and variance of \( q' \), not on the tail behavior. In the next chapter we show how the approximation developed in Theorem 2.1 can be used to determine the allocation fractions.
3

Service level constraints: fixed lead time case

3.1 Introduction

In Chapter 2, we formulated a model for a divergent \( N \)-echelon system. In this model, we have several control parameters. First, each stockpoint has an order-up-to-level. Each time when a stockpoint \( j \) places an order it attempts to raise its echelon inventory position to order-up-to-level \( S_j \). Second, sometimes the supplying stockpoint \( i \) say, cannot satisfy all the orders of its successors. Then material rationing is required. For that purpose we introduced an allocation function \( z_j \) for each successor \( j \in V_i \). In this chapter, we restrict to the linear allocation function as defined in equation (2.7). Using this allocation function, the echelon inventory position (after rationing) equals the order-up-to-level \( S_j \) minus a fraction of the amount of products stockpoint \( i \) is short to satisfy all orders. So, in principle every stockpoint \( j \) has two control parameters: An order-up-to-level \( S_j \) and an allocation fraction \( q_j \) (except for the most upstream stockpoint for which \( q_j := 1 \) by definition). The objective of this chapter is to determine all the control parameters in the system, such that every end-stockpoint attains its specific target service level. The service level under consideration is the fill rate, since this is often used in practice. However, we like to emphasize that all the results derived in this chapter can easily be extended to the other service measures (see Section 2.4.1). In this chapter, we address several methods to obtain the control parameters.

In the literature, the determination of these control parameters has received considerable attention. Eppen & Schrage [1981] introduced the Fair Share (FS) rationing policy for a 2-echelon system without central stock. This rationing policy ensures that the stockout probabilities at the end-stockpoints are equal. Extensions of the results of Eppen & Schrage are given by Federgruen & Zipkin [1984]. An excellent overview on rationing policies minimizing holding and penalty costs is given by Federgruen [1993]. Most papers reviewed by Federgruen discuss 2-echelon systems and it is not clear whether the results derived can be easily extended to arbitrary \( N \)-echelon systems, when taking into account computational considerations. This subject is addressed in Chapters 5 and 6, in which we derive the allocation functions minimizing holding and penalty costs in an arbitrary \( N \)-echelon system (even with non-identical holding and penalty costs). However, then the control parameters in such a system are a consequence of the cost structure chosen. Therefore, they cannot be applied to the situation discussed.
in this chapter, where we focus on target fill rates at end-stockpoints. In De Kok [1990] a
generalization of the rationing policy proposed by Eppen & Schrage [1981] is presented that enables to compute
the control parameters in a 2-echelon system without central stock taking into account fill rate con-
where the depot is allowed to hold stock. They introduced the concept of Consistent Appropriate Share (CAS) rationing. Verrijdt & De Kok [1996] present a modification of the heuristic approach
in De Kok [1990] to cope with significantly differing target fill rates. Verrijdt & De Kok [1995] show
that the results in De Kok [1990] can be generalized to arbitrary divergent N-echelon systems where
only end-stockpoints are allowed to hold stocks. Recently, Van der Heijden [1997b] introduced the
Balanced Stock (BS) rationing policy for a divergent N-echelon system without intermediate stocks.
The aforementioned rationing policies have not been extended yet to general N-echelon distribution systems where all intermediate and end-stockpoints are allowed to hold stock. In this chapter, we make
such extensions of the analysis and we carry out an extensive numerical comparison of the different
rationing policies. The comparison is based on the difference between target fill rates and actual fill
rates, where the actual fill rates are computed by discrete event simulation.

The chapter is organized as follows. In Section 3.2, two rationing policies and its variants are in-
vestigated by considering a 2-echelon system. The application of both policies is extended to an N-
echelon system in Section 3.3. An extensive numerical study has been undertaken to get insight in
the performance of both policies. In Sections 3.4 and 3.5, we consider many instances of a 2-echelon
system and a 3-echelon system, respectively. Finally, we present our conclusions in Section 3.6.

3.2 Controlling a 2-echelon system

In the previous chapter, we derived how to compute the service level at an end-stockpoint given the
control parameters. In the literature, several heuristics have been developed for 2-echelon systems
to determine the control parameters such that the service level constraints at the end-stockpoints are
satisfied. In this section, we present these heuristics, where the service level under consideration is the
fill rate. As already mentioned, all the results derived in this section can easily be extended to the other
service measures (see Section 2.4.1).

In Section 3.2.1, we describe several variants of the CAS rationing policy of De Kok, Lagodimos &
Seidel [1994]. In Section 3.2.2, we describe two variants of the BS rationing policy of Van der Heijden
[1997b]. Finally, in Section 3.2.3, we address the adaptation of the CAS rationing policy, which was
suggested by Diks & De Kok [1996a]. This adaptation differs from the CAS and BS rationing policy,
since it does not use the allocation function as defined in (2.7).

3.2.1 Consistent Appropriate Share rationing

Consider a central depot (denoted by index 1) supplying a number of end-stockpoints. Then the CAS
rationing policy of De Kok, Lagodimos & Seidel [1994] assumes that the order-up-to-level of each
end-stockpoint j equals

\[ s_j = \mu_{ec(1)} + g_j \sum_{i \in V_1} (S_i - \mu_{ec(i)}), \quad j \in V_1, \]  

(3.1)

where \( \mu_{ec(i)} \) denotes the expected demand at end-stockpoint j between the placement of an order at
stockpoint 1 and the earliest possible arrival time of products from the next order at stockpoint j. In
general, we define

\[ \mu_{ec(i)} := \sum_{n \in E_i} \left( \sum_{j \in V_{i+1} \text{ ech}(0)} L_j + I_n + R \right) \mu_n. \]
3.2. Controlling a 2-echelon system

For instance, \( \mu_{ek(0)} \) in Figure 2.1 equals \((L_3 + L_5 + R)\mu_5 + (L_3 + L_4 + L_5 + R)\mu_6 + (L_3 + L_4 + L_5 + R)\mu_9 \).

Substitution of (3.1) in (2.8), and next substituting the result in (2.18) yields

\[
\beta_j = 1 - \frac{E[(D_{k(j)} - \mu_{ek(0)} - q_j O_1)^+ - E[(D_{k(j)} - \mu_{ek(0)} - q_j O_1)^+]_+]}{R \mu_j}, \quad j \in V_1,
\]

(3.2)

with \( O_1 := S_1 - \Delta_1 - (D_{k(j)} - \Delta_1)^+ - \sum_{a \in V_1} \mu_{ek(0)} = \sum_{a \in V_1} (S_a - \mu_{ek(0)}) - (D_{k(j)} - \Delta_1)^+ \).

In order to determine the control parameters such that the fill rate of end stockpoint \( j \) equals the predetermined target fill rate \( \beta^*_j \), we solve the following problem

\[
\begin{align*}
\beta_j &= \beta^*_j, & j &\in V_1, \\
\sum_{a \in V_1} q_a &= 1,
\end{align*}
\]

(3.3)

where \( \beta_j \) is given by (3.2). The problem above consists of \(|V_1| + 2\) decision variables, namely the allocation fractions \( q_j \), \( j \in V_1 \), the order-up-to-level \( S_1 \) and \( \Delta_1 \). The order-up-to-levels \( S_j \), \( j \in V_1 \), immediately results from (3.1) when we use that \( \sum_{a \in V_1} S_a = S_1 - \Delta_1 \). Unfortunately, problem (3.3) only consists of \(|V_1| + 1\) equations. Therefore in the remainder of this section we solve problem (3.3) for a given \( \Delta_1 \). This implies that the maximum physical stock at the depot is chosen on before hand. If \( \Delta_1 = 0 \) the depot will not hold any stock, i.e., when a product arrives at the depot it is immediately allocated to one of the end-stockpoints. If \( \Delta_1 = \infty \), the system decomposes into \(|V_1|\) single location systems working in parallel.

In the literature, several heuristics have been developed to solve (3.3) for a given \( \Delta_1 \). Below we discuss four heuristics, respectively indicated by CAS1, CAS2, CAS3 and CAS4. The first two heuristics were proposed by De Kok, Lagodimos & Seidel [1994] based on earlier work of De Kok [1990]. The latter two heuristics are discussed in Verrijdt & De Kok [1996]. We address these heuristics successively. 

**CAS1:**

(i) Initialize \( S_1 \) and \( \epsilon > 0 \).

(ii) Use (3.2) to determine the allocation fraction \( q_j \) for every end-stockpoint \( j \).

(iii) If \( \sum_{a \in V_1} q_a < 1 - \epsilon \) then decrease \( S_1 \) and return to step (ii),

if \( \sum_{a \in V_1} q_a > 1 - \epsilon \) then increase \( S_1 \) and return to step (ii).

The computational burden of this algorithm is related to step (ii) where we have to solve \(|V_1|\) equations.

De Kok, Lagodimos & Seidel [1994] solve each equation by using bisection, since they assumed that \( \beta_j \) is an increasing function with respect to \( q_j \). In Diks & De Kok [1996a], it is argued that this is only true for large values of \( \beta_j \). Therefore Diks & De Kok proposed a minor adaptation of the CAS rationing rule, such that, indeed, an increase of \( q_j \) guarantees an increase of the fill rate attained at end-stockpoint \( j \). In Section 3.2.3, we address this adaptation extensively.

**CAS2:**

(i) Determine for every end-stockpoint \( j \) the order-up-to-level \( S_j \) such that \( \beta_j = \beta^*_j \), assuming \( \Delta_1 = \infty \). This order-up-to-level can be determined from (2.18) after substituting \( \beta^*_j = S_j \).
(ii). In correspondence with (3.1) we define

\[ q_j := \frac{S_j - \mu_{\text{ech}(j)}}{\sum_{n \in V_1} (S_n - \mu_{\text{ech}(n)})} \quad j \in V_1. \]

(iii). Use (3.2) to determine for every end-stockpoint \( j \) the required order-up-to-level at the depot, denoted by \( S_1(j) \), such that \( \beta_j = \beta_j^* \).

(iv). Define

\[ S_1 := \frac{\sum_{n \in V_1} S_1[n]}{|V_1|}. \]

Since the allocation fractions are defined in step (ii), we only have one decision variable left (i.e., \( S_1 \)) to satisfy the remaining \( |V_1| \) service equations. So, unlike CAS1, the CAS2 heuristic approximates the solution of problem (3.3). Therefore, it is reasonable to expect that CAS1 outperforms CAS2 if \( \epsilon \) is sufficiently small.

As argued in Vervijdt & De Kok [1996], the CAS2 heuristic is justifiable when the differences between the values of \( S_1[j] \) in step (iii) for the different end-stockpoints are small. However, when we are dealing with different target fill rates the values of \( S_1[j] \) may differ more than desirable. By averaging over these values in step (iv) this leads to end-stockpoints \( j \) for which the attained fill rate is too small (\( S_1[j] > S_1^* \)), and to end-stockpoints \( j \) for which the attained fill rate is too high (\( S_1[j] < S_1^* \)). It was felt that by adjusting the allocation fractions the performance of CAS2 could be improved. Vervijdt & De Kok [1996] developed two methods for adjusting these allocation fractions, namely the 'extreme case' method and the 'group' method. In this chapter we refer to these two methods as CAS3 and CAS4, respectively. Both methods are an extension of CAS2. Hence, steps (i) to (iv) precede the steps below. Now, we describe these two methods successively.

CAS3:

(v). Determine a stockpoint \( m \in V_1 \) for which

\[ |S_1'[m] - S_1'| \leq |S_1'[j] - S_1'| \quad j \in V_1. \]

(vi). If \( S_1'[m] < S_1' \) then the adapted allocation fractions are defined by

\[ \tilde{q}_j := \begin{cases} q_j - \delta \quad & j = m \\ q_j + \delta \frac{q_j}{1 - q_m} \quad & j \neq m. \end{cases} \]

If \( S_1'[m] > S_1' \) then the adapted allocation fractions are defined by

\[ \tilde{q}_j := \begin{cases} q_j + \delta \quad & j = m \\ q_j - \delta \frac{q_j}{1 - q_m} \quad & j \neq m. \end{cases} \]

(vii). Return to step (iii) of heuristic CAS2 (after adapting \( \delta \)) until \( \delta \) minimizes

\[ S_1^{\text{max}} = \max \{ S_1[j] \mid j \in V_1 \}, S_1^{\text{min}} = \min \{ S_1[j] \mid j \in V_1 \}. \]

The parameter \( \delta \) determines to what extent the allocation fractions are increased or decreased.
3.2. Controlling a 2-echeion system

CAS4:

(iii) Divide the successors of stockpoint 1 into two groups A and B.

\[ A := \{ j \in V_1 | S_j > S_1 \} \quad \text{and} \quad B := \{ j \in V_1 | S_j \leq S_1 \}. \]

(vi) Define the adapted allocation fractions by

\[ \tilde{q}_j := \begin{cases} \frac{(1 - \delta)q_j}{1 + \delta - \frac{28}{(1 + \delta)q_j} \sum_{s \in A} q_s} & j \in A, \\ \frac{(1 + \delta)q_j}{1 + \delta - \frac{28}{(1 + \delta)q_j} \sum_{s \in B} q_s} & j \in B. \end{cases} \]

(vii) Return to step (iii) of heuristic CAS2 (after adapting \( \delta \)) until \( \delta \) minimizes

\[ \frac{S_{j_{\text{max}}} - S_{j_{\text{min}}}}{S_1 - \mu_{\text{min}}}. \]

Again, \( \{q_j\}_{j \in V_1} \) sum up to one.

3.2.2 Balanced Stock Rationing

In Van der Heijden [1997b], it is argued that by not defining the order-up-to-levels \( \{s_j\}_{j \in V_1} \) as in (3.1) we obtain more degrees of freedom, which can be used to better tune the control parameters. The algorithm of Van der Heijden works as follows. First, the allocation fractions \( \{q_j\}_{j \in V_1} \) are determined such that a surrogate for the expected amount of imbalance is minimized. Next, the order-up-to-levels \( \{s_j\}_{j \in V_1} \) are determined so as to guarantee the target fill rates at the end-stockpoints. Like in the CAS heuristics it is assumed that \( \Delta_1 = 0 \) is known in advance.

Van der Heijden proposes to determine the allocation fractions \( \{q_j\}_{j \in V_1} \) based on the system with \( \Delta_1 = 0 \) (even if \( \Delta_1 > 0 \)), since then the allocation fractions can be determined independently of the order-up-to-levels. The allocation fractions \( \{q_j\}_{j \in V_1} \) are chosen such that the mean imbalance at the stockless depot (i.e., \( \Omega_1^H \)) is minimized. From equation (2.24) it follows that \( \mu_{\Delta,1} \) does not depend on the allocation fractions \( \{q_j\}_{j \in V_1} \). Next, consider the effect of \( s_{\Delta,1} \) on the mean imbalance at the depot. Differentiation of (2.24) to \( s_{\Delta,1} \) proves that the mean imbalance is strictly increasing in \( s_{\Delta,1} \), so we have to minimize \( s_{\Delta,1} \). If we would choose the allocation fractions such that the mean imbalance at the depot is minimized we obtain

\[ \tilde{q}_j = \frac{\sigma_j^2}{2\sigma_1^2}, \quad j \in V_1. \]  

(3.4)

Unfortunately, these \( \{q_j\}_{j \in V_1} \) do not sum up to 1, but to \( \frac{1}{2} \). In order to get allocation fractions that minimize the mean imbalance and sum up to one, we apply the Lagrange-multiplier technique. Hence,

\[ \min \sum_{a \in V_1} \Omega_a^H \quad \Rightarrow \quad \min \sum_{a \in V_1} \Omega_a^H + c_1 \left( 1 - \sum_{a \in V_1} q_a \right). \]

s.t. \[ \sum_{a \in V_1} q_a = 1 \]

The coefficient \( c_1 \) is the Lagrange-multiplier. To solve this optimization problem we take the derivative of the objective function to \( q_j \). This yields

\[ \frac{\delta}{\delta q_j} \left( \sum_{a \in V_1} \Omega_a^H + c_1 \left( 1 - \sum_{a \in V_1} q_a \right) \right) = c_1 = 0, \quad j \in V_1. \]  

(3.5)
In Theorem 2.1 we presented an approximation for \( \Omega_{ij}^u \). From that theorem it follows that \( \Omega_{ij}^u \) only depends on the allocation fraction \( q_j \), and not on other allocation fractions. Hence, substitution of (2.24) in (3.5) yields

\[
d \left( \sigma_{\alpha_j}, \phi \left( \frac{\mu_{\alpha_j}}{\varphi_{\alpha_j}} \right) + \mu_{\alpha_j} \Phi \left( \frac{\mu_{\alpha_j}}{\varphi_{\alpha_j}} \right) \right) = \epsilon_i, \quad j \in V_i.
\]

After some elementary algebra we obtain that

\[
\frac{\phi \left( \frac{\mu_{\alpha_j}}{\varphi_{\alpha_j}} \right)}{\sigma_{\alpha_j}} T_i \left( 2q_j \sigma_i^2 - \sigma_j^2 \right) = \epsilon_i, \quad j \in V_i.
\]

The \( \epsilon_i \) is determined such that the allocation fractions sum up to one.

Van der Heijden [1997b] developed a heuristic to determine all the control parameters (given \( \Delta_i \)). We refer to this heuristic as BS1. An adaptation of this heuristic is proposed by Van Domselaar [1996], which is referred to as the BS2 heuristic. We address these heuristics successively.

**BS1:**

(i) Compute lower bounds \( q_j \) for \( q_j \) for \( j \in V_i \) using (3.4).

(ii) Use bisection to find \( \epsilon_i \) of (3.6) such that the allocation fractions sum up to one. In each step of the bisection, the corresponding values for \( q_j \) where \( q_j \neq \epsilon_i \) are found by another bisection, where \( q_j \) should be in the interval \( (q_j, 1) \).

(iii) Determine for every end-stockpoint \( j \) the order-up-to-level \( S_j \) such that \( \beta_j = \beta_j^* \). This order-up-to-level can be determined from (2.18) after substituting \( L_i = S_j - q_j (D_i + \Delta_i) \).

(iv) The order-up-to-level \( S_i \) follows from

\[
S_i = \sum_{k \in V_i} S_k + \Delta_i.
\]

**BS2:**

Instead of minimizing \( \Omega_i^u \), we could also choose to minimize \( \sum_{k \in V_i} \sigma_k^2 \) subject to \( \sum_{k \in V_i} q_k = 1 \).

The Lagrange multiplier technique yields

\[
q_j = \frac{\sigma_i^2}{2 \sigma_j^2} + \frac{1}{2|V_i|}, \quad j \in V_i.
\]

Both the BS1 and BS2 heuristics are tested in Sections 3.4 and 3.5.

### 3.2.3 Adapted Consistent Appropriate Share rationing

When using CAS rationing we know after substituting (3.1) in (2.8) that

\[
L_i^j = \mu_{\text{req}(j)} + q_j O_i, \quad j \in V_i,
\]

where \( O_i \) is defined as in (3.2). This \( O_i \) is the so-called projected systemwide net inventory introduced by De Kok, Lagodimos & Seidel [1994]. It represents the amount of products which has to be divided over the end-stockpoints after allocating \( \mu_{\text{req}(j)} \) to each end-stockpoint \( j \). CAS rationing always allocates a fixed fraction \( q_j \) of this amount \( O_i \) to stockpoint \( j \). Since \( O_i \) may be negative an increase of \( q_j \) does not necessarily cause an increase of \( \beta_j \). When the systemwide projected inventory at time \( t \) is negative an increase of \( q_j \) means that the amount of stock allocated to end-stockpoint \( j \) decreases. However, when at time \( t \) the projected net inventory is positive an increase of \( q_j \) results in an increase
of $\hat{L}^j$. In order to get a consistent rationing policy Diks & De Kok [1996a] suggested to adapt the CAS rationing policy slightly, such that an increase of $q_j$ results in an increase of $\beta_j$. This is accomplished by rationing such that

$$\hat{L}^j = \mu_{\text{ech}(j)} + q_j O_j^* - q_j'(-O_j)^*, \quad j \in V_1,$$  \hspace{1cm} (3.9)

where $q_j'$ is a monotonously decreasing function with respect to $q_j$. Clearly, for $\{q_j\}_{j \in V_1}$ we require $\sum_{j \in V_1} q_j = 1$. In the numerical study of Sections 3.4 for sake of convenience we define

$$q_j' := \frac{1 - q_j}{|V_1| - 1}, \quad j \in V_1.$$  \hspace{1cm} (3.10)

After subsequently substituting this definition of $q_j'$ in (3.9), and substituting the result in (2.18) we obtain

$$\beta_j = 1 - \frac{E\left[\left(D_j^L - \mu_{\text{ech}(j)} - q_j O_j - (q_j - q_j')(-O_j)^*\right)^+\right]}{R\mu_j},$$

$$\beta_j = 1 + \frac{E\left[\left(D_j^L - \mu_{\text{ech}(j)} - q_j O_j - (q_j - q_j')(-O_j)^*\right)^-\right]}{R\mu_j}, \quad j \in V_1.$$  \hspace{1cm} (3.10)

Note that $\beta_j$ given by (3.10) very much resembles (3.2). Especially, when the systemwide projected inventory $O_j$ is nonnegative in most of the periods. In practice usually large service levels are required. Therefore, indeed, $O_j$ must be nonnegative in most periods, and thus the impact of the adaptation of the CAS rationing policy probably has minor effects on the performance. In order to determine the allocation fractions $\{q_j\}_{j \in V_1}$ and $S_1$ (given $\Delta_1$) we use a similar heuristic as CAS1. We refer to this heuristic as ACAS.

**ACAS:**

(i) Initialize $S_1$, and $\epsilon > 0$.

(ii) Use (3.10) to determine the allocation fraction $q_j$ for every end-stockpoint $j$.

(iii) If $\sum_{j \in V_1} q_j < 1 - \epsilon$ then decrease $S_1$ and return to step (ii).

(iv) If $\sum_{j \in V_1} q_j > 1 - \epsilon$ then increase $S_1$ and return to step (ii).

In step (ii) of ACAS we use a bisection procedure to determine $q_j$. Unlike step (ii) of CAS1, this always yields a unique solution.

### 3.3 Controlling an N-echelon system

In Section 3.2, we concentrated on heuristics for 2-echelon systems. In literature one seldom finds extensions to more general $N$-echelon systems, although, in practice large production and distribution networks are frequently encountered. Therefore, generalization of the heuristics of the previous section is needed. In this section we extend each heuristic of the previous section, except for the CAS1 and ACAS rationing policy. There are several reasons for this. First, it is unclear how to extend the CAS1 and ACAS policy to $N$-echelon systems straightforwardly. Second, the numerical results in Section 3.4.2 concerning 2-echelon models indicate that it is not worth while to extend both policies.

Section 3.3.1 describes the generalization of the CAS2 heuristic (as well as the CAS3 and CAS4 heuristics). This generalization is proposed by De Kok [1994]. Section 3.3.2 describes the generalization of both the BS1 and BS2 heuristic.
3.3.1 Consistent Appropriate Share Rationing

The generalization of the CAS2 heuristic and its adaptations (CAS3 and CAS4) is rather straightforward if we use a decomposition approach. We start with the determination of the control parameters at the end-stockpoints, and then work our way up through the system. When using this decomposition approach the control parameters of a stockpoint, \( i \) say, are determined \textit{given} the control parameters of stockpoints downstream of stockpoint \( i \). We do not alter already determined allocation fractions in contrast with the order-up-to-levels. So the generalization of the CAS2 heuristic consists of the following steps:

\textbf{CAS2 (and CAS3/CAS4):}

\( \text{(i). } n := 1. \)

\( \text{(ii). } \) Determine for every end-stockpoint \( j \) the order-up-to-level \( S'_{j} \) such that \( \beta_{j} = \beta'_{j} \), assuming \( \Delta_{i} \) would be infinity for every \( i \in M \). This \( S'_{j} \) can be determined from (2.18) after substituting \( \bar{L}_{j} = S'_{j}. \)

\( \text{(iii). } n := n + 1. \)

\( \text{(iv). } \) Consider a stockpoint \( i \in W_{n} \). Define for every \( j \in V_{i}, \)

\[
q_{j} := \frac{S'_{j} - \mu_{ech(j)}}{\sum_{i \in V_{i}} (S'_{i} - \mu_{ech(i)})}.
\]

\( \text{(v). } \) Determine for every stockpoint \( k \in E_{i} \) with \( j \in V_{i} \), the order-up-to-level required at stockpoint \( i \), denoted by \( S[j, k] \), such that \( \beta_{k} = \beta'_{k}. \)

\( \text{(vi). } \) Define

\[
S_{i} := \frac{\sum_{j \in V_{i}} \sum_{k \in E_{i}} S[j, k]}{|E_{i}|}.
\]

\( \text{(vii). } \) In case of CAS3 or CAS4 we adapt the allocation fraction as suggested in Section 3.2.1 (cf. step \( (v) \) and \( (vi) \)). Next, we return to step \( (v) \) (after adapting \( \delta \)) until \( \delta \) minimizes

\[
S'_{i, \text{max}} = \frac{S'_{i} - \mu_{ech(i)}}{S'_{i}}
\]

where \( S'_{i, \text{max}} := \max\{S[j]|j \in V_{i}\} \) and \( S'_{i, \text{min}} := \min\{S[j]|j \in V_{i}\}. \)

\( \text{(viii). } \) Execute steps \( (iv)-(vii) \) for every stockpoint \( i \in W_{n}. \)

\( \text{(ix). } \) if \( n < N \) then return to step \( (iii) \). Otherwise, the order-up-to-level of the most upstream stockpoint \( S_{1} \) is defined as \( S'_{1}. \) From \( S_{1} \) and the allocation fractions determined in step \( (iv) \) we determine all the downstream order-up-to-levels from

\[
S_{j} = \mu_{ech(j)} + q_{j} (S_{i} - \Delta_{i} - \sum_{i \in V_{i}} \mu_{ech(i)}), \quad j \in V_{i}.
\]

In Verrijdt & De Kok [1998], the CAS2 heuristic was developed for the case where \( \Delta_{i} = 0 \) for all intermediate stockpoints \( i \). De Kok [1994] extended these results to the case where also the intermediate stockpoints may keep stock on hand.
3.4. Numerical experiment for 2-echelon models

3.3.2 Balanced Stock rationing

Expressions for the allocation fractions become more complicated for these $N$-echelon systems, since it is cumbersome to determine $\Omega_j^N$ for a stockpoint $j \in V_l$. As a simple approximation Van der Heijden [1997b] proposes to assume that the variation in the echelon inventory position of stockpoint $j$ just after rationing has only minor effect on the allocation fractions. In that case we can determine the allocation fractions $\{q_j\}_{j \in V_l}$ as we did in Section 3.2.2, after making the following substitutions:

$$
\mu_1 \rightarrow \sum_{n \in E_j} \mu_n, \quad \sigma^2_1 \rightarrow \sum_{n \in E_j} \sigma^2_n,
$$

$$
\mu_j \rightarrow \sum_{n \in E_j} \gamma_n, \quad \sigma^2_j \rightarrow \sum_{n \in E_j} \gamma^2_n.
$$

(3.11)

So the BS1 heuristic is as follows

**BS1:**

(i) Determine for every stockpoint $i \in M$ the lower bounds for $\{q_j\}_{j \in V_l}$ by

$$
\bar{q}_j = \frac{\sigma^2_j}{2\sigma^2}, \quad j \in V_l.
$$

(ii) Determine for every stockpoint $i \in M$ the value $c_i$ of the equation below such that the allocation fraction sum up to one.

$$
\phi \left( \frac{\mu_{o_i}}{\sigma_{o_i}} \right) \phi \left( \frac{\mu_{o_j}}{\sigma_{o_j}} \right) T_i (2q_j \sigma^2_j - \sigma^2_j) = c_i, \quad j \in V_l.
$$

In each step of the bisection, the corresponding values for $\{q_j\}_{j \in V_l}$ are found by another bisection, where $q_j$ should be in the interval $(\bar{q}_j, 1)$.

(iii) $n := 1$.

(iv) Determine for every end-stockpoint $j$ the order-up-to-level $S_j$ such that $\beta_j = \beta^*_j$. This order-up-to-level is determined from (2.18) after substitution of (2.13).

(v) $n := n + 1$;

(vi) Determine for every stockpoint $i \in W_k$ the order-up-to-level $S_i$ by

$$
S_i = \sum_{n \in V_l} S_n + \Delta_i.
$$

(vii) If $n < N$ then return to step (v).

Again the BS2 heuristic is identical to the BS1 heuristic, except for step (ii). The BS2 heuristic defines the allocation fractions of stockpoint $i$ similar to (3.7):

$$
q_i = \frac{\sigma^2_i}{2\sigma^2} + \frac{1}{2|V_l|}, \quad j \in V_l.
$$

3.4 Numerical experiment for 2-echelon models

We extensively tested all rationing policies as described in Section 3.2 by comparing analytical results to simulation results. That is, we analyze the performance of five variants of CAS rationing and two variants of BS rationing. We use the difference between target fill rate and actual fill rate achieved by a particular rationing policy as a performance measure. One policy is considered to be more accurate
than the other if the mean absolute deviation from the target fill rate is smaller over all test runs. Also we consider the maximum deviation between actual and target fill rate as a measure of robustness. The experimental design for 2-echelon models is described in the next section. The numerical results are presented and discussed in Section 3.4.2.

3.4.1 Experimental design for 2-echelon models

In our experiment we test 2-echelon models, in which a central depot (denoted by index 1) supplies products to two so-called service groups. A service group consists of a number of end-stockpoints with the same service, demand and lead time characteristics. The number of end-stockpoints in both service groups is the same. To normalize time and quantities, we made the following choices for all test runs:

- the review period equals \( R = 1 \).
- the mean demand per time unit for each end-stockpoint \( A \) equals \( E[D_{A_i}] = 10 \).

Furthermore, the one-period demands of all stockpoints are independent. Since the downstream lead times are usually small, we take \( L_1 = 1 \) as lead time between central depot and each end-stockpoint \( j \in V_1 \) in all test runs. Eight other parameters are varied in our experiment. We chose two different values for each parameter (see Table 3.1), except for the central stock level. As discussed in Section 3.2, the amount of central stock is a result of the choice of the parameter \( \Delta_1 \). From equation (2.8) it can be shown that the amount of central stock equals \( E[(\Delta_1 - D_{L_1}^1)^+] \), so it is convenient to express \( \Delta_1 \) in the mean system demand during the lead time \( L_1 \), say \( \Delta_1 = \alpha_1 E[D_{L_1}^1] \) for some constant \( \alpha_1 \). We have relatively much central stock if \( \alpha_1 > 1 \) (say \( \alpha_1 = 1.2 \)), relatively little central stock if \( \alpha_1 < 1 \) (say \( \alpha_1 = 0.8 \)) and no central stock if \( \alpha_1 = 0 \). Using these three values of the constant \( \alpha_1 \), we determine the appropriate value of \( \Delta_1 \) for each case.

<table>
<thead>
<tr>
<th>parameter</th>
<th>description</th>
<th>values in test runs</th>
</tr>
</thead>
<tbody>
<tr>
<td>( E[D_A] )</td>
<td>Number of end-stockpoints per service group</td>
<td>1, 3</td>
</tr>
<tr>
<td>( E[D_B] )</td>
<td>The mean demand per period at an end-stockpoint in service group ( A )</td>
<td>10</td>
</tr>
<tr>
<td>( c[D_A] )</td>
<td>The coefficient of variation of demand per period at an end-stockpoint in service group ( A )</td>
<td>0.4, 0.8</td>
</tr>
<tr>
<td>( c[D_B] )</td>
<td>The coefficient of variation of demand per period at an end-stockpoint in service group ( B )</td>
<td>0.4, 0.8</td>
</tr>
<tr>
<td>( \beta_A )</td>
<td>The target fill rate at an end-stockpoint in service group ( A ) (%)</td>
<td>90, 99</td>
</tr>
<tr>
<td>( \beta_B )</td>
<td>The target fill rate at an end-stockpoint in service group ( B ) (%)</td>
<td>90, 99</td>
</tr>
<tr>
<td>( L_1 )</td>
<td>Lead time from external supplier to the central depot</td>
<td>1.3</td>
</tr>
<tr>
<td>( \alpha_1 )</td>
<td>Constant, describing the level of stock at the central depot ( \Delta_1 := \alpha_1 E[D_{L_1}^1] )</td>
<td>0, 0.8, 1.2</td>
</tr>
</tbody>
</table>

Table 3.1. Parameter values in the experiment with 2-echelon models.

We tested all possible parameter combinations, yielding \( 3 \times 27 = 384 \) cases per rationing policy. The performance of the rationing policies for each case is tested by an extensive simulation of 200,000 time periods to ensure high simulation accuracy.
3.4. Numerical experiment for 2-echelon models

3.4.2 Results for 2-echelon models

The performance of each rationing policy, the variants of CAS and BS rationing, is shown in Figures 3.1–3.4. Because a deviation from the target service level has usually more serious consequences in the case of a high target service level, we separately give the rationing policy performance for each fill rate level (see Figures 3.1 and 3.2). Further, Figures 3.3 and 3.4 show the performance of each rationing policy depending on the central stock level. Note that rationing policy CAS1 did not converge in two cases. These cases are removed from the figures for CAS1 only.

The overall results show that BS rationing performs better than CAS rationing with respect to both average performance and worst case performance. The BS1 rationing policy performs best, but the simple variant as suggested by Van Donkelaar [1996] is also better than all variants of CAS rationing. Because BS rationing aims to reduce imbalance, the deviation from target fill rate is less than for CAS rationing. Note that the mean physical stock in the system is approximately equal for all rationing policies. Over all cases, the mean physical stock varies between 3.37 weeks (BS2 rationing) and 3.43

![Figure 3.1. Mean absolute deviation of the target fill rate per target fill rate.](image1)

![Figure 3.2. Maximum absolute deviation of the target fill rate per target fill rate.](image2)
Figure 3.3. Mean absolute deviation of the target fill rate per central stock level.

Figure 3.4. Maximum absolute deviation of the target fill rate per central stock level.

weeks (CAS1 rationing).

It is remarkable that the so-called improved variants of CAS rationing do not perform better than the basic CAS rationing rule by De Kok [1990]. In some cases improvement is obtained indeed as is shown in Verrijdt & De Kok [1996] and De Kok, Lagodimos & Seidel [1994]. However, this extensive test shows that worsening occurs as well in some other cases. As an example, consider the following case. A stockless central depot ($\Delta_1 = 0$) supplies two service group consisting of one end-stockpoint each. The supply lead time to the central depot equals $L_1 = 3$. The characteristics per service group are shown in Table 3.2.

Table 3.2 shows that all 'improved' CAS rationing policies yield highly imbalanced results. As a consequence of a poor choice for the allocation fractions and the order-up-to-levels, the actual fill rate is too high for service group A and far too low for service group B. Apparently the approximate solution of the system of nonlinear equations (3.3) deviates strongly from the real solution or the service level is very sensitive to the value of the rationing parameters, but the rationing parameters are
3.4. Numerical experiment for 2-echelon models

<table>
<thead>
<tr>
<th>service group</th>
<th>$E[D]$</th>
<th>$c[D]$</th>
<th>$\beta_{\text{avg}}$</th>
<th>$\beta_{\text{CAS1}}$</th>
<th>$\beta_{\text{CAS2}}$</th>
<th>$\beta_{\text{CAS3}}$</th>
<th>$\beta_{\text{CAS4}}$</th>
<th>$\beta_{\text{ACAS}}$</th>
<th>$\beta_{\text{BS1}}$</th>
<th>$\beta_{\text{BS2}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>10</td>
<td>0.8</td>
<td>99%</td>
<td>100.0%</td>
<td>99.3%</td>
<td>99.9%</td>
<td>100.0%</td>
<td>99.9%</td>
<td>98.9%</td>
<td>99.4%</td>
</tr>
<tr>
<td>$B$</td>
<td>30</td>
<td>0.8</td>
<td>90%</td>
<td>82.0%</td>
<td>90.7%</td>
<td>82.6%</td>
<td>81.7%</td>
<td>83.2%</td>
<td>89.8%</td>
<td>88.8%</td>
</tr>
</tbody>
</table>

Table 3.2. An example where 'improved' CAS rationing is worse than basic CAS rationing.

...accidentally better. Note that also here BS rationing is better than basic CAS rationing.

Next, we consider the performance of the allocation rule depending on target fill rate and central stock level. Firstly, Figures 3.1 and 3.2 show that fortunately all rationing policies perform better for high service levels than for low service levels. Extreme deviations from target occur mainly for $\beta = 90\%$ and for some rationing policies only. In the second place, Figures 3.3 and 3.4 show that all rationing policies perform better in the presence of much central stock. This is not surprising, because central stock diminishes imbalance.

Finally, some comments on the computation times. The time to simulate one case varies between several minutes up to about 20 minutes CPU time for specific cases on a Pentium-75 PC, whereas the time required to calculate the rationing parameters for one case equals less than 1 second. Figure 3.5 depicts the computation time performance of each rationing policy. Because the number of stockpoints in the system has a large impact on the computation time of an instance, we separately give the computation time performance for each number of stockpoints in the system (denoted by $N$ in Figure 3.5). In this figure we depicted for each rationing policy and for each $N$, the average, minimum and maximum computation time of all instances. From this we conclude that the fastest variant of the CAS rationing policy is the CAS2 variant. A little bit more computation time is required for CAS3 and CAS4, which approximately require the same amount of time. Note that CAS1 and ACAS require the most computation time, far more than both BS1 and BS2. As can be expected the BS2 rationing policy requires

<table>
<thead>
<tr>
<th>Policy</th>
<th>$N$</th>
<th>Average</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAS1</td>
<td>3</td>
<td>0.096</td>
<td>0.000</td>
<td>0.283</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.764</td>
<td>0.266</td>
<td>2.967</td>
</tr>
<tr>
<td>CAS2</td>
<td>3</td>
<td>0.011</td>
<td>0.000</td>
<td>0.834</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.051</td>
<td>0.000</td>
<td>0.167</td>
</tr>
<tr>
<td>CAS3</td>
<td>3</td>
<td>0.025</td>
<td>0.000</td>
<td>0.117</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.093</td>
<td>0.017</td>
<td>0.767</td>
</tr>
<tr>
<td>CAS4</td>
<td>3</td>
<td>0.034</td>
<td>0.000</td>
<td>0.233</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.116</td>
<td>0.016</td>
<td>0.850</td>
</tr>
<tr>
<td>ACAS</td>
<td>3</td>
<td>0.093</td>
<td>0.016</td>
<td>0.217</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.762</td>
<td>0.250</td>
<td>2.000</td>
</tr>
<tr>
<td>BS1</td>
<td>3</td>
<td>0.009</td>
<td>0.000</td>
<td>0.033</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.031</td>
<td>0.000</td>
<td>0.100</td>
</tr>
<tr>
<td>BS2</td>
<td>3</td>
<td>0.006</td>
<td>0.000</td>
<td>0.033</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.023</td>
<td>0.000</td>
<td>0.100</td>
</tr>
</tbody>
</table>

Figure 3.5. The CPU times of all 384 instances of a 2-echelon system on a SPARC station 5 (in seconds), where $N$ denotes the number of stockpoints in the system.
less computation time than the BS1 rationing policy.

3.5 Numerical experiment for 3-echelon models

In this section, we discuss the design and results of an experiment with 3-echelon models. We analyze only three variants of CAS rationing for the following reasons:

- Extension of rationing policy ACAS to a 3-echelon context is not straightforward. In principle it is possible, but the numerical results of the experiment with 2-echelon models shows that this is not worth while.

- The CAS1 allocation rule is similar to CAS3 and CAS4, because all these rules try to find an exact solution of the nonlinear system of equations (3.3). Because CAS1 rationing does not perform better than CAS3 and CAS4 in the 2-echelon experiment, it does not seem to be worth while to extend this approach to a 3-echelon setting as well.

Hence, we analyze both variants of BS rationing and only three variants of CAS rationing: CAS2, CAS3 and CAS4. The experimental design for 3-echelon models is described in the next subsection. The numerical results are presented and discussed in Section 3.5.2.

3.5.1 Experimental design for 3-echelon models

In our experiment we test 3-echelon models in which a central depot (denoted by index 1) supplies products to two so-called echelon groups (see Figure 3.6). An echelon group consists of a number of intermediate stockpoints that each deliver products to two service groups. Each service group consists of an equal number of identical end-stockpoints, but two service groups may be different.

When designing the experiment, some attention should be given to the values of $\Delta_i$, defining the maximum physical stock level of a stockpoint $j \in V_i$. From equation (2.12) it can be shown that the

![Figure 3.6. Echelon groups and service groups.](image-url)
3.5. Numerical experiment for 3-echelon models

mean amount of physical stock of an intermediate stockpoint \( j \in V_1 \) equals \( E[(\Delta_j - D_{j} - q_j(D_{j} - \Delta_j)^{+})^{+}] \), so it is convenient to define \( \Delta_j \) as

\[
\Delta_j := a_j \left( E[D_{j}^{+}] + q_j E[(D_{j} - \Delta_j)^{+}]^{-1} \right), \quad j \in V_1.
\] (3.12)

Because the allocation fractions \( \{q_j\}_{j \in V_1} \) are not known on input for a specific rationing policy, we plug in the approximation (3.7) of Van Dorselaar [1996]. Now, by using an appropriate choice for \( a_j \) in our experiment we obtain a reasonable value for \( \Delta_j \).

By definition we take the following parameters fixed for all test runs:

- the review period equals \( R = 1 \).
- the mean demand per period for each end-stockpoint in service group 1 within echelon group 1 equals 10.
- the lead time to every end-stockpoint \( k \) equal \( L_k = 1 \) (independent of the service- and echelon-group).

Further we impose the following restrictions within the experiment to keep the number of test runs within reasonable limits:

- the number of intermediate stockpoints is the same for both echelon groups in a single test run.
- the number of end-stockpoints per service group is identical for all service groups in a single test run.
- the lead time between central depot and each intermediate stockpoint \( j \in V_1 \) is the same.
- The value \( a_j \) is the same for every intermediate stockpoint \( j \in V_1 \). Hence from (3.12) also \( \Delta_j \) is the same for every intermediate stockpoint \( j \in V_1 \).

For the demand and service characteristics at each end-stockpoint we take the following values:

- the mean demand at a stockpoint in service group \( m \) of echelon group \( l \) equals \( E[D_{m}] = 10 \) or \( 30 \) (except \( E[D_{m}] := 10 \) by definition).
- the coefficient of variation of the demand at a stockpoint in service group \( m \) of echelon group \( l \) equals \( c[D_{m}] = 0.4 \) or \( 0.8 \).
- the target fill rate at a stockpoint in service group \( m \) of echelon group \( l \) equals \( \beta_{m} = 90\% \) or \( 99\% \).

When the set of experimental runs is carefully chosen, we only need 87 parameter combinations to analyze the 11 demand- and service parameters, instead of the \( 2^{11} = 2048 \) demand- and service parameter combinations for the 3-echelon experiment. Before we show how to obtain these 87 parameter combinations, we define the following parameter levels:

- category \( E[D_{m}] \): L(ow) = 10 and H(igh) = 30.
- category \( c[D_{m}] \): L(ow) = 0.4 and H(igh) = 0.8.
- category \( \beta_{m} \): L(ow) = 90\% and H(igh) = 99\%.

We use two stages to determine the set of parameter combinations to be used in the 3-echelon experiment. First, we state which target fill rate combinations are useful, given the demand characteristics within service groups and within echelon groups. Then, we select combinations of demand characteristics and combine these with all target fill rate combinations as defined in the first step.

**Step 1:** Selection of the combinations of the four target fill rate parameters \( \beta_{m} \).

We distinguish between the following situations:

- both echelon groups are the same.
the service groups within an echelon group are the same.

If the echelon groups are the same, we can skip those models that are obtained by interchanging the target service levels between echelon groups. If the service groups within an echelon group are the same, we skip those models that are obtained by interchanging the target service levels between service groups within the same echelon group. Therefore we only need the set of target fill rate combinations as shown in Table 3.3.

<table>
<thead>
<tr>
<th>Fill rate combinations $(\beta_{11}, \beta_{12})$ and $(\beta_{21}, \beta_{22})$</th>
<th>Echelon groups are the same</th>
<th>Echelon groups are different</th>
</tr>
</thead>
<tbody>
<tr>
<td>Service groups within an echelon group are the same</td>
<td>$(L, L)$ and $(L, L)$</td>
<td>$(L, L)$ and $(L, L)$</td>
</tr>
<tr>
<td></td>
<td>$(L, L)$ and $(H, H)$</td>
<td>$(L, L)$ and $(H, H)$</td>
</tr>
<tr>
<td></td>
<td>$(H, H)$ and $(H, H)$</td>
<td>$(H, H)$ and $(L, L)$</td>
</tr>
<tr>
<td></td>
<td>$(L, H)$ and $(L, H)$</td>
<td>$(L, H)$ and $(L, H)$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Service groups within an echelon group are different</th>
<th>$(L, L)$ and $(L, L)$</th>
<th>$(L, L)$ and $(L, L)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$(L, L)$ and $(H, H)$</td>
<td>$(L, L)$ and $(H, H)$</td>
</tr>
<tr>
<td></td>
<td>$(H, H)$ and $(H, H)$</td>
<td>$(H, H)$ and $(L, L)$</td>
</tr>
<tr>
<td></td>
<td>$(L, H)$ and $(L, H)$</td>
<td>$(L, H)$ and $(L, H)$</td>
</tr>
<tr>
<td></td>
<td>$(H, L)$ and $(H, L)$</td>
<td>$(H, L)$ and $(H, L)$</td>
</tr>
</tbody>
</table>

Table 3.3. Combinations of fill rates given the demand characteristics.

Step 2: Selection of the combinations of the seven demand characteristics $E[D_{1n}]$ (i ≠ 1 or m ≠ 1) and $c[D_{1n}]$.

For the mean demand per end-stockpoint, we select the following combinations:

(i). all end-stockpoints have low mean demand (all stockpoints with high mean demand give similar results, this is a matter of scaling).

(ii). the end-stockpoints in echelon group 1 all face low mean demand, while the end-stockpoints in echelon group 2 all face high mean demand (interchange of echelon groups yields similar models then).

(iii). the mean demand levels in both echelon groups are the same, but service group 1 faces high mean demand and service group 2 faces low mean demand (interchange of service groups within the same echelon group yields similar models then).

For each set of values for the mean demand, we choose all combinations of demand coefficients of variation, but we eliminate those models that can be transformed to each other by interchanging end-stockpoints within and/or between service groups. As a result, we obtain the parameter sets depicted in Table 3.4. This table shows 15 combinations of demand parameters and the relevant number of fill rate combinations. In this way, we end up with 87 parameter combinations instead of $2^{15} = 2048$.

For the remaining parameters we make the following choices:

- two values for $\delta_1$, defined by $\Delta_1 := a_1 E[D_{1n}]$ for $a_1 = 0$ and $a_1 = 1.2$.
- two values for $\delta_j$, defined by $\Delta_j := a_j E[D_{j,n}] + a_j E((D_{j,n} - \Delta_1)^+)$ for $a_j = 0$ and $a_j = 1.2$.
- number of stockpoints:
3.5. Numerical experiment for 3-echelon models

<table>
<thead>
<tr>
<th>(μ₁₁, μ₁₂) and (μ₂₁, μ₂₂)</th>
<th>(c₁₁, c₁₂) and (c₂₁, c₂₂)</th>
<th>echelon groups</th>
<th>service groups within echelon groups</th>
<th>number of fill rate combinations</th>
</tr>
</thead>
<tbody>
<tr>
<td>(L, L) and (L, L)</td>
<td>(L, L) and (L, L)</td>
<td>same</td>
<td>same</td>
<td>4</td>
</tr>
<tr>
<td>(L, L) and (H, H)</td>
<td>(L, L) and (H, H)</td>
<td>different</td>
<td>same</td>
<td>5</td>
</tr>
<tr>
<td>(L, H) and (H, H)</td>
<td>(L, H) and (H, H)</td>
<td>same</td>
<td>same</td>
<td>4</td>
</tr>
<tr>
<td>(L, H) and (L, H)</td>
<td>(L, H) and (L, H)</td>
<td>different</td>
<td>different</td>
<td>6</td>
</tr>
</tbody>
</table>

| (L, L) and (H, H)     | (L, L) and (L, L)      | different     | same                                | 5                             |
| (L, L) and (H, H)     | (L, L) and (L, L)      | different     | same                                | 5                             |
| (H, H) and (H, H)     | (H, H) and (L, L)      | different     | same                                | 5                             |
| (L, H) and (L, H)     | (L, H) and (L, H)      | different     | different                           | 8                             |

| (L, H) and (L, H)     | (L, L) and (L, L)      | same          | different                           | 6                             |
| (L, L) and (H, H)     | (L, H) and (H, H)      | different     | different                           | 8                             |
| (H, H) and (H, H)     | (H, H) and (H, H)      | same          | different                           | 6                             |
| (L, H) and (L, H)     | (L, H) and (L, H)      | same          | different                           | 6                             |
| (L, L) and (H, H)     | (L, H) and (L, H)      | different     | different                           | 8                             |

Table 3.4. Combinations of E[Dᵹ] and c[Dᵹ].

(i) one end-stockpoint per service group and one intermediate stockpoint per echelon group.

(ii) three end-stockpoints per service group and one intermediate stockpoint per echelon group.

(iii) three end-stockpoints per service group and two intermediate stockpoints per echelon group.

- lead times:
  (i) $L_1 = 1$ and $L_f = 1$ for every stockpoint $j \in V_1$.
  (ii) $L_1 = 3$ and $L_f = 1$ for every stockpoint $j \in V_1$.
  (iii) $L_1 = 3$ and $L_f = 4$ for every stockpoint $j \in V_1$.

In total we now have $7^3 \times 2^2 \times 2^3 = 3132$ test runs for each rationing policy. This is still a large amount of numerical effort, but it is acceptable. The performance of the rationing policies for each case is tested by a simulation of 100,000 time periods.

3.5.2 Results for 3-echelon models

The performance of each rationing policy, the three variants of CAS rationing and the two variants of BS rationing, is shown in Figures 3.7–3.10 below. Again we give separate results per target fill rate (Figures 3.7 and 3.8) and per upstream stock level (Figures 3.9 and 3.10).

The results of the 3-echelon experiment are a logical extension of the results of the 2-echelon experiment. Again, BS rationing performs better than CAS rationing and the original BS rationing performs best. It is remarkable that the performance of the various rationing policies is not worse than for 2-echelon models. Apparently there is no accumulation of approximation errors. For CAS rationing, the errors seem even to compensate each other slightly. The performance of BS rationing is however slightly worse than for 2-echelon models, probably because of the fact that an additional approximation is made when establishing the rationing parameters. The effect of central and intermediate stocks
Service level constraints: fixed lead time case

**Figure 3.7.** Mean absolute deviation of the target fill rate per target fill rate.

**Figure 3.8.** Maximum absolute deviation of the target fill rate per target fill rate.

is neglected and only taken into account when calculating the order-up-to levels. Also it is remarkable that the so-called improved variants of CAS rationing do not perform better than the basic CAS allocation rule by De Kok [1990]. Note that also here the mean physical stock in the system is approximately equal for all rationing policies. Overall, the mean physical stock varies between 5.24 times the mean period review demand (CAS2 rationing) and 5.36 times the mean period review demand (BS1 rationing).

Next, we consider the performance of the allocation rule depending on target fill rate and central stock level. Figures 3.7 and 3.8 show that fortunately all rationing policies perform better for high service levels than for low service levels. Figures 3.9 and 3.10 show that all rationing policies perform better in the presence of much upstream stock, because imbalance is reduced.

Finally, the computation time performance of each rationing policy is depicted in Figure 3.11. Again, we separately give the computation time performance for each number of stockpoints in the system (denoted by \( N \) in Figure 3.11). In this figure, we depicted for each rationing policy and for
each \( N \), the average, minimum and maximum computation time of all instances. From this we conclude that the fastest variant of the CAS rationing policy is the CAS2 variant. The CAS3 and CAS4 require approximately the same amount of computation time, and for some instances this may even be around 13 seconds! Far less computations time is needed for the BS rationing policy. The BS2 policy requires the least computation time.

3.6 Conclusions

In this chapter, we have extensively examined various rationing policies for the control of the divergent \( N \)-echelon model as presented in Chapter 2. Both models with and without intermediate stock are analyzed. Where necessary, the policies as available in the literature are extended to a general \( N \)-echelon context with intermediate stocks.

The most important result is that BS rationing performs better than CAS rationing, both on aver-
Table 3.11. The CPU times of all 3132 instances of a 3-echelon system on a SPARC station 5 (in seconds), where $N$ denotes the number of stockpoints in the system.

<table>
<thead>
<tr>
<th>Policy</th>
<th>$N$</th>
<th>Average</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAS2</td>
<td>6</td>
<td>0.057</td>
<td>0.016</td>
<td>0.167</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.446</td>
<td>0.100</td>
<td>2.183</td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>1.213</td>
<td>0.200</td>
<td>6.550</td>
</tr>
<tr>
<td>CAS3</td>
<td>6</td>
<td>0.110</td>
<td>0.033</td>
<td>0.434</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.879</td>
<td>0.200</td>
<td>4.500</td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>2.385</td>
<td>0.450</td>
<td>12.90</td>
</tr>
<tr>
<td>CAS4</td>
<td>6</td>
<td>0.121</td>
<td>0.033</td>
<td>0.733</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.898</td>
<td>0.167</td>
<td>4.350</td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>2.439</td>
<td>0.383</td>
<td>13.00</td>
</tr>
<tr>
<td>BS1</td>
<td>6</td>
<td>0.025</td>
<td>0.000</td>
<td>0.100</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.082</td>
<td>0.016</td>
<td>0.400</td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>0.173</td>
<td>0.066</td>
<td>0.767</td>
</tr>
<tr>
<td>BS2</td>
<td>6</td>
<td>0.017</td>
<td>0.000</td>
<td>0.084</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.064</td>
<td>0.000</td>
<td>0.383</td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>0.137</td>
<td>0.033</td>
<td>0.817</td>
</tr>
</tbody>
</table>

age and worst case. Within the class of CAS rationing policies, it is remarkable that the most simple approach of De Kok [1990] performs best. The variants of CAS rationing which have been derived in the past years do not appear to be better than the simple, original approach by De Kok [1990]. The most serious errors occur in cases of high imbalance, especially

- for relatively low service levels,
- in situations with little or no central or intermediate stocks.

Overall, the original BS rationing policy (denoted by BS1) is the best rationing policy in the test. Although the BS1 rationing policy is not very difficult to implement, the simple BS rationing variant of Van Donselaar (denoted by BS2) is even more simple and can easily be used in spreadsheet applications. Hence from a practical point of view we suggest to use BS2 rationing. Another advantage of BS rationing is the fact that the determination of the allocation fractions is decoupled from the determination of the order-up-to levels. Because of this, BS rationing can probably be used more easily for model extensions, such as the introduction of random lead times, order points or lot sizing. The extension to random lead times is addressed in Chapter 4. The other extensions are subjects for further research.
4

Service level constraints: random lead times

4.1 Introduction

In Chapter 2, we formulated a model for a divergent $N$-echelon system under echelon stock control. In Chapter 3, we developed several algorithms to determine the control parameters such that customer service level requirements are satisfied. A prerequisite for applying these algorithms is that every lead time is fixed. In real-world supply chains this assumption is reasonable for the lead times to end-stockpoints, since they represent transportation times in most cases. The assumption is less reasonable for lead times to intermediate stockpoints, since they often represent stochastic manufacturing throughput times. These manufacturing throughput times are stochastic due to, e.g., capacity constraints, machine-breakdowns, and batching decisions. We have to be aware that multi-echelon systems represent a supply chain for a set of products, but which share resources together with other sets of products. This sharing of resources is the major cause for randomness of lead times.

So far no analysis is available for multi-echelon systems with echelon stock policies and random lead times with the exception of the paper of Zipkin [1991] who deals with continuous review $(S-1,S)$ models. Therefore, in this chapter we extend the analysis performed in Chapter 3 by incorporating random lead times. Seidel & De Kok [1990] noted that such an extension is straightforward for the lead times to end-stockpoints. However, the more practical situation of random lead times to intermediate stockpoints cannot be incorporated straightforwardly. When we deal with random lead times in the model as formulated in Chapter 2 we have to be more specific about the replenishment mechanism of the order-up-to-policies. The most upstream stockpoint follows a periodic review order-up-to-policy, i.e., every $R$ periods this stockpoint places an order at an external supplier to raise the echelon inventory position to its order-up-to-level. However, a more downstream stockpoint raises its echelon inventory position to the order-up-to-level upon arrival of a replenishment at its successor. Hence, the ordering decisions are taken at random points in time. The rate at which ordering decisions are taken is obviously equal to $1/R$. We assume that the lead times to a stockpoint are identically distributed random variables, but that subsequent replenishment orders do not cross. This implies that subsequent lead times are correlated. In this chapter, we model this (auto)correlation assuming that the lead times constitute an AR(1) process. Simulation experiments show that this modeling assumption yields good
approximations for performance measures even if the lead time process is not AR(1). Finally, when a stockpoint has insufficient stock to satisfy the orders of its successors we apply BS2 rationing. In this chapter, we restrict ourselves to this rationing policy for several reasons. First, the results in Chapter 3 indicate that BS rationing outperforms CAS rationing. Second, in BS rationing the determination of the order-up-to-levels is decoupled from the determination of the allocation fractions. Third, the variant of Van Den Bergen (BS2) yields a simple and explicit expression. Finally, like in Chapter 3 we assume that the so-called maximum stock levels at intermediate stockpoints (i.e., \( N_i \) for \( i \in M \)) are known before applying the algorithm. The maximum stock level is the difference between the order-up-to-level at the stockpoint and the sum of the order-up-to-levels at its successors.

In this chapter, we concentrate on the determination of the control parameters, i.e., the order-up-to-levels and the parameters of the BS rationing policy, such that the customer service level constraints are satisfied. Like the previous chapter the service level under consideration is the fill rate, since this is often used in practice. We develop a heuristic algorithm to compute the control parameters. We like to emphasize that this algorithm can easily be extended to other types of service level constraints. Extensive testing of the algorithm (using discrete event simulation) indicate that the algorithm performs quite well. Thereupon we use the algorithms to investigate whether the incorporation of autocorrelation of the lead times into the analysis is important for the accuracy of the algorithm. Furthermore, we investigate the managerial issue of the impact of randomness of lead times on stock investments.

This chapter is organized as follows. First, in Section 4.2, we address several methods to deal with the no-crossing random lead times. In the Sections 4.3–4.5, we derive the analysis of the divergent multi-echelon system with random lead times. In Section 4.3, we start with the 2-echelon model without central stock. The analysis is extended to 2-echelon systems with central stocks in Section 4.4, and to general \( N \)-echelon systems in Section 4.5. A summary of the algorithm is given in Section 4.6. We validate our heuristic algorithm by comparison to simulation results. In order to simulate the multi-echelon inventory system we need to generate the lead times in each stockpoint such that subsequent orders do not cross. For that purpose we develop a simple model in Section 4.7. Next, in Section 4.8, we validate the performance of the algorithm both for 2-echelon and for 3-echelon systems. Some sensitivity analysis is presented in Section 4.9. Finally, we give our conclusions in Section 4.10.

4.2 Modeling random lead times

The fact that the lead time in an inventory system is rarely constant is widely recognized. It is important to account for lead time variability when analyzing multi-echelon inventory systems, since ignoring it may lead to a high cost and a poor performance (cf. Gross & Soriani [1969]). An important issue when incorporating stochastic lead times is whether successive lead times are independent, since in that case orders can cross. A cross over occurs when a quantity that was ordered in a latter period arrives before one that was ordered in an earlier period. In practice successive lead times are usually dependent (cf. Yano [1987]), therefore the "no crossing" assumption needs to be incorporated in the model. Various modeling approaches have been used to circumvent this complication. When the interval between successive orders is large enough, the probability of cross over is negligible and can be omitted (cf. Hadley & Whitin [1963]). Friedman [1984] circumvents the problem by assuming that there is always only one order outstanding. In Sphicas [1982] and Sphicas & Nash [1984] the problem of cross overs is eliminated by assuming that unit demands are non-interchangeable. It is clear that these approaches are not generally valid.

To our knowledge Kaplan [1970] was the first which explicitly incorporated the no crossing assumption. A mechanism of the arrival of orders is used, to ensure that orders never cross, while in
4.3. Analysis of 2-echelon systems without central stock

General lead times will be dependent. Kaplan (see also Nahmias [1979] and Anupindi, Morton & Pentico [1996]) developed this mechanism for a periodic review system, which places an order at the beginning of every period. This mechanism is based on a stationary, discrete-time Markov process \( \{U(t)\} \) with nonnegative integer states, representing the number of outstanding orders after a possible delivery at time \( t \), with

\[
Pr(U(t + 1) = j | U(t) = i) = \begin{cases} 
\pi_j & j = 0, 1, \ldots, i \\
1 - \sum_{k=0}^{i-1} \pi_k & j = i + 1.
\end{cases}
\] (4.1)

The situation where at time \( t + 1 \) no delivery is made corresponds with \( j = i + 1 \). Note that (4.1) assumes that the probability of arrival of an outstanding order is independent of the number of outstanding orders. Nahmias [1979] showed that for this system:

\[
U(t + 1) = \min(U'(t + 1), U(t) + 1),
\]

where \( \{U'(t)\} \) is a sequence of i.i.d. random variables with \( \pi_j := Pr(U'(t) = j) \). Echardt [1984] derives the distribution of the lead time \( L \), for this specification, namely

\[
Pr(L = i) = \begin{cases} 
\pi_0 & i = 0 \\
(1 - \pi_0)(1 - \pi_0 - \pi_1) \cdots (1 - \sum_{j=0}^{i-1} \pi_j) \prod_{j=0}^{i-1} \pi_j & i = 1, 2, \ldots, m,
\end{cases}
\]

where \( m \) denotes the maximum lead time. Like Zipkin [1986], we think it is hard to imagine a physical process giving rise to this scenario. Moreover, many parameters need to be estimated from data, and furthermore, Zipkin has shown that not any lead time distribution (e.g., the geometric distribution) can be attained. Zipkin extended the arrival mechanism of Kaplan [1970] by relaxing (4.1) (along with the discrete time assumption). Furthermore, it is not clear how to choose an order arrival process that fits certain lead time characteristics (e.g., the mean and variance as observed in practice). However, still the lead time of an order is independent of the time it was placed. In our opinion this is not very realistic in practical situations. E.g., in production systems lead time variability typically results from a shortage of capacity during certain periods. Hence, in these systems the time of placing an order has a large impact on the lead time attained. For this purpose we develop a simple model in Section 4.7 which incorporates this aspect. Furthermore, it is a more direct approach to model the lead time process (which does not result from the arrival mechanism).

Hecess & De Klein [1995] introduced another model to deal with stochastic lead times. They distinguish between a start-up and a follow-up order. If at an order epoch \( t \) all orders placed previously have arrived or will arrive before \( t + l_0 \) (where \( l_0 \) denotes the minimal lead time) the order is a start-up order. For a start-up order \( Pr(L = l_j) := \pi_j \), where the discrete lead time \( L \) is defined at the points \( l_0, l_1, \ldots, l_{m-1} \) (in ascending order). On the other hand when at an order epoch \( t \) outstanding orders will not arrive before \( t + l_0 \), then this order is a follow-up order. A follow-up order will be delivered simultaneously with the start-up order not yet arrived. Note that this model incorporates the aspect of the lead time depending on the time the order is placed. However, their model requires a large amount of information (i.e., all the possible lead times \( l_j \) of the start-up order and its probabilities \( \pi_j \)) whereas our model (see Section 4.7) only requires the first two moments of the lead time.

4.3 Analysis of 2-echelon systems without central stock

In this section, we derive mathematical expressions to calculate the system control parameters and the performance measures for a 2-echelon system without central stock. We consider one central depot
(having index 1) supplying a number of end-stockpoints \( j \in V_1 \). In Section 4.4 we add the option that the central depot is allowed to hold stock. Extension of the results to general \( N \)-echelon systems is discussed in Section 4.5.

In Section 4.3.1, we describe the calculation of the order-up-to-levels \( \{S_j\}_{j \in V_1} \) by assuming that the allocation fractions \( \{q_j\}_{j \in V_1} \) are known. It appears that some aggregate demand characteristics are required for the calculations. These are derived in Section 4.3.2. In Section 4.3.3, we derive an expression for the allocation fractions \( \{q_j\}_{j \in V_1} \), such that a surrogate expression for the mean imbalance at the central depot is minimized. It appears that these allocation fractions are only based on demand and lead time characteristics, and hence are independent of fill rate requirements and order-up-to-levels. This is attractive, both from an analytical and a computational point of view. Together, Sections 4.3.1–4.3.3 give the control parameters \( \{S_j\}_{j \in V_1} \) and \( \{q_j\}_{j \in V_1} \). Finally, we show how to approximate the mean physical stock in Section 4.3.4.

### 4.3.1 Determination of the order-up-to-levels

In this section, we derive how to compute the fill rate \( \beta_j \) of an end-stockpoint \( j \in V_1 \) given all the control parameters. Before addressing the derivation, we introduce \( L_{11} \) and \( L_{12} \), which denotes two successive lead times of stockpoint \( t \). We need this notation in order to analyze the impact of the interdependency of two successive lead times (due to the no crossing assumption) on the behavior of the stock levels.

Without loss of generality, suppose that at time 0 the central depot issues a replenishment order that raises the echelon inventory position to the level \( S_t \). This order arrives after a random lead time, \( L_{11} \) say. Then every end-stockpoint \( j \) inspects its echelon inventory position and places a replenishment order at the depot. Since we consider a stockless depot (i.e., \( \Lambda_t = 0 \)) no end-stockpoint is able to raise the echelon inventory position to its order-up-to-level. Hence, the complete echelon stock of the depot is allocated to the end-stockpoints. Using allocation rule (2.7), and substituting \( \Lambda_t = 0 \) in equation (2.8) yields

\[
J^j_{t_{11}} = S_j - q_j D^j_{t_{11}}, \quad j \in V_1.
\]

The products which are allocated to end-stockpoint \( j \) at time \( L_{11} \), arrive at time \( L_{11} + L_{1j} \). Then from the sample path technique it follows that

\[
J^j_{t_{11} + L_{1j}} = S_j - q_j D^j_{t_{11} + L_{1j}} - D^j_{t_{1j} + L_{1j} + L_{1j}}, \quad j \in V_1.
\]

The next order arrives at end-stockpoint \( j \) at time \( R + L_{11} + L_{12} \). Just prior to this arrival, the echelon stock of end-stockpoint \( j \) equals

\[
J^j_{t_{11} + L_{1j} + L_{1j}} = J^j_{t_{11} + L_{1j}} - D^j_{t_{1j} + L_{1j} + L_{1j}}, \quad j \in V_1.
\]

Now consider the replenishment cycle \( (L_{11} + L_{1j}, R + L_{12} + L_{1j}) \) of end-stockpoint \( j \). We have that

Amount backordered (begin cycle) \( = (D^j_{t_{11} + L_{1j}} + q_j D^j_{t_{1j}} - S_j)^+ \),

Amount backordered (end cycle) \( = (D^j_{t_{1j} + L_{1j} + L_{1j}} + q_j D^j_{t_{1j}} - S_j)^+ \).

Now the fill rate \( \beta_j \) is computed by one minus the ratio of the expected amount backordered at end-stockpoint \( j \) during this replenishment cycle and the expected demand at this end-stockpoint during the cycle. Hence,

\[
\beta_j = 1 - \frac{E \left[ (D^j_{t_{11} + L_{1j}} + q_j D^j_{t_{1j}} - S_j)^+ - (D^j_{t_{1j} + L_{1j} + L_{1j}} + q_j D^j_{t_{1j}} - S_j)^+ \right]}{R \mu_j}
\]

(4.2)
4.3 Analysis of 2-echelon systems without central stock

Since we consider stationary demand equation (4.2) is simplified to

$$
\beta_j = 1 - \frac{E \left[ \left( (D'_{L_j} + R_{j}) + q_j D'_{L_{j-1}} - S_j \right)^+ - (D'_{L_j} + q_j D'_{L_{j-1}} - S_j)^- \right]}{R_{\mu_j}},
$$

(4.3)

where $R_{\mu_j} := R + L_{12} - L_{11}$ denotes the interarrival time of two subsequent orders of stockpoint 1. Since each end-stockpoint $j$ inspect its echelon inventory position immediately upon an order arrival at stockpoint 1, the review period of end-stockpoint $j$ is identically distributed as $R_{\mu_j}$ (random variable with mean $R$). To compute $\beta_j$ from (4.3) we fit mixed Erlang distribution to the first two moments of the stochastic components in the numerator, $X_{ij} := D'_{L_{j+i}} + q_j D'_{L_{j}}$ and $Y_{ij} := D'_{L_{j+i}} + q_j D'_{L_{j}}$. Then, the numerator is computed by using (2.20). For the details on the fitting procedure and the computation of the numerator we refer to Section 2.4.1.

So, given all the control parameters we are able to compute the fill rate $\beta_j$ attained at each end-stockpoint $j$. While the purpose of this chapter is to determine the control parameters such that the fill rate attained at each end-stockpoint $j$ equals its target value $\beta^*_j$. To accomplish this we proceed as follows. Suppose the allocation fractions are known. By performing bisection on $S_j$, we determine the order-up-to-levels $S_j$ satisfying $\beta_j = \beta^*_j$. In this way the order-up-to-levels are calculated one-by-one for each end-stockpoint $j$. Next, the order-up-to-level of the central depot results from

$$
S_1 = \sum_{i \in V_1} S_i.
$$

Besides the allocation fractions $\{q_j\}_{j \in V_1}$ we also need to know the first two moments of the aforementioned stochastic components $X_{ij}$ and $Y_{ij}$ in order to determine the order-up-to-levels as described above. First, we determine the first two moments of $X_{ij}$ and $Y_{ij}$. We find that

$$
E[X_{ij}] = E[D'_{L_{j+i}}] + q_j E[D'_{L_{j}}],
$$

(4.4a)

$$
\text{var}(X_{ij}) = \text{var}(D'_{L_{j+i}}) + q_j^2 \text{var}(D'_{L_{j}}) + 2q_j \text{cov}(D'_{L_{j+i}}, D'_{L_{j}}),
$$

(4.4b)

$$
E[Y_{ij}] = E[D'_{L_{j+i}}] + q_j E[D'_{L_{j}}],
$$

(4.5a)

$$
\text{var}(Y_{ij}) = \text{var}(D'_{L_{j+i}}) + q_j^2 \text{var}(D'_{L_{j}}) + 2q_j \text{cov}(D'_{L_{j+i}}, D'_{L_{j}}).
$$

(4.5b)

Now the problem is the derivation of the mean and variance of $X_{ij}$ and $Y_{ij}$. We see that we need several demand characteristics to evaluate (4.4) and (4.5), among which two covariances. In Section 4.3.2 we show how these characteristics are obtained. Next, in Section 4.3.3 we determine the allocation fractions (independently from the order-up-to-levels).

4.3.2 Calculation of the demand characteristics

From Section 4.3.1, we see that we need the following demand characteristics to calculate the control parameters $(S_j)_{j \in V_1}$:

(i). the mean and variance of $D'_{L_j}$ for each end-stockpoint $j \in V_1$.

(ii). the mean and variance of $D'_{L_1}$.

(iii). the mean and variance of $D'_{L_{j+i}}$ for each end-stockpoint $j \in V_1$.

(iv). $\text{cov}(D'_{L_{j+i}}, D'_{L_j})$ for each end-stockpoint $j \in V_1$.

(v). $\text{cov}(D'_{L_{j}}, D'_{L_{j-1}})$ for each end-stockpoint $j \in V_1$. 


Below, we derive expressions for the five demand characteristics.

First, we derive the mean and variance of $D_{t_j}$ for each end-stockpoint $j$ by conditioning on $L_j$ (cf. Silver & Peterson [1985]).

\[
E[D_{t_j}] = \mu_j E[L_j]. \\
\text{var}(D_{t_j}) = \sigma_j^2 E[L_j] + \mu_j^2 \text{var}(L_j).
\] (4.6a) (4.6b)

Second, the mean and variance of $D_{t_j}^{I}$ is obtained by using similar equations. Note that since we assumed that the demand at the different end-stockpoints is independent:

\[
\mu_i = \sum_{s \in V_s} \mu_s \quad \text{and} \quad \sigma_i^2 = \sum_{s \in V_s} \sigma_s^2.
\]

Note that correlation between demand at different end-stockpoints could easily be included here.

Third, the mean and variance of $D_{L_i+R_i}$ for each end-stockpoint $j$ should be calculated. To do this, we should carefully specify the time interval of the demand. From the derivation of (4.2) it follows that we are interested in the distribution of the demand at end-stockpoint $j$ during the time interval $[L_{12}, R + L_{12} + L_{23}]$. In order to determine this distribution we need to know how $L_{11}$ and $L_{12}$ relate to each other. For the analysis we assume that the relation between $L_{11}$ and $L_{12}$ is described by an AR(1) process. Although, one might assume another process we chose the AR(1) process for its simple correlation structure. From this assumption we have that

\[
L_{12} := \rho_1 L_{11} + \eta_{11}.
\] (4.7)

where $\bar{L}_1 := L_1 - E[L_1]$. The noise process $[\eta_{11}]$ is a series of mutually independent random variables with $E[\eta_{11}] = 0$ and $\text{var}[\eta_{11}] = (1 - \rho_1^2)\text{var}(L_1)$. As a consequence of (4.7) some interesting properties can be derived

\[
E[L_{12} | L_{11}] = \rho_1 L_{11} + (1 - \rho_1) E[L_1]. \\
cov[L_{11}, L_{12}] = \rho_1 \text{var}(L_1).
\] (4.8a) (4.8b)

Note from (4.8b) that $\rho_1$ equals the coefficient of correlation of two successive lead times of stockpoint 1. Although, the lead time process of an AR(1) process may lead to crossing orders, for the analysis we need a tractable correlation structure between successive lead times. In the numerical experiment presented in Section 4.8 it appears that this modeling assumption yields good approximations for performance measures even if the lead time process is not AR(1). From the properties (4.8) we are able to derive the mean and variance of $D_{L_i+R_i}$ (see Lemma 4.1).

**Lemma 4.1.**

\[
E[D_{L_i+R_i}] = \mu_j (R + E[L_j]). \\
\text{var}(D_{L_i+R_i}) = \sigma_j^2 (R + E[L_j]) + \mu_j^2 \{\text{var}(L_j) + 2(1 - \rho_1)\text{var}(L_1)\}.
\]
4.3. Analysis of 2-echelon systems without central stock

Proof. The proof is given by conditioning on the interval length $L_j + R_1$.

$$E[D_{i,j+R_1}^j] = E[D_{i,j+R_1}^j] = \int_0^\infty E[D_{i,j+R_1}^j] dF_{i,j+R_1}(l) = \int_0^\infty l \mu_j dF_{i,j+R_1}(l)$$

$$= \mu_j E[L_j + R_1] = \mu_j \{E[L_j] + R_1\}.$$  

$$E[D_{i,j+R_1}^j]^2 = \int_0^\infty E[D_{i,j+R_1}^j]^2 dF_{i,j+R_1}(l) = \int_0^\infty (\sigma_j^2 + \lambda_j^2 \mu_j^2) dF_{i,j+R_1}(l)$$

$$= \sigma_j^2 \{E[L_j] + R_1\} + \mu_j^2 \{\text{var}[L_j] + R + L_1 + L_2\} + \text{var}[L_j] + \text{var}[(E[L_j] + R_1)]$$

$$= \sigma_j^2 \{E[L_j] + R_1\} + \mu_j^2 \{\text{var}[L_j] + \text{var}[L_1 + L_2] + (E[L_j] + R_1)^2\}$$

$$= \sigma_j^2 \{E[L_j] + R_1\} + \mu_j^2 \{\text{var}[L_j] + 2(1 - \rho_1) \text{var}[L_1]\}.$$  

Hence,

$$\text{var}[D_{i,j+R_1}^j] = E[D_{i,j+R_1}^j]^2 - E[D_{i,j+R_1}^j]$$

$$= \sigma_j^2 \{E[L_j] + R_1\} + \mu_j^2 \{\text{var}[L_j] + 2(1 - \rho_1) \text{var}[L_1]\}.$$  

Fourth, an expression for $\text{cov}[D_{i,j+R_1}^j, D_{i,j}^j]$ is derived in Lemma 4.2.

**Lemma 4.2.**

$$\text{cov}[D_{i,j+R_1}^j, D_{i,j}^j] = -(1 - \rho_1) \mu_j \text{var}[L_1].$$  

(4.9)

Proof. We derive an expression for $\text{cov}[D_{i,j+R_1}^j, D_{i,j}^j]$ by carefully specifying the time intervals of the demand. As already stated, $D_{i,j+R_1}^j$ is the demand at end-stockpoint $j$ in the time interval $(L_{11}, R + L_1, L_2)$. Whereas $D_{i,j}^j$ relates to the demand in all end-stockpoints during the time interval $[0, L_{11}]$. Hence, by conditioning on $L_{11}$, we obtain that

$$\text{cov}[D_{i,j+R_1}^j, D_{i,j}^j] = E[D_{i,j+R_1}^j] E[D_{i,j}^j] - E[D_{i,j+R_1}^j] E[D_{i,j}^j]$$

$$= \mu_j \int_0^\infty \{E[L_j] + R + \lambda_j(l - E[L_1])\} \text{var}[L_1]$$

$$= -\mu_j \text{var}[L_1]$$

By specifying the time intervals (as in the proof of Lemma 4.2) we immediately see that $D_{i,j+R_1}^j$ and $D_{i,j}^j$ are negatively correlated. If the first lead time $L_{11}$ is relatively long, the total demand in $(0, L_{11})$
will be relatively high, but also \((L_{11}, R + L_{12} + L_{13})\) is relatively short and hence the demand at end-stockpoint \(j\) during this interval will be relatively low. If we ignore this negative correlation, the variance of \(X_{ij}\) is overestimated, resulting in higher stock levels than actually necessary, see equation (4.3). We will return to this subject at our numerical analysis in Section 4.9.

Fourth, one can easily see that \(\text{cov}(D_{i_1}^{T_1}, D_{i_2}^{T_1}) = 0\), since \(D_{i_2}^{T_1}\) relates to the demand in time interval \((L_{11}, L_{11} + L_{12})\), whereas \(D_{i_1}^{T_1}\) relates to the demand in time interval \((0, L_{11})\).

### 4.3.3 Calculation of the allocation fractions

From the results in Section 4.3.1, it follows that we can tune the order up-to-levels to the target fill rates \(f_{i,j}\) for any arbitrary set of allocation fractions \(q_{i,j}\), provided that the balance assumption is not violated. From the results of Chapter 2, we decided to determine the allocation fractions \(q_{i,j}\) such that the expected imbalance is minimized. We achieve this by using the BS rationing policy, taking into account the stochastic, correlated lead times.

Following Section 2.5, we define the amount of imbalance at end-stockpoint \(j\) by \(\Omega_{i,j}^{U} := E[-r_{i,j}]\).

Assuming a balanced situation at the previous allocation epoch, we find that

\[
q_{i,j} = q_j (D_{i,-R-L_{11}}^{T_1} - D_{i,-L_{11}}^{T_1} + D_{i,-R-L_{12}}^{T_1}, \quad j \in V_1.
\]

Next, as in Theorem 2.1, we approximate \(\Omega_{i,j}^{U}\) by

\[
\Omega_{i,j}^{U}(t) \approx \sigma_{i,j} \left( \frac{\mu_{i,n}}{\sigma_{i,n}} \right) + \mu_{i,n} \Phi \left( \frac{\mu_{i,n}}{\sigma_{i,n}} \right), \quad j \in V_1,
\]

with

\[
\mu_{i,n} := -R \mu_j,
\]

\[
\sigma_{i,n}^2 := 2q_j E[T_1] \sigma_j^2 + (R - 2q_j E[T_1]) \sigma_j^2 + 2 \text{var}[L_1] (1 - \rho_j) \mu_j \sigma_j \mu_j \sigma_j.
\]

with \(T_1 = \min(R, L_1)\).

In Section 3.2.2, we addressed two variants of the BS rationing policy. First, we minimize the approximation of \(\Omega_{i,j}^{U}\) by using a numerical method. Second, we minimize (4.10b) subject to \(\sum_{i \in V_1} q_n = 1\), where \(\sum_{i \in V_1} q_n\) is used as a surrogate expression for the imbalance. Numerical tests in Section 3.2 revealed that the first approach is somewhat better, but the second approach is considerably simpler. Because our numerical experiments indicated that the difference between the two approaches decreases if random lead times are introduced, we restrict ourselves here to the second approach. Taking into account the variance and autocorrelation of the lead time process of stockpoint \(1\), we determine \(q_{i,j}\) by minimizing (4.10b) subject to \(\sum_{i \in V_1} q_n = 1\). The Lagrange-multiplier technique yields

\[
q_{i,j} = \frac{1}{2|V_1|} \left( \frac{E[T_1] \sigma_j^2 + \mu_j (1 - \rho_j) \text{var}[L_1] \left( \frac{2 \mu_j \sigma_j}{|V_1|} \right)}{2 E[T_1] \sigma_j^2 + 2 \mu_j^2 (1 - \rho_j) \text{var}[L_1]} \right), \quad j \in V_1.
\]

From (4.11) it can be shown that \(0 < q_{i,j} < 1\) for all \(j \in V_1\) and that \(\sum_{i \in V_1} q_n = 1\). Note that \(E[T_1] = E[\min(R, L_1)]\) should be calculated to evaluate (4.11). This can be done by fitting e.g., a mixed Erlang distribution to the first two moments of \(L_1\).
4.4. Analysis of 2-echelon systems with central stock

4.3.4 Physical stock

We can find an expression for the mean physical stock per end-stockpoint \( j \) by using the following approximation

\[
\psi_j \approx \frac{E[\text{stock at start of replenishment cycle}]}{2} + \frac{E[\text{stock at end of replenishment cycle}]}{2}
\]  

(4.12)

which can be written as

\[
\psi_j \approx S_j - q_j E[L_j] \mu_j - \left( E[L_j] + \frac{R}{2} \right) \mu_j
\]

\[
+ \frac{E \left( D_t^j + q_j D_t^j - S_j \right)}{2} + E \left( D_t^j + q_j D_t^j - S_j \right)
\]

(4.13)

The two expectations in the numerator of (4.13) represent the mean shortage at the start and at the end of a replenishment cycle, which are already calculated in (4.3). The approximation given by (4.13) is not exact because of two reasons. First, we use (4.12). In the paper of Van der Heijden & De Kok [1977], the accuracy of this approximation (4.12) is tested for a single location inventory system. For large fill rates (say larger than 90%) the accuracy is excellent, since the probability that the echelon stock at the end of a replenishment cycle is positive is sufficiently large. Unfortunately, the approximation yields inaccurate results for small fill rates. This can easily be corrected by including an estimate for the expected physical stock halfway the replenishment cycle. By applying Simpson's rule for numerical integration, the approximation accuracy improves drastically. Second, we neglect the fact that the behavior of the physical stock during a large replenishment cycle differs from the behavior during a short replenishment cycle. Due to negative correlation between two successive replenishment cycles we typically get that a short replenishment cycle has more physical stock per period than a large replenishment cycle. So the mean physical stock per period should be weighed by the lengths of the replenishment cycles. In Section 4.4.4, we present a way to accomplish this by presenting an approximation for the mean physical stock at the depot in a divergent 2-echelon systems with central stock. How to do this for end-stockpoints remains a topic for further research. Some preliminary numerical experiments reveal that the accuracy of (4.13) is dominated by this latter reason.

4.4 Analysis of 2-echelon systems with central stock

In this section, we extend the results of Section 4.3 to the situation where the central depot is allowed to hold stock. The subjects are treated in the same order as in Section 4.3. In Section 4.4.1, we present how to determine the order-up-to-levels in order to satisfy the fill rate constraints at the end-stockpoints. In this determination, we assume that the allocation fractions are known. In Section 4.4.2, we derive several demand characteristics which are required to determine the order-up-to-levels. In Section 4.4.3, we derive an expression for the allocation fractions, such that a surrogate of the mean imbalance is minimized. Finally, in Section 4.4.4, we show how to approximate the mean physical stock at the various stockpoints.
4.4.1 Determination of the order-up-to-levels

Equation (4.3) can easily be modified such that the central depot holds central stock up to some specified level $\Delta_1$:

$$
\beta_j = \frac{E\left[(D_{ij}^l + R_j + q_j(D_{ij}^l - \Delta_1)^+ - S_j)^+ - (D_{ij}^l + q_j(D_{ij}^l - \Delta_1)^+ - S_j)^+\right]}{R \mu_j}.
$$

(4.14)

Again, $\beta_j = \beta_j^*$ is solved numerically using two-moment approximations for both terms in the numerator, $X_{ij} := D_{ij}^l + R_j + q_j(D_{ij}^l - \Delta_1)^+$ and $Y_{ij} := D_{ij}^l + q_j(D_{ij}^l - \Delta_1)^+$. For the mean and variance of $X_{ij}$ and $Y_{ij}$ we now find that

$$
E[X_{ij}] = E[D_{ij}^l + R_j + q_jE[(D_{ij}^l - \Delta_1)^+]].
$$

(4.15a)

$$
\text{var}[X_{ij}] = \text{var}[D_{ij}^l + R_j + q_j\text{var}[(D_{ij}^l - \Delta_1)^+]] + 2q_j \text{cov}[D_{ij}^l, (D_{ij}^l - \Delta_1)^+].
$$

(4.15b)

$$
E[Y_{ij}] = E[D_{ij}^l] + q_jE[(D_{ij}^l - \Delta_1)^+].
$$

(4.16a)

$$
\text{var}[Y_{ij}] = \text{var}[D_{ij}^l] + q_j^2\text{var}[(D_{ij}^l - \Delta_1)^+] + 2q_j \text{cov}[D_{ij}^l, (D_{ij}^l - \Delta_1)^+].
$$

(4.16b)

We see that we need several other demand characteristics to evaluate (4.15)–(4.16), among which two complex covariances. In the next section we show how these characteristics are obtained.

4.4.2 Calculation of the demand characteristics

We need to obtain additional demand characteristics to calculate the control parameters $\{S_j\}_{j \in V_i}$:

(i) the mean and variance of $(D_{ij}^l - \Delta_1)^+$,

(ii) $\text{cov}[D_{ij}^l, (D_{ij}^l - \Delta_1)^+]$ for each end-stockpoint $j \in V_i$,

(iii) $\text{cov}[D_{ij}^l, (D_{ij}^l - \Delta_1)^+]$ for each end-stockpoint $j \in V_i$.

First, we determine the mean and variance of $(D_{ij}^l - \Delta_1)^+$ by fitting a mixed Erlang distribution to the first two moments of $D_{ij}^l$, and next applying (2.20).

Second, we derive an expression for $\text{cov}[D_{ij}^l, (D_{ij}^l - \Delta_1)^+]$ [see Lemma 4.3].

**Lemma 4.3.**

$$
\text{cov}[D_{ij}^l, (D_{ij}^l - \Delta_1)^+] = -\mu_j(1 - \rho_l) \text{cov}[L_1, g(L_1)].
$$

(4.17)

with $g(L_1) := E[(D_{ij}^l - \Delta_1)^+|L_1]$. Proof: Again by considering the time intervals we obtain that

$$
\text{cov}[D_{ij}^l, (D_{ij}^l - \Delta_1)^+] = E[D_{ij}^l, (D_{ij}^l - \Delta_1)^+] - E[D_{ij}^l]E[(D_{ij}^l - \Delta_1)^+]
$$

$$
= \int_0^\infty E[D_{ij}^l, (D_{ij}^l - \Delta_1)^+] dF_{L_1}(l)
$$

$$
= \int_0^\infty \mu_j E[L_1] + R_j + \rho_l l \{ E[(D_{ij}^l - \Delta_1)^+] - E[(D_{ij}^l - \Delta_1)^+] + 1 \} dF_{L_1}(l)
$$

$$
= -\mu_j(1 - \rho_l) \int_0^\infty (l - E[L_1]) \{ E[(D_{ij}^l - \Delta_1)^+] - E[(D_{ij}^l - \Delta_1)^+] + 1 \} dF_{L_1}(l)
$$

$$
= -\mu_j(1 - \rho_l) \text{cov}[L_1, g(L_1)].
$$

\qed
4.4. Analysis of 2-echelon systems with central stock

Note that function $g$ in (4.17) is a function of the random variable $L_1$. In general it is difficult to evaluate expression (4.17). Therefore we propose to approximate this function $g$ by a first-order Taylor expansion around $E[L_1]$. Note that a Taylor expansion of a function $g$ around $a$ equals

$$g(x) = \sum_{m=0}^{\infty} \frac{g^{(m)}(a)}{m!} (x - a)^m,$$

where $g^{(m)}$ denotes the $m$th derivative of function $g$. Hence, the first-order Taylor expansion around $a$ equals

$$g(x) \approx g(a) + g'(a)(x - a).$$

Taking the center $a$ equal to $E[L_1]$, and substituting this in the expression above yields

$$g(L_1) \approx g(E[L_1]) + g'(E[L_1])(L_1 - E[L_1]).$$

Substitution of the expression above in $\text{cov}[L_1, g(L_1)]$ yields

$$\begin{align*}
\text{cov}[L_1, g(L_1)] & \approx \text{cov}
\left[
L_1, g(E[L_1]) + g'(E[L_1])(L_1 - E[L_1])
\right] \\
& = \text{cov}[
L_1, g'(E[L_1])L_1
]
\end{align*}$$

$$= \text{var}[L_1] \cdot g'(E[L_1]).$$

(4.18)

Substitution of (4.18) in (4.17) yields

$$\text{cov}[D_{t_1}, g'(E[L_1])] = -\{1 - \rho_1\} \cdot \mu_j \cdot \text{var}[L_1] \cdot g'(E[L_1]).$$

(4.19)

Although some approximation for $g'(\cdot)$ is possible, this function gets increasingly complicated if the number of echelons in the system increase, see Section 4.5. Therefore we calculate $g'(\cdot)$ using a numerical derivative:

$$g'(E[L_1]) \approx \frac{g(E[L_1] + \epsilon) - g(E[L_1])}{\epsilon},$$

(4.20)

where $\epsilon$ is some small number (we took $\epsilon = 0.001 \cdot E[L_1]$).

Third, we need an expression for $\text{cov}[D_{t_1}, (D_{t_1} - \Delta_t)^+]$. Using the same reasoning as in Section 4.3.2, we know that this covariance is zero.

4.4.3 Calculation of the allocation fractions

The derivation of allocation fractions minimizing the imbalance becomes more complicated in the presence of central stock. Therefore we propose to use the allocation fractions as obtained from the situation with a stockless central depot (see equation (4.11)). In Chapter 3, we showed that this approach yields satisfactory results for deterministic lead times.

4.4.4 Physical stock

In this section, we derive approximate expressions for the mean amount of physical stock held at each stockpoint. Let us first concentrate on the central depot. Note that the central stock remains constant between the arrival of two successive replenishment orders, so at first sight we only need the mean physical stock at the central depot just after allocation, $E[(\Delta_t - D_{t_1})^+]$. However, the time period between two order arrivals is a random variable now, which is correlated with the stock level just after allocation. If the first replenishment order lead time is relatively short, then:

- the time between two successive order arrivals is relatively long, and
- the total demand during the first lead time is relatively low, so that the central stock just after allocation is relatively high.
Hence we should account for the length of the time between order arrivals $R_1 = R + L_{12} - L_{11}$ as well. To obtain a tractable expression for the mean physical stock at the central depot we apply the renewal reward theorem. This yields

$$\psi_1 \approx \frac{E[(R + L_{12} - L_{11})(\Delta_1 - D_{L_{11}})^+]}{R}.$$  \hspace{1cm} (4.21)$$

Probably the expression above is an approximation, since the renewal reward theorem requires independent, subsequent replenishment cycles; however, in this case these are dependent. To obtain a more tractable expression for (4.21) we consider Theorem 4.1.

**Theorem 4.1.** The mean physical stock at the central depot (having index 1) can be approximated by

$$\psi_1 \approx E[(\Delta_1 - D_{L_{11}})^+] - \frac{1 - \rho_1}{R} \text{cov}[(L_1, (\Delta_1 - D_{L_{11}})^+)].$$  \hspace{1cm} (4.22)$$

**Proof.** To prove this theorem we only need to rewrite the numerator of (4.21).

$$E[(R + L_{12} - L_{11})(\Delta_1 - D_{L_{11}})^+]$$

$$= \text{RE}[(\Delta_1 - D_{L_{11}})^+] E[L_1 ](\Delta_1 - D_{L_{11}})^+] + \int_0^\infty E[L_2 | (\Delta_1 - D_{L_{11}})^+] dF_{L_2}(t)$$

$$= \text{RE}[(\Delta_1 - D_{L_{11}})^+] E[L_1 ](\Delta_1 - D_{L_{11}})^+] + \int_0^\infty E[L_2 | (\Delta_1 - D_{L_{11}})^+] dF_{L_2}(t)$$

$$= \text{RE}[(\Delta_1 - D_{L_{11}})^+] E[L_1 ](\Delta_1 - D_{L_{11}})^+] + \int_0^\infty (\rho_1 + (1 - \rho_1)E[L_1 ]E[(\Delta_1 - D_{L_{11}})^+]) dF_{L_2}(t)$$

$$= \text{RE}[(\Delta_1 - D_{L_{11}})^+] E[L_1 ](\Delta_1 - D_{L_{11}})^+] + \int_0^\infty (1 - \rho_1)E[L_1 ]E[(\Delta_1 - D_{L_{11}})^+] + (1 - \rho_1)E[L_1 ]E[(\Delta_1 - D_{L_{11}})^+]$$

$$= \text{RE}[(\Delta_1 - D_{L_{11}})^+] E[L_1 ](\Delta_1 - D_{L_{11}})^+]$$

Hence, substitution of the result above in (4.21) completes the proof. □

Still it is rather cumbersome to compute $\psi_1$, since we need to compute $\text{cov}[(L_1, (\Delta_1 - D_{L_{11}})^+)]$. We suggest to approximate this covariance as follows

$$\text{cov}[(L_1, (\Delta_1 - D_{L_{11}})^+)] \approx E[L_1 ](\Delta_1 - D_{L_{11}})^+] - E[L_1 ]E[(\Delta_1 - D_{L_{11}})^+]$$

$$= \int_0^\infty [E[(\Delta_1 - D_{L_{11}})^+] | L_1 | dF_{L_1}(t) - E[L_1 ]E[(\Delta_1 - D_{L_{11}})^+] = \text{cov}[(L_1, \tilde{g}(L_1))]$$

$$= \text{cov}[(L_1, \tilde{g}(L_1))],$$

with $\tilde{g}(L_1) := E[(\Delta_1 - D_{L_{11}})^+ | L_1 ]$. Analogously to (4.18) one can prove that $\text{cov}[(L_1, \tilde{g}(L_1)] = g(E[L_1 ])$ var$[L_1 ]$. Substitution of this result in (4.22) yields

$$\psi_1 \approx E[(\Delta_1 - D_{L_{11}})^+] - \frac{1 - \rho_1}{R} \text{var}[L_1 ] \tilde{g}(E[L_1 ]).$$  \hspace{1cm} (4.23)$$

Note that $\tilde{g}(E[L_1 ])$ can be approximated similarly to (4.20).

For the mean physical stock in the end-stockpoints, only a small modification of expression (4.13) is required. We find that

$$\psi_j \approx S_j - q_j E[(D_{L_{1i}} - \Delta_i)^+] - \left(E[L_j ] + \frac{R}{2}\right)\mu_j$$

$$+ E[(D_{L_{1i}} + q_j (D_{L_{1i}} - \Delta_i)^+ - S_j)^+] + E[(D_{L_{1i} + \Delta_i} + q_j (D_{L_{1i}} - \Delta_i)^+ - S_j)^+]$$

$$= \frac{1}{2}.$$
4.5 Analysis of general \( N \)-echelon systems

In this section, we extend our results to general \( N \)-echelon models. For sake of convenience, the expressions are given for 3-echelon systems. Modification to more than three echelons is straightforward.

4.5.1 Determination of the order-up-to-levels

We consider an end-stockpoint \( k \in V_j \) with \( j \in V_i \). Then from Chapter 3 it follows that the following modification of (4.14) is required:

\[
\beta_k = 1 - \frac{E[(X_{ijk} - S_j)^+]}{R \mu_k}, \tag{4.24}
\]

with

\[
X_{ijk} := D_{L_k + R_j}^i + q_k \left( D_{L_j}^i - \Delta_j + q_j (D_{L_j}^i - \Delta_j)^+ \right)^+, \tag{4.24a}
\]

\[
Y_{ijk} := D_{L_k}^i + q_k \left( D_{L_j}^i - \Delta_j + q_j (D_{L_j}^i - \Delta_j)^+ \right)^+. \tag{4.24b}
\]

Analogous to \( R_1 \), we define \( R_i \) as the random interarrival time of two subsequent orders at stockpoint \( i \). Hence, we have that \( R_i := R + L_{i2} - L_{i1} + L_{j2} - L_{j1} \). Again, we solve the fill rate constraint \( \beta_k = \beta_k^* \) numerically using two-moment approximations for \( X_{ijk} \) and \( Y_{ijk} \). Eliminating zero covariances, we find for the mean and variance of \( X_{ijk} \) and \( Y_{ijk} \):

\[
E[X_{ijk}] = E[D_{L_k + R_j}^i] + q_k \left[ \left( D_{L_j}^i - \Delta_j + q_j (D_{L_j}^i - \Delta_j)^+ \right)^+ \right], \tag{4.25a}
\]

\[
\text{var}(X_{ijk}) = \text{var}(D_{L_k + R_j}^i) + q_k^2 \text{var} \left[ \left( D_{L_j}^i - \Delta_j + q_j (D_{L_j}^i - \Delta_j)^+ \right)^+ \right] + 2q_k \text{cov} \left[ D_{L_k + R_j}^i, \left( D_{L_j}^i - \Delta_j + q_j (D_{L_j}^i - \Delta_j)^+ \right)^+ \right]. \tag{4.25b}
\]

\[
E[Y_{ijk}] = E[D_{L_k}^i] + q_k \left[ \left( D_{L_j}^i - \Delta_j + q_j (D_{L_j}^i - \Delta_j)^+ \right)^+ \right], \tag{4.26a}
\]

\[
\text{var}(Y_{ijk}) = \text{var}(D_{L_k}^i) + q_k^2 \text{var} \left[ \left( D_{L_j}^i - \Delta_j + q_j (D_{L_j}^i - \Delta_j)^+ \right)^+ \right]. \tag{4.26b}
\]

4.5.2 Calculation of the demand characteristics

We need the following additional demand characteristics to calculate the control parameters \( \{S_{k_j}, k \in V_j \} \):

(i). the mean and variance of \( D_{L_k + R_j}^i \) for each end-stockpoint \( k \in V_j \).

(ii). the mean and variance of \( \left( D_{L_j}^i - \Delta_j + q_j (D_{L_j}^i - \Delta_j)^+ \right)^+ \).

(iii). \( \text{cov} \left[ D_{L_k + R_j}^i, \left( D_{L_j}^i - \Delta_j + q_j (D_{L_j}^i - \Delta_j)^+ \right)^+ \right] \) for each end-stockpoint \( k \in V_j \).

First, the mean and variance of \( D_{L_k + R_j}^i \) is obtained analogously to the 2-echelon system. Now we need the demand characteristics in a slightly different time interval, namely \( [L_{i1} + L_{j1}, R + L_{i1} + L_{j2} + L_{k2}] \) instead of \( [L_{i1}, R + L_{i1} + L_{k2}] \) as for the 2-echelon system. Again, by conditioning on the
interval length \( R + L_{12} + L_{1} - L_{11} + L_{12} - L_{11} \) we obtain

\[
E[D_{l_{1} + \tau_{j}}] = \mu_{l}(R + E[L_{1}]),
\]

\[
\text{var}(D_{l_{1} + \tau_{j}}) = \sigma_{l}^{2}(R + E[L_{1}]) + \mu_{l}^{2} \text{var}(L_{1}) + 2\mu_{l}^{2}(1 - \rho_{l}) \text{var}(L_{1}) + (1 - \rho_{l}) \text{var}(L_{j}).
\]

(4.27a)

(4.27b)

Extension to general \( N \)-echelon systems is straightforward.

Second, the approximation of the mean and variance of \( (D_{l_{j}} - \Delta_{j} + q_{j}(D_{l_{1}} - \Delta_{1})^{\ast})^{\ast} \) is a straightforward extension of the 2-echelon system. First, we approximate \( D_{l_{1}}^{\ast} \) by a mixed Erlang distribution. Next, we calculate the first two moments of \( (D_{l_{1}} - \Delta_{1})^{\ast} \) as described in Section 2.4.1. Using these two moments, we approximate \( D_{l_{1}}^{\ast} + q_{j}(D_{l_{1}} - \Delta_{1})^{\ast} \) by a mixed Erlang distribution. Finally, we obtain the desired mean and variance of \( (D_{l_{j}} - \Delta_{j} + q_{j}(D_{l_{1}} - \Delta_{1})^{\ast})^{\ast} \) from (2.20). It is clear that this approach can be extended to general \( N \)-echelon systems.

Third, we derive an expression for \( \text{cov} \left[ D_{l_{1} + \tau_{j}}^{\ast}, (D_{l_{j}} - \Delta_{j} + q_{j}(D_{l_{1}} - \Delta_{1})^{\ast})^{\ast} \right] \) (see Lemma 4.4).

\[
\text{cov} \left[ D_{l_{1} + \tau_{j}}^{\ast}, (D_{l_{j}} - \Delta_{j} + q_{j}(D_{l_{1}} - \Delta_{1})^{\ast})^{\ast} \right] = -(1 - \rho_{j}) \mu_{l} \text{cov}[L_{1}, g_{1}(L_{1})] - (1 - \rho_{j}) \mu_{l} \text{cov}[L_{j}, g_{j}(L_{j})].
\]

(4.28)

with

\[
g_{1}(L_{1}) := E \left[ \left( D_{l_{j}}^{\ast} - \Delta_{j} + q_{j}(D_{l_{1}}^{\ast} - \Delta_{1})^{\ast} \right)^{\ast} \bigg| L_{1} \right],
\]

\[
g_{j}(L_{j}) := E \left[ \left( D_{l_{j}}^{\ast} - \Delta_{j} + q_{j}(D_{l_{1}}^{\ast} - \Delta_{1})^{\ast} \right)^{\ast} \bigg| L_{j} \right].
\]

**Proof.** The proof is similar to the proofs given in Lemma 4.2 and 4.3.

\[
\text{cov} \left[ D_{l_{1} + \tau_{j}}^{\ast}, (D_{l_{j}} - \Delta_{j} + q_{j}(D_{l_{1}} - \Delta_{1})^{\ast})^{\ast} \right] = E \left[ D_{l_{1} + \tau_{j}}^{\ast} E \left[ (D_{l_{j}} - \Delta_{j} + q_{j}(D_{l_{1}} - \Delta_{1})^{\ast})^{\ast} \right] \right] - E \left[ D_{l_{1} + \tau_{j}}^{\ast} \right] \text{var} \left[ (D_{l_{j}} - \Delta_{j} + q_{j}(D_{l_{1}} - \Delta_{1})^{\ast})^{\ast} \right]
\]

Again by considering the time intervals it follows that \( E[D_{l_{1} + \tau_{j}}^{\ast}] = E[D_{l_{1} + \tau_{j}}] \). By conditioning on both \( L_{11} \) and \( L_{11} \) we obtain that

\[
E[D_{l_{1} + \tau_{j}}^{\ast}] = \int_{0}^{\infty} \int_{0}^{\infty} \mu_{l}(R + L_{1} + (1 - \rho_{l})E[L_{1}] - m) dF_{l_{1}}(l) dF_{l_{1}}(m) + \rho_{l} m + (1 - \rho_{l})E[L_{1}] - m) dF_{l_{1}}(l) dF_{l_{1}}(m)
\]

\[
\int_{0}^{\infty} \int_{0}^{\infty} \mu_{l}(-1 - \rho_{l})E[L_{1} - m - E[L_{1}]) dF_{l_{1}}(l) dF_{l_{1}}(m).
\]
4.5. Analysis of general $N$-echelon systems

Using the definition of the covariance and the above equation it follows that

$$
cov \left[ D_{L_l}^l, \left( D_{L_j}^j - \Delta_j + q_j (D_{L_i}^i - \Delta_i)^+ \right)^+ \right] = \\
= \int_0^\infty \int_0^\infty \mu_k \left[ -(1 - \rho_j) (l - E[L_l]) - (1 - \rho_j) (m - E[L_j]) \right] \left( E \left[ \left( D_{L_l}^l - \Delta_j + q_j (D_{L_i}^i - \Delta_i)^+ \right)^+ \left| L_{l1} = l, L_{l2} = m \right] \right) \left( E \left[ \left( D_{L_j}^j - \Delta_j + q_j (D_{L_i}^i - \Delta_i)^+ \right)^+ \left| L_{j1} = m \right] \right) \right) \, dF_{L_l}(l) \, dF_{L_j}(m) \]

$$

Similarly to (4.17), expression (4.28) can be approximated by using first-order Taylor expansions. This yields

$$
cov \left[ D_{L_l}^l, \left( D_{L_j}^j - \Delta_j + q_j (D_{L_i}^i - \Delta_i)^+ \right)^+ \right] = -(1 - \rho_j) \mu_k \left( E[L_l] \right) \left( E[L_j] \right) \left( q_j \right) \left( \sigma_j \right) \approx -(1 - \rho_j) \mu_k \left( E[L_l] \right) \left( E[L_j] \right) \left( q_j \right) \left( \sigma_j \right). \quad (4.29)
$$

Also these calculations can be extended straightforwardly to general $N$-echelon systems.

4.5.3 Calculation of the allocation fractions

Similar to Section 3.3.2, the allocation fractions are calculated according to (4.11), ignoring the effect of lead times in earlier stages. For the intermediate stockpoints, we also use (4.11) making the appropriate substitutions for $\mu_1, \sigma_1^2, \mu_j, \text{ and } \sigma_j^2$ (see (3.11)).

4.5.4 Physical stock

First, note that expression (4.23) for the mean physical stock in the most upstream stockpoint remains valid. Second, for the mean physical stock in an intermediate stockpoint $j$ we find analogously to Section 4.4.4:

$$
\psi_j \approx \frac{E \left[ (R + L_{j2} - L_{j1} + L_{j3} - L_{j4}) (\Delta_j - D_{L_j}^j + q_j (\Delta_i - D_{L_i}^i)^+)^+ \right]}{R}.
$$
Theorem 4.2. The mean physical stock at stockpoint $j \in V_i$ can be approximated by
\[
\phi_j \approx E \left[ \left( \Delta_j - D_{L_j}^i - q_j \left( D_{L_1}^i - \Delta_1 \right)^+ \right)^+ \right] \\
- \frac{1 - \rho_j}{R} \text{cov} \left[ L_{1j}, \left( \Delta_j - D_{L_j}^i - q_j \left( D_{L_1}^i - \Delta_1 \right)^+ \right)^+ \right] \\
- \frac{1 - \rho_j}{R} \text{cov} \left[ L_{1j}, \left( \Delta_j - D_{L_1}^i - q_j \left( D_{L_1}^i - \Delta_1 \right)^+ \right)^+ \right].
\]
(4.30)

Proof. The proof of this theorem is similar to the proof of Theorem 4.1.
\[
E \left[ (R + L_{j1} - L_{j1} + L_{11} - L_{11}) \left( \Delta_j - D_{L_{j1}}^i - q_j (\Delta_1 - D_{L_{11}}^i)^+ \right)^+ \right] \\
= R E \left[ \left( \Delta_j - D_{L_{j1}}^i - q_j (\Delta_1 - D_{L_{11}}^i)^+ \right)^+ \right] \\
- E \left[ L_{11} \left( \Delta_j - D_{L_{j1}}^i - q_j (\Delta_1 - D_{L_{11}}^i)^+ \right)^+ \right] - E \left[ L_{11} \left( \Delta_j - D_{L_{11}}^i - q_j (\Delta_1 - D_{L_{11}}^i)^+ \right)^+ \right] \\
+ E \left[ L_{11} \left( \Delta_j - D_{L_{j1}}^i - q_j (\Delta_1 - D_{L_{11}}^i)^+ \right)^+ \right] + E \left[ L_{11} \left( \Delta_j - D_{L_{11}}^i - q_j (\Delta_1 - D_{L_{11}}^i)^+ \right)^+ \right].
\]
(4.31)
The last term in the equation above can be rewritten as follows
\[
E \left[ L_{j1} \Delta_j - D_{L_{j1}}^i - q_j (\Delta_1 - D_{L_{11}}^i)^+ \right]^+ \\
= \int_{0}^{\infty} E \left[ L_{j1} \left( \Delta_j - D_{L_{j1}}^i - q_j (\Delta_1 - D_{L_{11}}^i)^+ \right)^+ \right] \left. \right| L_{j1} = l \quad dF_{j1}(l) \\
= \int_{0}^{\infty} E \left[ L_{j1} \right] E \left[ \left( \Delta_j - D_{L_{j1}}^i - q_j (\Delta_1 - D_{L_{11}}^i)^+ \right)^+ \right] \left. \right| L_{j1} = l \quad dF_{j1}(l) \\
= \int_{0}^{\infty} \left( \rho_j + (1 - \rho_j) E \left( L_{j1} \right) \right) E \left[ \left( \Delta_j - D_{L_{j1}}^i - q_j (\Delta_1 - D_{L_{11}}^i)^+ \right)^+ \right] \left. \right| L_{j1} = l \quad dF_{j1}(l) \\
= \rho_j E \left[ L_{j1} \right] \left( \Delta_j - D_{L_{11}}^i - q_j (\Delta_1 - D_{L_{11}}^i)^+ \right)^+ \\
+ (1 - \rho_j) E \left[ L_{j1} \right] E \left[ \left( \Delta_j - D_{L_{j1}}^i - q_j (\Delta_1 - D_{L_{11}}^i)^+ \right)^+ \right].
\]
(4.32a)

Analogous to (4.32a) it holds that
\[
E \left[ L_{11} \left( \Delta_j - D_{L_{j1}}^i - q_j (\Delta_1 - D_{L_{11}}^i)^+ \right)^+ \right] = \rho_j E \left[ L_{11} \left( \Delta_j - D_{L_{11}}^i - q_j (\Delta_1 - D_{L_{11}}^i)^+ \right)^+ \right] \\
+ (1 - \rho_j) E \left[ L_{11} \right] E \left[ \left( \Delta_j - D_{L_{j1}}^i - q_j (\Delta_1 - D_{L_{11}}^i)^+ \right)^+ \right].
\]
(4.32b)

Substituting (4.32) in (4.31), and next using some elementary algebra proves the theorem. \(\Box\)

It is rather cumbersome to derive exact expressions for both covariances in (4.30), therefore we approximated both covariances similarly to (4.22). This results in
\[
cov \left[ L_{11}, \left( \Delta_j - D_{L_{j1}}^i - q_j \left( D_{L_1}^i - \Delta_1 \right)^+ \right)^+ \right] \approx \text{var} \left[ L_{11} \right] \tilde{g}(E[L_1]).
\]
\[
cov \left[ L_{j1}, \left( \Delta_j - D_{L_{j1}}^i - q_j \left( D_{L_1}^i - \Delta_1 \right)^+ \right)^+ \right] \approx \text{var} \left[ L_{j1} \right] \tilde{g}(E[L_{j1}]).
\]
4.6. Summary of the algorithm

with

\[ g_i(L_i) := E \left[ \left( \Delta_j - D_{i,j} - q_j \right) (D_{i,j} - \Delta_i)^+ \right] L_i, \]
\[ \tilde{g}_i(L_j) := E \left[ \left( \Delta_j - D_{i,j} - q_j \right) (D_{i,j} - \Delta_i)^+ \right] L_j. \]

Substitution of the approximations of both covariances in (4.30) yields

\[ \psi_j \approx E \left[ \left( \Delta_j - D_{i,j} - q_j \right) (D_{i,j} - \Delta_i)^+ \right] \frac{1 - \rho_j}{R} \text{var}(L_j) \tilde{g}_j(E[L_j]) \left( 1 - \frac{1 - \rho_j}{R} \text{var}(L_j) \tilde{g}_j(E[L_j]) \right). \]

Similar to the preceding sections, we make the following extension for the mean physical stock in the end-stockpoints:

\[ \psi_k \approx \mu_k \left( \frac{(X_{1,\mu} - S_k)^+}{2} + \frac{(Y_{1,\mu} - S_k)^+}{2} \right) + \left( E[L_k] + \frac{R}{2} \right) \mu_k, \]

where \( E[(X_{1,\mu} - S_k)^+] \) and \( E[(Y_{1,\mu} - S_k)^+] \) denote the expected shortage at the end and at the start of a replenishment cycle respectively, see equation (4.24).

4.6 Summary of the algorithm

The purpose of our algorithm is to calculate the allocation fractions and the order-up-to-levels for every stockpoint in the system, such that (different) target fill rates for the end-stockpoints are attained using minimal inventory imbalance. For sake of convenience, we explain the algorithm by considering a 3-echelon system as in the previous section. The extension of this algorithm to more than three echelons is straightforward. We need the following information as a starting point for our algorithm:

- the length of the review period \( R \).
- the mean and standard deviation of the demand at each end-stockpoint \( k \in E: \mu_k \) and \( \sigma_k^2 \).
- the target fill rates \( \mu_k \) for each end-stockpoint \( k \).
- the mean, variance and first-order autocorrelation of the lead time for each stockpoint \( i \in M \cup E: E[L_i], \text{var}(L_i) \) and \( \rho_i \).
- the maximum inventory level \( \Delta_i \) for each intermediate stockpoint \( i \in M \).

Our algorithm consists of the following steps:

1. Calculate the following demand characteristics, working from the end-stockpoints in the upstream direction:
   (a) the mean and variance of the demand at each intermediate stockpoints \( i \in M: \mu_i \) and \( \sigma_i^2 \).
   (b) the mean and variance of \( D_{i,j}^k \) for each stockpoint \( i \in M \cup E \) (see equation (4.6)).
   (c) the mean and variance of \( D_{i,j}^k \) for each end-stockpoint \( k \in E \) (see equation (4.27)).
   (d) the covariances required for the evaluation of the fill rate expressions (4.24), like
      \[ \text{cov} \left[ D_{i,j}^{k+\epsilon}, \left( D_{i,j}^{k+\epsilon} - \Delta_j + q_j(D_{i,j}^{k+\epsilon} - \Delta_i)^+ \right) \right], \]
      where \( \epsilon \in V_j \) and \( j \in V_1 \). These can be obtained by (4.29).
(e) similarly, the covariances required to evaluate intermediate stock levels, like
\[
\text{cov}\left[L_i, \left(\Delta_j - D_{ij} - q_j (D_{ij} - \Delta_j)^+\right)^+\right], \text{cov}\left[L_i, \left(\Delta_j - D_{ij} - q_j (D_{ij} - \Delta_j)^+\right)^+\right]
\]
and \(\text{cov}(L_i, (\Delta_1 - D_{ij})^+)\), see (4.23) and (4.30).

(ii). Determine the allocation fractions \(\{q_i\}_{i \in W}\) for each intermediate stockpoint \(i \in M\) from equation (4.11), after making the appropriate substitutions for \(\mu_i, \sigma_i^2, \mu_j\) and \(\sigma_j^2\).

(iii). Determine the order-up-to-level \(S_i\) for each end-stockpoint \(i\):

(a) Approximate the distributions of \(X_{ij}\) and \(Y_{ij}\) as given by (4.24) by mixed Erlang distributions as described in Section 2.4.1. The mean and variance of \(X_{ij}\) and \(Y_{ij}\) are given by (4.25) and (4.26).

(b) Use bisection to find from (4.24) the order-up-to-level \(S_i\) that match the target fill rate \(P_i^*\).

(iv). Determine the order-up-to-level \(S_i\) for each intermediate stockpoint \(i \in M\) from \(S_i = \Delta_i + \sum_{i \in W} S_{ii}\) working from the end-stockpoints in the upstream direction.

(v). Determine the mean physical stock per end-stockpoint from equation (4.33) using the expected shortage at the start and at the end of a replenishment cycle as found in step 3.

(vi). Finally, determine the mean physical stock per intermediate stockpoint from equations (4.23) and (4.30).

4.7 Model to generate lead times without crossing

A special point of attention is the stochastic process by which successive lead times are generated. We stated that orders may not cross. This is reasonable, but it complicates simulation. If we simply generate independent lead times, we cannot prevent that orders cross in general. Therefore, in this section we develop a simple model for a lead time process in which orders do not cross. We show that the combination of a single server queue and a deterministic pipeline yields a versatile and simple lead time process. We use results from queuing theory to choose the parameters of this process such that the successive throughput times (approximately) match a prespecified lead time mean and variance.

Although the approach is general applicable, we focus on divergent systems with a periodic review, order-up-to policy. The results of this section facilitates building a simulation model of multi-echelon inventory systems, to be used for scenario analysis and/or validation of approximate analytical methods. In fact, we developed this approach for the latter purpose, see Section 4.8.

The remainder of this section is structured as follows. First, we consider a single location inventory system. Section 4.7.1 addresses the model and approach for this situation. The analysis of the model is discussed in Section 4.7.2. In Section 4.7.3, we show how the lead time model can be applied to the divergent multi-echelon system treated in this chapter. An important assumption is required to apply the lead time model to downstream stockpoints. We verify the impact of this assumption by a simulation study.

4.7.1 Single location system

Consider a single location inventory system in which the lead time \(L\) is a random variable with mean \(\mu_L\) and variance \(\sigma_L^2\). Denote by \(A_n\) the time period between issuing the \((n-1)^{th}\) and the \(n^{th}\) replenishment order. \(A_n\) can be a random variable, e.g., under continuous review, or deterministic, e.g., under periodic review. Since it is assumed that orders cannot cross, the subsequent lead times faced in the inventory system are correlated for most lead time distributions. Therefore it becomes rather cumbersome to
4.7. Model to generate lead times without crossing

model a lead time process for which the mean and variance coincide with the target values $\mu_L$ and $\sigma^2_L$, respectively.

To illustrate this we consider a single location inventory system under periodic review. Figure 4.1 depicts the arrival times of two subsequent orders. At time $t$ an order is placed at the supplier with lead time $L_1$. This order arrives at $t + L_1$. At time $t + R$ the stockpoint again inspects the inventory position, and places an order with lead time $L_2$. This order arrives at $t + R + L_2$. Since subsequent orders do not cross, there has to hold $R + L_2 \geq L_1$. If the lead time distribution of $L$ is such that this always holds, the successive lead times can be independent. An example of such a distribution is a uniform distribution on $[\alpha, \alpha + \beta]$ with $\alpha > 0$ and $0 \leq \beta \leq R$. For the more general case where not always $R + L_2 \geq L_1$, it is obvious that $L_1$ depends on $L_2$. In the remainder of this section we develop an approach to model the lead time process of interrelated lead times, such that the mean and variance of the lead time equals the predetermined $\mu_L$ and $\sigma^2_L$, respectively.

![Figure 4.1. Arrival process of orders.](image)

**Model and approach**

The lead time is modeled by the sojourn time in a $G1/G1/1$-queue plus a deterministic pipeline time. The arrival of a customer at this queue corresponds to the stockpoint placing an order, and since the server adopts the FCFS-discipline, customers (i.e. orders) cannot cross. For our convenience we introduce the following notation:

$B$  The service time of a customer in the queuing system.

$W$  The waiting time of a customer in the queuing system.

$L_s$ The sojourn time of a customer in the queuing system, $L_s = W + B$.

The lead time is not modeled solely by a queuing system for the following reason. In practice the mean lead time can be large in contrast with its small variance. In order to obtain a lead time process for which the sojourn time has a large mean and a small variance, a very high utilization ($> 0.999$) is required. So, using the sojourn time of the $G1/G1/1$-queue solely is inadequate. Therefore we suggest to model the lead time as the sum of the sojourn time in a $G1/G1/1$-queue plus a fixed time in a pipeline (see Figure 4.2).

![Figure 4.2. Model of the lead time process.](image)

By setting this fixed time properly we avoid lead time processes with an extremely high utilization. Thus, the lead time $L$ is given by

\[ L := L_s + L_D. \]
where \( L_D \) denotes a fixed nonnegative time and the random variable \( L_S \) denotes the stochastic part of the lead time (in general it depends on preceding lead times). As one possible interpretation one might think of \( L_D \) as a fixed handling & transportation time, and \( L_S \) as the production time which is subjected to capacity constraints.

### 4.7.2 Analysis

The sojourn time of the \( G1/G1/1 \)-queue plus the deterministic pipeline time has mean \( E[W] + E[B] + L_D \) and variance \( \text{var}[W] + \text{var}[B] \). So, we have to find the parameters of the lead time process, such that \( E[L] = E[W] + E[B] + L_D \) and \( \text{var}[L] = \text{var}[B] + \text{var}[W] \) for given \( E[L] \) and \( \text{var}[L] \). The process by which replenishment orders are generated is given, so the remaining degrees of freedom are \( L_D \) and the service distribution \( B \). Using only the first two moments of \( B \) to analyze the \( G1/G1/1 \)-queue, this means that we have to find the right values for \( L_D \), \( E[B] \) and \( \text{var}[B] \). So we have one degree of freedom. We deal with this as follows. A suitable way of splitting \( L \) into the fixed \( L_D \) and the random \( L_S \), is just by taking an extreme distribution for \( B \). Then from queuing theory it follows that the sojourn time in a \( G1/M/1 \)-queue is exponentially distributed as well. So \( c[L_S] = 1 \).

After some elementary algebra it follows

\[
L_D = (1 - c[L]) E[L], \quad c[L] \leq 1. \tag{4.34}
\]

From (4.34) it is obvious that this way of splitting \( L \) into the fixed \( L_D \) and the random part \( L_S \) is only valid for \( c[L] \leq 1 \). However, lead times with \( c[L] > 1 \) are exceptional. So, by using the splitting rule of (4.34) the service process easily follows from the analysis of the \( G1/M/1 \)-queue. The service time \( B \) is exponentially distributed with

\[
E[B] = E[L_S] \left(1 - A^*(\frac{1}{E[L_S]})\right),
\]

where \( A^*(s) \) denotes the Laplace-Stieltjes transform of \( A \). Note that for the mixed Erlang distribution addressed in Section 2.4.1 holds

\[
A^*(s) = \theta_1 \left(\frac{\lambda_1}{\lambda_1 + s}\right)^\eta_1 + \theta_2 \left(\frac{\lambda_2}{\lambda_2 + s}\right)^\eta_2.
\]

For the specific case where the orders arrive periodically we have a \( D/M/1 \)-queue. Then \( B \) is exponentially distributed with

\[
E[B] = E[L_S] \left(1 - e^{-\frac{1}{E[L_S]}}\right). \tag{4.35}
\]

Diks & Van der Heijden [1997] also give another method to split the lead time into the fixed \( L_D \) and the random part \( L_S \). First, they developed an algorithm to determine \( E[B] \) and \( \text{var}[B] \) for a given value of \( L_D \). Next, they determine \( L_D \) such that the first-order autocorrelation equals \( \rho \). Numerical experiments reveal that for some range of \( \rho \) values there are two possibilities to split \( L \), whereas for another range there is no possibility at all! Since their method is approximate and time-consuming we suggest to use the exact and simple splitting rule (4.34).

### 4.7.3 Divergent multi-echelon systems

In this chapter, we analyze a divergent system using a push mechanism, i.e., upon arrival of an order at a stockpoint, the products are shipped towards its successors. The \( D/M/1 \)-analysis performed in Section 4.7.2 applies to the most upstream stockpoint, since it uses a periodic review policy. For a
4.7. Model to generate lead times without crossing

downstream stockpoint holds that the arrival process is the departure process of its upstream stockpoint. Hence, the interarrival times generally will be correlated. As an approximation we suggest to model the lead time process in a downstream stockpoint by the sojourn time in a $G1/M/1$-queue plus a fixed handling & transportation time. So we disregard the dependency between the interarrival times. In the remainder of this section we examine the approximation of assuming independent arrivals of orders at downstream stockpoints.

For our convenience, we describe how to model the lead time process of the downstream stockpoint in a 2-serial system (see Figure 4.3). The extension to larger serial systems, or even divergent systems, is straightforward since for all cases the arrival process of the $G1/M/1$-queue equals the departure process of its predecessor. Suppose splitting rule (4.34) is used in both stockpoints. Stockpoint 1 inspects the inventory at the start of every period, i.e., $R = 1$, and upon arrival of such order stockpoint 2 inspects its inventory position and places an order. For stockpoint 1 we choose $E[L_1] \in [0.5, 0.75, 1, 1.5, 2, 3]$, and $\epsilon[L_1] \in [0.25, 0.5, 0.75]$. For stockpoint 2 we set $E[L_2] = 1$ and $\text{var}[L_2] = 0.5$. We only varied the lead time parameters of stockpoint 1, since these influence the behavior of the departure process of stockpoint 1, and hence the arrival process of stockpoint 2.

![Figure 4.3. The 2-serial system.](image)

We use a two-moment approximation for the interdeparture time process. It is clear that the mean interdeparture time equals the mean interarrival time at stockpoint 1. The variance of the interdeparture time of stockpoint 1 can be determined from the three approximations for the squared coefficient of variation of the interdeparture times of a $G1/G/1$-queue given by Buzacott & Shanthikumar [1993]. To examine the impact of neglecting the dependency of arrivals at stockpoint 2 we simulated $E[L_2]$ and $\text{var}[L_2]$ of stockpoint 2 for the aforementioned 18 instances (simulated values are denoted by $\hat{E}[L_2]$ and $\hat{\text{var}}[L_2]$, respectively). This is done for the three approximations derived in Buzacott & Shanthikumar [1993] (see Table 4.1). For every instance we simulated 400,000 periods. Approximation 1 and 2 results in lead times for which both the mean and the variance are too low, while approximation 3 generally results in lead times for which both the mean and the variance are too high. It turns out that approximation 1 is always worse than approximation 2. Furthermore, the deviation of the mean lead time of approximation 2 and 3 are almost equal. However, approximation 3 results in a much better performance of the lead time variance. Hence, from this experiment we suggest to use approximation 3. As can be seen from Table 4.1 the performance of the approximations depend on the utilization, $\rho$, say, of the queue of stockpoint 1. When $\rho$ is small customers hardly interact with each other (since almost every customer arrives at an empty queue of stockpoint 1). When $\rho$ is large, almost every customer sees a non-empty queue. Hence most of the times the interdeparture time of two successive customers equals a service time. Since the service times are independent, also the interarrival times at the queue of stockpoint 2 are independent. Table 4.1 shows that the performance of the approximations are the worse when $\rho$ is around 0.8. However, for small $\rho$ and high $\rho$ the approximations yield very satisfactory results.

Finally, we note that it takes quite some time before the simulator 'stabilizes'. This issue was already discussed in Vinson [1972]. Take for example the instance where $E[L_1] = 3$ and $\epsilon[L_1] = 0.75$. Since the parameters of the queuing system of stockpoint 1 are determined by the exact expression (4.35) we know that $\hat{\text{var}}[L_1]$ has to converge to $\text{var}[L_1] = 0.5025$. In Table 4.2 this variance is depicted for several different sizes of the simulation run. That it converges very slowly is due to the high
Service level constraints: random lead times

Table 4.1. The (simulated) mean and variance of the lead time observed at stockpoint 2 ($E[L_2] = 1$ and $\text{var}[L_2] = 0.5$), for the three approximations of Buzacott & Shanthikumar[1993].

<table>
<thead>
<tr>
<th>Stockpoint 1</th>
<th>Approx. 1</th>
<th>Approx. 2</th>
<th>Approx. 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E[L_2]$</td>
<td>$\hat{E}[L_2]$</td>
<td>$\hat{\text{var}}[L_2]$</td>
<td>$\hat{E}[L_2]$</td>
</tr>
<tr>
<td>0.25</td>
<td>0.998</td>
<td>0.494</td>
<td>0.998</td>
</tr>
<tr>
<td>0.50</td>
<td>0.989</td>
<td>0.474</td>
<td>0.991</td>
</tr>
<tr>
<td>0.75</td>
<td>0.981</td>
<td>0.457</td>
<td>0.987</td>
</tr>
<tr>
<td>1.00</td>
<td>0.994</td>
<td>0.483</td>
<td>0.994</td>
</tr>
<tr>
<td>1.50</td>
<td>0.973</td>
<td>0.443</td>
<td>0.981</td>
</tr>
<tr>
<td>2.00</td>
<td>0.975</td>
<td>0.448</td>
<td>0.982</td>
</tr>
<tr>
<td>3.00</td>
<td>0.972</td>
<td>0.444</td>
<td>0.979</td>
</tr>
<tr>
<td>6.00</td>
<td>0.969</td>
<td>0.442</td>
<td>0.975</td>
</tr>
</tbody>
</table>

Table 4.2. $\hat{\text{var}}[L_1]$ for several cycle times.

<table>
<thead>
<tr>
<th>Cycle time</th>
<th>$\hat{\text{var}}[L_1]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50,000</td>
<td>5.655</td>
</tr>
<tr>
<td>100,000</td>
<td>5.421</td>
</tr>
<tr>
<td>200,000</td>
<td>5.339</td>
</tr>
<tr>
<td>300,000</td>
<td>5.273</td>
</tr>
<tr>
<td>400,000</td>
<td>5.077</td>
</tr>
</tbody>
</table>

correlation between successive lead times. When this correlation is not so large the convergence is much quicker.

4.8 Numerical validation

The algorithm of Section 4.6 that we developed is approximate, therefore we establish the accuracy of our approximations using extensive numerical experimentation, both for 2-echelon systems (experimental design in Section 4.8.1, results in Section 4.8.2) and for 3-echelon systems (experimental design in Section 4.8.3, results in Section 4.8.4). We use the difference between target fill rate and actual (simulated) fill rate as a measure of accuracy. Next to the mean absolute deviation from the target
4.8. Numerical validation

fill rate, we also consider the maximum deviation as a measure of robustness.

A special point of attention is the stochastic process by which successive lead times are generated in the simulation model. We assumed that orders may not cross. This is reasonable in practice, but it complicates simulation. If we simply generate independent lead times we cannot prevent that orders cross in general. Therefore we use the technique as described in Section 4.7. First, we split the lead time in a deterministic part and a stochastic part according to splitting rule (4.34). Next, we determine the parameters of the queuing model such that the sojourn time of a customer plus a fixed time has mean $E[L]$ and variance $\text{var}[L]$.

4.8.1 Experimental design for 2-echelon models

In the first experiment we test 2-echelon systems, in which one central depot (having index 1) supplies products to two so-called service groups. For sake of convenience we refer to these groups as service group A and B respectively. A service group consists of a number of local stockpoints with the same service, demand and lead time characteristics. Both service groups consist of three end-stockpoints, so we have six end-stockpoints in total. To normalize time and quantities, we made the following choices for all test runs:

- the review period equals $R = 1$.
- the mean demand per time unit for each local stockpoint in service group A equals $E[D_A] = 10$.

Unfortunately, the number of parameters that can be varied in our experiment is still quite large. To keep the size of the experiment within reasonable limits, we take the following parameters fixed:

- The mean lead time to the central depot equals $E[L_1] = 3$. Reason for this is that most upstream (production) lead times are usually larger than a review period. Besides, the experiments for deterministic lead times in Chapter 3 revealed that the accuracy of the approximation was better for $L_1 = 1$ than for $L_1 = 3$, so the latter should be tested.
- The downstream lead times are usually small, because these lead times represent usually order picking, handling and transport times. Therefore we take $E[L_j] = 1$ in all test runs.

Eight other parameters are varied in our experiment. We choose two different values for each parameter (see Table 4.3), except for the variation in the depot lead time $L_1$.

<table>
<thead>
<tr>
<th>parameter</th>
<th>description</th>
<th>values in test runs</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E[D_A]$</td>
<td>The mean demand per period at an end-stockpoint in group B</td>
<td>10, 30</td>
</tr>
<tr>
<td>$c(D_A)$</td>
<td>The coefficient of variation of demand per period at an end-stockpoint in group A</td>
<td>0.4, 0.8</td>
</tr>
<tr>
<td>$c(D_B)$</td>
<td>The coefficient of variation of demand per period at an end-stockpoint in group B</td>
<td>0.4, 0.8</td>
</tr>
<tr>
<td>$c(L_1)$</td>
<td>The coefficient of variation of $L_1$</td>
<td>0, 0.25, 0.5</td>
</tr>
<tr>
<td>$c(L_j)$</td>
<td>The coefficient of variation of $L_j$</td>
<td>0, 0.5</td>
</tr>
<tr>
<td>$\beta_A$</td>
<td>The target fill rate at an end-stockpoint in group A (%)</td>
<td>90, 99</td>
</tr>
<tr>
<td>$\beta_B$</td>
<td>The target fill rate at an end-stockpoint in group B (%)</td>
<td>90, 99</td>
</tr>
<tr>
<td>$a_1$</td>
<td>Constant, describing the level of stock at the central depot</td>
<td>0, 1.2</td>
</tr>
</tbody>
</table>

Table 4.3. Parameter values in the experiment with 2-echelon systems.

Since the variation of $L_1$ is largest in practice, we should carefully examine the effect of this variance. To choose the maximum amount of central stock $\Delta_1$, we proceed as follows. Equation (4.23) shows
that the amount of central stock heavily depends on \((\Delta_1 - D_1^*)^+\). Therefore it is convenient to express \(\Delta_1\) in the mean system demand during the lead time \(L_1\), say for some constant \(a_1\). We have significant central stock if \(a_1 > 1\), so we choose \(a_1 = 1.2\). Also, we consider the situation with a stockless central depot, i.e., \(a_1 = 0\). Using the value of the constant \(a_1\), we determine the appropriate value of \(\Delta_1\) for each case.

We tested all possible parameter combinations, yielding \(3 \times 27 = 384\) cases. The performance of the algorithm is tested by an extensive simulation of 75,000 time periods for each case to ensure high simulation accuracy.

### 4.8.2 Results for 2-echelon models

The accuracy of our algorithm is shown in the Tables 4.4–4.6 below. This accuracy is expressed as mean and maximum absolute deviation from target in per cents. Because a deviation from the target service level usually has more serious consequences in case of a high target service level, we separately give the rationing policy performance for each fill rate level in Table 4.4. Further, we show the performance for stockless and stock holding central depot separately in Table 4.5. Reason for this is that we may expect a better performance in the case of a stockless central depot, since we do not need a Taylor expansion then, cf. equations (4.9) and (4.19).

<table>
<thead>
<tr>
<th>Target fill rate ((\hat{\beta}^*))</th>
<th>Mean absolute deviation</th>
<th>Maximum absolute deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>90%</td>
<td>0.52%</td>
<td>1.80%</td>
</tr>
<tr>
<td>99%</td>
<td>0.41%</td>
<td>1.06%</td>
</tr>
<tr>
<td>ALL</td>
<td>0.45%</td>
<td>1.80%</td>
</tr>
</tbody>
</table>

**Table 4.4. Fill rate accuracy per target fill rate.**

<table>
<thead>
<tr>
<th>Central stock level</th>
<th>Mean absolute deviation</th>
<th>Maximum absolute deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a_1 = 0)</td>
<td>0.39%</td>
<td>1.06%</td>
</tr>
<tr>
<td>(a_1 = 1.2)</td>
<td>0.54%</td>
<td>1.80%</td>
</tr>
<tr>
<td>ALL</td>
<td>0.45%</td>
<td>1.80%</td>
</tr>
</tbody>
</table>

**Table 4.5. Fill rate accuracy per central stock level.**

Finally, we show the accuracy of the approximation for the system stock in Table 4.6, expressed as the mean and maximum relative deviation from simulation results. Here the system stock is defined as the sum of the mean physical stock in the central depot and the end-stockpoints plus the pipeline stock between central depot and end-stockpoints. The pipeline stock from the external supplier to the central depot is not included, because this is usual for external account.

<table>
<thead>
<tr>
<th>Central stock level</th>
<th>Mean absolute deviation</th>
<th>Maximum absolute deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a_1 = 0)</td>
<td>0.4%</td>
<td>1.6%</td>
</tr>
<tr>
<td>(a_1 = 1.2)</td>
<td>2.2%</td>
<td>5.0%</td>
</tr>
<tr>
<td>ALL</td>
<td>1.3%</td>
<td>5.0%</td>
</tr>
</tbody>
</table>

**Table 4.6. Stock accuracy per central stock level.**
4.8. Numerical validation

The overall results show that the accuracy is sufficient for practical applications, although the accuracy for the stock levels is clearly better if \( a_1 = 0 \). Of course, zero central stock is easy to 'approximate'. For more detailed numerical analysis we refer to Section 4.9.

4.8.3 Experimental design for 3-echelon models

To design an experiment for 3-echelon models, we proceed from the design that is developed in Chapter 3 for the situation with deterministic lead times. We consider a system consisting of one central depot, supplying 4 intermediate stockpoints. Each intermediate stockpoint supplies 6 end-stockpoints, so the system consists of 24 end-stockpoints totally. This is the largest system considered in Chapter 3, showing the highest approximation errors in the case of deterministic lead times. Because the accuracy is usually better for smaller systems, we chose this large 3-echelon system with 1+4+24=29 stockpoints.

To keep the number of test runs within reasonable limits, we take the following parameters fixed:

- the review period equals \( R = 1 \);
- the lead time to the central depot has mean \( E[L_1] = 3 \) and coefficient of variation \( c[L_1] = 0.5 \);
- the lead time to each intermediate stockpoint \( j \) from its supplier 1 has mean \( E[L_j] = 1 \) the lead time to each end-stockpoint \( k \) from its supplier \( j \) has mean \( E[L_k] = 1 \).

For the other parameters, we selected the values as shown in Table 4.7. For the demand and service characteristics, we used the experimental design as described in Section 3.5.1, where it is shown that 87 parameter sets are sufficient to cover this part of the design. We combined these 87 sets with all possible combinations of \( c[L_1] \), \( c[L_j] \), \( a_1 \) and \( a_j \), resulting in \( 24 \times 87 = 1392 \) test runs. The accuracy of our approximation is tested by a simulation of 25,000 time periods for each case.

<table>
<thead>
<tr>
<th>parameter</th>
<th>description</th>
<th>values in test runs</th>
</tr>
</thead>
<tbody>
<tr>
<td>( E[D_k] )</td>
<td>The mean demand per period at an end-stockpoint ( k )</td>
<td>10, 30</td>
</tr>
<tr>
<td>( c[L_1] )</td>
<td>The coefficient of variation of demand per period at an end-stockpoint ( k )</td>
<td>0.4, 0.8</td>
</tr>
<tr>
<td>( c[L_j] )</td>
<td>The coefficient of variation of lead time ( L_1 )</td>
<td>0, 0.5</td>
</tr>
<tr>
<td>( c[L_k] )</td>
<td>The coefficient of variation of lead time ( L_j )</td>
<td>0, 0.5</td>
</tr>
<tr>
<td>( \beta^* )</td>
<td>The target fill rate at an end-stockpoint (%)</td>
<td>90, 99</td>
</tr>
<tr>
<td>( a_1 )</td>
<td>Constant, describing the level of stock at the central depot</td>
<td>0, 1.2</td>
</tr>
<tr>
<td>( a_j )</td>
<td>Constant, describing the level of stock at intermediate stockpoint ( j )</td>
<td>0, 1.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>parameter</th>
<th>description</th>
<th>values in test runs</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a_j = \Delta_j / E[D_k] )</td>
<td>( a_j = \Delta_j / E[D_k] )</td>
<td>0, 1.2</td>
</tr>
</tbody>
</table>

Table 4.7. Parameter values in the experiment with 3-echelon systems.

4.8.4 Results for 3-echelon models

Similar to Section 4.8.2, we present results on the accuracy of our algorithm for the 3-echelon systems in the Tables 4.8-4.10.

We give the fill rate accuracy both per target fill rate (Table 4.8) and per intermediate stock level (Table 4.9). Although the results for the 3-echelon models are worse than for the 2-echelon models, we think that our method is still accurate enough for practical applications. This is especially true for systems with stockless intermediate stockpoints, such as the hierarchical planning procedure as described by De Koke (1990).
Table 4.8. Fill rate accuracy per target fill rate.

<table>
<thead>
<tr>
<th>Target fill rate ($p^*$)</th>
<th>Mean absolute deviation</th>
<th>Maximum absolute deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>90%</td>
<td>1.14%</td>
<td>2.48%</td>
</tr>
<tr>
<td>99%</td>
<td>0.75%</td>
<td>1.60%</td>
</tr>
<tr>
<td>ALL</td>
<td>0.95%</td>
<td>2.48%</td>
</tr>
</tbody>
</table>

Table 4.9. Fill rate accuracy per intermediate stock level.

<table>
<thead>
<tr>
<th>Intermediate stock levels</th>
<th>Mean absolute deviation</th>
<th>Maximum absolute deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(a_1, a_2) = (0, 0)$</td>
<td>0.78%</td>
<td>1.69%</td>
</tr>
<tr>
<td>$(a_1, a_2) = (0, 1.2)$</td>
<td>0.96%</td>
<td>2.48%</td>
</tr>
<tr>
<td>$(a_1, a_2) = (1.2, 0)$</td>
<td>1.01%</td>
<td>2.01%</td>
</tr>
<tr>
<td>$(a_1, a_2) = (1.2, 1.2)$</td>
<td>1.04%</td>
<td>2.40%</td>
</tr>
<tr>
<td>ALL</td>
<td>0.95%</td>
<td>2.48%</td>
</tr>
</tbody>
</table>

Table 4.10. Stock accuracy per intermediate stock level.

4.8.5 Computational effort

It appears that the approximate method is fast. The CPU time using a Pentium 100 MHz PC equals 0.13 seconds per case on average for the 2-echelon systems and 0.30 seconds on average for the 3-echelon systems. Little computational effort is important, because often control rules have to be established for hundreds or thousands of products, which in addition are not independent.

4.9 Sensitivity analysis

Now that a tool is available, it can be used for some sensitivity analysis to get some insight in the effects of lead time variation and autocorrelation. We focus on the following two questions:

(i) Which effect has lead time variation on the amount of stock required to obtain prespecified target service levels?

(ii) Is it really important to include correlations in the approximate method? Life becomes considerably easier if we ignore it, so can’t we keep it simple?
4.9. Sensitivity analysis

We will answer these questions by more detailed analysis of the results for 2-echelon models only to keep the results clear.

To start with the first question, we take a look at Table 4.11 in which the average amount of system stock is shown for various combinations of lead time variation. To make the various cases comparable, we expressed the average amount of system stock in mean demand of all end-stockpoints during one review period.

<table>
<thead>
<tr>
<th>c[L_1]</th>
<th>c[L_1] = 0</th>
<th>c[L_1] = 0.25</th>
<th>c[L_1] = 0.5</th>
<th>ALL</th>
</tr>
</thead>
<tbody>
<tr>
<td>c[L_1] = 0</td>
<td>3.4</td>
<td>3.7</td>
<td>3.5</td>
<td></td>
</tr>
<tr>
<td>c[L_1] = 0.25</td>
<td>4.1</td>
<td>4.4</td>
<td>4.3</td>
<td></td>
</tr>
<tr>
<td>c[L_1] = 0.5</td>
<td>5.4</td>
<td>5.6</td>
<td>5.5</td>
<td></td>
</tr>
<tr>
<td>ALL</td>
<td>4.3</td>
<td>4.6</td>
<td>4.4</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.11. Effect of lead time variation on the average amount of system stock (in mean demand of all end-stockpoints during one review period) required to satisfy the fill rate constraints.

We see that especially the upstream lead time variation has a significant impact on the amount of stock required to reach the target fill rates. The effect of the lead time variation between central depot and intermediate stockpoints is less, which can be partly explained by the fact that $E[L_1] = 3$ and $E[L_2] = 1$, so $var[L_1]$ is larger than $var[L_2]$. Anyway, we see that we can not ignore the lead time variation, because in practice significant lead time variation is present in the upstream part of the network.

Now the question arises what has more impact, demand variation or lead time variation? Usually the attention is focused on demand variation, but is this justified? This question is answered by looking at Table 4.12. Here we consider the most important source of lead time variation $c[L_1]$ and the various combinations of demand variation in the two service groups.

<table>
<thead>
<tr>
<th>$c[L_1]$</th>
<th>0.4, 0.4</th>
<th>0.4, 0.8</th>
<th>0.8, 0.8</th>
<th>ALL</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2.8</td>
<td>3.5</td>
<td>4.3</td>
<td>3.5</td>
</tr>
<tr>
<td>0.25</td>
<td>3.6</td>
<td>4.3</td>
<td>4.9</td>
<td>4.3</td>
</tr>
<tr>
<td>0.5</td>
<td>4.9</td>
<td>5.5</td>
<td>6.1</td>
<td>5.5</td>
</tr>
<tr>
<td>ALL</td>
<td>3.8</td>
<td>4.4</td>
<td>5.1</td>
<td>4.4</td>
</tr>
</tbody>
</table>

Table 4.12. Effect of lead time variation and demand variation on the average amount of system stock (in mean demand of all end-stockpoints during one review period) required to satisfy the fill rate constraints.

Table 4.12 shows that in our experiment the effect of lead time variation even dominates the effect of demand variation on the stock levels required to achieve the target fill rates! Although this can be parameter dependent, it is clear that both supply and demand variation should be taken into account when determining safety stock levels in multi-echelon divergent systems. This conclusion is in line with the results obtained by Gross & Soriani (1969) for a single location stockpoint.

The second question is important to judge whether our method can be simplified considerably by excluding the correlations, such as given by the equations (4.9), (4.19) and (4.28). We recalculated all the control parameters for all 384 cases, ignoring all these correlations. Also, we simulated the 384 cases with the modified control parameters. In Table 4.13 below we compare the accuracy of the simplified method to the original results. As can be seen from the formulae, this difference is only relevant if $c[L_1] > 0$. 

Performance measure | $c(L_1)$ | original | ignore correlations |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean absolute deviation from target fill rate</td>
<td>0.25</td>
<td>0.45%</td>
<td>1.39%</td>
</tr>
<tr>
<td>from target fill rate</td>
<td>0.5</td>
<td>0.67%</td>
<td>1.54%</td>
</tr>
<tr>
<td>Maximum absolute deviation from target fill rate</td>
<td>0.25</td>
<td>1.41%</td>
<td>4.03%</td>
</tr>
<tr>
<td>from target fill rate</td>
<td>0.5</td>
<td>1.80%</td>
<td>3.45%</td>
</tr>
</tbody>
</table>

Table 4.13. The effect of neglecting correlations.

As expected, we see that the performance of our method is better if we take into account the correlation effects. The price paid is that of considerable additional complexity.

4.10 Conclusions

In this chapter, we developed an algorithm to analyze multi-echelon divergent systems under periodic review echelon order-up-to-policies with both stochastic demand and lead times. Validation of our method by extensive comparison to simulation results, both for 2-echelon and for 3-echelon systems, shows that our algorithm is sufficiently accurate for practical applications. This is in particular true for situations where the intermediate stockpoints do not carry stocks and are only used as allocation points. An example is the hierarchical planning procedure as described in Chapter 1. Using our method, it is easy to include stochastic production lead times in this procedure, yielding accurate results.

Our method can be simplified by ignoring the (complicating) correlations involved, but at some loss of accuracy, as is shown by numerical experimentation. Further we showed the importance of including lead time variation in the model. Although frequently the attention is focused on demand variation, we showed that the effect of lead time variation may be larger than the effect of demand variation on the stock levels required to obtain prespecified target fill rates.
5

Cost minimization: optimal allocation functions

5.1 Introduction

In Chapter 2, we formulated a model description for a divergent multi-echelon inventory system. In this model each stockpoint employs an echelon order-up-to-policy. This means that every time a stockpoint \( j \) inspects its inventory it places an order to raise its echelon inventory position to the order-up-to-level \( S_j \). When its predecessor (denoted by index \( i \)) has insufficient material to satisfy the orders of its successors then material rationing is required. For that purpose we introduced the allocation function \( z_j(x) \). In Chapters 3 and 4, we assumed that the allocation function \( z_j(x) \) is given by

\[
z_j(x) := S_j - q_j \left( \sum_{i \in V_j} x_i \right), \quad j \in V.
\]

This simplifies the determination of the control parameters considerably, since besides the order-up-to-levels \( [S_j]_{i \in V_j} \) we only need to determine the allocation fractions \( [q_j]_{i \in V_j} \). In Chapters 3 and 4, we developed several algorithms to determine all these parameters such that the service level constraints at the end-stockpoints are satisfied. A prerequisite to apply these algorithms is that the mean amount of stock held at intermediate stockpoints (or more precise \( \Delta [i \in T_M] \)) is known. By varying the mean amount of stock at each intermediate stockpoint and using standard Newtonian search methods one could find the cost-optimal mean stock levels subject to the service level constraints. Only for a 2-echelon system such a study has been undertaken. For more than two echelons it becomes rather awkward and little insight could be obtained from such a study.

So far we concentrated on determining the control parameters such that the service level constraints at the end-stockpoints are satisfied. In this chapter, we focus on cost-optimal control policies. Our analysis will reveal that these cost-optimal policies can be found without the use of a Newtonian search method. Let us now address several important contributions in the field of determining cost-optimal control policies in a multi-echelon environment. The initial research on this field is generally attributed to Clark & Scarf (1960), who studied an \( N \)-echelon serial system operating under periodic review ordering policies. They applied the concept of echelon stock (cf. Clark (1958)) to prove that the optimal control policies for the \( N \)-echelon serial system with discounted penalty and holding costs, are characterized by \( N \) so-called echelon order-up-to-levels.
Another important contribution is the paper of Eppen & Schrage [1981], where also the problem of rationing is studied. Eppen & Schrage analyzed a divergent two-echelon system in which a central depot supplies multiple end-stockpoints. Some restrictive assumptions are made. First, the depot cannot hold any stock. Second, the proportional holding and penalty costs at the end-stockpoints are identical. Third, the demand per period at each end-stockpoint is normally distributed. Fourth, the lead time of the end-stockpoints are identical. Finally, each time an allocation is made, the depot receives enough material from the supplier to be able to allocate the material to each end-stockpoint so that an equal fraction is achieved on an appropriately chosen demand functions. This assumption is referred to as the allocation assumption. Under the aforementioned assumptions Eppen & Schrage derive an optimal order-up-to-policy at the depot, assuming no setup costs. In case of fixed setup costs, an approximately optimal policy is derived. Federgruen & Zipkin [1984] extend the model and results considered by Eppen & Schrage. Holding and penalty costs do not have to be identical across the end-stockpoints and period demands at the end-stockpoints do not have to be normally distributed, but may belong to a larger class of demand distributions (e.g., Erlang and gamma distribution). The allocation of material at the depot is determined by solving the myopic allocation problem. It minimizes the expected costs in the period that the allocation actually takes effect, ignoring costs in all subsequent periods. Federgruen & Zipkin used dynamic programming to prove that these myopic allocations are optimal given the allocation assumption. For an excellent overview concerning dynamic programming techniques to analyze multi-echelon systems we refer to the survey paper of Federgruen [1993].

In this chapter, more research is done to extend the existing results on these models in order to generalize them to divergent N-echelon systems, and to obtain insight in the structural properties of the cost-optimal control policy. Therefore, in this chapter we analyze the same model as formulated in Chapter 2, but in addition we introduce the following cost structure. Holding costs are incurred at every stockpoint just before an order arrival, and in case of an end-stockpoint penalty costs are incurred when this stockpoint is in a backlog position. Note that intermediate stockpoints cannot be in a backlog position. The objective is to determine the control parameters such that the average costs per period in the long run is minimized. The model can be regarded as an extension of those by Langenhoff & Zijm [1990] and Van Houtum & Zijm [1991b]. Van Houtum & Zijm proved exact decomposition results for a divergent 2-echelon system given the balance assumption. A detailed review of this work, together with other related papers, is given by Van Houtum, Inderfurth & Zijm [1996]. In this chapter we prove that decomposition is exact for a divergent N-echelon system given the balance assumption. Decomposition reduces the complex control problem to: (1) determining the optimal order-up-to-level at each stockpoint, and (2) determining the optimal allocation functions at each intermediate stockpoint. We used the result of Gong, De Kok & Ding [1994] to prove that the first problem coincides with the classical newsvendor problem. Similar newsvendor-type results are derived in Rogers & Tsaihakitarian [1991]. They considered the optimization problem of minimizing the penalty costs in a divergent 2-echelon system, subjected to a budget constraint on the total holding costs. The second problem is solved by applying the Lagrangian multiplier technique. From this we obtain insight in the structure of the cost-optimal allocation functions. An algorithm is developed to actually determine these functions. Based on several properties of the optimal allocation functions we classify them into four classes.

The remainder of the chapter is organized as follows. In Section 5.2, we briefly describe the model under consideration. In Section 5.3, we decompose the system such that we obtain a tractable expression for the average holding and penalty costs per period as a function of the employed control policy. Also some important properties of this cost function are derived. In Section 5.4, we use these properties
5.2 Model description

We consider the same divergent multi-echelon inventory system as introduced in Chapter 2 (see e.g. Figure 2.1). So every $R$ periods the most upstream stockpoint (having index 1) issues a replenishment order. The replenishment order arrives after $L_1$ periods, where $L_1$ is fixed. Then the physical stock at stockpoint 1 is allocated immediately to its successors. There are two possibilities:

(i). The physical stock is sufficient to raise the echelon inventory position of each successor to its order-up-to-level. Then the required amounts are sent to the successors and excess stock is kept at stockpoint 1 to be allocated in the next occasion.

(ii). The physical stock is not sufficient to reach the order-up-to-levels. Then the available physical stock is allocated to its successors by using the allocation functions $\{z_j\}_{j \in J_1}$.

Besides the assumptions introduced in Section 2.3 we also make the balance assumption in the remainder of this chapter. This balance assumption guarantees that the rationing policy always allocates nonnegative stock quantities. For an extensive discussion on these assumptions we refer to the Sections 2.3 and 2.5.

For presentational purposes we refer to the length of one review period as one period ($R = 1$), and assume that the lead time of each stockpoint $i$ is a multiple integer of $R$. In Section 5.6.1 we prove that this latter assumption can be relaxed. By assuming that $L_i$ is a multiple integer of $R$, we can synchronize the order arrivals and the times that the costs are incurred (just before the order arrivals). At the end of each period both penalty and holding costs are incurred. The penalty costs equal $p_i$ for each backordered product at end-stockpoint $i$. The holding costs equal $h_i + \sum_{j \in J_i} h_{ij}$ for a product at stockpoint $i$ or in transfer to one of its successors. Note that $h_i$ can be regarded as an additional holding cost due to value added in stockpoint $i$. No fixed ordering costs are assumed. The objective of the analysis is to determine a cost-optimal replenishment policy. That is a policy, which minimizes the expected total costs on the long run.

5.3 Analysis cost function

In this section, we present an average cost analysis for the divergent $N$-echelon system, which is based on the analysis in Langenhouff & Zijm [1990] and Van Houtum & Zijm [1991a]. First, we derive an expression for the cost function, which yields the average system costs as a function of the employed control policy (Section 5.3.1). To minimize the cost function we first derive important properties of: (1) the allocation function (Section 5.3.2), and (2) the cost function (Section 5.3.3). These properties are used in Section 5.4 to characterize the optimal control policy given decomposition. Also we apply these properties in Section 5.5 to prove that decomposition is exact (given the balance assumption), i.e., it results in cost-optimal control policies. Besides the notation introduced in Section 2.2 we introduce some additional notation in Table 5.1 with respect to the analysis of the system.
Cost minimization: optimal allocation functions

\( \psi_i \) All control parameters downstream of stockpoint \( i \), i.e.,
\[ \emptyset \cup_{j \in \mathcal{V}_i} (z_j, S_j, \psi_j) \quad i \in \mathcal{M}. \]

\( D_i(S_i, \psi_i) \) The expected total costs in \( \{\text{ech}(i)\} \) at the end of an arbitrary period given the echelon order-up-to-level \( S_i \) and \( \psi_i \). (If \( \mathcal{W}_i = \emptyset \) we suppress \( \psi_i \).

\( F_{i}^{(L)}/f_{i}^{(L)} \) cdf/pdf of demand at all end-stockpoints in \( E_i \) during \( L \) periods (if \( L = 1 \) we suppress the index).

\[
\begin{cases}
\left( F_{i}^{(L)}(x) \right) & x < 0 \\
\left( F_{i}^{(L)}(x + \Delta) \right) & x \geq 0.
\end{cases}
\]

\( p_i \) Minimum penalty costs of each stockpoint \( j \in E_i \), i.e., \( \min_{j \in E_i} p_j \).

\( h_i \) Minimum value added at a successor of stockpoint \( i \), i.e., \( \min_{j \in \mathcal{V}_i} h_j \).

\( \mathcal{U}_i \) All successors of stockpoint \( i \), which have the same minimum penalty costs as stockpoint \( i \), i.e., \( \{ j \in \mathcal{V} | p^*_j = p_i \} \).

\( \mathcal{V}_i \) All successors of stockpoint \( i \) for which the minimum penalty is added to a product, i.e., \( \{ j \in \mathcal{V} | h^*_j = h_i \} \).

\textbf{Table 5.1.} Notation with respect to the analysis of the system.

5.3.1 Cost function

At the end of each period holding costs and penalty costs are incurred. The holding costs per product at stockpoint \( i \) or towards one of its successors equals \( h_i + \sum_{a \in U_i} h_a \), and the penalty costs per backlogged product at end-stockpoint \( j \) equals \( p_j \). Then the total costs incurred at time \( t \) (at the end of a period) equal

\[
\sum_{i \in \mathcal{M}} (h_i + \sum_{a \in U_i} h_a) (J_i^t - \sum_{j \in \mathcal{V}_i} J_j^t) + \sum_{i \in \mathcal{M}} (h_i + \sum_{a \in U_i} h_a) (J_i^t)^+ + p_i (-J_i^t)^+ .
\]

In Lemma 5.1 we rewrite the expression above to obtain an expression which associates costs to the echelon stock at each stockpoint.

**Lemma 5.1.** When at the end of an arbitrary period, at time \( t \) say, the echelon stock of a stockpoint \( i \) equals \( J_i^t \), the total costs incurred at the end of this period equal

\[
\sum_{i \in \mathcal{M}} (h_i J_i^t) + (p_i + h_i + \sum_{a \in U_i} h_a) (-J_i^t)^+ = \sum_{i \in \mathcal{M}} h_i J_i^t.
\]

**Proof.** In order to prove this theorem we use that if \( i \in \mathcal{M} \) then

\[
(h_i + \sum_{a \in U_i} h_a) J_i^t = \sum_{j \in \mathcal{V}_i} (h_i + \sum_{a \in U_i} h_a) J_j^t = \sum_{j \in \mathcal{V}_i} \sum_{a \in U_j} h_a J_j^t .
\]

From (5.1) and the property that for the most upstream stockpoint \( 1 \) we have \( U_1 = \emptyset \), it follows that

\[
\sum_{i \in \mathcal{M}} (h_i + \sum_{a \in U_i} h_a) J_i^t = \sum_{i \in \mathcal{M} \cup \{1\}} \sum_{a \in U_i} h_a J_i^t .
\]

\[\text{5.2}\]
5.3. Analysis cost function

From the definition of the inventory costs \( h_i \) and penalty costs \( p_i \), it easily follows that the total costs equals

\[
\sum_{i \in E} \left( (h_i + \sum_{n \in \mathcal{L}_i} h_n)(J_i')^+ + p_i(-J_i')^+ \right) + \sum_{i \in M} (h_i + \sum_{n \in \mathcal{L}_i} h_n)(J_i - \sum_{j \in \mathcal{L}_i} J_j') =
\]

\[
\sum_{i \in \mathcal{L}_i} \left( (h_i + \sum_{n \in \mathcal{L}_i} h_n)J_i' + (h_i + \sum_{n \in \mathcal{L}_i} h_n + p_i)(-J_i')^+ \right) + \sum_{i \in M} h_i J_i'
\]

\[
\sum_{i \in \mathcal{L}_i} \sum_{n \in \mathcal{L}_i} h_n J_i' - \sum_{i \in M} (h_i + \sum_{n \in \mathcal{L}_i} h_n) \sum_{j \in \mathcal{L}_i} J_j' \quad \text{(5.2)}
\]

\[
\sum_{i \in \mathcal{L}_i} \left( (h_i + \sum_{n \in \mathcal{L}_i} h_n)J_i' + (h_i + \sum_{n \in \mathcal{L}_i} h_n + p_i)(-J_i')^+ \right) + \sum_{i \in M} h_i J_i' + \sum_{i \in \mathcal{L}_i} \sum_{n \in \mathcal{L}_i} h_n J_i' - \sum_{i \in \mathcal{L}_i} \sum_{n \in \mathcal{L}_i} h_n J_i'.
\]

Rewriting the expression above completes the proof. \( \square \)

Lemma 5.1 constitutes the basis for the decomposition. This lemma implies that costs \( h_i J_i' \) are incurred at each stockpoint \( i \) (independent of the sign of \( J_i' \)), whereas at an end-stockpoint extra costs are incurred when it is in a backlog position. These extra costs are penalty costs \( p_i \) per backlogged product, and holding costs \( h_i + \sum_{n \in \mathcal{L}_i} h_n \) per backlogged product. These extra holding costs are incurred, since the echelon stock of an intermediate stockpoint \( j \) (for which \( i \in E_j \)) does not coincide with the number of products in \( \text{ech}(j) \) due to the backorders at the end-stockpoints \( i \in E_j \).

In order to relate the inventory of a stockpoint at the beginning of a period with the costs incurred at the end of this period we define the one-period cost function \( C_i(x) \).

\[ C_i(x) := \text{The end of period expected costs incurred at stockpoint } i, \text{ when at the beginning of this period the echelon stock of stockpoint } i \text{ is increased to } x. \]

To evaluate \( C_i(x) \) we assume that the cdf of the total demand at all end-stockpoints in \( E_i \) during \( L \) periods is given by \( F_i^{(L)} \) (if \( L = 1 \), we suppress the index). From Lemma 5.1 it follows that

\[ C_i(x) := \begin{cases} 
\int_0^x h_i(x - u) \, dF_i(u) + \int_x^\infty (h_i + \sum_{n \in \mathcal{L}_i} h_n + p_i)(u - x) \, dF_i(u) & i \in E \\
\int_0^x h_i(x - u) \, dF_i(u) & i \in M.
\end{cases} \]

Before computing the expected costs of echelon \( i \) at the end of an arbitrary period, denoted by \( D_i(S_i, \Psi_i) \), we have to decide how to ration the available stock when a stockpoint has insufficient stock to satisfy all demand. Suppose that at the beginning of an arbitrary period (just before rationing), stockpoint \( i \) has an echelon stock of \( x \) products. Each successor \( j \in \mathcal{L}_i \) wants to raise its echelon inventory position to \( S_j \). Hence, if \( x \geq \sum_{n \in \mathcal{L}_i} S_n \), the echelon inventory position of stockpoint \( j \) just after rationing becomes \( S_j \), and the remainder \( x - \sum_{n \in \mathcal{L}_i} S_n \) is retained at stockpoint \( i \). However, if \( x < \sum_{n \in \mathcal{L}_i} S_n \), we have to deal with the main difficulties of divergent multi-echelon systems: how should we ration the available stock over these successors? For that purpose we introduced the allocation function \( z_j(x) \) in Section 2.4. This means that the rationing policy allocates the products such that the echelon inventory position of stockpoint \( j \) after rationing equals \( z_j(x) \). Since no products are retained at stockpoint \( i \), we have

\[ \sum_{n \in \mathcal{L}_i} z_n(x) = x. \quad (5.3) \]
Unlike the previous Chapters 3 and 4 we like to determine the optimal allocation function, and thus do not require a specific form of \( z_j(x) \). What we mean with optimal is explained in the next section. This means that when a stockpoint \( i \) rations the available echelon stock \( x \) over its successors \( j \in V_i \), the echelon inventory position of every stockpoint \( j \) just after rationing is at least as large as it was just before rationing. The next lemma enables us to determine the expected costs in echelon \( i \) per period, denoted by \( D_i(S_i, \Psi_i) \), provided that stockpoint \( i \) is always able to raise its echelon inventory position to \( S_i \) and that the control parameters downstream of stockpoint \( i \) are given by \( \Psi_i \).

**Lemma 5.2.** For replenishment policy \((S_i, \Psi_i)\) the expected costs in echelon \( i \) per period equals

\[
D_i(S_i, \Psi_i) = \int_0^\infty C_i(S_i - u) \, dF_i^{(i)}(u) + \sum_{j \in V_i} \left[ \int_b^{\infty} D_j(S_j, \Psi_j) \, dF_j^{(j)}(u) + \int_{S_j}^{\infty} D_j(z_j(S_i - u), \Psi_j) \, dF_j^{(j)}(u) \right].
\]  

(5.4)

**Proof.** Consider a replenishment policy \((S_i, \Psi_i)\). Suppose that at the beginning of an arbitrary review period \( t \) the echelon inventory position of stockpoint \( i \) is raised to \( S_i \) and the total demand in the periods \( t \) up to and including period \( t + L_i - 1 \) equals \( u \). Hence at the end of period \( t + L_i - 1 \) (just after order arrival) the echelon stock of stockpoint \( i \) equals \( S_i - u \). Therefore at time \( t + L_i \) the expected costs incurred for echelon \( i \) equals \( C_i(S_i - u) \).

If at the beginning of period \( t + L_i \) it holds that \( S_i - u < \sum_{j \in V_i} S_j \), every stockpoint \( j \in V_i \) raises its echelon inventory position to \( S_j \). Therefore at time \( t + L_i + L_j \) the expected costs incurred for echelon \( j \) equals \( D_j(S_j, \Psi_j) \).

However, if at the beginning of period \( t + L_i \) we have \( S_i - u < \sum_{j \in V_i} S_j \), every stockpoint \( j \in V_i \) raises its echelon inventory position to \( z_j(S_i - u) \). Therefore at time \( t + L_i + L_j \) the expected costs incurred for stockpoint \( j \) equals \( D_j(z_j(S_i - u), \Psi_j) \).

By conditioning on \( u \) and using the above mentioned relations the theorem follows. \( \square \)

In the following analysis, we assume that \( F_i(x) \) is a differentiable cdf for \( x \geq 0 \). Hence, from Lemma 5.2 it follows that \( D_i(S_i, \Psi_i) \) is also a differentiable function with respect to \( S_i \). Substitution of the definition of \( C_i(x) \) in (5.4) yields

\[
D_i(S_i) = h_i(S_i - (L_i + 1)u_i) + \int_{S_i}^{\infty} \left( h_i + \sum_{j \in U_i} h_j + p_j(u - S_i) dF_j^{(j + 1)}(u) \right), \quad i \in E
\]

(5.5a)

\[
D_i(S_i, \Psi_i) = h_i(S_i - (L_i + 1)u_i) + \sum_{j \in V_i} \left[ D_j(S_j, \Psi_j) + \int_0^{\infty} \left( D_j(z_j(S_i - u), \Psi_j) - D_j(S_j, \Psi_j) \right) dF_j^{(j + 1)}(u) \right], \quad i \in M.
\]

(5.5b)

The objective of this chapter is to determine the cost-optimal control policy \((\hat{S}_i, \hat{\Psi}_i)\) which minimizes \( D_i(S_i, \Psi_i) \) (where index \( i \) denotes the most upstream stockpoint). To achieve this we propose system decomposition, as in Langenhofer & Zijm [1990] and Van Houtum & Zijm [1991b]. Applying system decomposition means that after the determination of the downstream control parameters, we first determine the allocation functions \((z_j)_{j \in V_i}\) minimizing \( D_i(S_i, \Psi_i) \) (given \((S_j, \Psi_j)_{j \in V_i}\)), and next determine the order-up-to-level \( S_j \) which minimizes \( D_i(S_i, \Psi_i) \) (given \( \Psi_i \)). In Section 5.3.2 we derive
5.3. Analysis cost function

several properties of the allocation functions minimizing \( D_s(S_j, \Psi_j) \). These properties enable us to develop an algorithm in Section 5.4.1 to determine these allocation functions. Furthermore, we make a classification of the allocation functions based on their properties. In Section 5.3.3 we concentrate on the determination of a tractable expression for \( \frac{\partial D_s(S_j, \Psi_j)}{\partial S_j} \). From this derivative we determine in Section 5.4.2 the order-up-to-level which minimizes \( D_s(S_j, \Psi_j) \). Also we derive some additional properties of this cost function. These are required for the proof that decomposition is exact (see Section 5.5).

5.3.2 Properties of the allocation function

For our convenience we introduce some definitions.

**Definition 5.1.** \( z_j \) is optimal (given \( \Psi_j \)) when \( D_j(z_j(x), \Psi_j) \) is minimized for all \( x \).

\( S_j \) is optimal (given \( \Psi_j \)) when \( S_j = \arg \min \{ S | \frac{\partial D_j(S, \Psi_j)}{\partial S} = 0 \} \).

**Definition 5.2.** \( \Psi_j := \bigcup_{j \in V} (z_j, S_j, \Psi_j) \) is locally optimal when for every \( j \in V \):

(i) the allocation function \( z_j \) is optimal (given \( \Psi_j \)).

(ii) the order-up-to-level \( S_j \) is optimal (given \( \Psi_j \)).

**Definition 5.3.** \( \Psi_j \) is optimal, if \( \Psi_j \) is locally optimal for every \( j \in \text{ech}(i) \).

In Definition 5.1, we defined the order-up-to level \( S_j \), however, it is unclear whether there exists such an optimal \( S_j \). Moreover, if indeed \( S_j \) exists it is unclear whether \( S_j \) is a minimum or a maximum. In Section 5.3.3, we prove that \( S_j \) exists and is a global minimum. Next, in Lemma 5.3, we present a very important property of the optimal allocation function \( z_j \), which is used throughout this chapter.

**Lemma 5.3.** The optimal allocation functions \( \{ z_j \}_{j \in V} \) minimize \( D_j(S_j, \Psi_j) \) if there exists a function \( \lambda_j(x) \) such that

\[
\frac{\partial D_j(S_j, \Psi_j)}{\partial S_j} = \lambda_j(x) \quad \text{for every} \ j \in V \text{ and } x \in \mathbb{R}.
\]  

(5.6)

**Proof:** We give an outline of the proof. When the echelon stock of stockpoint \( i \), say \( x \), is not sufficient to raise the echelon inventory positions of all successors \( j \in V \) to their echelon order-up-to-level \( S_j \), the echelon stock \( x \) is rationed over these successors. Lemma 5.2 implies that in order to do this optimally we have to solve

\[
\min_{j \in \mathcal{R}} \sum_{j \in \mathcal{R}} D_j(z_j(x), \Psi_j) \quad \text{s.t.} \quad \sum_{j \in \mathcal{R}} z_j(x) = x,
\]

where \( \mathcal{R} \) represents the set of all possible allocation functions.

Using the Lagrange multiplier technique yields that the allocation functions \( \{ z_j \}_{j \in V} \) can only be optimal when (5.6) holds.

**Lemma 5.3** proves that the derivative of \( D_j(S_j, \Psi_j) \) with respect to \( S_j \) in the point \( S_j = z_j(x) \) is independent of stockpoint \( j \). This is very important in order to characterize the cost optimal control policy. With Lemma 5.3, we are able to derive several properties of \( \{ z_j \}_{j \in V} \). For example, in Lemma 5.4, we prove that every optimal allocation function \( z_j \) is non-decreasing.

**Lemma 5.4.** If for every stockpoint \( j \in V \) the cost function \( D_j(S_j, \Psi_j) \) is convex with respect to \( S_j \), then there exists a set of optimal allocation functions \( \{ z_j \}_{j \in V} \) such that

\[
\frac{d^2 z_j(x)}{dx^2} \geq 0 \quad \text{for all} \ x, \ j \in V.
\]
Specifically, if for every stockpoint $j \in V_i$ the cost function $D_j(S_j, \Psi_j)$ is linearly decreasing for $S_j < x_j^*$ and strictly convex for $S_j \geq x_j^*$, then for the set of optimal allocation functions it holds that

$$\frac{d\tilde{z}_j(x)}{dx} > 0 \quad \text{for} \quad x \geq x_j^*, \quad j \in V_i,$$

where $x_j^*$ equals 0 for each end-stockpoint $i$, and equals $\min\{x|\tilde{z}_j(x) \geq x_j^*\}$ for $j \in V_i$ for each intermediate stockpoint $i$.

Proof. Let $D_j(S_j, \Psi_j)$ be convex with respect to $S_j$ for every $j \in V_i$. Assume for every set of optimal allocation functions $\{\hat{z}_j\}_{j \in V_i}$ there exists an $m \in V_i$ such that $d\tilde{z}_m(x)/dx < 0$ for some $x$. Then there exists an $\epsilon > 0$ such that $\tilde{z}_m(x+\epsilon) < \tilde{z}_m(x)$. Let us distinguish between two cases:

- There exists a stockpoint $k \in V_i$ such that

$$\frac{\partial D_k(S_k, \Psi_k)}{\partial S_k} |_{S_k = \hat{z}_k(x+\epsilon)} > \frac{\partial D_k(S_k, \Psi_k)}{\partial S_k} |_{S_k = \hat{z}_k(x)}.$$

From Lemma 5.3 it follows that

$$\frac{\partial D_j(S_j, \Psi_j)}{\partial S_j} |_{S_j = \hat{z}_j(x+\epsilon)} > \frac{\partial D_j(S_j, \Psi_j)}{\partial S_j} |_{S_j = \hat{z}_j(x)} \quad \text{for every} \quad j \in V_i.$$

Since $D_j(S_j, \Psi_j)$ is convex with respect to $S_j$ we know that $\tilde{z}_j(x+\epsilon) > \tilde{z}_j(x)$ for every $j \in V_i$. This contradicts our earlier made assumption.

- For every successor $j \in V_i$ holds

$$\frac{\partial D_j(S_j, \Psi_j)}{\partial S_j} |_{S_j = \hat{z}_j(x+\epsilon)} \leq \frac{\partial D_j(S_j, \Psi_j)}{\partial S_j} |_{S_j = \hat{z}_j(x)}.$$  \hfill (5.7)

Suppose there exists a stockpoint $k \in V_i$ such that

$$\frac{\partial D_j(S_j, \Psi_j)}{\partial S_j} |_{S_j = \hat{z}_j(x+\epsilon)} < \frac{\partial D_j(S_j, \Psi_j)}{\partial S_j} |_{S_j = \hat{z}_j(x)}.$$  \hfill (5.7*)

Then applying Lemma 5.3 using the convexity of $D_j(S_j, \Psi_j)$ yields $\tilde{z}_j(x+\epsilon) < \tilde{z}_j(x)$ for every $j \in V_i$. So $\sum_{j \in V_i} \tilde{z}_j(x+\epsilon) < \sum_{j \in V_i} \tilde{z}_j(x)$. This contradicts (5.3), which states $\sum_{j \in V_i} \tilde{z}_j(x+\epsilon) = x + \epsilon > x = \sum_{j \in V_i} \tilde{z}_j(x)$. Hence, equation (5.7) reduces to

$$\frac{\partial D_j(S_j, \Psi_j)}{\partial S_j} |_{S_j = \hat{z}_j(x+\epsilon)} = \frac{\partial D_j(S_j, \Psi_j)}{\partial S_j} |_{S_j = \hat{z}_j(x)} \quad \text{for} \quad j \in V_i.$$

Next, we construct a set of optimal allocation functions $\{\tilde{z}_j\}_{j \in V_i}$, for which $\tilde{z}_j(x+\epsilon) \geq \tilde{z}_j(x)$ for every $j \in V_i$. This contradicts our earlier made assumption that for every set of optimal allocation functions $\{\tilde{z}_j\}_{j \in V_i}$ there exists an $m \in V_i$ and $\epsilon > 0$ such that $\tilde{z}_m(x+\epsilon) < \tilde{z}_m(x)$. Let us define the allocation functions $\tilde{z}_j$ identical to $\hat{z}_j$ for every $j \in V_i$, except in $x + \epsilon$:

$$\tilde{z}_j(x+\epsilon) := \begin{cases} \hat{z}_j(x) & j \in A_i, \\ \hat{z}_j(x) + q_j \epsilon & j \in B_i \end{cases}$$

with

$$A_i := \{ j \in V_i | \tilde{z}_j(x+\epsilon) \leq \hat{z}_j(x) \}, \quad B_i := \{ j \in V_i | \tilde{z}_j(x+\epsilon) > \hat{z}_j(x) \},$$

and

$$q_j := \sum_{k \in A_i} (\tilde{z}_k(x+\epsilon) - \hat{z}_k(x)) > 0, \quad j \in B_i.$$
5.3. Analysis of cost function

From $\xi_n(x + \epsilon) < \xi_n(x)$ and (5.3) we conclude that both $A_i$ and $B_i$ are non-empty sets. We prove that $\xi_j$ is an optimal allocation function in $x + \epsilon$. First, it can be shown that $\sum_{j \in V_i} \xi_j(x + \epsilon) = x + \epsilon$.

Next, we prove that equation (5.6) is satisfied in $x + \epsilon$.

(1) We have

$$\frac{\partial D_j(S_i, \Psi_j)}{\partial S_j} |_{S_j = \xi_j(x + \epsilon)} = \frac{\partial D_j(S_j, \Psi_j)}{\partial S_j} |_{S_j = \xi_j(n)} , \quad j \in A_i.$$  

(2) We have

$$\frac{\partial D_j(S_i, \Psi_j)}{\partial S_j} |_{S_j = \xi_j(x + \epsilon)} = \frac{\partial D_j(S_j, \Psi_j)}{\partial S_j} |_{S_j = \xi_j(x)} , \quad j \in B_i.$$  

Since $D_j(S_i, \Psi_j)$ is convex with respect to $S_j$ and $\xi_j(x) < \xi_j(x + \epsilon) < \xi_j(x + \epsilon)$ we know from (5.7)

$$\frac{\partial D_j(S_i, \Psi_j)}{\partial S_j} |_{S_j = \xi_j(x + \epsilon)} = \frac{\partial D_j(S_j, \Psi_j)}{\partial S_j} |_{S_j = \xi_j(x)}.$$  

So, $(\xi_j)_{j \in V_i}$ is a set of optimal allocation functions. However, $\xi_j(x + \epsilon) \geq \xi_j(x)$ for all $j \in V_i$ contradicts our earlier made assumption.

Assume for every stockpoint $j \in V_i$ the cost function $D_j(S_j, \Psi_j)$ is linear decreasing for $S_j < x_j^*$ and strictly convex for $S_j \geq x_j^*$. Let $x_j^* := \min \{ x | \xi_j(x) \geq x_j^* \}$ for $j \in V_i$. Now we prove that $\xi_j(x)$ is strictly increasing for $x \geq x_j^*$. Assume for a stockpoint $m \in V_i$ holds $d\xi_m(x)/dx \leq 0$ for a $x \geq x_j^*$. Then there exists an $\epsilon > 0$ for which $\xi_m(x + \epsilon) \leq \xi_m(x)$. Since $D_m(S_m, \Psi_m)$ is a convex function, we obtain

$$\frac{\partial D_m(S_m, \Psi_m)}{\partial S_m} |_{S_m = \xi_m(x + \epsilon)} \leq \frac{\partial D_m(S_m, \Psi_m)}{\partial S_m} |_{S_m = \xi_m(x)}.$$  

Using Lemma 5.3 and (5.8) we obtain,

$$\frac{\partial D_j(S_j, \Psi_j)}{\partial S_j} |_{S_j = \xi_j(x + \epsilon)} \leq \frac{\partial D_j(S_j, \Psi_j)}{\partial S_j} |_{S_j = \xi_j(x)} \quad \text{for every} \quad j \in V_i.$$  

Since $x \geq x_j^*$ and $\xi_j(x)$ is non-decreasing in $x$ it follows $\xi_j(x) \geq x_j^*$. The cost function $D_j(S_j, \Psi_j)$ is strictly convex in $S_j = \xi_j(x) \geq x_j^*$. Hence, $\xi_j(x + \epsilon) \geq \xi_j(x)$. So, $\sum_{j \in V_i} \xi_j(x + \epsilon) \leq \sum_{j \in V_i} \xi_j(x)$. This contradicts (5.3), which states $\sum_{j \in V_i} \xi_j(x + \epsilon) = x + \epsilon > x = \sum_{j \in V_i} \xi_j(x)$. Hence $d\xi_m(x)/dx > 0$ for $x \geq x_j^*$.

Besides the fact that Lemma 5.4 states a very important property of an optimal allocation function (i.e., it is a non-decreasing function), this lemma is also required to prove that the cost function $D_j(S_j, \Psi_j)$ is convex with respect to $S_j$ (see Theorem 5.4). Another important property of an optimal allocation function is given in the next lemma.

Lemma 5.5. If for every stockpoint $j \in V_i$, the cost function $D_j(S_j, \Psi_j)$ is convex with respect to $S_j$, then for a local optimal $\Psi_i = \bigcup_{j \in V_i} \{ \xi_j, S_j \}$ it holds that $\xi_j(S_i) = S_j$.

Proof: This proof is by contradiction. Assume $D_j(S_j, \Psi_j)$ is convex with respect to $S_j$ for every $j \in V_i$. Suppose there exists a stockpoint $k \in V_i$ for which $\xi_k(S_i) < S_k$. From the definition of $S_k$ and the convexity of $D_k(S_k, \Psi_k)$, we conclude

$$\frac{\partial D_k(S_k, \Psi_k)}{\partial S_k} |_{S_k = \xi_k(S_k)} < \frac{\partial D_k(S_k, \Psi_k)}{\partial S_k} |_{S_k = S_k} = 0.$$  


Hence, from Lemma 5.3 it follows that \( \lambda_i(\sum_{i \in V} \hat{S}_i) < 0 \). Since \( \sum_{i \in V} \hat{S}_i(\sum_{i \in V} \hat{S}_i) = \sum_{i \in V} \hat{S}_i \), there has to be a stockpoint, \( m \) say, for which \( \hat{z}_m(\sum_{i \in V} \hat{S}_i) > \hat{S}_m \). Using similar arguments as above it can be shown that \( \lambda_i(\sum_{i \in V} \hat{S}_i) > 0 \). This leads to a contradiction. Hence, \( \hat{z}_i(\sum_{i \in V} \hat{S}_i) > \hat{S}_i \).

Analogous we can prove \( \hat{z}_i(\sum_{i \in V} \hat{S}_i) \leq \hat{S}_i \). Thus \( \hat{z}_i(\sum_{i \in V} \hat{S}_i) = \hat{S}_i \). \( \square \)

If the echelon stock of stockpoint \( i \) drops just below \( \sum_{i \in V} \hat{S}_i \), say \( \sum_{i \in V} \hat{S}_i - \epsilon \), then rationing is required. Hence, the echelon inventory position of a successor \( j \in V_i \) (just after rationing) equals \( \hat{z}_j(\sum_{i \in V} \hat{S}_i - \epsilon) \). However, if the echelon stock of stockpoint \( i \) equals or exceed \( \sum_{i \in V} \hat{S}_i \), then the echelon stock position of successor \( j \) (just after rationing) equals \( \hat{S}_j \). From Lemma 5.5 we know that \( \lim_{\epsilon \to 0} \hat{z}_j(\sum_{i \in V} \hat{S}_i - \epsilon) = \hat{S}_j \), so the continuity of the number of products allocated to successor \( j \) is guaranteed. Furthermore, this lemma is used throughout the chapter. It plays an important role in the proof of Theorem 5.4 (to prove the convexity of the cost function).

### 5.3.3 Properties of the cost function

So far we concentrated on characterizing the optimal allocation function. In this section, we focus on obtaining properties of the cost function \( D_i(S_i, \Psi_i) \). To determine the optimal order-to-level \( \hat{S}_i \), we need a tractable expression for \( \frac{\partial D_i(S_i, \Psi_i)}{\partial S_i} \). In Theorem 5.1, we derive such an expression. With the result presented in Theorem 5.2 we can simplify this expression considerably (see Theorem 5.3). Finally, in Theorem 5.4, we derive several properties (e.g., convexity) of the cost function to characterize the behavior of an optimal allocation function (Section 5.4.2), and to prove that system decomposition is exact (Section 5.5).

**Theorem 5.1.** For every stockpoint \( i \in M \) with a locally optimal \( \Psi_i \) it holds that

\[
\frac{\partial D_i(S_i, \Psi_i)}{\partial S_i} = h_i + \int_0^\infty \lambda_i(S_i - u) dF_i^{(L)}(u), \quad \text{for } \Delta_i < 0,
\]

\[
\frac{\partial D_i(S_i, \Psi_i)}{\partial S_i} = h_i + \int_0^\infty \lambda_i(S_i - u - \Delta_i) d\left( F_i^{(L)} \right)_\Delta(u), \quad \text{for } \Delta_i \geq 0,
\]

where \( \lambda_i(x) \) is defined in (5.6).

**Proof.** From (5.5) we obtain

\[
\frac{\partial D_i(S_i, \Psi_i)}{\partial S_i} = \frac{\partial}{\partial S_i} \left[ h_i(S_i - (L_i + 1)\mu_i + \sum_{j \in V_i} D_j(S_j, \Psi_j) + \int_{\Delta_i}^{\infty} D_j(\hat{z}_j(S_i - u), \Psi_j) - D_j(S_j, \Psi_j) dF_i^{(L)}(u) \right]
\]

\[
= h_i + \sum_{j \in V_i} \int_{\Delta_i}^{\infty} \frac{d\hat{z}_j(x)}{dx} \bigg|_{x=S_i-u} dF_i^{(L)}(u).
\]

Since \( \Psi_i \) is locally optimal, \( \hat{z}_j \) \( j \in V_i \) satisfy (5.6). Applying Lemma 5.3 yields

\[
\frac{\partial D_i(S_i, \Psi_i)}{\partial S_i} = h_i + \int_0^\infty \lambda_i(S_i - u) \sum_{j \in V_i} \frac{d\hat{z}_j(x)}{dx} \bigg|_{x=S_i-u} dF_i^{(L)}(u).
\]

Since equality (5.3) holds for every allocation function, using this property yields

\[
\frac{\partial D_i(S_i, \Psi_i)}{\partial S_i} = h_i + \int_0^\infty \frac{d\hat{z}_i(S_i - u)}{dx} dF_i^{(L)}(u). \quad (5.9)
\]
5.3. Analysis cost function

When \( \Lambda_t < 0 \) the proof of the theorem is trivial from (5.9), since \( F_{i}^{(\Lambda)}(u) = 0 \) when \( u < 0 \). For the case \( \Lambda_t \geq 0 \), writing \( u^* \) for \( u \sim \Delta_t \), after substitution of \( u^* \) in (5.9) we obtain

\[
\frac{\partial D_t(S_i, \Psi_t)}{\partial S_i} = \frac{h_t}{i} + \int_{0}^{\infty} \lambda_t(S_i - u^* - \Delta_t) dF_{i}^{(\Lambda)}(u^* + \Delta_t) = \frac{h_t}{i} + \int_{0+}^{\infty} \lambda_t(S_i - u^* - \Delta_t) d\left(F_{i}^{(\Lambda)}\right)(u^*).
\]

Since \( \Psi_t \) is locally optimal, from Lemma 5.5 we know that \( \lambda_t(S_i - \Delta_t) = 0 \).

From results obtained by Gong, De Kok & Ding [1994], we felt that \( \frac{\partial D_t(S_i, \Psi_t)}{\partial S_i} \) presented in Theorem 5.1 could be simplified to a more insightful expression. Gong, De Kok & Ding proved that the optimal order-up-to-level \( S_i \) of a stockpoint \( i \) in a serial system satisfies a newsvendor-type equation. Since we suspected that this result could be extended to divergent systems, we derive an expression for the non-stockout probability of an end-stockpoint in Theorem 5.2. It turns out that, indeed, the expression for the non-stockout probability has a lot in common with the expression of \( \frac{\partial D_t(S_i, \Psi_t)}{\partial S_i} \) presented in Theorem 5.1.

**Theorem 5.2.** Let \( \alpha_t^i(S_i) \) denote the non-stockout probability of an end-stockpoint \( k \) in a divergent echelon system, in which the most upstream stockpoint \( i \) has an echelon order-up-to-level \( S_i \). If for every allocation function in stockpoint \( i \) it holds that \( \varepsilon_i \sum_{i \neq S_i} S_n = S_i \), then

\[
\alpha_t^i(S_i) = \begin{cases} 
F_{i}^{(\Lambda+1)}(S_i) & i \in E, \\
\int_{0}^{\infty} \alpha_t^i(z_j(S_i - u)) dF_i^{(\Lambda)}(u) & i \in M, \ \Lambda_t < 0 \\
\int_{0}^{\infty} \alpha_t^i(z_j(S_i - u - \Lambda_t)) d\left(F_i^{(\Lambda)}\right)(u) & i \in M, \ \Lambda_t \geq 0,
\end{cases}
\]  

(5.10)

for every \( j \in V_i \) and \( k \in E_j \).

**Proof.** This proof is by induction on \( i \). When \( i \in E \) it is trivial.

Suppose equality (5.10) holds for a divergent system with most upstream stockpoint \( j \) (induction assumption). Then the non-stockout probability of an end-stockpoint \( k \) equals \( \alpha_t^j(S_j) \), where \( S_j \) equals the order-up-to-level of this stockpoint \( j \). Next, consider a divergent system with most upstream stockpoint \( i \), and \( j \in V_i \). Let \( \alpha_t^i \) denote the non-stockout probability of this stockpoint \( k \) as a result of the rationing decision at the beginning of period \( t \).

When at the beginning of an arbitrary period \( t \) the echelon stock of stockpoint \( i \) is less than the sum of all the order-up-to-levels of its successors, rationing is necessary. This means that every successor \( j \in V_i \) gets \( z_j(S_i - D_{i-L_j}^t) \) instead of order-up-to-level \( S_j \). From the induction assumption we obtain that for \( j \in V_i \) and \( k \in E_j \)

\[
S_i - D_{i-L_j}^t < \sum_{n \in V_i} S_n \implies \alpha_t^i(S_i) = \alpha_t^j(z_j(S_i - D_{i-L_j}^t)).
\]  

(5.11)

However, if at time \( t \) the echelon stock at stockpoint \( i \) is sufficient, all its successors raise their echelon inventory positions to their order up-to-levels. From the induction assumption, we obtain that for \( j \in V_i \) and \( k \in E_j \)

\[
S_i - D_{i-L_j}^t \geq \sum_{n \in V_i} S_n \implies \alpha_t^i(S_i) = \alpha_t^j(S_j).
\]  

(5.12)
Since the demand process is stationary we may suppress the index $i$ in $a^j_{i}(S_i)$. From (5.11) and (5.12) it can be verified that for $j \in V_i$ and $k \in E_j$,

$$\Delta_i < 0 \implies a^j_i(S_i) = \int_0^{\Delta_i} a^j_i(z_j(S_i - u)) dF^{(L)}_i(u).$$

(5.13)

$$\Delta_i \geq 0 \implies a^j_i(S_i) = \int_0^{\Delta_i} a^j_i(z_j(S_i - u)) dF^{(L)}_i(u) + \int_{\Delta_i}^{\infty} a^j_i(z_j(S_i - u)) dF^{(L^k)}_i(u).$$

(5.14)

Rewriting (5.14), using the assumption that $z_j(\sum_{i \in V_i} S_i) = S_j$, yields that for $j \in V_i$ and $k \in E_j$

$$\Delta_i > 0 \implies a^j_i(S_i) = \int_0^{\Delta_i} a^j_i(z_j(S_i - \Delta_i - u)) d\left(F^{(L^k)}_i\right)^{\Delta_i}(u).$$

(5.15)

From (5.13) and (5.15) it follows that equality (5.10) also holds for $i$.

From Theorems 5.1 and 5.2, we are able to derive an explicit expression for $\frac{\partial D_i(S_i, \hat{S}_i)}{\partial S_i}$ when $\hat{S}_i$ is optimal.

**Theorem 5.3.** For every $D_i(S_i, \hat{S}_i)$ with optimal $\hat{S}_i$ it holds that

$$\frac{\partial D_i(S_i, \hat{S}_i)}{\partial S_i} = -(\sum_{m \in V_i} h_x + p_j) + (h_x + \sum_{m \in V_i} h_x + p_j) a^j_i(S_i) \text{ for every } k \in E_i.$$

(5.16)

**Proof.** This theorem is proved by induction on $i$. First, consider an end-stockpoint $i \in E$. Then by differentiating (5.5a) to $S_i$ we obtain

$$\frac{dD_i(S_i)}{dS_i} = h_i - (h_i + \sum_{m \in V_i} h_x + p_j) (1 - F^{(L^k+1)}_i(S_i)).$$

(5.17)

Since $a^j_i(S_i) = F^{(L^k+1)}_i(S_i)$ (see (5.10)) we know that (5.16) is valid for $i \in E$.

Next, consider a stockpoint $i \in M$. Assume the theorem holds for every successor $j \in V_i$ (induction assumption). Let $S_i = \sum_{j \in V_i} \hat{S}_j$. From Theorem 5.1 it follows

$$\frac{\partial D_i(S_i, \hat{S}_i)}{\partial S_i} = h_i + \int_0^{\Delta_i} \lambda_i(S_i - u) dF^{(L)}_i(u).$$

Substitution of the definition of $\lambda_i(S_i - u)$ in the equation above yields

$$\frac{\partial D_i(S_i, \hat{S}_i)}{\partial S_i} = h_i + \int_0^{\Delta_i} \frac{\partial D_j(S_j, \hat{S}_j)}{\partial S_j} \bigg|_{S_j = (S_i - u)} dF^{(L^k)}_i(u) \text{ for every } j \in V_i.$$

Using the induction assumption we have

$$\frac{\partial D_i(S_i, \hat{S}_i)}{\partial S_i} = h_i + \int_0^{\Delta_i} -(\sum_{m \in V_i} h_x + p_j) + (h_x + \sum_{m \in V_i} h_x + p_j) a^j_i(z_j(S_i - u)) dF^{(L^k)}_i(u), \quad k \in E_j.$$

(5.18)

Theorem 5.2 directly completes the proof for $S_i = \sum_{j \in V_i} \hat{S}_j$. The proof is completely analogous for $S_i \geq \sum_{j \in V_i} \hat{S}_j$.

Van Heutum & Zijn [1991a] derived a similar expression for $\frac{\partial D_i(S_i, \hat{S}_i)}{\partial S_i}$ (for a serial system), although, they used a convolution of several distributions instead of the non-stockout probability $a^j_i(S_i)$. From Theorem 5.2, it is immediately clear that the optimal order-up-to-level satisfies a newsboy-type equation. In Section 5.4.2, we elaborate on this.
5.3. Analysis cost function

Finally, in Theorem 5.4 we derive some additional properties of the cost function \( D_i(S_i, \hat{\psi}_j) \). These yield insight in the behavior of the optimal allocation function (Section 5.4.1), and are required to prove the optimality of the decomposition (Section 5.5).

**Theorem 5.4.** For each end-stockpoint \( i \in E \), and for each intermediate stockpoint \( i \) we define \( x_i^* := 0 \), and for each intermediate stockpoint \( i \) we define \( x_i^* := \min \{ x \mid \hat{\psi}_j(x) \geq x_i^* \text{ for every } j \in V_i \} \). Then the cost function \( D_i(S_i, \hat{\psi}_j) \) with optimal \( \hat{\psi}_j \) is

(i) linear decreasing with slope \( -(\sum_{a \in U_i} h_a + p_i^*) \) for \( S_i < x_i^* \).

(ii) convex with respect to \( S_i \). Specifically, if for every end-stockpoint \( k \in E \), the demand function \( F_k(x) \) is strictly increasing for \( x \geq 0 \), then \( D_i(S_i, \hat{\psi}_j) \) is a strict convex function of \( S_i \geq x_i^* \).

(iii) \( \partial D_i(S_i, \hat{\psi}_j) / \partial S_i \) tends to \( h_i \) as \( S_i \) goes to infinity.

**Proof.** This proof is by induction on \( i \).

First, consider an end-stockpoint \( i \in E \), and define \( x_i^* := 0 \). Result (i) immediately results from (5.17), since \( F_i^{(L+1)}(x) = 0 \) for \( x \leq 0 \).

Result (ii) immediately follows from differentiating (5.17) to \( S_i \). This yields

\[
\frac{d^2 D_i(S_i)}{dS_i^2} = (h_i + \sum_{a \in U_i} h_a + p_i) f_i^{(L+1)}(S_i) \geq 0.
\]

Finally, we note that when \( x \) goes to infinity, \( F_i^{(L+1)}(x) \) increases to 1, which proves (iii).

Next, consider a stockpoint \( i \in M \), and define \( x_i^* := \min \{ x \mid \hat{\psi}_j(x) \geq x_i^* \text{ for every } j \in V_i \} \). Assume the theorem holds for every successor \( j \in V_i \) (induction assumption). First we prove (i). Let \( x < x_i^* \).

Then, there exists a stockpoint \( m \in V_i \) for which \( \hat{\psi}_m(x) \leq x_i^* \). Using the induction assumption and Lemma 5.3 we obtain that \( \lambda_i(x) = -\left( \sum_{a \in U_i} h_a + p_i^* \right) \). From the definition of \( \lambda_i(x) \) we have

\[
\lambda_i(x) = \left. \frac{\partial D_i(S_i, \hat{\psi}_j)}{\partial S_j} \right|_{S_j = \hat{\psi}_j(x)} = -(\sum_{a \in U_i} h_a + p_i^*), \quad x < x_i^*, \quad j \in V_i.
\] (5.18)

Substitution of \( \lambda_i(x) \) in the formulas of Theorem 5.1 yields

\[
\frac{\partial D_i(S_i, \hat{\psi}_j)}{\partial S_i} = h_i - \left( \sum_{a \in U_i} h_a + p_i^* \right) = -(\sum_{a \in U_i} h_a + p_i^*), \quad S_i < x_i^*.
\]

Next, we prove \( p_i^* = p_i^* \). Suppose there exists a \( j \in V_i \) for which \( p_i^* > p_j^* \). Using the fact that \( D_j(S_j, \hat{\psi}_j) \) is convex with respect to \( S_j \) and the induction assumption we have

\[
\frac{\partial D_i(S_j, \hat{\psi}_j)}{\partial S_j} \geq -(\sum_{a \in U_j} h_a + p_j^*).
\]

But, since \( p_i^* > p_j^* \) and \( U_m = U_j \) we also know

\[
\frac{\partial D_j(S_j, \hat{\psi}_j)}{\partial S_j} \geq -(\sum_{a \in U_j} h_a + p_j^*).
\]

This contradicts (5.18). Thus, a necessary condition for \( m \) is

\[
p_i^* \leq p_j^* \quad \text{for every } \quad j \in V_i.
\]
Hence, $p_n^* = p_j^*$, which proves (i).

Next, we prove that $D_i(S_n, \xi_j)$ is convex with respect to $S_i$. From the induction assumption we have that $D_i(S_j, \xi_j)$ is convex with respect to $S_j$ for every $j \in V_i$. Then from Lemma 5.5 we know that 

$$
\frac{\partial D_i(S_j, \xi_j)}{\partial S_j} = \sum_{n \in V_j} \Delta_i(S_j, \xi_j) = \sum_{n \in V_j} S_n = S_j.
$$

Using this result, after substituting $x = \sum_{n \in V_i} S_n$ in (5.6), yields $\lambda_i(\sum_{n \in V_i} S_n) = 0$. This property and (5.9) yields

$$
\frac{\partial^2 D_i(S_j, \xi_j)}{\partial S_j^2} = \int_{x = \sum_{n \in V_i} S_n}^{\infty} \frac{d^2 F_i^{(S_j)}(u)}{dS_j} \frac{\partial^2 D_i(S_j, \xi_j)}{\partial S_j^2} dF_i^{(S_j)}(u) \text{ for every } j \in V_i.
$$

From Lemma 5.4 (using the induction assumption) it immediately follows that $\frac{\partial^2 D_i(S_j, \xi_j)}{\partial S_j^2} \geq 0$. Specifically, let $x \geq x_j^*$. It can be shown that $\sum_{n \in V_j} S_n \leq S_j^*$. Thus, there exists an $u \geq 0$ such that $x - \sum_{n \in V_j} S_n \leq u \leq x - x_j^*$. Then from Lemma 5.4 and the induction assumption it follows that $D_i(S_j, \xi_j)$ is strictly convex for $S_j \geq x_j^*$.

Finally, we prove that when $S_i$ goes to infinity $\partial D_i(S_n, \xi_j)/\partial S_i$ tends to $h_i$. From the induction assumption and Lemma 5.3 it follows

$$
(h_i + \sum_{n \in V_i} h_n + p_n^*) \leq \lambda_i(x) \leq h_i^*.
$$

(5.19)

From (5.9) and (5.19) it follows

$$
h_i - (h_i + \sum_{n \in V_i} h_n + p_n^*) \int_{B_i}^{\infty} dF_i^{(S_j)}(u) \leq \frac{\partial D_i(S_n, \xi_j)}{\partial S_i} \leq h_i^* \int_{B_i}^{\infty} dF_i^{(S_j)}(u).
$$

When $S_i$ tends to infinity both the lower and upper bound converge to $h_i$. \qed

Note that from Theorem 5.4 it follows that if $h_i$ is positive, a minimum of the function $D_i(S_n, \xi_j)$ exists for $S_i \geq x_j^*$. Specifically, if at every stockpoint $k \in E_i$, the demand function $F_j(x)$ is strictly increasing in $x \geq 0$ the uniqueness of this minimum is guaranteed. However, if $h_i$ equals 0 the minimum is attained in infinity.

5.4 Optimal control policy

From the results in the previous section we are able to determine the cost-optimal control policy. In Section 5.4.1 and 5.4.2, we present how to determine the optimal allocation function and the optimal order-up-to-level, respectively.

5.4.1 Optimal allocation function

Consider the situation where we already determined the optimal $\xi_j$ for $j \in V_i$. Next, we determine the optimal allocation functions $\xi_j$. From Theorem 5.3 we know that

$$
\frac{\partial D_i(S_j, \xi_j)}{\partial S_j} = -(\sum_{n \in V_j} h_n + p_n) + (h_i + \sum_{n \in V_i} h_n + p_n) \lambda_i(S_j), \quad k \in E_j.
$$

Substitution of the equality above in Lemma 5.3 yields (after some elementary algebra)

$$
\lambda_i(x) + \sum_{n \in V_i} h_n + p_n
$$

$$
\alpha_j(x) = \frac{\lambda_i(x) + \sum_{n \in V_i} h_n + p_n}{h_k + \sum_{n \in V_k} h_n + p_k}, \quad j \in V_i, \quad k \in E_j.
$$
5.4. Optimal control policy

The $\lambda_i(x)$ is determined such that $\sum_{k \in E_i} \ell_k(x) = x$. This can be done by a one-dimensional bisection procedure where $-(h_i + \sum_{k \in E_i} h_k + p_i^r) \leq \lambda_i(x) \leq h_i^*$ (see equation (5.19)). So the whole procedure works as depicted in Figure 5.1.

(i) Initialize $x$, and $\epsilon > 0$.

(ii) Choose $-h_i + \sum_{k \in E_i} h_k + p_i^r \leq \lambda_i(x) \leq h_i^*$.

(iii) Determine $\tilde{\ell}_j(x)$ from

$$\lambda_i(x) + \sum_{k \in E_i} h_k + p_i^r$$

$$a_j^0(\tilde{\ell}_j(x)) = \frac{h_k + \sum_{k \in E_i} h_k + p_i^r}{k \in E_j, j \in V_i, k \in E_j},$$

(iv) if $\sum_{k \in E_i} \tilde{\ell}_k(x) < x - \epsilon$ then increase $\lambda_i(x)$ and return to step (ii).

if $\sum_{k \in E_i} \tilde{\ell}_k(x) > x - \epsilon$ then decrease $\lambda_i(x)$ and return to step (ii).

Figure 5.1. Algorithm to determine the optimal allocation functions $[\tilde{\ell}_j]$.

The way to adapt $\lambda_i(x)$ as suggested in step (iv) of Figure 5.1 is valid, due to the convexity of the cost function $D_j(S_j, \Psi_j)$ for $j \in V_i$. Unfortunately, we have to apply the procedure above for every $x$. Hence, it is not suited for practical purposes. However, it does provide insight in the behavior of these optimal allocation functions.

From the results of the previous section, we derive some interesting properties of $[\tilde{\ell}_j]$ for $j \in V_i$. These properties are formalized in Theorem 5.5. This states that if the echelon stock of stockpoint $i$ just before rationing, $x$ say, is less than or equal to $x_i^*$, the optimal allocation policy ensures that only the echelons $j$ with with the lowest penalty cost ($p_i^r$) get a small (or even negative) echelon inventory position after rationing. If $x$ is very large, the optimal allocation policy ensures that the major part of $x$ is allocated to the successor $j$ with the lowest added value ($h_i^*$). After presenting this theorem we elaborate on the behavior of an optimal allocation function, by giving some examples of the behavior of such a function.

Theorem 5.5. For every optimal set of allocation functions it holds that

(i) $\tilde{\ell}_j(x) \in [\delta_j, \bar{\delta}_j]$ for $x \leq x_i^* = \sum_{k \in V_i} x_i^* + \sum_{k \in V_i \setminus \bar{V}_i} \bar{\delta}_j$ and $j \in V_i \setminus \bar{V}_i$.

(ii) $\tilde{\ell}_j(x) < \bar{\delta}_j$ for every $x$ and $j \in V_i \setminus \bar{V}_i$, where

$$\delta_j := \arg \min_x \left\{ \frac{\partial D_j(S_j, \Psi_j)}{\partial S_j} = -\left( \sum_{k \in E_j} h_k + p_i^r \right) \right\} \text{ for } j \in V_i \setminus \bar{V}_i,$$

$$\bar{\delta}_j := \arg \max_x \left\{ \frac{\partial D_j(S_j, \Psi_j)}{\partial S_j} = -\left( \sum_{k \in E_j} h_k + p_i^r \right) \right\} \text{ for } j \in V_i \setminus \bar{V}_i,$$

$$\bar{\delta}_j := \arg \min_x \left\{ \frac{\partial D_j(S_j, \Psi_j)}{\partial S_j} = h_i^* \right\} \text{ for } j \in V_i \setminus \bar{V}_i.$$
Proof. (i). Suppose a stockpoint $i$ has to ration its echelon stock $x \leq x^*$ over its successors. Consider a successor, $m$, say, for which $p^*_m > p^*_i$. So, $m \in V \setminus V_i$. We prove (i) by showing that $\hat{e}_m(x) \geq \Sigma^i_m$ and $\hat{e}_m(x) \leq \Sigma^i_m$, respectively. For both cases the proofs are given by contradiction.

- Assume $\hat{e}_m(x) < \Sigma^i_m$. From the convexity of $D_m(S_m, \hat{\Psi}_m)$ with respect to $S_m$, and the definition of $\Sigma^i_m$ we obtain

$$\frac{\partial D_m(S_m, \hat{\Psi}_m)}{\partial S_m} \bigg|_{S_m = \Sigma^i_m} < \frac{\partial D_m(S_m, \hat{\Psi}_m)}{\partial S_m} \bigg|_{S_m = \Sigma^i_m} = -\left( \sum_{n \in U_m} h_n + p^*_n \right).$$

From Lemma 5.3 we obtain

$$\lambda_i(x) > -(h_i + \sum_{n \in U_i} h_n + p^*_i). \quad (5.20)$$

Next, we consider a stockpoint, $j$, say, for which $p^*_j = p^*_i$. So, $j \in V_i$. From the convexity of $D_j(S_j, \hat{\Psi}_j)$ and Theorem 5.4 it follows

$$\frac{\partial D_j(S_j, \hat{\Psi}_j)}{\partial S_j} \bigg|_{S_j = \Sigma^i_j} \geq -\left( \sum_{n \in U_j} h_n + p^*_j \right) = -\left( \sum_{n \in U_j} h_n + p^*_j \right).$$

Again using Lemma 5.3, we obtain

$$\lambda_i(x) \leq -(h_i + \sum_{n \in U_i} h_n + p^*_i). \quad (5.21)$$

Equation (5.20) and (5.21) lead to a contradiction. Hence $\hat{e}_m(x) \geq \Sigma^i_m$.

- Assume $\hat{e}_m(x) > \Sigma^i_m$. Similar to (5.20) it can be shown that

$$\lambda_i(x) < -(h_i + \sum_{n \in U_i} h_n + p^*_i). \quad (5.22)$$

After substitution of (5.3) in (5.22), we have that for every stockpoint $j \in V \setminus V_i$

$$\frac{\partial D_j(S_j, \hat{\Psi}_j)}{\partial S_j} \bigg|_{S_j = \Sigma^i_j} = \left( \sum_{n \in U_j} h_n + p^*_j \right) \bigg|_{S_j = \Sigma^i_j}.$$

Since $D_j(S_j, \hat{\Psi}_j)$ is convex with respect to $S_j$, we know that $\hat{e}_j(x) > \Sigma^i_j$. Using this and equality (5.3) we obtain

$$\sum_{n \in U_j} e_n(x) = x - \sum_{n \in V \setminus V_j} e_n(x) < x - \sum_{n \in V \setminus V_j} e_n(x) = x - \sum_{n \in U_j} h_n + \sum_{n \in U_j} p^*_j \leq \sum_{n \in U_j} x^*_i.$$

So, there exists a stockpoint $j \in V_j$ for which $\hat{e}_j(x) < x^*_j$. From Theorem 5.4 (i), we know that for this stockpoint $j$

$$\frac{\partial D_j(S_j, \hat{\Psi}_j)}{\partial S_j} \bigg|_{S_j = \Sigma^i_j} = -\left( \sum_{n \in U_j} h_n + p^*_j \right) = -\left( \sum_{n \in U_j} h_n + p^*_j \right).$$

Again using Lemma 5.3, we obtain

$$\lambda_i(x) < -(h_i + \sum_{n \in U_i} h_n + p^*_i). \quad (5.23)$$

Equation (5.22) and (5.23) lead to a contradiction. Hence, $\hat{e}_m(x) \leq \Sigma^i_m$. So $\hat{e}_m(x) \in [\Sigma^i_m, \Sigma^i_m]$. 


(ii). Suppose some stockpoint \( i \) has to ration its echelon stock \( x \) (with \( x \leq x_i^* \)) over its successors. Consider such a successor, \( m \) say, for which \( h_m > h_i^* \). So, \( m \in V_i \setminus \overline{V}_i \). Assume that \( z_m(x) \geq \overline{x}_m \). Since \( D_m(S_m, \overline{x}_m) \) is convex with respect to \( S_m \) we have
\[
\frac{\partial D_m(S_m, \overline{x}_m)}{\partial S_m} \bigg|_{S_m = \overline{x}_m} \geq \frac{\partial D_m(S_m, \overline{x}_m)}{\partial S_m} \bigg|_{S_m = z_m} = h_i^*.
\]
So from Lemma 5.3 it follows that \( \lambda_i(x) \geq h_i^* \). From Theorem 5.4 it is clear that for a stockpoint \( j \in \overline{V} \), holds \( \partial D_j(S_j, \overline{x}_j) / \partial S_j < h_j = h_i^* \). Then Lemma 5.3 leads to a contradiction, since we obtain \( \lambda_i(x) < h_i^* \). This proves \( z_m(x) < \overline{x}_m \). \( \square \)

In case the cdf of the customer demand is strictly increasing for every end-stockpoint, it can be shown that \( \overline{x}_j = x_j^* \). So in that case \( \hat{z}_j(x) \) equals \( x_j^* \) for \( j \not\in V_i \) and \( x \leq x_j^* \). In the more general case where there is no strict monotonicity of the customer demand cdf, it can be shown that there exists a set of optimal allocation functions such that \( \hat{z}_j(x) \) equals \( x_j^* \) for \( j \not\in V_i \) and \( x \leq x_j^* \). In the remainder of this chapter, we consider this specific set.

From Theorem 5.5, it is possible to distinguish between four classes of allocation functions. Figure 5.2 depicts an example of an allocation function for each class.

\[ \hat{z}_j(x) \quad \overline{x}_j \quad x \]
\[ \hat{z}_j(x) \quad \overline{x}_j \quad x \]

(a) \( j \in V_i \), \( j \not\in \overline{V}_i \).

(b) \( j \not\in V_i \), \( j \not\in \overline{V}_i \).

\[ \hat{z}_j(x) \quad \overline{x}_j \quad x \]
\[ \hat{z}_j(x) \quad \overline{x}_j \quad x \]

(c) \( j \in V_i \), \( j \in \overline{V}_i \).

(d) \( j \not\in V_i \), \( j \in \overline{V}_i \).

**Figure 5.2.** The behavior of an allocation function for the four different classes.

Note that in Figures 5.2 (a) and (b) the amount of products allocated to successor \( j \) tends to a limit when \( x \) goes to infinity. The allocation functions in Figures 5.2 (c) and (d) do not have an upper bound. To understand the behavior of \( z_j(x) \), suppose stockpoint \( i \) has to allocate \( x \) products. Since high penalty costs are incurred for every backordered product, we would like to allocate as much as possible to...
successor $j$. However, holding costs are also incurred which assures that not too much stock is kept at the various stockpoints. It is clear that there is a trade off between penalty and holding costs. If the amount $x$ which needs to be allocated is large, then the allocation decisions are mainly based on the holding costs, since the penalty costs are very small. Specifically, if $x$ goes to infinity the reduction of the penalty costs of echelon $j$ by allocating an extra product to $j$ equals 0, while the additional holding costs are $h_j$. Therefore, an extra product will be allocated to those successors with minimal added value.

Also note that in Figures 5.2 (b) and (d) the amount of products allocated to successor $j$ is fixed when $x$ is sufficiently small. This does not hold in Figures 5.2 (a) and (c). If $x$ is very small (or even negative) there is an incentive to allocate the major part to the successors $j$ which (indirectly) supply end-stockpoints with high penalty costs ($j \not\in \mathcal{E}_s$). The actual amount allocated to a successor $j \not\in \mathcal{E}_s$ is such that the marginal costs in $\epsilon(j)$ equals the marginal costs of a successor $j \in \mathcal{E}_s$.

5.4.2 Optimal order-up-to-level

From Theorem 5.3 it follows that the optimal order-up-to-level $\hat{S}_i$ has to satisfy

$$\alpha'_t(\hat{S}_i) = \frac{\sum_{a \in \delta_i} p_a + p_t}{h_a + \sum_{a \in \delta_i} h_a + p_a} \text{ for every } k \in E_i. \quad (5.24)$$

Note that condition (5.24) resembles the classical newsvendor result, which prescribes the optimal critical ratio for a single location inventory system to be $p/(p + h)$.

In the next chapter we prove that if we restrict to a specific class of linear allocation functions then the newsvendor-type formula obtained in (3.24) also yields the optimal order-up-to-level. Hence, equation (5.24) yields the optimal order-up-to-level for a larger class of allocation functions.

5.5 Optimality of decomposition approach

We now apply the results from the previous sections to develop an optimization scheme for $(\hat{S}_i, \hat{\Psi}_i)$ where index $i$ denotes the most upstream stockpoint of the $N$-echelon system. We show that the problem of finding this optimal $(\hat{S}_i, \hat{\Psi}_i)$ can be decomposed into solving one-dimensional problems subst-

(i). $n := 1$.

(ii). $\begin{array}{ll}
\bullet & \text{Initialize } \hat{\Psi}_i := \emptyset \text{ for each stockpoint } i \in E. \\
\bullet & \text{Determine the optimal order-up-to-level } \hat{S}_i \text{ for each stockpoint } i \in E.
\end{array}$

(iii). $n := n + 1$.

(iv). Determine for each stockpoint $i \in \mathcal{E}_n$:

$\begin{array}{ll}
\bullet & \text{The optimal allocation functions } (\xi_j)_{j \in \delta_i} \text{ for every } x. \\
\bullet & \hat{\Psi}_i := \bigcup_{j \in \delta_i} (\xi_j, \hat{S}_j, \hat{\Psi}_j) \text{ for } i \in \mathcal{E}_n. \\
\bullet & \text{The optimal order-up-to-level } \hat{S}_i.
\end{array}$

(v). If $n < N$ then return to step (iii).

Figure 5.3. Decomposition approach.
5.5. Optimality of decomposition approach

sequently. In order to prove that the approach depicted in Figure 5.3 yields the optimal policy we introduce Definition 5.4 which is used in Lemma 5.6.

**Definition 5.4.** Let \( s_0 := S_i \cup \bigcup_{j \in V_i} \{ z_j, S_j, \Psi_j \} \) denotes an arbitrary policy. Then \( s_0 \) is defined as the control policy in which for every stockpoint \( i \in \mathcal{W}_{n+1} \):

(i). The order up-to-levels and allocation functions upstream of \( i \) are the same as in the policy \( s_0 \).

(ii). The order-up-to-level of stockpoint \( i \) is the same as in the policy \( s_0 \) and the allocation functions of stockpoint \( i \) are the same as in the optimal policy \( \Psi_i \).

(iii). The order-up-to-level and the allocation functions downstream of \( i \) are the same as in the optimal policy \( \Psi_i \).

\[ \square \]

**Lemma 5.6.** Let \( g_n(s_n) \) denote the expected total costs of the stockpoints in \( \bigcup_{j \in V_i} W_j \) when the multi-echelon system is controlled by policy \( s_n \). Then,

\[ g_n(s_n) \leq g_n(s_{n-1}), \quad n = 1, 2, \ldots, N. \]

**Proof:** Suppose \( s_0 \) is an arbitrary replenishment policy. Let us consider a stockpoint \( i \in \mathcal{W}_{n+1} \). For our convenience we introduce some additional notation. Let \( g'_n(s_n) \) denote the expected total costs in \( \bigcup_{j \in V_i} \text{ech}(j) \), due to the replenishment orders placed by policy \( s_n \) at time \( r \). The allocation function from stockpoint \( i \) to successor \( j \) when applying policy \( s_n \) and \( s_{n-1} \) is denoted by \( \bar{z}_i \) and \( z_j \), respectively. The order-up-to-level of \( j \) when applying policy \( s_n \) and \( s_{n-1} \) is denoted by \( \bar{S}_j \) and \( S_j \). We distinguish between \( \sum_{j \in V_i} S_j < \sum_{j \in V_i} \bar{S}_j \) and \( \sum_{j \in V_i} S_j > \sum_{j \in V_i} \bar{S}_j \).

- \( \sum_{j \in V_i} S_j < \sum_{j \in V_i} \bar{S}_j \):

If the echelon stock of stockpoint \( J_i \) is sufficient to raise the echelon inventory positions of all successors \( j \in V_i \) to their order-up-to-level, the expected costs of echelon \( j \) equals \( D_j(\bar{S}_j, \Psi_j) \) and \( D_j(S_j, \Psi_j) \) for replenishment policy \( s_n \) and \( s_{n-1} \), respectively. Hence, from the fact that \( \bar{S}_j \) minimizes \( D_j(S_j, \Psi_j) \) for \( S_j \) we obtain

\[ J_i \geq \sum_{j \in V_i} \bar{S}_j \implies g'_n(s_n) = \sum_{j \in V_i} D_j(\bar{S}_j, \Psi_j) \leq \sum_{j \in V_i} D_j(\bar{S}_j, \Psi_j) = g'_n(s_{n-1}). \]

However, if the echelon stock of stockpoint \( J_i \) is not sufficient, all the echelon stock is allocated to the successors \( j \in V_i \). Hence, from the fact that \( \bar{z}_i \) minimizes \( D_j(\bar{z}_i(x), \Psi_j) \) we obtain

\[ J_i < \sum_{j \in V_i} S_j \implies g'_n(s_n) = \sum_{j \in V_i} D_j(\bar{z}_i(x), \Psi_j) \leq \sum_{j \in V_i} D_j(\bar{z}_i(x), \Psi_j) = g'_n(s_{n-1}). \]

Finally, we analyze the situation where \( \sum_{j \in V_i} S_j < J_i < \sum_{j \in V_i} \bar{S}_j \). Then policy \( s_n \) ration all the echelon stock \( J_i \) over the stockpoints \( j \in V_i \) while policy \( s_{n-1} \) raises the echelon inventory positions of stockpoints \( j \in V_i \) to their order-up-to-level \( \bar{S}_j \), and the remainder \( J_i - \sum_{j \in V_i} \bar{S}_j \) is retained at stockpoint \( i \).

The allocation function \( \bar{z}_j \), defined by

\[ \bar{z}_j(x) := S_j - q_j \left( \sum_{j \in V_i} S_j - x \right) \quad \text{with} \quad q_j := \frac{S_j - S_i}{\sum_{k \in V_i} (S_k - S_i)}, \]
has the property that \( S_j < \xi_j(\tilde{J}_j) < \tilde{S}_j \) if \( S_j < \tilde{S}_j \), and \( \tilde{S}_j < \xi_j(\tilde{J}_j) < S_j \) if \( S_j > \tilde{S}_j \). From this property and the fact that \( D_j(S_j, \psi_j) \) is convex with respect to \( S_j \) (with the minimum in \( S_j = \tilde{S}_j \)) it follows

\[
\sum_{j \in V} D_j(\xi_j(\tilde{J}_j), \psi_j) \leq \sum_{j \in V} D_j(S_j, \psi_j).
\]

(5.25)

Since \( \xi_j \) is at least as good as \( \tilde{z}_j \), and equation (5.25) it follows that

\[
\sum_{j \in V} S_j < \tilde{J}_j < \sum_{j \in V} \tilde{S}_j \implies g'(s_n) = \sum_{j \in V} D_j(\tilde{z}_j(\tilde{J}_j), \psi_j) \leq \sum_{j \in V} D_j(\xi_j(\tilde{J}_j), \psi_j) \leq \sum_{j \in V} D_j(S_j, \psi_j) = g'(s_{n-1}).
\]

\[
\sum_{j \in V} S_j \geq \sum_{j \in V} \tilde{S}_j.
\]

Analogously, we can prove that

\[
\tilde{J}_j \geq \sum_{j \in V} \tilde{S}_j \implies g'(s_n) = \sum_{j \in V} D_j(\tilde{S}_j, \psi_j) \leq \sum_{j \in V} D_j(S_j, \psi_j) = g'(s_{n-1}).
\]

\[
\tilde{J}_j \leq \sum_{j \in V} \tilde{S}_j \implies g'(s_n) = \sum_{j \in V} D_j(\tilde{S}_j, \psi_j) \leq \sum_{j \in V} D_j(z_j(\tilde{J}_j), \psi_j) \leq g'(s_{n-1}).
\]

\[
\sum_{j \in V} \tilde{S}_j < \tilde{J}_j < \sum_{j \in V} S_j \implies g'(s_n) = \sum_{j \in V} D_j(\tilde{S}_j, \psi_j) \leq \sum_{j \in V} D_j(\xi_j(\tilde{J}_j), \psi_j) = g'(s_{n-1}).
\]

Note that when \( \sum_{j \in V} S_j = \sum_{j \in V} \tilde{S}_j \), the convexity of \( D_j(S_j, \psi_j) \) with respect to \( S_j \) (with the minimum in \( S_j = \tilde{S}_j \)) immediately results in \( g'(s_n) \leq g'(s_{n-1}) \).

So for an arbitrary period \( t \) and an arbitrary stockpoint \( i \in W_{n+1} \) it holds that \( g'(s_n) \leq g'(s_{n-1}) \).

From Lemma 5.6, we prove in Theorem 5.6 that decomposition is exact (given the balance assumption).

**Theorem 5.6.** The decomposition approach yields the cost-optimal control policy, i.e., minimizing the long-run average costs, provided that the balance assumption is satisfied.

**Proof.** This proof is very similar to the proofs in Van Houwelingen [1990] and Van Houwelingen & Zijm [1991b]. Consider an arbitrary replenishment policy \( s_N \). For both \( s_n \) and \( s_{n-1} \) the behavior of the stock at all stockpoints in \( \cup_{j=n+1}^{N} W_j \) is identical. Hence, for \( n = 1, \ldots, N \) holds

\[
g_N(s_n) = g_N(s_{n-1}) \leq \cdots \leq g_N(s_0).
\]

Using (5.26) and Lemma 5.6 yields

\[
g_N(s_N) = g_N(s_{N-1}) \leq \cdots \leq g_N(s_0).
\]

Because replenishment policy \( s_N \) does not have costs larger than an arbitrary policy, specifically the optimal policy, we conclude that \( s_N \) is cost-optimal.

\[
\square
\]

### 5.6 Extensions

In this section, we like to address two model extensions. First, in Section 5.6.1, we show to what extent the results obtained in this chapter can be used for the model where the lead times are not a multiple integer of the review period \( R \). Second, in Section 5.6.2, we show to what extent the results obtained in this chapter can be used for the model where the added value \( h_i \) is not only incurred for the products in or downstream of stockpoint \( i \), but also for the products in the pipeline towards stockpoint \( i \).
5.6. Extensions

5.6.1 Arbitrary fixed lead times

In this section, we give an outline of the proof that the results obtained in the previous sections can also be applied to the model with lead times which are not multiple integers of the review period \( R \). To illustrate how the results can be extended we consider an example. Consider a 3-echelon system (see Figure 5.4 (a)) in which the most upstream stockpoint 1 supplies a stockpoint 2, which supplies 2 end-stockpoints (stockpoint 3 and 4). The times at which orders arrive are depicted in Figure 5.4 (b).

\[
\begin{align*}
\text{Stockpoint 1} & \quad \text{Stockpoint 2} & \quad \text{Stockpoint 3} & \quad \text{Stockpoint 4} \\
& \quad \quad \quad \quad \Delta 1 & \quad \quad \quad \quad \Delta 1 & \quad \quad \quad \quad \Delta 1 \\
& \quad \quad \quad \quad \Delta 2 & \quad \quad \quad \quad \Delta 2 & \quad \quad \quad \quad \Delta 2 \\
& \quad \quad \quad \quad \Delta 3 & \quad \quad \quad \quad \Delta 3 & \quad \quad \quad \quad \Delta 3 \\
& \quad \quad \quad \quad \Delta 4 & \quad \quad \quad \quad \Delta 4 & \quad \quad \quad \quad \Delta 4 \\
0 & \quad R & \quad 2R & \quad \text{time}
\end{align*}
\]

(a) System layout. (b) Times at which orders arrives.

**Figure 5.4. Example of a divergent 3-echelon system (fig. a) and its arrivals (fig. b).**

By definition the orders of stockpoint 1 arrive at times 0, \( R \), \( 2R \), \( \ldots \). Like Section 5.2 holding costs \( h_1 \) are incurred for the amount of products in stockpoint 1 just before an order arrival at stockpoint 1. The costs incurred at time 0 equals

\[ h_1 (J_{0}^1 - J_{0}^2). \]

Immediately after the arrival of an order at stockpoint 1, an order is placed by stockpoint 2. This order arrives at time \( L_2 \). Since the lead time is not an integer multiple of \( R \), the arrival of this order does not coincide with an order arrival at stockpoint 1. Let \( \delta_1 \) denotes the time that stockpoint 1 has to wait after an order arrival at its predecessor till an order arrival at stockpoint 2 (e.g., in Figure 5.4 (b) \( \delta_1 \) equals 0.75\( R \)). By definition \( \delta_1 := 0 \). The orders arrive at stockpoint 2 at times \( \delta_1, R + \delta_1, 2R + \delta_1, \ldots \). Just before an arrival holding costs \( h_1 + h_2 \) are incurred per product. Hence the costs incurred at time \( \delta_2 \) equals

\[ (h_1 + h_2)(J_{\delta_2}^2 - J_{\delta_2}^3 - J_{\delta_2}^4). \]

Finally, in end-stockpoint 3 the orders arrive at times \( \delta_2 + \delta_3, R + \delta_2 + \delta_3, 2R + \delta_2 + \delta_3, \ldots \). Besides holding costs \( h_1 + h_2 + h_3 \) per product, also penalty costs \( p_3 \) are incurred per backlogged product. Hence, the costs incurred at stockpoint 3 at time \( \delta_3 + \delta_4 \) equals

\[ (h_1 + h_2 + h_3)(J_{\delta_3+\delta_4}^3)^+ + p_3(-J_{\delta_3+\delta_4}^4)^+. \]

Similarly, the costs incurred at stockpoint 4 at time \( \delta_4 + \delta_5 \) equals

\[ (h_1 + h_2 + h_3)(J_{\delta_4+\delta_5}^4)^+ + p_4(-J_{\delta_4+\delta_5}^5)^+. \]
By adding up all the aforementioned costs we obtain
\[ \sum_{i \in \mathcal{M}} \left( h_i + \sum_{n \in \mathcal{N}} h_n \right) \left( J_i^t - \sum_{j \in \mathcal{V}_i} J_i^t \right) + \sum_{i \in \mathcal{E}} \left( h_i + \sum_{n \in \mathcal{N}} h_n \right) \left( J_i^t \right)^+ + p_i \left( -J_i^t \right)^+ \]  
(5.27)
where \( t_i := 0 \), and \( t_j := t_k + \delta_j \) for every \( j \in \mathcal{V}_i \). Note that the costs of (5.27) equals the sum of the costs incurred at all the black signs in Figure 5.4 (b). Rewriting (5.27) (using the proof of Lemma 5.1) yields
\[ \sum_{i \in \mathcal{M}} h_i J_i^t + \sum_{i \in \mathcal{E}} \left( h_i J_i^t + \left( p_i + h_i + \sum_{n \in \mathcal{N}} h_n \right) \left( -J_i^t \right)^+ \right) + \sum_{i \in \mathcal{R} \cup \mathcal{K}} \left( h_i J_i^t - J_{i_{(\text{pen})}} \right) \]  
(5.28)
Note that \( J_{i_{(\text{pen})}} \) equals the customer demand at end-stockpoints \( k \in \mathcal{E}_i \) during \( (1_{\text{pen}}, l_i) \). Hence, rewriting (5.28) yields
\[ \sum_{i \in \mathcal{M}} h_i J_i^t + \sum_{i \in \mathcal{E}} \left( h_i J_i^t + \left( p_i + h_i + \sum_{n \in \mathcal{N}} h_n \right) \left( -J_i^t \right)^+ \right) - \sum_{i \in \mathcal{R} \cup \mathcal{K}} \left( \sum_{n \in \mathcal{N}_i} h_n \right) D_i \]  
(5.29)
The result in expression (5.29) coincides with Lemma 5.1. Holding costs \( h_i \) are incurred per product at each stockpoint \( i \) (just before an order arrival at \( i \)), whereas at an end-stockpoint extra costs are incurred when it is in a backlog position. These extra costs are penalty costs \( p_i \) per backlogged product, and holding costs \( h_i + \sum_{n \in \mathcal{N}_i} h_n \) per backlogged product. By comparing Lemma 5.1 with (5.29) we note that expression (5.29) has an additional term. This term, however, is irrelevant from the perspective of determining the cost-optimal control parameters. The average costs per replenishment cycle are smaller due to this extra term, but the optimal order-up-to-levels and allocation functions are the same for both cases. Since Lemma 5.1 constitutes the basis of the decomposition all the results derived in Sections 5.4 and 5.5 remain valid.

### 5.6.2 Different cost structure

So far we assumed that the holding costs per product in stockpoint \( i \) or in transfer towards one of its successors equals \( h_i + \sum_{n \in \mathcal{N}_i} h_n \), where \( h_i \) can be regarded as an additional holding cost due to the value added in stockpoint \( i \). In this cost structure the value is added to a product immediately after arrival at a stockpoint. In practice one could also imagine cost structures where the value is added just before shipping the product to one of its successors. In that case the results developed in this monograph still can be applied. To explain this, we show how to assign costs to a stockpoint such that we can decompose the system.

Using similar arguments as in (5.1) and (5.2) we know that
\[ \sum_{i \in \mathcal{M} \cup \mathcal{K}} \left( h_i + \sum_{n \in \mathcal{N}_i} h_n \right) I_i^t = \sum_{i \in \mathcal{M}} \left( \sum_{n \in \mathcal{N}_i} h_n \right) J_i^t. \]
From this it follows that the costs which are incurred at time \( t \) equals
\[ \sum_{i \in \mathcal{M}} \left( h_i + \sum_{n \in \mathcal{N}_i} h_n \right) I_i^t - \sum_{i \in \mathcal{E}} \left( h_i + \sum_{n \in \mathcal{N}_i} h_n \right) J_i^t + p_i \left( -J_i^t \right)^+ \]
\[ = \sum_{i \in \mathcal{M}} \left( h_i + \sum_{n \in \mathcal{N}_i} h_n \right) I_i^t - \sum_{i \in \mathcal{E}} \left( h_i + \sum_{n \in \mathcal{N}_i} h_n \right) J_i^t + p_i \left( -J_i^t \right)^+ \]
\[ = \sum_{i \in \mathcal{M}} h_i I_i^t + \sum_{i \in \mathcal{E}} \left( h_i I_i^t + \left( p_i + h_i + \sum_{n \in \mathcal{N}_i} h_n \right) \left( -J_i^t \right)^+ \right). \]
5.7. Conclusions

Substitution of $I_i = J_{i+k_i} - D_{i+k_i}$ yields

$$
\sum_{i \in \mathcal{E}} h_i I_i l_i + \sum_{i \in \mathcal{E}} \left( h_i J_{i+k_i} - (h_i + \sum_{n \in \mathcal{N}_i} h_n + p_i)(-J_i)^+ \right) + \sum_{i \in \mathcal{M}_i \in \mathcal{E}} h_i D_{i+k_i}^c.
$$

Since we consider stationary demand the expected systemwide costs per period for this cost structure is identical to the expected systemwide cost per period of the ‘original’ cost structure except for a constant value $\sum_{i \in \mathcal{M}_i \in \mathcal{E}} h_i L_{i+k_i}$. Since this value does not affect the cost-optimal control policy we can apply the results obtained in this chapter to determine this control policy.

5.7 Conclusions

The objective of this chapter is to determine a cost-optimal replenishment policy (i.e., a policy which minimizes the expected mean holding and penalty costs per review period) for the divergent multi-echelon inventory system as formulated in Chapter 2. We proved that the decomposition approach as in Langenhoff & Zijm [1990] can be extended to divergent $N$-echelon systems given the balance assumption. Hence, the complex multi-dimensional problem of determining the cost-optimal policy reduces to the problem of determining: (1) the optimal order-up-to-level at every stockpoint, and (2) the optimal allocation functions to its successors.

We used the result of Gong, De Kok & Ding [1994] to prove that the first problem coincides with the classical newsboy problem. The second problem is solved by applying the Lagrangian-multiplier technique. From this we obtain insight in the structure of the optimal allocation functions. An algorithm is developed to actually determine these functions. Based on several properties of the optimal allocation functions we classify these functions into four classes (see Figure 5.2). It is cumbersome and time-consuming to determine the set of optimal allocation functions. Therefore, there is a need for a more practically applicable approach to determine a replenishment policy which is almost cost-optimal but easy and fast to determine. In the next chapter, we develop such an approach.
Cost minimization: optimal allocation functions
6

Cost minimization: linear allocation functions

6.1 Introduction

In Chapters 3 and 4, we developed heuristics to determine the control parameters of the system as formulated in Chapter 2 such that the target service levels constraints are satisfied. Instead of this 'service related approach', in Chapter 5 we concentrated on determining the control parameters such that the expected total costs per review period are minimized. We considered a cost structure where only penalty and holding costs are incurred at the end of each review period. The penalty costs equals \( p_i \) for each backlogged product at end-stockpoint \( i \). The holding costs equals \( h_i + \sum_{n \in \text{level } i} h_n \) for each product in stockpoint \( i \) or in transfer to one of its successors. To determine the order-up-to-levels and allocation functions minimizing the expected costs per review period we applied the decomposition approach as introduced by Langenhouff & Zijm [1990]. In the previous chapter, we proved that this approach is exact given the balance assumption. It decomposes the complex multi-dimensional problem into several single-dimensional problems. However, it still remains rather cumbersome and time-consuming to determine the optimal allocation function \( f_i(x) \), since for every \( x \) we have to apply the algorithm as described in Figure 5.1. Hence, this algorithm is not suited for practical purposes. Therefore the subject of this chapter is to develop a practically applicable algorithm which is fast and yields good results (not necessarily optimal). One way of not having to apply the algorithm of Figure 5.1 for every \( x \) is by restricting it to linear allocation functions. In that way, it is much easier and less time-consuming to determine the allocation function. In this chapter, we restrict to a specific class of linear allocation functions. In Section 6.2, we introduce this specific class and argue the appropriateness of this class.

The chapter is organized as follows. In Section 6.2, we present an average cost analysis for the divergent N-echelon system as formulated in Chapter 5, given that each allocation function is within the specific class. In Section 6.3, we develop an algorithm to determine all the control parameters such that the expected total costs per review period is minimized. Besides the balance assumption, we also need that all stockpoints have a positive added value. In Section 6.4, we try to obtain insight in the performance of the algorithm by constructing an experimental design of 500 instances in which we compare the results obtained by the algorithm with those obtained by simulation for each instance. We conclude that for most instances the performance of the algorithm yields very good results. In Section
6.1 We explain why the algorithm of Section 6.3 cannot be used for systems in which some stockpoints do not add any value to the product. By applying the algorithm to an example we illustrate where an extension of the algorithm is required, and how this extension works. In Section 6.6, we formalize this extension. In Section 6.7, we validate the performance of the extension of the algorithm by applying it to the same instances as presented in Section 6.4 (except that every intermediate stockpoint does not add any value). In Section 6.8 we give a first step in deriving an upper bound for the error we make in this chapter by restricting to the specific class of linear allocation functions instead of the optimal allocation functions derived in Chapter 5. Finally, in Section 6.9, we give a few concluding remarks.

6.2 Analysis

In Chapter 5, we proved that the decomposition approach yields the cost-optimal control policy given the balance assumption. For practical purposes, however, it is rather cumbersome to determine the optimal allocation functions. Therefore we propose to use linear allocation functions instead of the optimal ones. In this chapter, we restrict to a specific class of linear allocation functions. In Section 6.2.1, we define this class and give a motivation for this restriction.

The decomposition approach presented in Chapter 5 has proven to be a fruitful approach to determine all the control parameters in the system. Therefore we suggest to apply this approach again, however, in this chapter the control parameters are determined under the restriction that each allocation function is within the specific class (see Section 6.2.1). Like in Chapter 5 we refer to the length of a review period as one period (\( R := 1 \)), and the lead times are multiples of \( R \). From the arguments stated in Section 5.6.1 it can be seen that this is not really a restriction.

6.2.1 Specific class of linear allocation functions

In this chapter, we restrict to a specific class of linear allocation functions. We assume that for each stockpoint \( j \in V \): the allocation function is given by

\[
\begin{align*}
\zeta_j(x) &= S_j - q_j \left( \sum_{n \in V_j} S_n - x \right) \quad \text{with} \quad \sum_{n \in V_j} q_n = 1, \quad q_n > 0 \quad \text{for all} \quad n \in V_j. \quad (6.1)
\end{align*}
\]

There are several reasons to restrict to this specific class. First, many properties of the optimal allocation functions also hold for the class of linear allocation functions given by (6.1). Specifically, the properties derived in both Lemma 5.4 and 5.5 are satisfied.

Second, allocation function (6.1) is a reasonable and easy way to ration the material, and has already been used in many papers, see for instance the papers on CAS and BS rationing policies addressed in Chapter 3. In order to determine the allocation function \( \zeta_j(x) \), we only need to determine the allocation fraction \( q_j \), besides the order-up-to-levels \( \{ S_j \}_{j \in V} \). Hence,

\[
\psi_j = \bigcup_{j \in V_j} \{ \zeta_j, S_j, \psi_j \} = \bigcup_{j \in V_j} \{ q_j, S_j, \psi_j \}.
\]

Third, for 2-echelon systems with identical holding and penalty costs at the end-stockpoints the optimal allocation function is within this class. Let us give a proof for this. Consider a divergent 2-echelon system, where the most upstream stockpoint (denoted by index 1) supplies a number of end-stockpoints. Let us assume that: (1) \( h_j := h \), and (2) \( p_j := p \) for every end-stockpoint \( j \in V \). Then from Theorem 5.2 and the algorithm depicted in Figure 5.1 it follows that

\[
F_j^{(s, t)}(\zeta_j(x)) = \frac{\lambda_j(x) + h_1 + p}{h_1 + h + p}, \quad j \in V_1, \quad (6.2)
\]
6.2. Analysis

Suppose the demand at each end-stockpoint is normally distributed. Then equation (6.2) can be rewritten to

\[
\Phi \left( \frac{\tilde{z}_j(x) - (L_j + 1)\mu_j}{\sigma_j \sqrt{L_j + 1}} \right) = \frac{\lambda_j(x) + h_1 + p}{h + h_1 + p}, \quad j \in V_1.
\]

Hence,

\[
\tilde{z}_j(x) = (L_j + 1)\mu_j + \sigma_j \sqrt{L_j + 1} \Phi^{-1} \left( \frac{\lambda_j(x) + h_1 + p}{h + h_1 + p} \right), \quad j \in V_1. \tag{6.3}
\]

Since the optimal allocation functions \( \{\tilde{z}_j\}_{j \in V_1} \) satisfy property (5.3) we obtain that

\[
\lambda_j(x) = (h + h_1 + p) \Phi \left( \frac{x - \sum_{a \in V_1} (L_a + 1)\mu_a}{\sum_{a \in V_1} \sigma_a \sqrt{L_a + 1}} \right) - (h_1 + p). \tag{6.4}
\]

Substitution of (6.4) in (6.3) yields

\[
\tilde{z}_j(x) = (L_j + 1)\mu_j + \sigma_j \sqrt{L_j + 1} \Phi^{-1} \left( \frac{h_1 + p}{h + h_1 + p} \right) \left( x - \sum_{a \in V_1} (L_a + 1)\mu_a \right), \quad j \in V_1. \tag{6.5}
\]

In order to prove that \( \tilde{z}_j \) given by (6.5) is an allocation function within the class of allocation functions satisfying (6.1), we proceed as follows. First, we determine the cost-optimal order-up-to-level \( \tilde{S}_j \) for every stockpoint \( j \in V_1 \) by minimizing \( D_j(S) \) for \( S_j \). Second, we show that by defining the allocation fractions appropriately the allocation function \( z_j(x) \) given by (6.1) (with \( S_j := \tilde{S}_j \) for \( j \in V_1 \)) coincides with (6.5). To determine \( \tilde{S}_j \) we use the newsboy-style equation (5.24). This yields

\[
\tilde{S}_j = (L_j + 1)\mu_j + \sigma_j \sqrt{L_j + 1} \Phi^{-1} \left( \frac{h_1 + p}{h + h_1 + p} \right), \quad j \in V_1. \tag{6.6}
\]

Substitution of (6.6) in (6.1) yields

\[
z_j(x) = (L_j + 1)\mu_j + q_j \left( x - \sum_{a \in V_1} (L_a + 1)\mu_a \right)
+ \Phi^{-1} \left( \frac{h_1 + p}{h + h_1 + p} \right) \left( \sigma_j \sqrt{L_j + 1} - q_j \sum_{a \in V_1} \sigma_a \sqrt{L_a + 1} \right), \quad j \in V_1. \tag{6.7}
\]

Next, by defining

\[
q_j := \frac{\sigma_j \sqrt{L_j + 1}}{\sum_{a \in V_1} \sigma_a \sqrt{L_a + 1}}, \quad j \in V_1,
\]

we note that (6.7) coincides with (6.5). The optimal allocation function (6.5) coincides with the allocation function of the well-known FS rationing policy of Eppen & Schrage [1981].

Finally, in Section 6.8 we consider 2-echelon systems with non-identical holding and penalty costs at the end-stockpoints. For these systems we provide some evidence that the error introduced by restricting to this specific class is small.
6.2.2 Analysis cost function

In the remainder of this section, we derive conditions for the order-up-to-levels and allocation fractions minimizing the expected total costs per review period. Furthermore, we introduce the following definition.

Definition 6.1. \( \Phi_i \) is quasi-optimal if for every successor \( j \in \mathcal{V}_i \):
(i) \( \hat{S}_j \) is optimal, i.e., \( \hat{S}_j \) is argmin \{ \( \mathbb{S} \) : \( \partial D_j(S, \Phi_j) / \partial S = 0 \) \}.
(ii) \( \Phi_j \) is quasi-optimal.

In Theorem 6.1, we derive a tractable expression for \( \partial D_i(S, \Phi_i) / \partial S_i \), by using Theorem 5.2, which enables us to determine a quasi-optimal control policy.

Theorem 6.1. Let \( \alpha_i^*(S_i) \) denotes the non-stockout probability of an end-stockpoint \( k \in \mathcal{E}_i \) in a divergent echelon system, in which the order-up-to-level of the most upstream stockpoint \( i \) equals \( S_i \). Then,

\[
\frac{\partial D_i(S_i)}{\partial S_i} = h_i - (h_i + \sum_{a \in \mathcal{E}_i} h_a + p_i) (1 - \alpha_i^*(S_i)) \quad i \in \mathcal{E}_i. \tag{6.8a}
\]

For an intermediate stockpoint \( i \in \mathcal{W}_k \) with quasi-optimal \( \Phi_i \) holds

\[
\frac{\partial D_i(S_i, \Phi_i)}{\partial S_i} = h_i + \sum_{a \in \mathcal{E}_i} q_{a,i} \left( h_{a,i} + \sum_{a \in \mathcal{E}_i} q_{a,i} \left( \cdots + \sum_{b \in \mathcal{E}_i} q_{b,a} \left( h_{b,a} + \sum_{b \in \mathcal{E}_i} q_{b,a} \left( \cdots \right) \right) \right) \right). \tag{6.8b}
\]

Proof. The proof is by induction on \( i \). If \( i \) is an end-stockpoint equation (6.8a) immediately results from differentiating (5.5b) to \( S_i \). For an intermediate stockpoint \( i \) we obtain from (5.5b)

\[
\frac{\partial D_i(S_i, \Phi_i)}{\partial S_i} = \left. \frac{\partial}{\partial S_i} \left( h_i(S_i) - (L_i + 1) \mu_i + \sum_{j \in \mathcal{V}_i} D_j(S_j, \Phi_j) + \int_{S_i - \sum_{a \in \mathcal{E}_i} \hat{S}_a}^{\infty} \frac{dz_j(x)}{dx} \left| \frac{\partial D_j(S_j, \Phi_j)}{\partial S_j} \right| \right) \right|_{S_i = \hat{S}_i} \int_{S_i - \sum_{a \in \mathcal{E}_i} \hat{S}_a}^{\infty} dF_i(z_j) \tag{6.8a}
\]

Differentiation of \( z_j(x) \) to \( x \) yields \( q_j \). Hence,

\[
\frac{\partial D_i(S_i, \Phi_i)}{\partial S_i} = h_i + \sum_{j \in \mathcal{V}_i} q_j \int_{\hat{S}_j - \sum_{a \in \mathcal{E}_i} \hat{S}_a}^{\infty} \left. \frac{\partial D_j(S_j, \Phi_j)}{\partial S_j} \right|_{S_j = \hat{S}_j} dF_i(S_j) \tag{6.8b}
\]

We distinguish between \( \hat{S}_i < \sum_{a \in \mathcal{E}_i} \hat{S}_a \) and \( \hat{S}_i \geq \sum_{a \in \mathcal{E}_i} \hat{S}_a \). In the former case we obtain

\[
\frac{\partial D_i(S_i, \Phi_i)}{\partial S_i} = h_i + \sum_{j \in \mathcal{V}_i} q_j \int_{0}^{\infty} \left. \frac{\partial D_j(S_j, \Phi_j)}{\partial S_j} \right|_{S_j = \hat{S}_j} dF_i(S_j) \tag{6.8c}
\]

In the latter case we obtain

\[
\frac{\partial D_i(S_i, \Phi_i)}{\partial S_i} = h_i + \sum_{j \in \mathcal{V}_i} q_j \int_{0}^{\infty} \left. \frac{\partial D_j(S_j, \Phi_j)}{\partial S_j} \right|_{S_j = \hat{S}_j} dF_i(S_j) \tag{6.8d}
\]
6.2. Analysis

In order to prove the theorem we note that

\[ \frac{\partial D_j(S_i, \hat{\Psi}_j)}{\partial S_j} \bigg|_{S_j = 0} = \frac{\partial D_j(S_i, \hat{\Psi}_j)}{\partial S_j} \bigg|_{S_j = 0} = 0. \]

Next, we use Theorem 5.2. Since the class of linear allocation functions is given by equation (6.1) we know that the condition to apply (5.10) is satisfied. To complete this proof we use induction and equation (5.10).

From Theorem 6.1 we prove convexity properties of the cost function \( D_i(S_i, \hat{\Psi}_i) \) with respect to \( S_i \) (see Corollary 6.1). From these properties and Theorem 6.1 we are able to derive the existence, the unicity and the nature of the extreme point \( \hat{S}_i \) for which \( \partial D_i(S_i, \hat{\Psi}_i) \)/\( \partial S_i = 0 \).

**Corollary 6.1.** The cost function \( D_i(S_i, \hat{\Psi}_i) \) with quasi-optimal policy \( \hat{\Psi}_i \) is convex with respect to \( S_i \). Specifically, if \( F_i(x) \) is strictly increasing for \( x \geq 0 \) then \( D_i(S_i, \hat{\Psi}_i) \) is strictly convex with respect to \( S_i \).

**Proof.** Taking the derivative of (6.8) to \( S_i \) yields

\[ \frac{\partial^2 D_i(S_i, \hat{\Psi}_i)}{\partial S_i^2} = \left( \begin{array}{c} (h_i + \sum_{x \in E} h_x + p_i) \frac{d\alpha_i^j(S_i)}{dS_i} \\ \sum_{i \in M} q_{i-1} \sum_{i \in M} q_{i-\ldots} \sum_{i \in M} q_{i+1} (h_i + \sum_{x \in E} h_x + p_i) \frac{d\alpha_i^j(S_i)}{dS_i} \\ \end{array} \right) \quad i \in E. \]

Using the monotonicity of the non-stockout probability completes the proof.

Consider a divergent multi-echelon system with positive penalty costs at the end-stockpoints. We conclude that if \( h_i \) is positive then there exist a finite \( S_i \) such that \( D_i(S_i, \hat{\Psi}_i) \) with quasi-optimal \( \hat{\Psi}_i \) is minimized for \( S_i \). Specifically, if \( F_i(x) \) is strictly increasing for \( x \geq 0 \) the unicity of this minimum is guaranteed. These results follow from Corollary 6.1 and the observation that if \( S_i \) tends to minus infinity then \( D_i(S_i, \hat{\Psi}_i) \) converges to \( -\sum_{x \in E} h_x - \sum_{i \in M} q_i \sum_{i \in M} q_{i-1} \sum_{i \in M} q_{i+1} p_i \), and if \( S_i \) tends to infinity then \( D_i(S_i, \hat{\Psi}_i) \) converges to \( h_i \). So, if \( h_i \) equals 0 the order-up-to-level \( \hat{S}_i \) equals infinity. Unfortunately, this results in an allocation function \( \hat{q}_i(x) \) which is intractable. This is the reason why we distinguish between two types of stockpoints:

(i). Stockpoint \( i \) adds a positive value to the product, i.e., \( h_i > 0 \) (results in a finite \( \hat{S}_i \)).

(ii). The stockpoint does not add any value to the product, i.e., \( h_i = 0 \) (results in an infinite \( \hat{S}_i \)).

Due to the intractability of the allocation function for the latter type we first address the divergent N-echelon systems where in each stockpoint some value is added to the product (see Sections 6.3-6.4). Later, we address how to deal with N-echelon systems in which for some stockpoints no value is added to the product (see Sections 6.5-6.7).

From (6.8) we derive the following **optimality conditions** for a quasi-optimal control policy (given the linear allocation functions of (6.1)):

The order-up-to-level \( \hat{S}_i \) of an end-stockpoint \( i \) satisfies

\[ h_i - (h_i + \sum_{x \in E} h_x + p_i)(1 - \alpha_i^j(\hat{S}_i)) = 0. \tag{6.9a} \]

The order-up-to-level \( \hat{S}_i \) of an intermediate stockpoint \( i \) and its downstream allocation fractions satisfy

\[ h_i + \sum_{i \in M} q_{i-1} \left( h_i + \sum_{i \in M} q_{i-\ldots} \left( \ldots + \sum_{i \in M} q_{i+1} \left( h_i - (h_i + \sum_{x \in E} h_x + p_i)(1 - \alpha_i^j(\hat{S}_i)) \right) \ldots \right) \right) = 0. \tag{6.9b} \]
where the allocation fractions of each stockpoint in ech(i) sum up to one, and are positive.

In Chapter 5 we proved that the order-up-to-level of a stockpoint \( i \) in the cost-optimal control policy satisfies a newsboy-style expression:

\[
a_k^*(\delta_i) = \frac{\sum_{a \in \epsilon i} h_a + p_k}{h_k + \sum_{a \in \epsilon i} h_a} \quad \text{for every} \quad k \in E_i. \tag{6.10}
\]

In this chapter we approximate the optimal allocation functions by linear allocation functions. Therefore (6.10) does not necessarily have to hold any longer. However, when (6.10) is satisfied for each stockpoint, then the optimality conditions of (6.9) are also satisfied.

6.3 Algorithm

In this section, we develop an algorithm to determine all the control parameters of the divergent multi-

echelon system. As already mentioned in the previous section, we first restrict ourselves to divergent

\( N \)-echelon systems where in every stockpoint some value is added to the product.

Figure 6.1 depicts procedure \textit{Main}, which determines the order in which the control parameters are computed.

\begin{verbatim}
procedure Main
begin
  n := 1;
  while n < N do
    begin
      for i \in W_n do
        begin
          if h_i > 0 then
            begin
              for k \in E_i do \( \delta_k^i := (\sum_{a \in \epsilon i} h_a + p_k)/(h_k + \sum_{a \in \epsilon i} h_a + p_k) \);
              ComputeParameters(i);
              Adaptation(i);
            end
          end
        end
      n := n + 1;
    end
end
\end{verbatim}

\textbf{Figure 6.1. Decomposition algorithm.}

This order results from the decomposition approach as explained in Section 5.5. This means that we first determine the order-up-to-levels at the end-stockpoints. Next, we continue with the determination of the control parameters of stockpoints with LLC \( = 2 \), LLC \( = 3 \), respectively, until we reach the most upstream stockpoint in the system. As can be seen in the procedure \textit{ComputeParameters} in Figure 6.2 we distinguish between end-stockpoints and intermediate stockpoints, since for an end-stockpoint we only need to compute the order-up-to-level, whereas for an intermediate stockpoint we also need to determine the allocation fractions.
6.3. Algorithm

procedure ComputeParameters(i)
begin
if i ∈ E then determine $\hat{S}_i$ by service-equation $\alpha'_i(\hat{S}_i) = \hat{d}'_i$ else
ComputeLocalParameters($\hat{S}_i, (q_j)_{i \in V}$)
end

Figure 6.2. Procedure to determine all the control parameters of stockpoint $i$.

For an end-stockpoint $i$, we determine the order-up-to-level from

$$\alpha'_i(\hat{S}_i) = \hat{d}'_i := \frac{\sum_{n \in V_i} h_n + p_i}{h_i + \sum_{n \in \ell_i} h_n + p_i}, \quad i \in E. \quad (6.11)$$

For an intermediate stockpoint $i$ we determine the order-up-to-level $\hat{S}_i$ and the allocation fractions $(q_j)_{i \in V}$ from the procedure ComputeLocalParameters (see Figure 6.3).

procedure ComputeLocalParameters($\hat{S}_i, (q_j)_{i \in V}$)
begin
initialize $S_i$, $\epsilon$, and stepsizes;
repeat
$\Delta_i := S_i - \sum_{n \in V_i} \hat{S}_n$;
for $j \in V, do$
begin
for $k \in E_j, do$ determine $q_{j(k)}$ from $\alpha'_i(\Delta_i, q_{j(k)}) = \hat{d}'_i$
$q_j := \sum_{k \in E_j} q_{j(k)}/|E_j|$
end
if $\sum_{n \in V} q_n \leq 1 - \epsilon$ then $S_i := S_i + \text{stepsizes}$;
if $\sum_{n \in V} q_n \geq 1 + \epsilon$ then $S_i := S_i - \text{stepsizes}$;
until $|\sum_{n \in V} q_n - 1| < \epsilon$
$\hat{S}_i := S_i$
end

Figure 6.3. Procedure to determine the order-up-to-level $\hat{S}_i$ and the allocation fractions $(q_j)_{i \in V}$ of an intermediate stockpoint $i$, when this stockpoint and all its successors have positive added values.

This procedure is based on the CAS1 heuristic as described in Section 3.2.1. When LLC(i)=2 the aforementioned heuristic coincides completely with the procedure ComputeLocalParameters. This means that we start with some initial value for $S_i$. Then we compute $\Delta_i := S_i - \sum_{n \in V_i} \hat{S}_n$. Next, the allocation fraction $q_j$ is determined from

$$\alpha'_i(\Delta_i, q_j) = \hat{d}'_j := \frac{\sum_{n \in \ell_j} h_n + p_j}{h_j + \sum_{n \in \ell_j} h_n + p_j}, \quad j \in E_i. \quad (6.12)$$

Since we require $\sum_{n \in V} q_n = 1$ we adapt $S_i$ until this additional constraint is also satisfied. Due to the monotonicity of the non-stockout probability we may use a bisection scheme for $S_i$. Note that in
the determination of \( q_i \) from (6.12) we use the already determined order-up-to-levels \( \{ \hat{S}_j \}_{j \in V_i} \). The heuristic can not be extended straightforwardly for a stockpoint \( i \) with LLC(\( i \)) > 2. To illustrate this we consider the following example. Suppose the decomposition algorithm already computed the allocation fractions \( q_u, q_w, \) and the order-up-to-levels \( S_u, S_w, S_y \) and \( S_z \) of the system as depicted in Figure 2.1. In order to obtain a quasi-optimal control policy we have to determine \( q_u, q_w \) and \( S_y \) such that the service level equations given by (6.12) are satisfied for each end-stockpoint, and \( q_u + q_w = 1 \). It is clear that in general the existence of such \( q_u, q_w \) and \( S_y \) satisfying these constraints is not guaranteed, since the number of constraints (four) exceeds the number of variables (three). In order to still use the decomposition approach we suggest to define \( q_y \) as in procedure ComputeLocalParameters (see Figure 6.3). This method is only justifiable when the differences between the values of \( q_{j,k} \) for the end-stockpoints \( k \in E_j \) are small. Because otherwise averaging these allocation fractions \( q_{j,k} \) implies that for some end-stockpoints the defined value \( q_y \) is too large and consequently the resulting service performance is too low, or the defined \( q_y \) is too small and consequently the resulting service performance is too large. This probably results in a bad performance of the control policy, since (6.9) does no longer hold.

After the determination of the order-up-to-level of an intermediate stockpoint, it may occur that this level is less than the sum of the order-up-to-levels of its successors. In that case, we adapt the order-up-to-level of every successor as depicted in the procedure Adaptation (see Figure 6.4).

```plaintext
procedure Adaptation(i)
begin
if \( \hat{S}_i < \sum_{j \in V_i} \hat{S}_j \) then
begin
for \( j \in V_i \) do
begin
\( \hat{S}_j := z_j(\hat{S}_j); \)
Adaptation(j)
end
end
end
```

**Figure 6.4.** Procedure to adapt the order-up-to-levels.

This procedure is based on the following theorem.

**Theorem 6.2.** Consider a multi-echelon system with most upstream stockpoint \( I \). The allocation functions are defined as in (6.1). If \( S_i < \sum_{j \in V_i} S_j \) then we redefine the order-up-to-level of each successor \( j \in V_i \) as follows \( S_j := z_j(S_i) \). For this adapted control policy holds
(i) The total expected costs in the multi-echelon system do not alter.
(ii) The non-stockout probabilities at the end-stockpoints do not alter.

**Proof.** (i) For our convenience we denote the order-up-to-level and the allocation function after the adaptation by \( \hat{S} \) (e.g., \( z_j(x) \) becomes \( \hat{z}_j(x) \)). An important property of the adapted order-up-to-levels is

\[
\sum_{j \in V_i} \hat{S}_j = \sum_{j \in V_i} z_j(S_i) = S_i. \tag{6.13}
\]
6.4. Numerical results

From the definition of $z_j(x)$ (see (6.11)) and (6.13) it follows

$$z_j^*(x) = S_j^* - q_j \left( \sum_{i \in V_i} S_i^k - x \right) = z_j(S_j) - q_j(S_j - x) = z_j(x), \quad j \in V_1. \quad (6.14)$$

So this adaptation of $S_j$ into $S_j^*$ does not influence the allocation function. Then result (i) follows immediately by substitution of (6.13) and (6.14) in (5.5).

(ii) From (5.10) we obtain an expression for the non-stockout probability of an end-stockpoint when $\Delta_1 < 0$. Due to (6.14) the non-stockout probability does not alter after adaptation. □

6.4 Numerical results

In this section, the performance of the algorithm of Section 6.3 is tested by considering 500 instances of the divergent 3-echelon system as depicted in Figure 6.5. The lead time of each intermediate stockpoint is drawn from a uniform distribution on $[1, \ldots, 8]$, and the lead time of an end-stockpoint is drawn from a uniform distribution on $[1, \ldots, 5]$. The mean demand and squared coefficient of variation per review period at an end-stockpoint is drawn from a uniform distribution on $[10, 25]$ and $[0.5, 1.5]$, respectively. Every stockpoint adds some value to the product. The amount added is uniformly distributed on $[0.1, 3]$.

![Diagram of a divergent 3-echelon inventory system]

**Figure 6.5.** Divergent 3-echelon inventory system.

Let $H_k$ denotes the holding costs of one product at end-stockpoint $k$ per review period, i.e., $H_k = h_k + \sum_{i \in I_k} h_i$. For a quasi-optimal control policy which satisfies (6.10) for all $i \in M \cup E$ we have that the non-stockout probability attained at end-stockpoint $k$ equals

$$\alpha_k = \frac{p_k}{H_k + p_k}, \quad k \in E. \quad (6.15)$$

Rewriting (6.15) yields

$$p_k = \frac{\alpha_k}{1 - \alpha_k} H_k, \quad k \in E. \quad (6.16)$$

In practice the service non-stockout probability required at an end-stockpoint usually is large, therefore we draw $\alpha_k$ from a uniform distribution on $[0.85, 0.99]$. So by defining $p_k$ as in (6.16) we know that if the algorithm would be exact then the non-stockout probability of end-stockpoint $k$ equals $\alpha_k$. 
Unfortunately, the algorithm of Section 6.3 approximates the quasi-optimal control policy, since (1) the algorithm uses an approximate step in procedure *ComputeLocalParameters* by averaging the allocation quantities \( q_{j(t)} \), and (2) the non-stockout probabilities which are computed in the algorithm...
6.5. Extension of the decomposition algorithm

use the method as described in Section 2.4.1. Therefore the non-stockout probability in stockpoint \( k \) resulting from the algorithm, denoted by \( q^*_k \), generally differs from \( q_k \). Next, every instance is simulated with the control parameters obtained by the algorithm. In case the imbalance would not affect the non-stockout probability attained at stockpoint \( k \), it would be equal to \( q^*_k \). However, usually the phenomenon of imbalance does have some effect on the stockout probability. Hence, stockpoint \( k \) attains a non-stockout probability of \( q^*_k \).

Figures 6.6 (a) and (b) depict the absolute differences \( q^*_k - q_k \), \( q^*_k - q^*_k \) and \( q^*_k - q_k \) for an end-stockpoint \( k \) with \( a_k \leq 0.95 \) and \( a_k > 0.95 \), respectively. Note that \( q^*_k - q_k \) represents the 'Algorithmic error' due to averaging \( q_{(k)} \) in procedure ComputeLocalParameters and using the approximate method as described in Section 2.4.1 to compute the non-stockout probabilities, \( q^*_k - q^*_k \) represents the 'Imbalance error' due to the violation of the balance assumption, and \( q^*_k - q_k \) represents the 'Total error'. We distinguish between these two cases \( a_k \leq 0.95 \) and \( a_k > 0.95 \), since an absolute error of 0.01 is acceptable in case \( a_k \) is not to large (e.g., 0.85), although, when \( a_k \) is large (e.g., 0.99) such an error is intolerable. Comparing these two figures indicates that when \( a_k \) is large the absolute errors diminish. Furthermore, we conclude that the algorithm works very well. Most of the difference between \( a_k \) and \( q^*_k \) is caused by imbalance.

6.5 Extension of the decomposition algorithm

Unfortunately, the algorithm addressed in Section 6.3 can only be applied if every stockpoint adds some positive value to the product. The case where some of the stockpoints does not add any value is also interesting from a practical point of view. Therefore, in the remainder of this chapter we analyze divergent multi-echelon systems where for some stockpoints no value is added to the product. Typically, such situations occur when components or subassemblies are shipped from the out-bound stockpoint of a supplier to the in-bound stockpoint of a customer (see Figure 6.7). Note that such a situation occurs in many real-world supply chains.

![Supply chain with alternating of stockpoints with and without added value.](image)

For these situations the decomposition algorithm of Section 6.3 cannot be applied. By considering an example, it becomes clear where an extension of the algorithm is required. Consider the distribution system as depicted in Figure 6.8.

First, the algorithm determines the order-up-to-level of every end-stockpoint by solving (6.11). Since \( h_1 = h_5 = 0 \) it follows from (6.11) that \( \delta^*_1 = \delta^*_5 = 1 \). Hence, the order-up-to-levels of end-stockpoints 3 and 5 are infinite, i.e., \( S_3 = S_5 = \infty \). On the other hand \( h_4 \) and \( h_6 \) are positive, therefore \( \delta^*_4 \) and \( \delta^*_6 \) are less than 1. Hence, the order-up-to-levels of stockpoint 4 and 6 equals some finite value \( S_4 \) and \( S_6 \), respectively.

Second, the algorithm considers stockpoint 2 (with low level code 2). From (6.12) we determine \( \delta^*_2 \) and \( \delta^*_2 \). Since stockpoint 2 does not add any value to the product we have that \( \delta^*_2 = \delta^*_2 \) and \( \delta^*_2 = \delta^*_2 \). Hence, \( S_2 = \infty \). Since the algorithm uses the linear allocation functions of (6.1), it is not clear how
to determine appropriate allocation fractions $q_4$ and $q_6$.

Finally, we consider the most upstream stockpoint 1 (with low level code 3). Again, the target non-stockout probabilities $\delta_k^4$ for $k \in \{3, \ldots, 6\}$ are defined by (6.12). Since stockpoint 1 has at least one successor with an infinite order-up-to-level, it follows that upon an order arrival at stockpoint 1, these products are immediately ordered by stockpoints 2 and 3. This means that stockpoint 1 never holds any stock, so $S_1 := \hat{S}_1 = \bar{S}_1$. Using similar arguments we have $S_2 = \hat{S}_2 + \bar{S}_2$. So $\Delta_1 = \Delta_3 = 0$. Using sample path arguments the remaining control parameters need to be determined such that

$$\delta_k^4 = \begin{cases} \text{Pr}(\hat{S}_k - q_k \ell_{k+1} - D_{k+1}^t \geq 0) & k \in \{3, 4\} \\ \text{Pr}(\hat{S}_k - q_k (\ell_{k+1} + q_k \ell_k) - D_{k+1}^t \geq 0) & k \in \{5, 6\}. \end{cases}$$

(6.17)

From equation (6.17) with $k = 4$, the allocation fraction $q_4$ is determined. The cost-optimal conditions do not impose any constraints on $q_2$ and $q_3$. We have one degree of freedom in choosing these allocation fractions $q_2$ and $q_3$, since the allocation fractions of stockpoint 1 have to sum up to 1. We suggest to use this degree of freedom to minimize the expected imbalance at stockpoint 1, since the decomposition approach requires that there is no imbalance at every stockpoint in the system. The BS rationing policy tries to establish this. Like in Section 3.2.2 we distinguish between two variants, referred to as BS1 and BS2. In the BS1 rationing policy the allocation fractions are determined such that the mean imbalance at stockpoint $i$ (i.e., $\Omega^i$) is minimized. In Section 3.2.2 we determine all allocation fractions such that an approximate expression of the mean imbalance is minimized. However, in this chapter some of the allocation fractions may already be chosen in order to minimize the expected total costs (e.g., in this example $q_4$ is already known). Let the set $A_i$ denoting those successors of stockpoint $i$ for which the allocation fractions result from the cost minimization, while the other successors are denoted by $B_i$. Note that

$$A_i := \{j \in V_i | h_j > 0\}, \quad B_i := \{j \in V_i | h_j = 0\}.$$

So, for every successor $j \in A_i$, the allocation fraction $q_j$ is determined by the algorithm of Section 6.3. The BS1 rationing policy determines the allocation fractions $q_j$ for $j \in B_i$ such that

$$\frac{dE[\Omega^j]}{dq_j} = \frac{\phi(\delta_j^4)}{\sigma_{\delta_j}^2} T_i (2q_j \sigma_j^2 - \sigma_j^2) = c_j, \quad j \in B_i,$$

(6.18)
6.6. Algorithm

We use bisection to find \( c_i \) such that the allocation fractions \( \{q_j\}_{j \in B_i} \) sum up to \( 1 - \sum_{n \in A_i} q_n \). In each step of the bisection, the corresponding values for \( \{q_j\}_{j \in B_i} \) are found by another bisection, where

\[
q_j \in \begin{cases} 
[0, 0.5 - \frac{\sigma_j^2}{(2\sigma^2)}) \sum_{n \in A_i} q_n > 0.5 \\
[0.5 - \frac{\sigma_j^2}{(2\sigma^2)}, 1] \sum_{n \in A_i} q_n \leq 0.5 
\end{cases}, \quad j \in B_i
\]

Another variant of the BS rationing policy is referred to as BS2. Instead of minimizing the mean imbalance, we suggest to minimize \( \sum_{j \in B_i} \sigma^2 q_j \) subject to \( \sum_{n \in A_i} q_n = 1 - \sum_{n \in A_i} q_n \). The Lagrange-multiplier technique yields

\[
q_j = \frac{\sigma_j^2}{2\sigma^2} + \frac{1}{|B_i|} \left( 1 - \sum_{n \in A_i} q_n - \sum_{n \in B_i} \frac{\sigma_j^2}{2\sigma^2} \right), \quad j \in B_i,
\]  

6.6 Algorithm

In Section 6.5 the extension of the decomposition algorithm is explained by applying it to the system depicted in Figure 6.8. In this section we formalize the extension. Note that only the procedure \texttt{ComputeParameters} needs to be adapted, while the procedures \texttt{Main}, \texttt{ComputeLocalParameters} and \texttt{Adaptation} remain valid. Figure 6.9 depicts the adapted version of the procedure \texttt{ComputeParameters}.

\begin{verbatim}
procedure ComputeParameters(i)
begin
  if \( i \in \mathcal{E} \) then determine \( \hat{s}_i \) from \( \alpha'(\hat{s}_i) = \hat{a}_i \) else
  begin
    if \( B_i = \emptyset \) then ComputeLocalParameters(\( \hat{s}_i, \{q_j\}_{j \in \mathcal{E} \setminus \mathcal{A}_i} \) else
    begin
      \( \Delta_i := 0; \)
      for \( j \in \mathcal{A}_i \) do
      begin
        for \( k \in \mathcal{E}_j \) do determine \( q_{j/k} \) from \( \alpha'_i(\Delta_i, q_{j/k}) = \hat{a}_i \);
        \( q_j := \sum_{k \in \mathcal{E}_j} q_{j/k} / |\mathcal{E}_j| \);
      end
      for \( j \in B_i \) do determine \( q_j \) by BS rationing such that \( \sum_{n \in B_i} q_n = 1 - \sum_{n \in A_i} q_n \);
      \( \hat{s}_i := \sum_{j \in B_i} \hat{s}_j \);
    end
  end
end
\end{verbatim}

\textbf{Figure 6.9.} Procedure to determine all the control parameters of an intermediate stockpoint \( i \) and all the unknown parameters of its downstream stockpoints.
By closely examining procedure \textit{ComputeParameters} we observe that when stockpoint \( i \) has positive added value, it not only determines the order-up-to-level \( \hat{S}_i \) and the allocation fractions \( \{q_i\}_{i \in V} \), but also the control parameters of the downstream stockpoints \( j \in \text{ech}(i) \) which: (1) does not add any value, and (2) all stockpoints on the path from \( i \) to \( j \) do not add any value. This is accomplished by the fact that the procedure is recursive. The computational effort of this procedure is dominated by the effort to solve several one-dimensional service equations. These equations are solved by a bisection scheme on the unknown parameter, e.g., in equation (6.17) we solve \( q_k^* = q_k^1 \) by bisection on \( q_k \in (0, 1) \). Like in procedure \textit{ComputeLocalParameters} we use the technique of averaging the \( q_k^* \) when the number of the stockpoints in \( E_j \) exceeds one. As already mentioned in Section 6.3 this is only justifiable when the differences between the values of \( q_k^* \) for \( k \in E_j \) are small.

### 6.7 Numerical results

In this section, the performance of the extension of the decomposition algorithm of Section 6.3 is tested by considering the same 500 instances of the 3-echelon distribution system as depicted in Figure 6.5. Only in this experiment, we assume that all end-stockpoints (except for the most upstream stockpoint) have no added value. Hence, without loss of generality \( H_k := 1 \) for \( k \in E \). So from (6.16) we know that by defining \( \rho_k := \alpha_k/(1 - \alpha_k) \) for \( k \in E \) the non-stockout probability attained at end-stockpoint \( k \) would be equal to \( \alpha_k \) if the algorithm were exact. From Section 6.4, however, we know that the algorithm uses two approximate steps.

Figures 6.10 (a) and (b) depict the absolute differences \( \alpha_k^4 - \alpha_k, \alpha_k^3 - \alpha_k^4 \) and \( \alpha_k^5 - \alpha_k \) for an end-stockpoint \( k \) with \( \alpha_k \leq 0.95 \) and \( \alpha_k > 0.95 \), respectively. Note that \( \alpha_k^4 - \alpha_k \) represents the 'algorithmic error' (Alg. error) due to the two approximate steps, and \( \alpha_k^5 - \alpha_k^4 \) represents the 'imbalance error' (Imm. error) due to the violation of the balance assumption. Like in Section 6.4 we distinguish between the two cases \( \alpha_k \leq 0.95 \) and \( \alpha_k > 0.95 \).

For both cases the 'algorithmic error' of BS1 and BS2 are almost identical and very small. This would advocate to use BS2 since it is much easier than BS1, which requires a nested bisection scheme. However, it turns out that the variability of the 'imbalance error' is smaller when applying BS1 instead of BS2. Comparing Figures 6.10 (a) and (b) suggests that the effect becomes stronger for \( \alpha_k \) not to large. Furthermore, this figure shows that the 'algorithmic error' is considerably smaller than the 'imbalance error'. Hence in order to improve the performance of the algorithm it is probably more efficient to focus our attention to the latter error. A way to reduce this error is by keeping some stock in every intermediate stockpoint \( i \) for which the algorithm suggests not to hold any stock, i.e., \( \Delta_i = 0 \). This means that we increase \( \Delta_i \) in order to keep some products in stock. Like in Section 3.5.1, we defined \( \Delta_i \) as in (3.12). In Figure 6.11 this adaptation of the algorithm is denoted by \( \Delta > 0 \), while the original algorithm is denoted by \( \Delta = 0 \). Figure 6.11 depicts the mean of the expected costs per period for all 500 instances, for both BS1 and BS2. It turns out that differences in costs between BS1 and BS2 are negligible. Furthermore, the difference between the expected total costs per period computed by the analysis and resulting from the simulation is much smaller in the case \( \Delta > 0 \) than \( \Delta = 0 \). This is due to the reduction of the imbalance in the case \( \Delta > 0 \). Finally, Figure 6.11 shows that the expected total costs per period by adapting the algorithm increases.
6.7. Numerical results

Figure 6.10. Performance of algorithm.
6.8 Error made by linearizing allocation functions

In Chapter 5, we derived several properties of the optimal allocation functions. From these properties we developed an algorithm (see Figure 5.1) to actually obtain these functions. Usually it is rather time-consuming and cumbersome to determine these functions. Note that there are some exceptions (e.g., the 2-echelon system with identical holding and penalty costs at the end-stockpoint(s)). Hence, for practical purposes we restricted ourselves to a specific class of linear allocation functions. By this restriction the expected costs incurred in the system per review period will probably be larger. To obtain insight in the magnitude of this increase we analyze the difference between the expected costs per review period in a system when applying the optimal allocation functions \(\{\xi\}_i\) and when applying the linear allocation functions \(\{\xi\}_i\) given by (6.1). For presentational purposes we consider the 2-echelon system as depicted in Figure 6.12. Without loss of generality we assume that \(p_2 \leq p_3\). When

Figure 6.12. A divergent 2-echelon system with two end-stockpoints.
the order-up-to-level in each stockpoint is identical for both cases, then from Lemma 5.2 it follows that
\[ D_j(S_i, \Psi_i) - D_j(S_i, \bar{\Psi}_i) = \sum_{j \in V_i} \int_0^{\infty} D_j(z_j(S_i - u)) - D_j(\tilde{z}_j(S_i - u)) \, dF^{(z_j)}(u), \] (6.20)
with \( \Psi_i := \bigcup_{j \in V_i} (z_j, \tilde{z}_j) \) and \( \bar{\Psi}_i := \bigcup_{j \in V_i} (\tilde{z}_j, \tilde{\tilde{z}}_j) \). To obtain an upperbound for (6.20) we distinguish between \( \Delta_i \leq u < u^* \) and \( u \geq u^* \), where \( u^* := \Delta_i + \max_{j \in V_i} \left( \frac{\tilde{z}_j}{\tilde{s}_j} \right) \). So,
\[ D_j(S_i, \Psi_i) - D_j(S_i, \bar{\Psi}_i) = \sum_{j \in V_i} \int_0^{u^*} D_j(z_j(S_i - u)) - D_j(\tilde{z}_j(S_i - u)) \, dF^{(z_j)}(u) + \sum_{j \in V_i} \int_{u^*}^{\infty} D_j(z_j(S_i - u)) - D_j(\tilde{z}_j(S_i - u)) \, dF^{(\tilde{z}_j)}(u). \] (6.21)
In order to obtain an upperbound for (6.20) we derive an expression for the latter term of (6.21), whereas the former term remains a topic for further research. For developing such expression for this latter term we need the values of \( z_j(S_i - u) \) and \( \tilde{z}_j(S_i - u) \). First, from (6.1) it follows that
\[ z_j(S_i - u) = S_j + q_j(\Delta_i - u), \quad j \in V_i. \] (6.22a)
Second, from Theorem 5.5 we know that for \( u \geq S_i \)
\[ \tilde{z}_j(S_i - u) = \begin{cases} S_i - u - x_j^* & j = 2 \\ x_j^* & j = 3 \end{cases}, \quad j \in V_i. \] (6.22b)
Let us assume that \( u \geq u^* \). Then \( z_j(S_i - u) \leq 0 \). Thus, from (6.22a) and (5.3a) it follows that
\[ D_j(z_j(S_i - u)) = (h_1 + p_j)((L_j + 1)u_j - S_j - q_j(\Delta_i - u)), \quad j \in V_i. \]
Since \( u \geq u^* \) implies \( u \geq S_i \), we use (6.22b) to derive an expression for \( D_j(\tilde{z}_j(S_i - u)) \). Before substituting (6.22b) in \( D_j(\tilde{z}_j(S_i - u)) \) we prove that indeed \( u \geq S_i \). For each \( j \in V_i \) we have that \( u \geq u^* \geq \Delta_i + S_j/q_i \). After some algebra we obtain that \( q_j(\Delta_i - u_j) \geq S_j \) for each \( j \in V_i \). Summarizing over \( j \in V_i \) on both sides yields \( u - \Delta_i \geq \sum_{j \in V_i} S_j \). Using the definition of \( \Delta_i \) proves \( u \geq S_i \). From \( u \geq S_i \), (6.22b) and (5.3a) it follows that
\[ D_j(\tilde{z}_j(S_i - u)) = \begin{cases} (h_1 + p_j)((L_j + 1)u_j - S_j - u + x_j^*) & j = 2 \\ (h_1 + p_j)((L_j + 1)u_j - x_j^*) + (h_1 + h_j + p_j)E[(x_j^* - D_j^{(\tilde{z}_j)}))]^+ & j = 3 \end{cases}. \]
After some elementary algebra we obtain
\[ \sum_{j \in V_i} D_j(z_j(S_i - u)) - D_j(\tilde{z}_j(S_i - u)) = (p_3 - p_2)(x_j^* - S_j - q_j(\Delta_i - u)) \\ - (h_1 + h_3 + p_3)E[(x_j^* - D_j^{(\tilde{z}_j)}))]^+ \leq (p_3 - p_2)(x_j^* - S_j - q_j(\Delta_i - u)). \] (6.23)
Note that if \( p_2 = p_3 \) the upperbound (6.23) is tight, since then \( x_j^* = 0 \). From (6.23) it follows that
\[ \sum_{j \in V_i} \int_0^{\infty} D_j(z_j(S_i - u)) - D_j(\tilde{z}_j(S_i - u)) \, dF^{(z_j)}(u) \leq \int_0^{\infty} (p_3 - p_2)(x_j^* - S_j - q_j(\Delta_i - u)) \, dF^{(z_j)}(u) \]
\[ = (p_3 - p_2)(x_j^* - q_j(\Delta_i - u) + (p_3 - p_2)q_jE[(D_j^{(z_j)}(u))]^+ \]
\[ = (p_3 - p_2)(x_j^* - q_j(\Delta_i - u) + (p_3 - p_2)q_jE[(D_j^{(\tilde{z}_j)} - u^+))]^+ \]. (6.24)
Note that if \( S_2/\theta_2 \geq S_1/\theta_1 \), we can simplify this upperbound even more since \( z_3(S_1 - u^*) = 0 \), otherwise \( z_3(S_1 - u^*) < 0 \).

The upperbound given by (6.24) is small for most instances. Therefore we expect we that by restricting to the specific class of linear allocation functions given by (6.1) we do not introduce a large error. But to draw to this conclusion we also have to obtain some insight in the impact of the former term of (6.21) on the error. For that purpose we determine the optimal allocation functions and the optimal linear allocation functions for a small example. Consider the divergent 2-echelon system as depicted in Figure 6.12. Stockpoint 1 has a lead time \( L_1 = 3 \), and the added value equals \( h_1 = 7 \). Stockpoint 2 has a lead time \( L_2 = 2 \), the demand process has the following characteristics \( \mu_2 = 10 \) and \( \sigma_2 = 8 \), and the penalty costs equal \( p_2 = 15 \). Stockpoint 3 faces a lead time \( L_3 = 3 \), the demand process has the following characteristics \( \mu_3 = 20 \) and \( \sigma_3 = 16 \), and the penalty costs equal \( p_3 = 20 \).

Then Figure 6.13 (a) depicts the optimal allocation functions and the linear allocation functions (obtained by applying the algorithms of this chapter) for \( h_2 = 3 \) and \( h_3 = 5 \). Note that the behavior of stockpoint 2 and 3 coincides with the behavior depicted in Figure 5.2 (c) and (b), respectively. The dotted lines in Figure 6.13 (a) show the optimal order-up-to-levels \( \hat{S}_2 = 46.4 \) and \( \hat{S}_3 = 126.32 \). For \( x \leq \hat{S}_2 + \hat{S}_3 = 184.78 \) the behavior of the optimal allocation functions are irrelevant, since then each end-stockpoint can raise its echelon inventory position to the order-up-to-level and the remainder is retained at stockpoint 1. From (6.21) we know that the allocation function is evaluated around \( x = \bar{x} = S_1 - u \). Since \( u \) equals the lead time demand of stockpoint 1, we know that approximately 98% of the times \( u \in (0, 72) \). Since \( \bar{x} = \Delta_1 + \Delta_2 + \Delta_3 \) and \( \Delta_3 = 53 \) we have that 98% of the times \( x \in [139, 211] \). Hence, the range for \( x \in [139, 184] \) affects the cost function the most. From figure 6.13 (a) we conclude that the optimal allocation functions are very well approximated by the linear ones for \( x > 50 \). Therefore, we expect that the error made by linearizing the error is very small for this example. Also the upperbound derived in (6.24) is very small. Since \( x_2^* = 47.68 \) and \( a^* = 232.5 \) we obtain an upperbound of approximately 0.11.

Figure 6.13 (b) depicts the optimal allocation functions and the linear allocation functions (obtained by applying the algorithm developed in this chapter) for \( h_2 = 5 \) and \( h_3 = 6 \). Note that the behavior of stockpoint 2 and 3 coincides with the behavior depicted in Figures 5.2 (a) and (d), respectively. Using similar arguments as stated above we expect that the error made by linearizing the optimal allocation function as suggested in Chapter 6 is small.

It is difficult to derive a good upperbound for the first term of (6.21). This remains a topic for further research. But by considering a numerical example of the divergent 2-echelon system as depicted in Figure 6.12 we illustrated that the linear allocation functions derived in this chapter approximates the optimal allocation functions very well for the most important part. Therefore we expect that the error made by linearizing the allocation function is acceptable. An extensive numerical experiment is required to determine how the error relates to the parameters of the system. This is also a subject for further research.

6.9 Conclusions

In this chapter, we addressed the problem of determining the control parameters of the divergent multi-echelon system as formulated in Chapter 2 such that the expected holding and penalty costs per review period are minimized. In Chapter 5, we proved that a decomposition of the system is exact given the balance assumption. Hence, the complex multi-dimensional problem of determining these parameters reduces to the problem of determining (for every stockpoint): (1) the optimal order-up-to-policy, and (2) the optimal allocation functions to its successors. It is rather cumbersome and time-consuming to
Figure 6.13. The optimal allocation function $\tilde{z}_j(x)$ and the linear allocation function $z_j(x)$ for the divergent 2-echelon system as depicted in Figure 6.12 for $j = 2, 3$. 

Determine these optimal allocation functions. Therefore, we restricted ourselves to a specific class of linear allocation functions. An algorithm is developed to compute the order-up-to-level and its allocation fractions such that the expected total costs is minimized as much as possible. The algorithm is validated by a numerical experiment on a 3-echelon system, which shows very good results.
Cost minimization: linear allocation functions
Lateral transshipments

7.1 Introduction

So far we studied the basic model as formulated in Chapter 2 to determine the control parameters such that the service level constraints at the end-stockpoints are satisfied (Chapters 3 and 4), and the expected systemwide holding and penalty costs are minimized (Chapters 5 and 6). In this chapter, we study the same divergent multi-echelon system. But now we allow for lateral transshipments between the end-stockpoints. By these transshipments we can decrease the expected holding costs needed to operate the system, but still guarantee the target customer service levels. On the other hand by allowing these lateral transshipments extra transportation (and handling) costs are involved. So the appropriateness of using lateral transshipments depends on the trade off between the extra costs involved with the transshipments and the decrease in holding costs. The objective of this chapter is to determine the control parameters in a divergent multi-echelon system with lateral transshipments between the end-stockpoints, such that all these end-stockpoints satisfy their service level constraints. From this analysis we are able to compare the model without transshipments with the model with transshipments. This enables us to quantify the decrease in holding costs and the increase of transshipments costs when allowing the transshipments. Hence we can determine for which instances it is profitable to allow for these lateral transshipments.

The major part of this chapter is devoted to a divergent 2-echelon system consisting of a central depot (CD) supplying a number of retailers. Both the CD and every retailer control their stock by a periodic review echelon order-up-to-policy, i.e., every review period a replenishment order is issued that raises the echelon inventory position to its order-up-to-level. Lateral transshipments between the retailers are allowed. In fact, we assume that periodically the cumulative echelon stock of the retailers is rebalanced by these transshipments. To keep the analysis of this model tractable we assume that all the retailers have the same lead time. This is reasonable, since usually the transportation times between the central depot and every retailer are approximately identical. Furthermore, like Karmarkar & Patel (1977), Hosley & Heyman (1977), Cohen, Kleindorfer & Lee (1986), Tzagaras (1989), Tzagaras & Cohen (1992) and Diks & De Kok (1996a) we assume that these transshipments are instantaneous, i.e., all transshipment lead times have a zero duration. In practice this may correspond to shipping
stock overnight. In this way the inventory system experiences zero transshipment lead times, while in reality it takes some time to ship the stock.

The model with non-zero lead times are extremely hard to deal with analytically. To our knowledge only the paper of Jönsson & Silver [1987] gives an analysis of a divergent 2-echelon system with positive transshipment lead times. Such an analysis is possible due to their assumption that the review period exceeds the sum of the depot lead time and the retailer lead time, so that at most one order is outstanding at any point in time. In this chapter, however, we do not make this assumption, so that it is possible that more than one order is outstanding at the same time. As already noted by Diks & De Kok [1996a], this complicates the analysis considerably.

Diks & De Kok [1996a] analyzed a divergent 2-echelon system with lateral transshipments. They adopt the CAS1 rationing policy both to allocate the stock at the CD and to rebalance the stock at the retailers. However, the extensive numerical experiment performed in Chapter 3 indicates that BS rationing outperforms CAS rationing, at least in divergent multi-echelon systems without lateral transshipments. Having good hopes that this result can be extended to systems with lateral transshipments, we study the same model as Diks & De Kok [1996a] in this chapter except that at the CD and at the retailers we adopt the BS2 rationing policy.

Finally, we briefly like to address the paper of Tagaras [1989], which analyzes a 2-echelon distribution system with lateral transshipments. In our opinion the usefulness of their model is doubtful due to the restrictive assumptions (only two retailers are considered, and the CD has an infinite capacity with a zero replenishment lead time). His model is characterized by complete pooling in that if there is an economic incentive to transship one item, then the maximum amount will be sent.

This chapter is organized as follows. In Section 7.2, we describe the system under consideration. In Section 7.3, we present the allocation policy adopted at the CD and the rebalancing policy adopted at the retailers. For both policies we use the BS2 rationing policy. In Section 7.4, all the control parameters of the system are determined in case the cumulative echelon stock of the retailers is rebalanced every review period. In Section 7.5, we extend the results to divergent $N$-echelon systems. In Section 7.6, we present some numerical results of our model, and compare them with the model of Van der Heijden [1997b] (which uses BS rationing and transshipments are not allowed) and the model of Diks & De Kok [1996a] (which uses CAS rationing and transshipments are allowed). In Section 7.7, we briefly address an extension of the model such that the cumulative echelon stock of the retailers is not rebalanced every review period, but only when the cumulative echelon stock of the retailers drops below a certain level. Finally, we give a few concluding remarks in Section 7.8.

7.2 Model description

Consider a divergent 2-echelon system consisting of a central depot (CD) supplying a number of retailers (see Figure 7.1). The CD can place orders at an external supplier, which has sufficient capacity to deliver any order within a fixed lead time.

The inventory system is controlled by a periodic review policy. That is, every period the CD (denoted by index 1) issues a replenishment order that raises the echelon inventory position to its order-up-to-level $S_1$. This order arrives after a fixed lead time $L_1$, where $L_1$ is a multiple integer of the review period $R$. Without loss of generality we assume that $R := 1$. Therefore, in the remainder of this chapter we refer to a review period as a period. Upon arrival it is decided how to allocate the order to the CD and the retailers. When allocating stock there are two possibilities:

(i) The physical stock at the CD is sufficient to raise the echelon inventory position of each retailer
7.3 System dynamics

Figure 7.1. Schematic representation of a divergent 2-echelon system with lateral transshipments.

Let $j$ to its order-up-to-level $S_j$. Then the required amounts are sent to the successors and excess stock is kept at the CD to be allocated at the next occasion.

(ii). The physical stock is not sufficient to raise the echelon inventory position of each retailer $j$ to its order-up-to-level $S_j$. Then material rationing is required to allocate the available stock to the retailers appropriately.

It takes a fixed lead time of $L_j$ periods to ship products from the CD to retailer $j$. We assume that all retailers have the same lead time $l$ (where $l$ is an integer). Hence, $L_j = l$ for every $j \in V_i$.

Immediately after the arrival of shipments at the retailers, it is decided whether the cumulative echelon stock of all retailers should be rebalanced or not. In case of rebalancing, some products are shipped from retailers with excess inventory to those which are low on inventory. In this chapter the rebalancing policy corresponds to the BS2 rationing policy as introduced in Section 3.2.2.

The order in which the rationing policy at the CD and the rebalancing policy at the retailers are applied influences the material flow in the system. We use the following order:

(i). Arrival of shipments at CD and retailers.

(ii). Decide on whether to rebalance the cumulative echelon stock of all retailers or not. In case of rebalancing the BS2 rationing policy is used.

(iii). Retailers place their orders at the CD.

(iv). Decide on how to allocate the echelon stock of the CD by using the BS2 rationing policy.

(v). CD places order at the external supplier.

The objective of this chapter is to determine all the control parameters in the system such that every retailer attains a predetermined target fill rate. Besides attaining the target fill rates we also like to minimize the holding costs and the lateral transshipments costs.

7.3 System dynamics

In this section, we explain how the material flows through the system in case all the control parameters are known. This enables us to determine the behavior of the stock levels at both the CD and the retailers. From that we derive: (1) the fill rates attained at the retailers, and (2) the expected amount transshipped per period. Next, in Section 7.4 we determine how to set the control parameters such that every retailer attains its target fill rate at minimal expected costs.
For sake of clarity, we derive the system dynamics in this section under the assumption that every period the echelon stock of the retailers is rebalanced. In Section 7.7, we suggest how the results could be extended to the case in which not every period rebalancing takes place.

For the mathematical analysis we use some additional notation as listed below:

**Stockpoint status**

For all order arrival times \( t \geq 0 \) we define:

- \( J^1_i \): Echelon stock of a stockpoint \( i \in M \cup E \) at time \( t \) just before order arrival.
- \( \bar{J}^1_i \): Echelon stock of a stockpoint \( i \in M \cup E \) at time \( t \) just after order arrival, but before rebalancing.
- \( \check{J}^1_i \): Echelon stock of a stockpoint \( i \in M \cup E \) at time \( t \) just after rebalancing.
- \( r^1_i \): Amount of products arriving at retailer \( i \in E \) at time \( t \) due to the rebalancing.

**Control parameters**

- \( S_i \): Order-up-to-level of a stockpoint \( i \in M \cup E \) with respect to the echelon inventory position.
- \( S'_i \): Order-up-to-level of a retailer \( i \in E \) with respect to the echelon stock.
- \( \Delta_i \): Control parameter of a retailer \( i \in E \) defined by \( \Delta_i := S_i - S'_i \).

**Performance measures**

- \( T \): Expected amount of stock transshipped between retailers per period.
- \( T_i \): Expected amount of stock transshipped by a retailer \( i \in E \) per period.

### 7.3.1 Rationing policy at the CD

Consider the system at the beginning of an arbitrary period, say \( t - L_1 \). At this time the CD places an order at the external supplier to raise its echelon inventory position to \( S_i \). Since the lead time equals \( L_1 \), this order arrives at the beginning of period \( t \). So the echelon stock of the CD just after the arrival of this order equals

\[
J^1_i = S_i - D^1_{t-L_1}.
\]

If \( J^1_i \geq \sum_{s \in V_i} S_s \), then each retailer \( j \) is able to raise its echelon inventory position to order-up-to-level \( S_j \). Thus,

\[
D^1_{t-L_1} \leq \Delta_i \implies J^1_j = S_j, \quad j \in V_i. \tag{7.1}
\]

However, if \( J^1_i < \sum_{s \in V_i} S_s \), then the complete echelon stock of the CD is allocated to the retailers by using an appropriate rationing policy. For practical purposes we suggest to apply the linear allocation function given by (2.7). This means that

\[
D^1_{t-L_1} \geq \Delta_i \implies J^1_j = S_j - q_j(D^1_{t-L_1} - \Delta_i), \quad j \in V_i. \tag{7.2}
\]

The allocation fractions \( \{q_j\}_{j \in V_i} \) both satisfy \( \sum_{j \in V_i} q_j = 1 \) and \( q_j > 0 \) for each \( j \in V_i \). From (7.1) and (7.2) it follows that

\[
J^1_j = S_j - q_j(D^1_{t-L_1} - \Delta_i)^+, \quad j \in V_i. \tag{7.3}
\]

The amount of products which are allocated to retailer \( j \) at time \( t \) (at the CD) equals

\[
q^1_j = J^1_j - r^1_j, \quad j \in V_i.
\]

Recall that \( q^1_j < 0 \) may occur for some \( j \) and \( t \), which means that a negative amount of the physical stock at the CD is allocated to retailer \( j \) at time \( t \). In practice this is usually impossible. Therefore, in
7.3. System dynamics

This chapter we also assume the balance assumption. Note that in this chapter the balance assumption is even less restrictive than before since the echelon stock at the retailers is balanced every period.

In Section 3.2.2, the allocation fractions \( q_j \) for each retailer \( j \) are determined such that an approximate expression for the imbalance is minimized. This yields

\[
q_j = \frac{\sigma_j^2}{2\sigma_j^2} + \frac{1}{2|V_j|}, \quad j \in V_1.
\] (7.4)

Despite the fact that in the development of these allocation fractions we considered a divergent 2-echelon system without lateral transshipments we propose to use these allocation fractions. There are several reasons for this. First, it simplifies the analysis considerably. Second, the allocation fractions are easy to compute and therefore very suited for practical purposes. Third, we expect that these allocation fractions does not play a very important role due to the rebalancing every period.

7.3.2 Rebalancing policy at the retailers

After the allocation at the CD at time \( t \), it takes \( l \) periods to ship the products from the CD to the retailers. During these periods: (1) retailer \( j \) faces a customer demand of \( D_{j,t+1} \), and (2) the cumulative echelon stock of each retailer is rebalanced at time \( t+1 \), \( t+2 \), \ldots, \( t+l-1 \). Therefore the echelon stock at time \( t+l \) equals

\[
J_{j,t+l} = J_{j,t} - D_{j,t+1} + \sum_{i=t+1}^{t+l-1} t_i^j, \quad j \in V_1.
\] (7.5)

where \( t_i^j \) either denotes the amount of products which arrives at retailer \( j \) at time \( s \) due to rebalancing (if \( t_i^j \geq 0 \)), or denotes the amount of products which are shipped from retailer \( j \) to the other retailers at time \( s \) due to rebalancing (if \( t_i^j < 0 \)). By definition,

\[
t_i^j := J_i^j - J_i^j, \quad i \in E.
\] (7.6)

Since rebalancing only reallocates the available stock we know that \( \sum_{i \in V_1} t_i^j = 0 \) for every \( t \). From (7.3), (7.5) and the aforementioned property it can be shown that the cumulative echelon stock of all retailers at time \( t+l \) equals

\[
\sum_{s \in V_1} J_{s,t+l} = \sum_{s \in V_1} S_s - (D_{s-1,t+l} - \Delta s)^+ - D_{s,t+l}.
\] (7.7)

We suggest to use the BS2 rationing policy to reallocate these products to the retailers. For that purpose we introduce the control parameter \( S_j^t \) for each retailer \( j \). This \( S_j^t \) can be interpreted as the order-up-to-level of retailer \( j \) with respect to its echelon stock, in contrast with \( S_j \) which represents the order-up-to-level of retailer \( j \) with respect to its echelon inventory position. An important distinction between the allocation policy used at the CD and the rebalancing policy used at the retailers is the following. The allocation policy at the CD allows for retaining stock at the CD (therefore we distinguished between \( D_{s-1,t+l} \leq \Delta s \) and \( D_{s-1,t+l} > \Delta s \)). However, the rebalancing policy at the retailers cannot retain any stock since all the available products need to be allocated. In order to accomplish this we assume

\[
\sum_{s \in V_1} S_s = \sum_{j \in V_1} S_j^t.
\] (7.8)

Assumption (7.8) guarantees that \( \sum_{s \in V_1} J_{s,t} \leq \sum_{j \in V_1} S_j^t \) for any time \( t \), which means that rationing is always required (we do not have to distinguish between two cases anymore). So one might think of \( S_j^t \) as a control parameter in order to ration the stock appropriately, instead of the number of products retailer
j really requires to have after rebalancing. The amount of products short in order to raise the echelon stock of each retailer \( j \) to its order-up-to-level \( S'_j \) equals

\[
S'_j = \left( \sum_{n \in V_1} S_n - (D'_{t+1,t+1} - \Delta_t)^+ - D'_{t+1,t} \right) = (D'_{t-1,t} - \Delta_t)^+ + D'_{t+1,t}.
\]

The BS2 rationing policy suggests to divide this shortage over the retailers by the allocation fractions \( \{q_j\}_{j \in V_1} \), given by (7.4). Hence,

\[
J'_{i,j} = S'_j - q_j \left( (D'_{t-1,t} - \Delta_t)^+ + D'_{t+1,t} \right), \quad j \in V_1.
\] (7.9)

Using (7.9) we are able to derive the fill rate equation for each retailer \( j \)

\[
\beta_j = 1 - \frac{E[(D'_{t+1,t+1} - J'_{i,j})^+] - E[(J'_{i,j})^+]}{\mu_j}, \quad j \in V_1.
\] (7.10)

### 7.4 Determination of the control parameters

In this section, we determine the control parameters. This is done by decomposing the problem of the determination of these parameters into subproblems. First, we determine \( \Delta_j = S_j - S'_j \) for every retailer \( j \) by minimizing the expected total stock transshipped per period (Section 7.4.1). Second, assuming \( \Delta_1 \) is given, we determine \( \{S'_j\}_{j \in V_1} \) such that the customer fill rate constraints are satisfied (Section 7.4.2). Finally, we determine \( \Delta_1 \) such that the expected total holding costs per period are minimized (Section 7.4.3).

#### 7.4.1 Determination of the control parameters of the rationing policy

From (7.9) and (7.10) it follows that the fill rate \( \beta_j \) is independent of \( S_j \). These degrees of freedom are used to determine \( \Delta_j = S_j - S'_j \) such that the expected total stock transshipped per period (denoted by \( T \)) is minimized. Note that this corresponds to minimizing the expected transshipment costs per period in case the costs of shipping stock from one retailer to another are equal. The problem we have to solve can be formulated as a non-linear optimization problem with \( |V_1| \) variables

\[
\min_{\{\Delta_j\}_{j \in V_1}} T(\{\Delta_j\}_{j \in V_1}) \quad \text{s.t.} \quad \sum_{n \in V_1} \Delta_n = 0.
\] (7.11)

The objective function \( T \) can be computed as follows

\[
T(\{\Delta_j\}_{j \in V_1}) = \sum_{j \in V_1} T_j(\Delta_j) \quad \text{with} \quad T_j(\Delta_j) := E[(r'_{i,j})^+].
\]

In order to compute the expected amount of products arriving at retailer \( j \) due to the rebalancing at time \( t + 1 \), denoted by \( T_j(\Delta_j) \), we need to have a tractable expression for \( r'_{i,j} \). Substitution of (7.3) in (7.5), and substituting the result and (7.9) in (7.6) yields

\[
r'_{i,j} = D'_{t+1,t} - q_j D'_{t+1,t} - \Delta_j - \sum_{s \in V_1} r'_s, \quad j \in V_1.
\] (7.12)

Define \( \bar{r}'_j := D'_{t-1,t} - q_j D'_{t-1,t} - \bar{r}'_j \), then from (7.12) it follows that

\[
\sum_{s \in V_1} \bar{r}'_s = \Delta_j, \quad j \in V_1.
\] (7.13)
7.4. Determination of the control parameters

If \( \tau^l_t \) is known for \( l - 1 \) subsequent time periods, say \( s = t + 1, t + 2, \ldots, t + l - 1 \), then \( \tau^l_{s+1} \) immediately results from (7.13). Specifically, \( \tau^l_{s+1} = \tau^l_{s+1} \) for any integer value \( k \). Assuming \( \{\tau^l_t\} \) is a stationary process, i.e., \( E[\tau^l_t] = E[\tau^l_{s+1}] \) for any \( t \), then from (7.13) and the definition of \( \tau^l_t \) it follows that \( \tau^l_t = \Delta_t / l \). Hence,

\[
\tau^l_t = D^l_{l-1,t} - q_j D^l_{l-1,t} - \frac{\Delta_t}{l}, \quad j \in V_t. \tag{7.14}
\]

Before continuing our analysis, let us elaborate on how to choose the initial state of the system such that indeed \( \{\tau^l_t\} \) is a stationary process. Suppose the initial state of the system is characterized by \( \{\pi^l_0\}_{s=0} \), where \( \pi^l_0 \) denotes the echelon stock of retailer \( j \in V_t \) at time \( 0 \), and \( \pi^l_j (1 \leq t \leq l) \) denotes the amount of products arriving at retailer \( j \) in exactly \( t \) periods. The pipeline and on hand stock of the CD are irrelevant. At the end of the first period holds

\[
\begin{align*}
J^l_t &= \pi^l_0 + \pi^l_t - D^l_{0,t}, & j \in V_t, \\
\tilde{J}^l_t &= S^l_j - q_j \left( \sum_{s \in \mathcal{V}_t} S^l_s - \sum_{s \in \mathcal{V}_t} \pi^l_s + \pi^l_t + D^l_{0,t} \right), & j \in V_t.
\end{align*}
\]

Since \( \tau^l_t = \tilde{J}^l_t - J^l_t \) we obtain

\[
\tau^l_t = S^l_j - q_j \left( \sum_{s \in \mathcal{V}_t} S^l_s - \sum_{s \in \mathcal{V}_t} \pi^l_s + \pi^l_t + D^l_{0,t} \right) - \left( \pi^l_0 + \pi^l_t + D^l_{0,t} \right).
\]

By definition \( \tilde{J}^l_t = D^l_{0,1} - q_j D^l_{0,1} - \tau^l_t \). Hence,

\[
\tau^l_t = (\pi^l_0 + \pi^l_t) - q_j \sum_{s \in \mathcal{V}_t} \pi^l_s - S^l_j + q_j \sum_{s \in \mathcal{V}_t} S^l_s, \quad j \in V_t. \tag{7.15a}
\]

Let us now derive a similar expression for \( \tau^l_t \) when \( 1 < t \leq l \). For \( 1 < t \leq l \) we know that \( \tau^l_t = \tilde{J}^l_t - J^l_t = J^l_t - (\tilde{J}^l_{t-1} - D^l_{t-1,t} + \pi^l_t) \). Substitution of the definition of \( \tilde{J}^l_t := S^l_j - q_j \left( \sum_{s \in \mathcal{V}_t} S^l_s - \sum_{s \in \mathcal{V}_t} \pi^l_s \right) \) in the aforementioned expression yields

\[
\tau^l_t = q_j \left( \sum_{s \in \mathcal{V}_t} \pi^l_s + \sum_{s \in \mathcal{V}_t} S^l_s + D^l_{0,t} - \pi^l_t \right), \quad j \in V_t, \quad 1 < t \leq l.
\]

Note that \( \sum_{s \in \mathcal{V}_t} \pi^l_s - \sum_{s \in \mathcal{V}_t} S^l_s \) equals the amount of products arriving at all retailers at time \( t \) minus the demand at all retailers during \( [l-1, t) \). Hence,

\[
\tau^l_t = q_j \left( \sum_{s \in \mathcal{V}_t} \pi^l_s + D^l_{l-1,t} - \pi^l_t \right), \quad j \in V_t, \quad 1 < t \leq l.
\]

Substitution of \( \tau^l_t := D^l_{l-1,t} - q_j D^l_{l-1,t} - \tau^l_t \) in the definition of \( \tau^l_t \) yields

\[
\tau^l_t = \pi^l_t - q_j \sum_{s \in \mathcal{V}_t} \pi^l_s, \quad j \in V_t, \quad 1 < t \leq l. \tag{7.15b}
\]

From (7.15) it follows that in order to have \( \tau^l_t = \Delta_t / l \) for \( j \in V_t \) and \( 1 \leq t \leq l \), we have to determine \( \{\pi^l_t\} \) such that

\[
\begin{align*}
\pi^l_0 + \pi^l_t &= \mu_j + q_j \sum_{s \in \mathcal{V}_t} \left( \pi^l_s + \pi^l_t - \mu_j \right) + S^l_j - q_j \sum_{s \in \mathcal{V}_t} S^l_s, \\
\pi^l_t &= \mu_j + q_j \sum_{s \in \mathcal{V}_t} (\pi^l_s - \mu_j), \quad 1 < t \leq l.
\end{align*}
\]
An initial state satisfying the above equalities is
\[ \pi_t' := S'_j - q_j \sum_{s \in V_i} S'_s, \quad \pi_t' := \mu_j, \quad 1 \leq t \leq l. \]

Under the balance assumption this initial state yields stationarity. Unfortunately, due to imbalance we cannot guarantee the stationarity of \( \pi_t' \).

To compute the expected amount transshipped to retailer \( j \) per period (i.e., \( T_j(\Delta_j) \)), we use a two-moment approximation for \( \pi_t' \). Because \( \pi_t' \) has a probability distribution function on the entire interval \((-\infty, \infty)\) we approximate this distribution by a normal distribution. Since \( T_j(\Delta_j) = E[(\pi_t')^+], \) we obtain
\[ T_j(\Delta_j) \approx \sigma_{\pi_t'} \Phi \left( \frac{\mu_j}{\sigma_{\pi_t'}} \right) + \mu_j \Phi \left( \frac{\mu_j}{\sigma_{\pi_t'}} \right), \quad j \in V_1, \tag{7.16} \]

with
\[ \mu_{\pi_t'} := \mu_j - q_j \mu_1 - \frac{\Delta_j}{l} \quad \text{and} \quad \sigma_{\pi_t'} := \sqrt{(1 - 2q_j)\sigma_j^2 + q_j^2\sigma_1^2}. \tag{7.17} \]

Applying the Lagrange-multiplier technique on (7.11) yields
\[ \frac{dT_j(\Delta_j)}{d\Delta_j} = \lambda, \quad j \in V_1. \tag{7.18} \]

By differentiating (7.16) to \( \Delta_j \), and substituting the result in (7.18) yields
\[ \Phi \left( \frac{\mu_j}{\sigma_{\pi_t'}} \right) = \lambda, \quad j \in V_1. \tag{7.19} \]

Substituting (7.17) in (7.19), and rewriting the result yields
\[ \Delta_j = l(\mu_j - q_j \mu_1 - \Phi^{-1}(\lambda)\sigma_{\pi_t'}), \quad j \in V_1. \]

From \( \sum_{j \in V_1} \Delta_j = 0 \) it follows that \( \Phi^{-1}(\lambda) = 0 \). Hence,
\[ \Delta_j = S_j - S'_j = l(\mu_j - q_j \mu_1), \quad j \in V_1. \tag{7.20} \]

Substitution of (7.20) in (7.17) yields \( \mu_{\pi_t'} = 0 \). This means that each retailer \( j \) is in 'balance', since the mean amount of products which is shipped out of retailer \( j \) per period equals the mean amount of products which is shipped to retailer \( j \) per period. This is intuitively clear, since if \( \mu_{\pi_t'} \) would be less than 0, then the CD is systematically allocating to much products to retailer \( j \). Therefore upon arrival some of these products are immediately shipped to other retailers.

Substitution of \( \mu_{\pi_t'} = 0 \) in equation (7.16) yields a simple approximation for the expected amount of products which are transshipped to retailer \( j \) per period,
\[ T_j = \sqrt{\frac{1 - 2q_j \mu_j^2}{2\pi} + \frac{q_j^2}{2\pi} \sigma_j^2}, \quad j \in V_1. \tag{7.21} \]

### 7.4.2 Determination of the control parameters of the rebalancing policy

In this section, we elaborate on how the order-up-to-level \( S_1 \) and the control parameters \( \{S'_j\}_{j \in V} \) are determined given \( \Delta_i \). The control parameter \( S'_j \) is implicitly defined by target the fill rate \( \beta_j \). This can be seen by substituting (7.9) in (7.10), which yields
\[ \beta_j = 1 - \frac{E[(\Psi_j - S'_j)^+)] - E[(q_j \Psi_2 - S'_j)^+]}{\mu_j}, \quad j \in V_1. \tag{7.22} \]
7.5. Extension to divergent N-echelon system

with $\Psi_{ij} := q_j ((D_{ij}^s - \Delta_j)^+ + D_j^s) + D_j^i$ and $\Psi_3 := (D_{i}^s - \Delta_i)^+ + D_i^s$.

To determine $S_j$ we suggest to solve $\beta_j = \beta_j^*$ numerically. We use a similar approach as in Section 2.4.1. To apply this approach we first need to determine the mean and variance of both $\Psi_{ij}$ and $\Psi_3$.

Because the random variables $D_{ij}^s$, $D_j^i$ and $D_j^s$ of $\Psi_{ij}$ are independent, we know that the mean and variance of $\Psi_{ij}$ are

$$E[\Psi_{ij}] = q_j (E[(D_{ij}^s - \Delta_j)^+] + \mu_j) + \mu_j, \quad \text{var}[\Psi_{ij}] = q_j^2 (\text{var}[(D_{ij}^s - \Delta_j)^+] + \text{var}D_j^s)^2 + \sigma_j^2.$$

The mean and variance of $(D_j^s - \Delta_j)^+$ are determined by fitting a mixed Erlang distribution to the lead time demand $D_j^s$, and next applying formulae (2.20a) and (2.20b). Analogously, the mean and variance of $\Psi_3$ are determined. Next, we fit mixed Erlang distributions to the first two moments of the random variables $\Psi_{ij}$ and $\Psi_3$. Again, by applying the formulae (2.20a) and (2.20b) we compute $E[(\Psi_{ij} - S_j)^+]$ and $E[(\Psi_3 - S_j)^+]$. Now, $\beta_j = \beta_j^*$ is solved for $S_j$ by using bisection.

Finally, we determine the order-up-to-level at the CD by

$$S_i = \Delta_i + \sum_{j \in \mathcal{V}_i} S_j = \Delta_i + \sum_{j \in \mathcal{V}_i} S_j^s. \quad (7.23)$$

Since $S_j^s$ is known for each retailer $S_i$ immediately results from (7.23).

7.4.3 Determination of $\Delta_i$

In Section 7.4.1, we have determined $[\Delta_j]_{j \in \mathcal{V}_i}$ such that the expected total amount transshipped per period is minimized. From (7.20), it follows that every $\Delta_j$ is independent of $\Delta_i$.

Next, in Section 7.4.2, we have determined $[S_j]_{j \in \mathcal{V}_i}$ such that each retailer attains its target fill rate. From (7.22) it is obvious that these control parameters depend on $\Delta_i$. In this section, we determine how to set $\Delta_i$ such that the expected total holding costs at the end of every period (just before an order arrival) is minimized. Following Chapters 5 and 6, holding costs $h_i$ are incurred for every product at the CD or in one of the pipelines towards a retailer, and holding costs $h_i + h_j$ are incurred for a product at retailer $j$. For this cost structure, the expected total holding costs at the end of a period equals

$$HC(\Delta_i) := h_i E[S_i - D_{i,i+1}^s] + \sum_{j \in \mathcal{V}_i} h_j E\left[S_j - q_j ((D_{ij}^s - \Delta_j)^+ + D_j^i) - D_j^s\right]$$

$$+ \sum_{j \in \mathcal{V}_i} (h_i + h_j) E\left[-(S_j - q_j ((D_{ij}^s - \Delta_j)^+ + D_j^i) - D_j^s)^2\right].$$

$$= h_i \left(\Delta_i + \sum_{j \in \mathcal{V}_i} S_j^s - (L_i + 1) \mu_i\right) + \sum_{j \in \mathcal{V}_i} h_j \left(S_j^s - q_j \left(E[(D_{ij}^s - \Delta_j)^+] + \mu_j\right) - \mu_j\right)$$

$$+ \sum_{j \in \mathcal{V}_i} (h_i + h_j) E\left[-(S_j^s - q_j ((D_{ij}^s - \Delta_j)^+ + D_j^i) - D_j^s)^2\right].$$

A one-dimensional search over $\Delta_i$ yields the optimal value $\Delta_i$ which minimizes $HC(\Delta_i)$. Since in distribution systems usually no (or little) value is added to the product at retailer level ($h_i \approx 0$), empirically we find that $\Delta_i$ usually lies between 0 and 0.95$L_i\mu_i$. This means that no stock or very little stock is retained at the CD. This coincides with the results of other studies (cf. De Kok, Lagadimos & Seidel [1994]).

7.5 Extension to divergent N-echelon system

In this section, we extend the results to divergent N-echelon systems. We illustrate how to extend the results by considering the example in Figure 7.2, which depicts a single-item 3-echelon distribution

\[\text{Diagram or Figure 7.2} \]
system. First, there is a production facility which supplies a national distribution center (NDC). Second, this national depot center supplies two regional depots (RDC1 and RDC2). Finally, both regional depots supply two retailers. The retailers which are supplied by the same regional depot is referred to as a pooling group (cf. Lee [1987]). We make the assumption that the lead time of each retailer within a pooling group is identical. In this section, we only consider the model where lateral transshipments are allowed between the retailers of one pooling group. The model where also lateral transshipments between intermediate stockpoints (e.g., the regional depots) are allowed is a topic for further research.

The material flow in the upstream part of the network has already been analyzed in Chapter 2. At the beginning of a period, say \( t = L_1 \), the NDC raises its echelon inventory position to \( S_1 \). Since the lead time equals \( L_1 \), this order arrives at the beginning of period \( t \). So, the echelon stock of the NDC just after the arrival of this order equals

\[
J_1^1 = S_1 - D_{-L_1,t}^1.
\]

If \( J_1^1 \geq S_2 + S_3 \) then both regional depots are able to raise their echelon inventory position to its order-up-to-level. Thus,

\[
S_1 - D_{-L_1,t}^1 \geq S_2 + S_3 \implies I_1^j = S_j, \quad j \in \{2, 3\}.
\]  \hspace{2cm} (7.24)

Otherwise the complete echelon stock is allocated to both regional depots by using the linear allocation function of (2.7). Thus,

\[
S_1 - D_{-L_1,t}^1 < S_2 + S_3 \implies I_1^j = S_j - q_j \left( \sum_{n \in N_1} S_n - (S_1 - D_{-L_1,t}^1) \right), \quad j \in \{2, 3\}.
\]  \hspace{2cm} (7.25)

From (7.24) and (7.25) it follows

\[
I_1^j = S_j - q_j \left( D_{-L_1,t}^1 - \Delta_1 \right)^+, \quad j \in \{2, 3\}.
\]  \hspace{2cm} (7.26)

Next, we consider one of the regional depots, say RDC1. At the beginning of time \( t \) this regional depot places an order at the NDC to raise its echelon inventory position to \( S_2 \). However, since the regional depot is supplied by a stockpoint with a finite capacity, it is possible that this order can only be satisfied partially. This (partial) order arrives at the beginning of period \( t + L_2 \). Hence, at the beginning of period \( t + L_2 \) the echelon stock of the RDC1 equals

\[
J_{t+L_2}^1 = I_1^1 - D_{t+L_2}^2.
\]
7.5. Extension to divergent N-echelon system

If \( J_{i, t+1} \geq S_k + S_l \) then both retailers in pooling group 1 are able to raise the echelon inventory position to its order-up-to-level. Thus,

\[
I_{i+1, t+2}^i - D_{i+1, t+2}^i \geq S_k + S_l \implies I_{i+1, t+2}^k = S_j, \quad j \in \{4, 5\}. \tag{7.27}
\]

However, if \( J_{i, t+1} < S_k + S_l \) then the echelon stock of RDC1 is allocated to the retailers in pooling group 1 as follows

\[
I_{i+1, t+2}^i - D_{i+1, t+2}^i \leq S_k + S_l \implies I_{i+1, t+2}^m = S_j - q_j \left( \sum_{n \in V_2} S_n - (I_{i, t+1}^n - D_{i, t+1}^n) \right)^+, \quad j \in \{4, 5\}. \tag{7.28}
\]

From (7.27) and (7.28) it follows that

\[
I_{i+1, t+2}^i = S_j - q_j \left( \sum_{n \in V_2} S_n - (I_{i, t+1}^n - D_{i, t+1}^n) \right)^+, \quad j \in \{4, 5\}. \tag{7.29}
\]

Substitution of (7.26) in (7.29) yields

\[
I_{i+1, t+2}^i = S_j - q_j \left( D_{i+1, t+2}^i - \Delta_2 + q_2 \left( D_{i+1, t+2}^i - \Delta_2 \right)^+ \right)^+, \quad j \in \{4, 5\}. \tag{7.30}
\]

From equation (7.30), it follows that the cumulative echelon stock of all retailers in pooling group 1 at time \( t + L_2 + L_3 \) (recall \( L_4 = L_3 \)) equals

\[
\sum_{j \in V_2} I_{i+1, t+1}^j = \sum_{j \in V_2} I_{i+1, t+1}^j - D_{i+1, t+1}^j - D_{i+1, t+2}^j, \quad j \in V_2.
\]

Substitution of (7.30) in the expression above yields that

\[
\sum_{j \in V_2} I_{i+1, t+1}^j = \sum_{n \in V_2} S_n - \left( D_{i+1, t+2}^i - \Delta_2 + q_2 \left( D_{i+1, t+2}^i - \Delta_2 \right)^+ \right)^+ - D_{i+1, t+2}^j, \quad j \in V_2.
\]

Following the reasoning in Section 7.3.2 this stock is reallocated to the retailers by a BS2 rationing policy. Since \( \sum_{n \in V_2} S_n = \sum_{n \in V_2} S_n \), we obtain

\[
I_{i+1, t+1}^j = S_k - q_k \left( (D_{i+1, t+1}^j - \Delta_1 + q_1 (D_{i+1, t+1}^j - \Delta_1)^+ + D_{i+1, t+1}^j) \right)^+, \quad j \in V_2.
\]

Analogue to the analysis in Section 7.4.1 it can be shown that

\[
\Delta_1 = S_k - S_k = L_4 (\mu_k - \mu_k), \quad k \in E \quad \text{and} \quad \text{pre}(k) = j.
\]

Furthermore, the order-up-to-levels \( S_k^j \) are computed similarly to Section 7.4.2 by solving \( \beta_k = \beta_k^j \) for each retailer \( k \), where

\[
\beta_k = \frac{E[(\Psi_{1,k} - S_k)^+]}{\mu_k}, \quad j \in V_i \quad \text{and} \quad k \in V_j.
\]

with

\[
\Psi_{1,k} := q_k \left( (I_{i, t+1} - \Delta_1 + q_1 (I_{i, t+1} - \Delta_1)^+ + I_{i, t+1}^k) \right)^+ + I_{i, t+1}^k, \quad j \in V_i \quad \text{and} \quad k \in V_j,
\]

\[
\Psi_{2,j} := (D_{i, t+1} - \Delta_1 + q_1 (D_{i, t+1} - \Delta_1)^+ + D_{i, t+1})^+, \quad j \in V_i.
\]
7.6 Numerical results

In this section, we compare the results of the model presented in the previous sections (referred to as the TBS-model) with the very related models of Van der Heijden [1997b] and Dik & De Kok [1996a]. Van der Heijden analyzes the 2-echelon system without transshipments as presented in Section 3.2.2, where the CD adopts the BS rationing policy. In this section, we refer to this model as the BS-model. Dik & De Kok analyzes a 2-echelon stock with transshipments. Both the allocation policy at the CD and the rebalancing policy at the retailers are CAS rationing policies. In this section, we refer to this model as the TCAS-model. Let us briefly elaborate on their algorithm to determine all the control parameters. First, they determine simultaneously the allocation fractions of the CAS rationing policy and the order-up-to-level at the CD, such that the customer fill rate constraints are satisfied. Second, they determine the allocation fractions of the CAS rationing policy at the CD such that the expected total amount transshipped per period is minimized. Recall that the order-up-to-levels for a CAS rationing policy can easily be determined from (3.1).

Consider the example of Dik & De Kok [1996a] where one CD supplies 5 non-identical retailers (see Table 7.1). The depot lead time and the retailer lead time equals 4 and 1, respectively. Every retailer needs to attain the same target fill rate $\beta^*$.

<table>
<thead>
<tr>
<th>Retailer $j$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_j$</td>
<td>5</td>
<td>10</td>
<td>10</td>
<td>15</td>
<td>20</td>
</tr>
<tr>
<td>$\sigma_j^2$</td>
<td>25</td>
<td>60</td>
<td>60</td>
<td>225</td>
<td>560</td>
</tr>
</tbody>
</table>

Table 7.1. The demand characteristics of the 5 retailers in the system.

Table 7.2 depicts the safety stock defined by $S_j = (L_j + l + 1)\mu_j$, which is required to satisfy the customer fill rates constraints. We computed this safety stock for three different values of $\beta^*$ = 0.7, 0.85 and 0.99. Also we varied the mean amount of stock retained at the CD by considering $a_1 = 0$, 0.9 and 1.2, where $a_1 := \Delta_1/(L_1\mu_1)$. From Table 7.2 we conclude that by allowing transshipments the safety stock needed to operate the system decreases considerably (between 8% and 22%). When transshipments are allowed the TCAS-model appears to be slightly better than the TBS-model for $\beta^* = 0.7$, whereas the TBS-model appears to be better than the TCAS-model for $\beta^* = 0.99$. It is hard to draw a well-founded conclusion about which model is the best model. This can be explained by the fill rates attained at the retailers for each model (see Table 7.3). For instance, the TBS-model attains fill rates which exceeds the target fill rate for $\beta^* = 0.7$ (and therefore requires more safety stock), whereas the TBS-model attains lower fill rates than the TCAS-model for $\beta^* = 0.99$ (and therefore requires less safety stock).

<table>
<thead>
<tr>
<th>$\beta^*$</th>
<th>$a_1 = 0$</th>
<th>$a_1 = 0.9$</th>
<th>$a_1 = 1.2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_j$</td>
<td>$\sigma_j^2$</td>
<td>$\mu_j$</td>
<td>$\sigma_j^2$</td>
</tr>
<tr>
<td>0.70</td>
<td>45</td>
<td>35</td>
<td>38</td>
</tr>
<tr>
<td>0.85</td>
<td>103</td>
<td>89</td>
<td>87</td>
</tr>
<tr>
<td>0.99</td>
<td>275</td>
<td>259</td>
<td>230</td>
</tr>
</tbody>
</table>

Table 7.2. Safety stock in a divergent 2-echelon system with 5 non-identical retailers ($L_j = 4$ and $l = 1$).

Table 7.3 depicts the customer fill rates obtained by simulation. From this table we conclude that the performance of the heuristic to numerically solve the fill rate equation (see Section 7.4.2) is excellent. All the attained fill rates only slightly differs from the target fill rate (at most 0.012), except for
the fill rate attained at retailer 1 in the BS-model. This can be explained by the imbalance experienced in the BS-model.

<table>
<thead>
<tr>
<th>$a_1$</th>
<th>BS</th>
<th>TBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.748</td>
<td>0.702</td>
</tr>
<tr>
<td>2</td>
<td>0.703</td>
<td>0.701</td>
</tr>
<tr>
<td>3</td>
<td>0.703</td>
<td>0.701</td>
</tr>
<tr>
<td>4</td>
<td>0.706</td>
<td>0.698</td>
</tr>
<tr>
<td>5</td>
<td>0.708</td>
<td>0.698</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$a_1$</th>
<th>BS</th>
<th>TBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.70</td>
<td>0.724</td>
<td>0.697</td>
</tr>
<tr>
<td></td>
<td>0.694</td>
<td>0.697</td>
</tr>
<tr>
<td></td>
<td>0.694</td>
<td>0.697</td>
</tr>
<tr>
<td></td>
<td>0.696</td>
<td>0.696</td>
</tr>
<tr>
<td></td>
<td>0.695</td>
<td>0.699</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$a_1$</th>
<th>BS</th>
<th>TBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.870</td>
<td>0.842</td>
</tr>
<tr>
<td>2</td>
<td>0.849</td>
<td>0.844</td>
</tr>
<tr>
<td>3</td>
<td>0.848</td>
<td>0.844</td>
</tr>
<tr>
<td>4</td>
<td>0.845</td>
<td>0.845</td>
</tr>
<tr>
<td>5</td>
<td>0.840</td>
<td>0.845</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$a_1$</th>
<th>BS</th>
<th>TBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.70</td>
<td>0.994</td>
<td>0.989</td>
</tr>
<tr>
<td></td>
<td>0.991</td>
<td>0.990</td>
</tr>
<tr>
<td></td>
<td>0.992</td>
<td>0.988</td>
</tr>
<tr>
<td></td>
<td>0.990</td>
<td>0.989</td>
</tr>
<tr>
<td></td>
<td>0.990</td>
<td>0.989</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$a_1$</th>
<th>BS</th>
<th>TBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.994</td>
<td>0.989</td>
</tr>
<tr>
<td>2</td>
<td>0.991</td>
<td>0.990</td>
</tr>
<tr>
<td>3</td>
<td>0.992</td>
<td>0.988</td>
</tr>
<tr>
<td>4</td>
<td>0.990</td>
<td>0.989</td>
</tr>
<tr>
<td>5</td>
<td>0.990</td>
<td>0.989</td>
</tr>
</tbody>
</table>

Table 7.3. Customer fill rates attained in a divergent 2-echelon system with 5 non-identical retailers ($L_1 = 4$ and $l = 1$).

Table 7.4 depicts $\Omega^{y_1}$, i.e., the percentage that the allocation policy of the CD allocates a negative quantity to at least one of the retailers. From this table we conclude that, indeed, in the BS-model the imbalance is large (especially for $a_1 = 0$). A large decrease in imbalance is accomplished by allowing transshipments (especially for the TBS-model).

<table>
<thead>
<tr>
<th>$a_1$</th>
<th>BS</th>
<th>TBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.70</td>
<td>0.43</td>
<td>0.16</td>
</tr>
<tr>
<td>0.85</td>
<td>0.43</td>
<td>0.28</td>
</tr>
<tr>
<td>0.99</td>
<td>0.43</td>
<td>0.02</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$a_1$</th>
<th>BS</th>
<th>TBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>0.10</td>
<td>0.08</td>
</tr>
<tr>
<td>0.10</td>
<td>0.10</td>
<td>0.02</td>
</tr>
<tr>
<td>0.10</td>
<td>0.10</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Table 7.4. Imbalance at the CD in a divergent 2-echelon system with 5 non-identical retailers ($L_1 = 4$ and $l = 1$).

Table 7.5 depicts the mean amount transshipped per period. For $\beta^* = 0.7$ the expected total amount transshipped per period in the TBS-model is less than in the TCAS-model (for $\beta^* = 0.99$ the TBS-
model is slightly better). However, in our opinion the reduction of safety stock is more important than the reduction in transportation costs. The reason for this is that when the model rebalances every review period, then it is not so important anymore how much to transship. The frequency of transshipping plays a more important role in the total costs involved with transshipping than the amount transshipped, since usually these amounts are small. Also note that the expected total amount transshipped per period in the TBS-model does not depend on $\beta^*$ or $a_1$ (this coincides with approximation (7.21)), whereas in the TCAS-model it is both depend on $\beta^*$ and $a_1$.

The impact of the number of retailers on the reduction of the safety stock is analyzed in Table 7.6. It turns out that in the case $\beta^* = 0.7$ and $\beta^* = 0.85$ the results of the TCAS-model and the TBS-model are similar, but for $\beta^* = 0.99$ the TBS-model appears to outperform the TCAS-model. Especially, when the number of retailers is large. Then the reduction of the safety stock can even becomes 24% (compared to the safety stock needed in the BS-model).

<table>
<thead>
<tr>
<th>$a_1 = 0$</th>
<th>BS</th>
<th>TCAS</th>
<th>TBS</th>
<th>BS</th>
<th>TCAS</th>
<th>TBS</th>
<th>BS</th>
<th>TCAS</th>
<th>TBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta^*$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.70</td>
<td>2</td>
<td>14</td>
<td>13</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>28</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>0.70</td>
<td>3</td>
<td>18</td>
<td>16</td>
<td>16</td>
<td>20</td>
<td>18</td>
<td>39</td>
<td>37</td>
<td>37</td>
</tr>
<tr>
<td>0.70</td>
<td>4</td>
<td>23</td>
<td>19</td>
<td>19</td>
<td>24</td>
<td>21</td>
<td>50</td>
<td>46</td>
<td>46</td>
</tr>
<tr>
<td>0.70</td>
<td>5</td>
<td>27</td>
<td>21</td>
<td>21</td>
<td>28</td>
<td>24</td>
<td>61</td>
<td>55</td>
<td>56</td>
</tr>
<tr>
<td>0.85</td>
<td>2</td>
<td>34</td>
<td>32</td>
<td>32</td>
<td>36</td>
<td>33</td>
<td>33</td>
<td>46</td>
<td>43</td>
</tr>
<tr>
<td>0.85</td>
<td>3</td>
<td>46</td>
<td>41</td>
<td>40</td>
<td>48</td>
<td>42</td>
<td>42</td>
<td>64</td>
<td>57</td>
</tr>
<tr>
<td>0.85</td>
<td>4</td>
<td>57</td>
<td>49</td>
<td>47</td>
<td>60</td>
<td>50</td>
<td>50</td>
<td>83</td>
<td>72</td>
</tr>
<tr>
<td>0.85</td>
<td>5</td>
<td>69</td>
<td>57</td>
<td>54</td>
<td>72</td>
<td>59</td>
<td>58</td>
<td>102</td>
<td>85</td>
</tr>
<tr>
<td>0.99</td>
<td>2</td>
<td>95</td>
<td>89</td>
<td>87</td>
<td>103</td>
<td>92</td>
<td>92</td>
<td>105</td>
<td>92</td>
</tr>
<tr>
<td>0.99</td>
<td>3</td>
<td>127</td>
<td>116</td>
<td>109</td>
<td>141</td>
<td>119</td>
<td>119</td>
<td>150</td>
<td>125</td>
</tr>
<tr>
<td>0.99</td>
<td>4</td>
<td>158</td>
<td>143</td>
<td>130</td>
<td>178</td>
<td>146</td>
<td>143</td>
<td>194</td>
<td>158</td>
</tr>
<tr>
<td>0.99</td>
<td>5</td>
<td>189</td>
<td>172</td>
<td>150</td>
<td>216</td>
<td>174</td>
<td>168</td>
<td>239</td>
<td>195</td>
</tr>
</tbody>
</table>

Table 7.6. Safety stock in a divergent 2-echelon system with identical retailers with $\mu_1 = 10$ and $\sigma_1^2 = 80$ ($L_1 = 4$ and $t = 1$).

7.7 Extension of the model

A disadvantage of the model presented so far is that every period the echelon stock of the retailers is rebalanced. In principle the retailers should only rebalance, when it is really necessary. E.g., when some retailers face large demands resulting in low inventories, whereas others have excess inventory. In this section we address a possible way of extending the model as described in Section 7.3 and 7.4 such that not every review period rebalancing takes place, but only when it is necessary. How to determine the control parameters for this extended model is a topic for further research.

Suppose that the echelon stock of the retailers is only rebalanced when the cumulative echelon stock drops below a certain level $Q$. Let the binary random variable $X_{i,t}$ we indicate whether the stock is rebalanced ($X_{i,t} = 1$) or not ($X_{i,t} = 0$). Then from (7.7) it follows that

$$\delta_{i,t} := Pr(X_{i,t} = 1) = Pr\left(\sum_{n \neq i} S_n - (D_{i,n-1} - A_i)^+ + D_{i,Ai}^1 < Q\right).$$

(7.31)

Note that this $\delta_{i,t}$ is independent of $t$ since we assumed stationair customer demand. Therefore we define $\delta$ as the probability of rebalancing the cumulative echelon stock at the retailers at the end of an
arbitrary period. Thus, of \( \delta \) as a managerial parameter. Management probably determines the value of \( \delta \) based on the trade-off between the fixed costs of rebalancing and the decrease in costs due to making better use of the available stock in the system. Given this value of \( \delta \) it is immediately clear how to choose \( Q \) such that the frequency of transshipment matches the target set by management. From (7.31) we obtain

\[
Q = \sum_{s \in S} s \cdot F_{s}^{-1}(1 - \delta) \quad \text{with} \quad Z := (D_{s} - \Delta_{s})^{+} + D_{s}^{+}.
\]

In our opinion, this an appropriate way to extend the model such that not every review period the echelon stock of the retailers is rebalanced. There are several reasons for this. First, when the cumulative echelon stock of the retailers is small it is important that this stock is appropriately distributed among the retailers. The smaller this cumulative echelon stock becomes the larger the probability becomes of having one or more retailers facing backorders while others have excess stock. Second, this way of incorporating that the retailers do not rebalance their echelon stock every period is still analytically tractable. Finally, management can simply influence the frequency of transshipping by setting only one parameter.

7.8 Conclusions and further research

In this chapter, we considered a 2-echelon distribution system consisting of a central depot supplying a number of retailers. Every review period an instantaneous rebalancing of the cumulative echelon stock of the retailers takes place, by transshipping stock from one retailer to another. We use the BS2 rationing policy as presented in Section 3.2.2 to rebalance the stock. Also the CD adopts a BS2 rationing policy to allocate incoming stock to the retailers. In this chapter, we determine all the control parameters of the inventory system such that the target fill rates are satisfied. Next, we minimize the expected total amount of stock transshipped per period.

The policy derived in this chapter (referred to as TBS) is easy to implement and very robust. The allocation fractions of the rationing policy at the CD and of the rebalancing policy at the retailers are given by (7.4). The parameters \( \Delta_{s} \) are determined by (7.20). Only for the computations of the order-up-to-levels we used a heuristic. Numerical results indicate that the performance of this heuristic is excellent. The TBS-policy can easily be extended to divergent \( N \)-echelon systems (see Section 7.5). Also we compared the results of the TBS-policy with:

(i) the BS-policy of Van der Heijden [1997b], who analyzes a divergent 2-echelon system without lateral transshipments, where the CD adopts a BS2 rationing policy.

(ii) the TCAS-policy of Diks \\& De Kok [1998b], who analyzes the same model as analyzed in this chapter, except they adopt the CAS1 rationing policy instead of a BS2 rationing policy.

By allowing lateral transshipments the TCAS-model needs far less safety stock than the BS-policy. For the examples considered in Table 7.2 the mean reduction of safety stock is approximately 13%. The TBS-policy even accomplishes a reduction of 15%, however, the deviation between the actual and target fill rates are a bit larger than for the TCAS-model (see Table 7.3). Furthermore, the amount of imbalance for the TBS-policy is negligible (less than the TCAS-policy, and considerably less than the BS-policy; see Table 7.4).

Finally, in Section 7.7 we suggested a first step to extend the model such that rebalancing does not take place every review period, but only when the echelon stock drops below a critical level. More research should be done on this model, to analyze the trade-off between the decrease of the rebalancing set-up costs and the increase in safety stock needed to satisfy the customer fill rates.
Conclusions and suggestions for future research

The aim of this monograph was to develop algorithms which enables us to optimize the material flow in divergent multi-echelon inventory systems by determining the control parameters such that some prespecified objective is attained. We addressed two different objectives. First, in Chapters 3 and 4 the system was controlled such that some target customer service levels are attained (service measure objective). Second, in Chapters 5 and 6 the system was controlled such that the expected systemwide holding and penalty costs per review period are minimized (cost measure objective). For both objectives we developed several algorithms to determine all the control parameters. In the development of these algorithms, we focused on numerical tractability and applicability. The strength of our algorithms is that they can handle a general divergent network layout and heterogeneous end-stockpoints, but moreover it allows us to optimize the system. The need to develop algorithms which can be applied to large systems mainly arose from the analyses of convergent systems (e.g., assembly systems). Roosling [1989] and Langenhoff & Zijn [1990] independently demonstrated that under some conditions (e.g., no set-up costs) an assembly system can be remodelled as a series system for a specific class of assembly systems. By relaxing some of the conditions Visschers [1996] showed that an assembly system can be remodelled as a number of divergent multi-echelon systems.

All the algorithms in this monograph were developed under the so-called balance assumption, which assumes that no negative number of items are allocated to a stockpoint in case of material rationing. Due to this assumption, it was important to validate the performance of each algorithm. This was done by extensive numerical experiments where we compared the analytical results with the results obtained by simulation.

In Chapter 3, we compared the performance of two different rationing policies and its variants. First, De Kok, Lagodimos & Seidel [1994] introduced the Consistent Appropriate Share (CAS) rationing policy. Second, Van der Heijden [1997b] introduced the Balanced Stock (BS) rationing policy. We presented several algorithms to determine all the control parameters such that the fill rate constraints at the end-stockpoints are satisfied. An extensive numerical experiment was performed to obtain insight in the performance and differences between all the variants of CAS and BS rationing. The most important result was that BS rationing performs better than CAS rationing, both on average and worst case. Within the class of CAS rationing policies, it is remarkable that the simple, original
approach of De Klok [1990] performs best. The most serious errors occur in cases of high imbalance, especially for relatively low service levels and situations with little or no intermediate stocks. Within the class of BS rationing policies, the original variant developed by Van der Heijden [1997b] is the best rationing policy. However, from a practical point of view we prefer the BS variant of Van Donselaar [1996], because it gives an easy and explicit expression for the allocation fractions, and the performance is excellent.

In Chapter 4, we extended the algorithm based on BS rationing model by incorporating random lead times. Like in Chapter 3, we performed an extensive numerical experiment to validate our algorithm. In this experiment, we make use of a novel technique for generating a random lead time process without crossing. This technique models a random lead time by the sojourn time of a customer in a $G_1/G/1$-queue plus a fixed lead time. The experiment revealed that our algorithm is sufficiently accurate for practical applications. This is in particular true for situations with little or no intermediate stocks. Furthermore, we showed the importance of including lead time variation in the model. Although, frequently the attention is focussed on demand variation, we showed that the effect of lead time variation may be larger than the effect of demand variation on the stock levels required to obtain prespecified target fill rates.

An important drawback of the algorithms developed in Chapters 3 and 4 (considering the service measure objective) is the assumption that the mean amount of stock held at the intermediate stockpoints has to be known in advance. The algorithms developed in Chapters 5 and 6 (considering the cost measure objective) do not require this assumption. These algorithms try to compute all the control parameters such that the expected systemwide holding and penalty costs per review period are minimized. In Chapter 5, we prove that the decomposition along the lines of Clark & Scarf [1960] and Langenhouff & Zijm [1990] is also exact for divergent $N$-echelon systems given the balance assumption. Hence, the complex multi-dimensional problem of determining the cost-optimal policy reduces to the problem of determining: (1) the order-up-to-level at each stockpoint, and (2) the optimal allocation functions at each intermediate stockpoint. We proved that the first problem coincides with the classical newsboy problem. Furthermore, we obtained insight in the structure of the optimal allocation functions. An algorithm was developed to actually determine these functions. Based on several properties of the optimal allocation functions we classified these functions into four classes.

It is rather cumbersome and time-consuming to determine the optimal allocation functions. Hence, there is a need for a more practically useful approach. Therefore, in Chapter 6 we restricted to a specific class of linear allocation functions. Under this additional restriction we developed an easy and fast algorithm to determine the control parameters such that the expected systemwide holding and penalty costs per review period are minimized.

A possible way to decrease the holding costs needed to operate the system, but still guarantee target service levels at the end-stockpoints, is by allowing lateral transshipments. In the situation where some end-stockpoints have excess inventory while others face shortages, lateral transshipments have gained in popularity as the appropriate recourse action to avoid shortages. However, by allowing lateral transshipments extra transportation (and handling) costs are involved. In Chapter 7, we analyzed divergent multi-echelon systems in which lateral transshipments between the end-stockpoints are allowed. By comparing the model with transshipments to a similar model without transshipment, we illustrated that a mean reduction of safety stock of 15% is possible by allowing transshipments. Furthermore, the amount of imbalance becomes negligible. Finally, in Section 7.7 a first step was made to extend the model such that rebalancing does not take place every review period, but only when the echelon stock drops below a critical level. More research should be done on this model, to analyze the
trade off between the decrease of the rebalancing set-up costs and the increase in safety stock needed to satisfy the customer service levels.

Finally, we would like to address some interesting issues which results from this monograph. First, in Section 8.1, we discuss how the algorithms in Chapters 3 and 6 could be extended by incorporating service level constraints. Second, in Section 8.2, we give some interesting topics for further research.

8.1 Cost minimization given service level constraints

In this monograph, we considered two different objectives: the service measure objective and the cost measure objective. To apply the algorithms developed for the service measure objective we require to have information on the mean amount of stock at the intermediate stockpoints in advance. Similarly, to apply the algorithms developed for the cost measure objective we require to have information on the penalty costs in advance. In principle we would like to have an algorithm which does not require this information in advance. For instance, we need an algorithm which determines the control parameters such that the expected systemwide holding costs is minimized subjected to some customer service level constraints. An important contribution which may eventually result in such an algorithm is the generalization of the theorem of Van Houtum, Inderfurth & Zijm [1996] to divergent N-echelon systems.

Theorem 8.1. Let \((\hat{S}_i, \hat{\xi}_i)\) denotes a cost optimal policy with respect to the sum of the expected systemwide holding and penalty costs per period, and suppose that the modified fill rate attained at end-stockpoint \(j\) (denoted by \(\hat{\gamma}_j\)) equals \(\hat{\gamma}_j\). Then \((\hat{S}_i, \hat{\xi}_i)\) is also a cost optimal policy with respect to the expected systemwide holding costs per period, within the class of ordering policies where \(\gamma_j\) equals at least \(\hat{\gamma}_j\).

Proof. Let us just give an outline of the proof. Suppose \((\hat{S}_i, \hat{\xi}_i)\) denotes a cost optimal policy with respect to the sum of the expected systemwide holding and penalty costs per period, and suppose that the modified fill rate attained at end-stockpoint \(j\) (denoted by \(\gamma_j\)) equals \(\hat{\gamma}_j\). Consider the following optimization problem

\[
\min \text{ Expected systemwide holding costs } + \sum_{j \in E} p_j E[-(J_j)^+] \quad \text{s.t.} \quad \gamma_j = \hat{\gamma}_j, \quad j \in E. \tag{P1}
\]

From (2.19) and \(\gamma_j = \hat{\gamma}_j\) it can easily be shown that \(E[-(J_j)^+] = (1 - \hat{\gamma}_j)\mu_j\). Substitution of this result in the objective function of the optimization problem (P1) yields

\[
\min \text{ Expected systemwide holding costs } + \sum_{j \in E} p_j (1 - \gamma_j)\mu_j \quad \text{s.t.} \quad \gamma_j = \hat{\gamma}_j, \quad j \in E.
\]

Since \((1 - \hat{\gamma}_j)\mu_j\) is a constant, the policy \((\hat{S}_i, \hat{\xi}_i)\) not only solves the optimization problem above, but also

\[
\min \text{ Expected systemwide holding costs } \quad \text{s.t.} \quad \gamma_j = \hat{\gamma}_j, \quad j \in E.
\]

Then it is trivial that the policy \((\hat{S}_i, \hat{\xi}_i)\) also solves

\[
\min \text{ Expected systemwide holding costs } \quad \text{s.t.} \quad \gamma_j \geq \hat{\gamma}_j, \quad j \in E. \tag{P2}
\]

From Theorem 8.1, it follows that if we know the penalty costs \(\{p_j\}_{j \in E}\) such that the modified fill rate attained at end-stockpoint \(j\) equals some predefined target value \(\hat{\gamma}_j\), we can solve the problem
(P2) simply by solving (P1). Problem (P1) (without the modified fill rate constraints) is solved by the algorithms developed in Chapters 5 and 6. So, the only open question remains how to determine the penalty costs \( \{ \bar{P}_j \}_{j \in E} \) such that \( \gamma_j = \bar{P}_j \) for all \( j \in E \). In my opinion this is an important topic for further research, because if an algorithm would be able to solve problem (P2) for arbitrary values of \( \{ \bar{P}_j \}_{j \in E} \), this algorithm could also be used for high values of \( \bar{P}_j \) to approximate the solution of problem (P2) where the modified fill rate constraints are replaced by the more common fill rate constraints.

8.2 Other possible model extensions

In this monograph, we assumed that the inventory at each stockpoint is controlled by a periodic review echelon order-up-to-policy. This means that every \( R \) period each stockpoint inspects its inventory to raise the echelon inventory position to its order-up-to-level (note that for the random lead time case the review period of stockpoint \( i \), denoted by \( R_i \), is a random variable with mean \( R \)). We assumed that the review period \( R \) is the same for all stockpoints in the system. In practice, however, one may face a situation where the review period depends on the location in the system. For instance, in Van der Heijden [1992] it is argued that more downstream stockpoints more frequently inspect their echelon inventory position than more upstream stockpoints. A possible way to deal with this is analyzed in the very recent paper of Van der Heijden [1997a].

An other model extension very related to this is not to use a periodic review echelon order-up-to-policies at each stockpoint, but a periodic review order-point, order-up-to-policy or a periodic review order-point, order-quantity policy.

Finally, we note that also the extension from single-item multi-echelon systems to multi-item multi-echelon systems requires more research.
Bibliography


AXSÄTER, S., AND K. ROSLING [1993], Notes: Installation vs. echelon stock policies for multilevel inventory control, Management Science 39, 1274–1280.

BERTRAND, J.W.M., AND A.G. DE KOK [1995], Performance measurement and performance control in supply chain management, EUT 95-05, Department of Industrial Engineering and Management, Eindhoven University of Technology.


BURGIN, T. [1975], The gamma distribution and inventory control, Operational Research Quarterly 26, 507–525.


DIks, E.B., AND A.G. De Kok [1996b], Cost-effective control policies for multi-echelon distribution systems, Memorandum COSOR 96-36, Department of Mathematics and Computing Science, Eindhoven University of Technology, Submitted for publication.


Donselaar, K. Van [1996], Personal communication.

Donselaar, K. Van, AND J. Wijngaard [1987], Commonality and safety stocks, Engineering Costs and Production Economics 12, 197–204.


Ehrhardt, R. [1984], (s, S) policies for a dynamic inventory model with stochastic lead times, Operations Research 32, 121–132.


Bibliography


HEIDEN, M.C. VAN DER [1997a], Multi-echelon inventory control in divergent systems with shipping frequencies, P 5, Institute for Business Engineering and Technology Application, University of Twente.


HEUTS, R., AND J. DE KLEIN [1995], An (s, q) inventory model with stochastic and interrelated lead times, *Naval Research Logistics* 42, 839–859.


NAHMIA, S. [1979], Simple approximations for a variety of dynamic leadtime lost-sales inventory


VINSON, C.E. [1972], The cost of ignoring lead time unreliability in inventory theory, Decision Sciences 3, 87–105.
ZIPKIN, P. [1986], Stochastic leadtimes in continuous-time inventory models, Naval Research Logistics Quarterly 33, 763–774.
Samenvatting

Dit proefschrift is gewijd aan het optimaliseren van de materiaalcoördinatie in een divergente meerechelon voorraadssysteem. Een divergente meerechelon voorraadssysteem is een netwerk van voorraadpunten, waarin elk voorraadpunt wordt bevoorrad door precies één voorraadpunt, maar zelf meerdere voorraadpunten kan bevoorraden. Dit soort voorraadsystemen komt veelvuldig voor aan het einde van een logistieke keten. Bijvoorbeeld, een producent leverent een type product aan een centraal depot, die op zijn beurt dit product distribueert over een aantal warenhuizen. Voor de warenhuizen is het van groot belang om de voorraad van dit product op peil te houden. Te weinig voorraad zal immers uiteindelijk leiden tot het niet kunnen leveren aan klanten, die daardoor misschien het product bij de concurrent kopen. Hierdoor loop je niet alleen de winst mis die je had kunnen verdienen door het product te verkopen, maar ook de goodwill van een klant. Aan de andere kant zal te veel voorraad leiden tot onnodig hoge voorraadkosten. Om de juiste hoeveelheid voorraad te bepalen zal er dus een afweging gemaakt moeten worden tussen de voorraadkosten en de boetekosten (de kosten verbonden met het niet kunnen leveren). De hoeveelheid voorraad in een warenhuis hangt af van: (1) De regelmaat waarmee het warenhuis orders plaatst, (2) De grootte van deze orders, en (3) Het voorraad verloop in het centrale depot, wat op zijn beurt natuurlijk weer afhangt van de regelmaat waarmee het centrale depot orders plaatst bij de producent en de grootte van deze orders. In dit proefschrift zullen we een model ontwikkelen om inzicht te krijgen in het verloop van de voorraad in het gehele systeem.

Het doel van dit proefschrift is om algoritmen te ontwikkelen die de parameters bepalen van een voorraadbeheersingsstrategie van divergente meerechelon systemen zodanig dat aan een service- of kostenkriterium wordt voldaan. Zo'n strategie bepaalt het voorraadverloop in elk voorraadpunt van het systeem. In dit proefschrift beperken we onze tot de volgende beheersingsstrategie. Periodiek zal elk voorraadpunt zijn zogenoemde echelon voorraadpost positie inspecteren en op basis daarvan een order plaatsen om deze voorraadpositie te verhogen tot een vast ophoogniveau. Wanneer zijn leverancier over te weinig voorraad beschikt om alle orders direct te kunnen leveren, dan zal hij overgaan tot het verdelen van de beschikbare voorraad op basis van allocatiefuncties. Dat deel van de orders dat niet direct geleverd kan worden, zal nict nageleverd worden maar gaat verloren. Dit in tegenstelling tot de klantenvraag bij een eindvoorraadpunt, d.w.z. een voorraadpunt dat in contact staat met de markt (in het bovenstaande voorbeeld was dit het warenhuis). Kortom, de parameters die bepaalt dienen te worden zijn: Het ophoogniveau en de allocatiefuncties van elk voorraadpunt in het systeem. Deze parameters worden bepaald op basis van een van tevoren gekozen criterium. We beschouwen twee verschillende criteria: het service- en kostenkriterium. Allereerst, ontwikkelen we in de hoofdstukken 3 en 4 algoritmen om de parameters te bepalen zodanig dat de servicegraad in een eindvoorraadpunt voldoet aan voorafgestelde eisen. Bijvoorbeeld, we eisen dat de fractie van de klantenvraag die onmiddellijk geleverd kan worden gelijk is aan 95%. Ten tweede, ontwikkelen we in de hoofdstukken 5 en 6 algoritmen om de parameters te bepalen zodanig dat de totale verwachte kosten in het systeem geminimaliseerd worden. Bij de ontwikkeling van deze algoritmen concentreren we ons met name op de numerieke stabiliteit en de snelheid, zodat dat we uiteindelijk een sterk gereedschap krijgen om snel en betrouwbare voorraadssystemen te kunnen optimaliseren. De ontwikkelde algoritmiek in dit
Samenvatting

Proefschrift onderscheidt zich van de meeste technieken in de literatuur, doordat het in staat is de materiaalcoördinatie voor een grote klasse van systemen te optimaliseren (willekeurig grote divergente systemen met heterogene eindvoorraadpunten).

De beheersingsstrategie zoals die hierboven beschreven is, wordt uitgebreid besproken in hoofdstuk 2. In hoofdstuk 3 introduceren we namelijk een divergent meer-echelon systeem dat de basis vormt voor de hoofdstukken 3 t/m 6. Naast een discussie van alle modelaanmanen, karakteriseren we het stoektastische verloop van de voorraad in elk voorraadpunt m.b.v. de 'sample path' techniek. Met behulp van deze techniek zijn we onder andere in staat om de echelon voorraadpositie in elk voorraadpunt te bepalen, mits alle parameters a priori bekend zijn. De verdeling van de echelon voorraadpositie wordt gegeven door een onvolledige converginge van continue verdelingsfuncties. Een approximatie voor deze verdeling kan verkregen worden door gebruik te maken van de recursieve techniek die door De Kok ontwikkeld is. Deze techniek is een mengsel van Erlang verdelingen op de eerste twee momenten van een niet negatieve stochast. Dit alles stelt ons in staat om de werkelijke servicegraad in elk voorraadpunt te bepalen. Verschillende servicegraad definities worden behandeld.

In hoofdstuk 3 presenteren we verschillende algoritmen om alle parameters te bepalen op basis van het servicegraad criterium. De algoritmen zijn gebaseerd op de Consistent Appropriate Share (CAS) strategie, ontwikkeld door De Kok, Lagodimos & Seidel [1994], en de Balanced Stock (BS) strategie, ontwikkeld door Van der Heijden [1997b]. Naast het uitbreiden van de toepasbaarheid van beide strategieën voeren we een uitgebreid numeriek experiment uit om inzicht te krijgen in de prestatie en verschillen tussen verschillende variaties die in de loop van de tijd ontwikkeld zijn voor de CAS en BS strategie. Het experiment bestaat uit het vergelijken van de analytisch bepaalde servicegraad met de uit simulatie verkregen servicegraad voor vele instanties. Een van de belangrijkste conclusies van dit experiment is dat de BS strategie beter is dan de CAS strategie.

De algoritmen in hoofdstuk 3 zijn ontwikkeld onder de veronderstelling dat de levertijd van elk voorraadpunt deterministisch is. Aangezien de levertijd van een eindvoorraadpunt meestal overeenkomt met de transporttijd is het redelijk om deze levertijd als deterministisch te beschouwen. Daarom representeert de levertijd in een voorraadpunt meer stroomopwaarts in het systeem meestal de productie. Deze productie is stochastisch door, bijvoorbeeld, capaciteitsbeperkingen, het uitvallen van machines of het batchen van orders. Daarom breiden we in hoofdstuk 4 het basismodel uit door stochastische levertijden toe te laten. Zoals in hoofdstuk 3 ontwikkelen we een algoritme die de parameters bepaalt op basis van het servicegraad criterium. Vanwege de superieure prestatie van de BS strategie beperken we ons tot deze strategie. Om in dit model met stochastische levertijden gebruik te maken van de al eerder genoemde sample path techniek, moeten we veronderstellen dat geplaatste orders elkaar niet inhalen. Voor het testen van de prestatie van het door ons ontwikkelde algoritme wordt weer opnieuw een uitgebreid numeriek experiment uitgevoerd. Hieruit blijkt dat prestatie van het algoritme voldoende nauwkeurig is om het in de praktijk toe te passen. Bij de simulatie maken we gebruik van een nieuwe algoritme om een stochastisch levertijd proces te genereren waarin orders elkaar niet inhalen. Deze techniek modelleert de levertijd als de systeemtijd van een klant in een G1/G/1-wachtlijn plus een constante tijd.

Een tekortkoming in de toepasbaarheid van de algoritme is de klachten in de hoofdstukken 3 en 4 is dat de gemiddelde hoeveelheid voorraad in niet-eindvoorraadpunten a priori bekend dient te zijn. Daarom ontwikkelen we in hoofdstuk 5 een algoritme die deze informatie niet nodig heeft, maar daarregen bepaalt dit algoritme de parameters op basis van het servicegraad criterium, maar op basis van het kostencriterium. We beschouwen enkel voorraad- en boetekosten. Het algoritme is gebaseerd op het echelon stock concept van Clark [1958] en de decompositie aanpak van Langenhoff & Zijm.
[1990]. In hoofdstuk 5 bewijzen we dat het complexe probleem van het bepalen van alle parameters gedecomposeerd kan worden (onder de zogenaamde balansaamname). Door deze decompositie kan voor elk voorraadpunt afzonderlijk het ophogenniveau en de allocatiefuncties bepaald worden. We bewijzen dat het probleem van het bepalen van het ophogenniveau overeenkomt met het alom bekende newsvoy probleem. Voor het bepalen van de allocatiefuncties ontwikkelen we een algoritme. Op basis van een aantal eigenschappen classificeren we deze optimale allocatiefuncties in vier klassen.

Over het algemeen is het zeer lastig en tijdrovend om voor elk voorraadpunt de optimale allocatiefuncties te bepalen. Daarom is er behoefte aan een meer praktisch toepasbare aanpak. In hoofdstuk 6 beperken we ons daarom tot een specifieke klasse van lineaire allocatiefuncties. Voor deze klasse ontwikkelen we een algoritme die alle parameters eenvoudig en snel bepalen, zodanig dat de verwachte totale kosten vrijwel geminimaliseerd worden.

Een manier om de kosten voor het houden van voorraad te verminderen, terwijl toch aan het servicegraad criterium wordt voldaan, is door het toelaten van overslag tussen eindvoorraadpunt. Wanneer het regelmatig voorkomt dat sommige eindvoorraadpunt ruim voldoende voorraad hebben terwijl andere moeten naleveren kan het interessant zijn om deze naleveringen te voorkomen door overslag tussen eindvoorraadpunt. Het toelaten van deze overslag vermindert de totale hoeveelheid producten in het systeem dat nodig is om aan het servicegraad criterium te voldoen, dit zal leiden tot lagere voorraadkosten, daarentegen staan daar wel extra kosten (bv. transport- en administratiekosten) tegenover. In hoofdstuk 7 analyseren we een divergent meer-echelon voorraadssysteem waarin overslag tussen eindvoorraadpunt toegestaan is. Door dit model te vergelijken met het model zonder overslag, kunnen we bepalen of het winstgevend is om overslag tussen eindvoorraadpunt toe te laten.
Curriculum Vitae

Erik Diks was born on July 20, 1970, in Veldhoven, the Netherlands. He attended HAVO-education at the Brabant-Havo in Boxtel, where he graduated in June 1987. Next, he attended VWO-education (pre-university education) at the Jacob-Roelandlyceum in Boxtel, where he graduated in June 1989. In September 1989 he started his study in applied mathematics at the Eindhoven University of Technology. At this institute he specialized in operations research and statistics. In 1992 he was employed as a teaching assistant at the department of Mathematics and Computing Science. He received his Master’s Degree in applied mathematics in August 1993. His master’s thesis called *The change of traffic characteristics in ATM networks and the use of traffic shapers*, describes the results of research carried out at PTT Research Leidschendam under the supervision of dr.ir. J. van der Wal, dr. J.L. van den Berg and dr. J.A.C. Resting. In September 1993, he became a Ph.D.-student at the Eindhoven University of Technology. The study described in this monograph was carried out at the department of Mathematics and Computing Science under the supervision of prof.dr. A.G. de Kok. Besides his scientific work he gave courses on stochastic operations research to master’s degree students. Since January 1996 he is an executive committee member of the ‘Vereniging voor Wiskundig Ingenieurs Eindhoven’ (WIRF). From September 1997 he is employed as an operations research engineer at Baan Company, where he is involved with the development of Enterprise Resource Planning information systems.
Stellingen
behorende bij het proefschrift

Controlling Divergent Multi-Echelon Systems

van

Erik B. Diks

I
Beschouw een divergent voorraadssysteem met meerdere echelons, waarin elk voorraadspunt periodiek een bestelling plaatst bij zijn leverancier. Iedere bestelling arriveert na een constante levertijd en wordt gepland om ofwel aan de vraag van de afnemers te voldoen, ofwel om het lokale voorraadniveau op te hogen. Het deel van de klantenvracht dat niet direct gedeeld kan worden, zal worden nabestaan. Deze nabestellingen brengen wel boetekosten met zich mee. Naast boete-
kosten, worden er voorraadkosten geheven over alle producten in het systeem. De boete- en
voorraadkosten worden geheven vlak voor de aankomst van iedere order bij een voorraadspunt.
Onder de balansnaming reikt nu dat door decompositie van dit systeem de beheersingstra-
tagie bepaald kan worden, die de totale verwachte kosten per tijdseenheid minimaliseert (zie
Hoofdstuk 5 van dit proefschrift).

II
In [1] analyseert Federgruen een 2-echelon systeem met constante levertijden, met als doel de
beheersingstraagte te bepalen die de gemiddelde verwachte kosten minimaliseert. Federgruen
beweert dat deze analyse kan worden uitgebreid naar stochastische levertijden door in alle re-
levante expressies simpelweg de verwachting te nemen over de marginaal verdelingen van de
verschillende levertijden.

a. Bovenaanzichts bevordering kan echter getoetst worden indien er een nadere beschrijving is
van de momenten waarop de boete- en voorraadkosten geheven worden.
b. Indien deze kosten geheven worden net voor de aankomst van iedere order bij een voor-
raadspunt, dan is deze bevordering onjuist.

Logistics of production and inventory, Handbooks in Operations Research and Manage-
Zij $X$ een continue kansvariabele op $[0, \infty)$ met verwachting $\mu$ en variatiecoëfficiënt $c$. Zij $Y$ een kansvariabele met een hypereponentialle verdeling, waarbij $Y$ met kans $q$ exponentieel verdeeld is met parameter $\lambda_1$ en met kans $1-q$ exponentieel verdeeld is met parameter $\lambda_2$. Kansvariabele $Y$ heeft dezelfde verwachting en variatiecoëfficiënt als $X$, wanneer $Y$ als volgt wordt gekozen:

$$\lambda_1 := \frac{\alpha \mu + 1 - q}{\alpha \mu} \quad \text{en} \quad \lambda_2 := \alpha \lambda_1,$$

waarbij $q$ vrij gekozen kan worden uit een interval, dat afhankelijk is van $c^2$ (zie Figuur 1). Na de kiezen voor $q$ volgt de waarde van $\alpha$ uit Figuur 1, waarbij

$$\alpha_{1,2} := \frac{(1-q)(c^2 + 1) \pm \sqrt{(1-q)(c^2 - 1) \alpha}}{2 - q(c^2 + 1)} \quad \text{en} \quad \alpha_3 := \frac{3 - c^2}{4}.$$

Figuur 1: Mogelijke waarden voor $q$ en $\alpha$, gegeven $c^2$.

IV

Voor het dimensioneren van een ATM shaper is het noodzakelijk om meer informatie te hebben over het aankomstproces, dan alleen de eerste twee momenten van de duur van aan- en uit-periodes.

V

Beschouw een celstrook met constante bit-rate (CBR) die m knooppunten doorloopt. In elk knooppunt interfereren deze celstrook met andere celstrook. De tussenvertrekstijdverdeling van twee opeenvolgende cellen, zodat de CBR celstrook m knooppunten doorloopt heeft, is goed te benaderen wanneer:

1. De bezettingsgraad in elke knooppunt laag is, zie kleiner dan 0.8 en $m < 20$.
2. De bezettingsgraad in elke knooppunt groot is, zie groter dan 0.9.

The change of traffic characteristics in ATM networks, B-ISDN Teletraffic Modelling Symposium, Alcatel Bell, Antwerp, February 1995, 95-98(3).

VI

Beschouw een divergenter systeem met 2 echelon bestaande uit één centraal depot (CD) en een aantal warenhuizen. Elke warenhuis met een bepaalde graad van dienstverlening. Verder levert de producent van het CD elke orde na een constante levertijd, tenzij het CD de producent verzoekt om de orde te verschenen en de producent in staat is om aan dit verzoek te voldoen. Of het, vanuit een kostenperspectief, verstandig is om orden te verschenen hangt vooral af van:

1. de kostenstructuur en
2. de verschillende capaciteit van de producent.

with supply lead time flexibility, Technical report, Eindhoven University of Technology.

VII

Beschouw een GI/G/1-wachtrij systeem. De tussenaankomsttijd $A$ van twee opeenvolgende klanten, is verdeeld volgens verdelingsfunctie $F_A$ met verwachting $E_A$. De bedieningsduur $B$ van een klant heeft verwachting $E_B < E_A$. De systeemtijd $S$ van een klant is verdeeld volgens verdelingsfunctie $F_S$ met verwachting $E_S$ en variante $\sigma^2_S$. De correlatiecoëfficiënt $\rho$ van twee opeenvolgende systeemtijden is gelijk aan

$$
\rho = 1 - \frac{(\mu_A - \mu_B)\sigma^2 + \int_0^{\infty} \int_0^{\infty} x(1 - F_A(x + y)) dy dF_S(x)}{\sigma^2_S}.
$$

[5] E.B. Diks and M.C. van der Heijden [1997], Modeling stochastic lead times in multi-
echelon systems, Probability in the Engineering and Informational Sciences 11, 459-485.
VIII

Milieu is in de mode.

IX

Menige schertstelling wordt gevoed door frustraties die in de AIO periode zijn opgedaan.

X

Intelligentie en geduld zijn geen voldoende voorwaarden om de tv-serie *Inspector Morse* te kunnen waarderen.