Multiple sourcing in single- and multi-echelon inventory systems

I n a u g u r a l d i s s e r t a t i o n

zur Erlangung des akademischen Grades eines Doktors der Wirtschaftswissenschaften der Universität Mannheim

vorgelegt

von

Steffen Klosterhalfen

aus Mannheim

Dekan: Prof. Dr. Hans H. Bauer

Erstberichterstatter: Prof. Dr. Stefan Minner

Zweitberichterstatter: Prof. Dr. Moritz Fleischmann

Tag der mündlichen Prüfung: 31. Mai 2010

To my parents

and

Agathe

This page intentionally left blank

Acknowledgements

Even though it has been already four and a half years since I have started working on the research that finally resulted in this thesis, it does not feel that long to me. It is commonly known that time flies when you are having stress . . . or was it fun? For me it has been definitely a mixture of both, with the latter one prevailing. Although very demanding and straining at times, I have really enjoyed my time at the (Department of Logistics at the) University of Mannheim. The reason for this having been such an extraordinary and positive experience in my life is most importantly due to the people I have met, learned from, and worked together with during this period. Without these outstanding individuals I would have never been able to complete this work.

First and foremost, I would like to thank my supervisor, Prof. Dr. Stefan Minner. With his great enthusiasm for doing research and his creativity he was the one who aroused my own interest in the research world during my diploma thesis and encouraged me to embark on this exciting journey. I have benefited tremendously from his valuable comments, his innovative approaches to research problems, and his persistent pursuit of achieving results of highest quality. I am deeply thankful to Stefan for teaching me how to conduct research by setting such a great example himself: always viewing problems from different angles, critically analyzing results, and thus deriving sound conclusions.

Furthermore, I am most grateful to Prof. Dr. Moritz Fleischmann for reviewing my thesis. Beyond that I have sincerely appreciated the warm welcome at his chair and his readiness to answer questions and discuss problems of various nature at any time.

During my research stays at Eindhoven University of Technology and Boston University I have had the opportunity to get to know many interesting people and various distinguished researchers. Two of them I would like to emphasize here in particular: Prof. Dr. Gudrun Kiesmüller and Professor Sean Willems. It has been a fruitful and exceptional experience for me to collaborate with them on joint research projects. I owe my deepest gratitude to both of them for all the advice they have provided and the progress I have made thanks to them.

The close bond I have formed with several (former and present) colleagues at the University of Mannheim significantly enhanced my learning, on and off the academic track. I can deem myself extremely lucky for having been part of such a great team. My special thanks go to Dr. Jan Arnold, Daniel Dittmar, Gabriele Eberhard, Dr. David Francas, Dr. Jörn Grahl, Stefan Hahler, Dr. Mirko Kremer, Dr. Lena Kunze, Ruth Pfitzmann, Volker Ruff, Dr. Holger Stephan, Dr. Sandra Transchel, Dr. Joachim Vaterrodt, Yao Yang, and Caroline Zuber. Moreover, my colleagues and friends from the Dean's Office of the Business School and the Graduate School of Economic & Social Sciences have greatly contributed to an amazing work and social environment.

Most importantly, I would like to express my dearest gratitude to my family. During the entire research process, especially my parents and my sister have provided me with unconditional support and encouragement. Knowing that I could always rely on them gave me the strength I needed to achieve this goal. Thanks a million to all of you.

It is my habit to save the best for last ... and I won't break it here. In this case, the best is you, Agathe. I am eternally grateful for you having been there for me particularly during these stressful last couple of months, when I was finishing my thesis. You have been immensely understanding about me having so little time for you ... and you always are about all my other minor and major quirks. You are the best that could have ever possibly happened to me.

Mannheim, July 2010

Contents

List of Figures

List of Tables

List of Abbreviations

ANOVA	Analysis of variance
BS	Best single-sourcing strategy
COP	Constant-order policy
DIP	Dual-index policy
DP	Dynamic program
e.g.	for example
gam	gamma
geom	geometric
GS	guaranteed-service
HS	hybrid-service
i.e.	that is
i.i.d.	identically and independently distributed
LS	Lost sales
MDP	Markov Decision Process
nbin	negative binomial
norm	normal
n/s	not significant
Pois	Poisson
Pr	Probability
SIP	Single-index policy
SS	stochastic-service
TC	Total cost
TRC	Total relevant cost
vs.	versus
w.l.o.g.	without loss of generality

List of Symbols

Single-echelon model with dual sourcing

Demand

d_t	demand realization in period t
D_t	demand random variable of period t
$D(L)$	demand random variable over L periods
$\bar{D}$	maximum feasible demand realization
FD_t	filled demand in period t
$f(x)$	single-period demand probability density/mass function
$F(x)$	single-period demand cumulative distribution function
f_L	L-period demand probability density function
F_L	L-period demand cumulative distribution function
$\phi(x)$	standard normal probability density function
$\Phi(x)$	standard normal cumulative distribution function
$\mathbb{E}[D] = \mu$	single-period demand expectation
$\text{VAR}[D] = \sigma^2$	single-period demand variance
σ	single-period demand standard deviation
$CV = \frac{\sigma}{\mu}$	coefficient of variation
$\Gamma(\kappa, \theta)$	complete gamma function with shape parameter κ and scale parameter θ

Timespans

L	(replenishment) lead time
L^f	lead time when sourcing from the fast supplier
L^s	lead time when sourcing from the slow supplier
$L^\Delta = L^s - L^f$	lead-time difference

Cost, service, and performance measures

c^f	procurement cost per unit when sourcing from the fast supplier
c^s	procurement cost per unit when sourcing from the slow supplier
$c = c^f - c^s$	procurement cost difference or expediting premium per unit
h	(local) holding cost per unit and period
h^e	echelon holding cost per unit and period
b	backorder cost per unit and period
$\alpha^{target}, \beta^{target}, \gamma^{target}$	theoretical target service levels, which are exogenously given
α, β, γ	achieved service levels

Inventory control and replenishment variables

B	local order-up-to level
B^f	local order-up-to level for replenishments with the fast supplier
B^s	local order-up-to level for replenishments with the slow supplier
$\Delta = B^s - B^f$	difference in the local order-up-to levels
B^{LS}	order-up-to level in the single-sourcing lost-sales model
δ^f	fraction of demand sourced from the fast supplier
δ^s	fraction of demand sourced from the slow supplier
NS_t	net stock at the end of period t
OH_t	on-hand stock at the end of period t
BO_t	backorders at the end of period t
PI_t	pipeline inventory (i.e. in-transit to the stage) at the beginning of a period before any orders arrive
IP_t	inventory position at the beginning of period t before receipt and placement of any orders
IP_t^f	fast inventory position at the beginning of period t
IP_t^s	slow inventory position at the beginning of period t

O_t	overshoot at the beginning of period t after ordering from the fast supplier
Q	constant order quantity (under the COP)
Q_t^f	fast order quantity of period t
Q_t^s	slow order quantity of period t
CAP^f	capacity limit on the fast order
CAP^s	capacity limit on the slow order
SST	safety stock level
$k(SL)$	safety factor for a given service level SL

Markov Chain

p_{ij}	transition probability from state i to j
$\mathcal{P}$	transition probability matrix
$\mathbf{o} = [o_1, o_2, ...]$	steady-state overshoot distribution
$\mathbf{Z}_t$	state vector of period t
$\mathcal{SSP}$	set of all possible states (state space)

Markov Decision Process

a	action or decision
$\mathcal{DSP}(\mathbf{i})$	set of all possible actions/decisions in state $\mathbf{i}$ (decision space)
$(a)_\infty$	a policy
$r(\mathbf{i}, a)$	reward/cost of being in state $\mathbf{i}$ and choosing decision a
$AC((a)_\infty)$	average cost or gain of a given policy $(a)_\infty$
$p_{\mathbf{ij}}(a)$	transition probability from state i to j when decision a is chosen
$\pi_{\mathbf{i}}$	steady-state probability of being in state $\mathbf{i}$
$\mathbf{Y}_t$	state vector of period t
$DC(a)$	direct cost of decision a
$HB(x)$	expected holding and backorder cost function

Multi-echelon model with dual sourcing

The notation already introduced for the single-echelon system also applies for the multi-echelon system with an additional stage/stockpoint index.

Network

$\langle i,j \rangle$	network that runs from stage i to j
$a_{k,\langle i,j \rangle}$	$k \times \langle i,j \rangle$-matrix that shows which stages are part of the GS subnetwork/HS stage $\langle i,j \rangle$
$\mathcal{J}$	set of all feasible GS subnetworks/HS stages
$\mathcal{A}$	arc set of the network representation
(i,j)	arc from i to j
$LC(i)$	level code of stockpoint i
$\mathcal{N}_i$	subset of stockpoints that are connected to i on the subgraph with stockpoints that have a higher level as i
$\mathcal{DS}(i)$	dual-sourcing subnetwork i, i.e. subset of stockpoints in the subgraph from one dual-sourced stockpoint (e.g. j) to the next i
$\mathcal{FS}$	subset of stockpoints that are not part of any dual-sourcing subnetwork, including the final stockpoint
$\mathcal{DS}^j(i)$	subset of stockpoints in $\mathcal{DS}(i)$, which are predecessors of stockpoint i's supplier j (including j)
$\mathcal{N}_i^{\mathcal{DS}(j)}$	subset of all stockpoints in the $\mathcal{DS}(j)$ upstream of i
P^{type}	number of stockpoints with $type \in \{SI, DS\}$, i.e. single or dual sourcing, respectively
P_K^{MS}	number of multiple-sourced stockpoints with K suppliers

Demand

ρ_{ij}	demand correlation between the single-period demands of products at stages/stockpoints i and j
UD^f	single-period demand random variable of the fast supplier
UD^s	single-period demand random variable of the slow supplier

$[UD^f]_L$	L-period demand random variable of the fast supplier
$[UD^s]_L$	L-period demand random variable of the slow supplier
p^f	probability that an order is placed with the fast supplier

Timespans

$L_{j,i}$	(replenishment) lead time between stockpoints j and i
$T_i/T_{j,i}$	processing time of stage i/between stockpoints j and i
T_i^f	processing time when stockpoint i sources from the fast supplier
T_i^s	processing time when stockpoint i sources from the slow supplier
ST_i	(outgoing) service time of stage/stockpoint i
ST_i^f	service time of the fast supplier of stockpoint i
ST_i^s	service time of the slow supplier of stockpoint i
τ_i	coverage or net replenishment time of stage/stockpoint i
$\vec{\tau}_{\langle i,j \rangle}$	vector of net replenishment times in the network from stage i to j
M_i	maximum replenishment lead time of stage/stockpoint i

Cost, service, and performance measures

c_i^{OF}	operating flexibility cost per unit at stage i
$c_{j,i}^{add}$	cost added when an item moves from stockpoint j to i
c_i^{cum}	cumulative cost of an item at stockpoint i
ν	holding-cost/interest rate for the underlying base period
η	scalar that converts the COGS in the same time unit as the pipeline and safety stock cost
α_i^{SS}	α-service level at stage i, which is used for sizing S_i in the SS approach
OS_{pure}^{type}	optimality share of the respective approach ($type \in \{GS, SS\}$) within the two pure approaches, GS and SS

OS_{all}^{type}	optimality share of the respective approach ($type \in \{GS, SS, HS\}$) within all three approaches
RB^{type}	relative benefit of the respective approach ($type \in \{GS, SS, HS\}$)

Inventory control and replenishment variables

$\vec{B}_{\langle i,j \rangle}$	vector of local order-up-to levels in the subnetwork from stage i to j
S_i	echelon order-up-to level at stage i
$F^{[S_1,\ldots S_i]}$	echelon distribution function at i
$Q_{(j,i),t}$	(replenishment) quantity that i sources from j in period t
$\delta_{j,i}$	sourcing fraction, i.e. fraction of the total order quantity that i sources from supplier j

Optimization model

$x_{\langle i,j \rangle}^{type}$	indicator variable that is 1, if the GS subnetwork ($type = GS$) or HS stage ($type = HS$) from i to j is chosen or 0 otherwise
$c_{\langle i,j \rangle}^{type}$	cost of the on-hand stock in the GS subnetwork ($type = GS$) or HS stage ($type = HS$) from i to j

Dynamic program

z_k	one-dimensional state variable of stage k
$\mathcal{Z}_k$	state space of stage k
u_k	one-dimensional decision variable of stage k
$\mathbf{u}_k = (u_k^1, u_k^2, u_k^3)$	three-dimensional decision variable of stage k
$\mathcal{U}_k(z_k)$	decision space of stage k within a state z_k
C_k^{SI}	minimum safety stock cost for the subnetwork with stockpoint set $\mathcal{N}_k$ in case all items are delivered by a different supplier

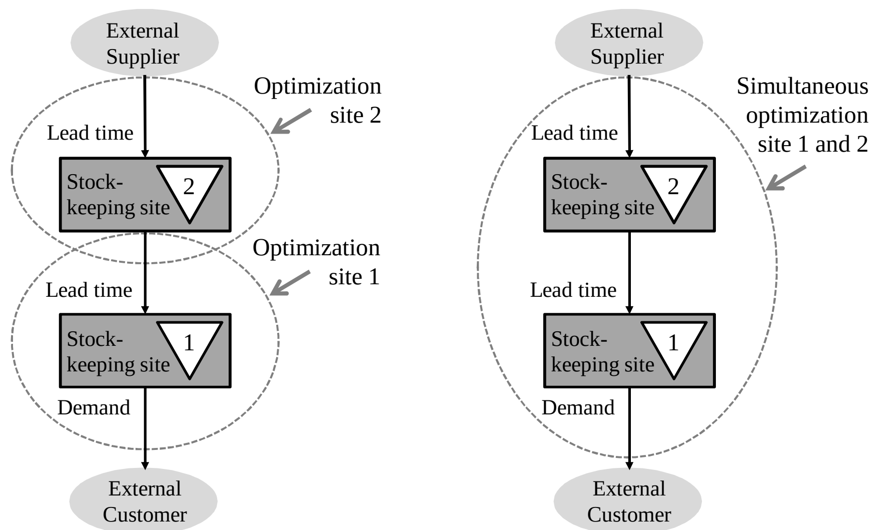
C_k^{DS}	minimum safety stock cost for the subnetwork with stockpoint set $\mathcal{N}_k$ in case an item is delivered by two suppliers
$C_k^{SI^{opt}}$	minimum of the sum of the safety stock cost for the subnetwork with stockpoint set $\mathcal{N}_k^{\mathcal{DS}(i)}$ and the total cost of the preceding dual-sourcing subnetworks
$C_k^{DS^{opt}}$	minimum total cost up to stockpoint k in case k is a dual-sourced stockpoint

1 Introduction

1.1 Motivation

According to a study of the Aberdeen Group (see Viswanathan (2007)), inventory management was ranked on top of the list of investments in application-oriented software for companies in 2007. Within inventory management, multi-tier or multi-echelon inventory optimization was the top priority. This software application area grew by 32% while the overall supply chain management space grew about 7% (see Trebilcock (2009b)). Although in 2008 the market for supply chain management software applications and services saw only a slight increase of 4% over 2007 according to AMR Research (see Trebilcock (2009a)), companies are still putting much emphasis on improving their inventory management activities. 91% of over 170 companies that were surveyed by the Aberdeen Group in 2009 indicated that they have made, or have been asked to provide, recommendations in the past six months to management on how to improve their inventory management processes (see Viswanathan (2009)). An effective inventory management is particularly important in times of economic downturn, like the current global recession. In order to contain cost and free working capital, inventories need to be reduced. On the other hand, there is the risk of losing business in case of insufficient inventories. For the solution of this cost-service trade-off, management more and more often employs advanced software tools to support their decisions (see Ellis et al. (2009)).

Traditional inventory management planning processes and software applications have only been capable of managing inventory at the individual site level. Even though a company might plan its inventory levels at several locations of its supply network centrally, the actual inventory optimization is done one location (or echelon) at a time (see Figure 1.1(a)). Such a sequential single-echelon approach completely neglects interdependencies between the sites, however. Thus, too much inventory might be held or inventory could end up in the wrong place, because aspects such as the following are not taken into account: Is it more costly to hold inventory at an upstream or downstream location? How does the order decision of a downstream location affect the demand process that the upstream location sees? Which level of service should the individual upstream locations provide to their internal customers such that the external customer demand can be satisfied according to the service target there?



(a) Sequential single-echelon approach (b) Multi-echelon approach

Figure 1.1: Sequential single-echelon vs. multi-echelon approach

In contrast, the inventory optimization software tools that have been developed over the past two decades take a holistic approach to inventory management. Such multi-echelon approaches consider all locations in the supply network simultaneously, from the external supplier to the end-customer, with the objective of minimizing total inventory cost in the entire system subject to the service, which is to be guaranteed towards the external customer (see Figure 1.1(b)). Thus, the shortcomings of the sequential single-echelon approach are counteracted. It is reported that 'it is not unusual for a global supply chain to see inventory levels reduced by as much as 15-25%' (Ellis et al. (2009)). There are two major drivers for these advances. First, information and computer technology has gone through great improvements making information across the entire supply network available for such a multi-echelon application and enabling the solution of very large and complex problems. Second, the advancement in multi-echelon inventory research in recent years has produced models that can capture and handle a broad variety of real-world problems and problem sizes (see, e.g., Willems (2008)). None of the software vendors such as LogicTools (which is now a part of IBM), Optiant®, SmartOps®, or ToolsGroup explain in detail the algorithms that are implemented in their tools. However, a closer look at the affiliated scientists and their scientific contributions in this area clearly suggests that these models and algorithms represent extensions of the two pioneering contributions to multi-echelon inventory research by Simpson (1958) and Clark and Scarf (1960).

Based on these two seminal papers on multi-echelon inventory models in base-stock/order-up-to level environments without lot-sizing, two competing research strands have developed over the years. Although they solve the same inventory optimization problem in their core, they make a different assumption with regard to the role of safety stock. The resulting consequences for the material flow in the system coined the terms full-delay and no-delay approaches (van Houtum et al. (1996)), or stochastic- and guaranteed-service approaches (Graves and Willems (2003)). In the stochastic-service (SS) approach, safety stock is assumed to be the only buffer against demand variability. The guaranteed-service (GS) approach, on the other hand, assumes that safety stock is sized to cover demand variability up to a certain level only, i.e. the normal or maximum reasonable demand. All variability exceeding this level is dealt with by other extraordinary countermeasures.

Both assumptions are quite strong. In reality, the truth probably lies somewhere in between. A single supply network might consist of stages with low flexibility as well as stages with high flexibility. For the former ones, the SS approach would be the appropriate one to use, whereas the latter ones are appropriate candidates for the application of the GS approach. In both the academic literature and the currently available software applications, the two approaches are mainly treated and implemented as mutually exclusive frameworks, however. Since both approaches yield different results due to the differing underlying assumptions, a practitioner (as a buyer of these products) faces the dilemma of choosing the appropriate approach and thus software tool for the inventory optimization of its supply networks. This thesis provides assistance in this respect by outlining both approaches in detail and comparing them in generic network settings. Even though this helps in selecting the better of the two approaches, still not all cost-saving potentials can be realized due to the choice of a single approach for the entire supply network, instead of a stage-wise decision within a single network. Therefore, the second contribution of this thesis lies in the development of an integrated framework, which combines the two pure approaches and thus enables a stage-wise choice.

Within each of the two pure multi-echelon frameworks various (additional) real-world aspects of inventory systems have been incorporated over the past two decades. These include non-stationary demand, forecasts, capacity constraints on the orders, to name a few. Another aspect has received less attention, even though it has become increasingly important for companies starting from around the turn of the century in the wake of 9/11 and a growing number of natural disasters that could cause disruptions in the supply flow: the sourcing strategy. The main decisions that characterize the sourcing strategy concern (see, e.g., Chopra and Meindl (2007)): (a) criteria for supplier identification (*How to establish the supplier base?*), (b) criteria for supplier selection (*How to pick suppliers from the base, who will receive an order from the company?*), and (c) procurement (*How much to order from each selected supplier?*). Whereas the first two aspects represent supply chain design decisions, the third one actually concerns inventory optimization and as such falls within the scope of this thesis. In most of the multi-echelon models it is implicitly assumed that the first two decisions result in the choice of a single supply source. Although such a single-sourcing strategy has been advocated to have many advantages, such as a stronger, long-term relationship with the supplier and the reduction of overheads required for managing multiple supplier relationships, it also has its risks in the form of total dependency of the functioning of the entire supply chain on a single source. Any disruption at the source could disable the supply chain (see, e.g., Lee and Wolfe (2003)). That is why companies turn to more flexible sourcing strategies, like dual or multiple sourcing, i.e. they rely on two or more supply options. These supply options can refer to different suppliers, e.g., an overseas plant and another one close by, or to different transportation modes, e.g., trucking, sea-, or air-shipping. This not only increases supply chain security, but also represents a means to effectively solve the cost-service trade-off with respect to inventory management. Examples are reported in the literature for Hewlett Packard (Beyer and Ward (2000)), Caterpillar (Rao et al. (2000)), or Océ (Scheller-Wolf et al. (2007)), amongst others. The supply options incur different costs, but also take up different amounts of time. Thus, one strategy could be to replenish the majority of items using the cheaper, but slower source, and to use the fast, but expensive source only in case of imminent stockouts caused by the volatility of demand. For example, HP produces the majority of its DeskJet printers in Singapore, because of that country's lower cost structure. Thus, the focus with this source is on cost efficiency. In addition, HP has manufacturing in Vancouver. This source ensures a fast response to the North American market and is utilized to guarantee high service (see Lee and Wolfe (2003)). Consequently, a dual-sourcing strategy enables a company to serve demand at low costs without compromising on service. In order to realize these potential gains from dual sourcing, an effective inventory control policy needs to be in place, which tells the company when to order from which source and how much.

In contrast to single-sourcing models, where optimal policy results are available for single-echelon as well as multi-echelon settings, the findings for dual-sourcing models are much more limited. Even for single-echelon models the optimal policy is just known for special cases. In more general settings, it can only be determined numerically by using complex mathematical models that require considerable amounts of computation time. This renders it inapt for the application in practice. That is why various non-optimal policies have been proposed in the literature that are easier to compute and manage. These include the single-index (SIP), constant-order (COP), dual-index (DIP), and order-splitting policy (OSP). The relative performance has been tested only of certain policies, however. Scheller-Wolf et al. (2007) provide a comparison of the SIP and DIP. The COP and OSP have not yet been taken into consideration. This thesis closes this gap by comparing the COP and the DIP. The OSP can be excluded from this single-echelon policy comparison due to its arguably inferior performance in the analyzed deterministic lead-time setting. This policy is usually studied under stochastic lead times as a simple and effective means to pool lead-time risk (see, e.g., Thomas and Tyworth (2006)). Thus, the thesis gives guidance to the practitioner as to which policy is an effective choice in specific supply system settings.

Even though the use of an effective dual-sourcing policy, which is found by the single-echelon analysis, might already save cost, a true supply chain inventory optimization approach would have to incorporate all locations of the supply network and thus take a multi-echelon perspective. Since at best only approximate approaches to this problem are available in the literature (see, e.g., the presented ideas in the final section of Graves and Willems (2005)) and therefore implemented in the available software applications, there is room for improvement in this area. As another contribution, this thesis aims at filling this void.

Consequently, the thesis contributes to the body of literature on dual-sourcing inventory control models in a single- and multi-echelon setting. In preparation of the multi-echelon dual-sourcing model development, the literature on multi-echelon inventory models with a single source of supply is extended, as well. The thesis centers around two major research topics:

(i) the detection of effective dual-sourcing inventory control policies in a single-echelon model, and

(ii) the integration of dual-sourcing into a multi-echelon inventory model.

1.2 Research questions

This thesis deals with inventory optimization in supply networks with multiple sourcing. The coordination of replenishment decisions, when two or more suppliers for the same item are available, is mostly studied in a single-echelon setting in the literature. This is also selected as the starting point of this thesis.

The first major research question that is addressed in this thesis is *what are effective dual-sourcing inventory control policies in a single-echelon setting*. In this context, the term 'effective' is understood in the sense of easily implementable and performing close to optimal. Although the optimal policy delivers the lowest cost, its computation might be rather complex, and thus render it less effective from a practicability point of view. Hence, non-optimal policies, which are simpler to compute and still show a satisfactory cost performance, can be regarded as being more effective. Consequently, in order to answer this major research question, several aspects need to be addressed. These make up the following subset of more specific research questions, which are analyzed in turn:

1. *How can the optimal inventory control policy for the studied single-echelon dual-sourcing problem be determined?*
2. *What are simple non-optimal policy alternatives and how can their optimal parameters be computed?*
3. *How do these (non-optimal) inventory control policies perform?*

The second major research question that is investigated in this thesis is *how to integrate dual sourcing into a multi-echelon inventory model*. The integration of dual sourcing into a multi-echelon setting requires good acumen of multi-echelon inventory models with a single sourcing option as a starting point. Here, the literature on multi-echelon inventory models without lot-sizing distinguishes between models following the stochastic-service or guaranteed-service framework, in general. It is not clear from the available contributions, whether any of these approaches is superior to the other. Therefore, it seems worthwhile to first analyze and compare both approaches, before the integration of dual sourcing is addressed. If one of the approaches was generally superior to the other, the extension to dual sourcing would only have to be done for this approach. Hence, the answer to the main research question is multilayered again. That is why it is broken down into several more specific research questions, which eventually provide an answer to the major one. These include:

1. *Given the characteristic assumptions and features of the two competing multi-echelon inventory optimization model strands in the literature, i.e. the stochastic- and guaranteed-service framework, is one of them generally superior to the other?*

   If this is not the case, in which settings does each approach perform well?
2. *Depending on the outcome of the first question, is a mutually exclusive implementation of a single multi-echelon approach for the entire supply network reasonable?*

Put differently, do situations exist where a combination of both approaches provides additional benefits and how can such an integrated approach be designed?

3. *Provided that none of the approaches is generally superior to the other, how can dual sourcing be accommodated in the guaranteed-service approach?*

At first glance, the second specific research question might not seem to be directly related to the ultimate goal of this thesis, which is the integration of dual sourcing into a multi-echelon inventory model. Although this question refers to an extension of the single-sourcing multi-echelon frameworks, the answer to this question fosters a better understanding of the two multi-echelon models and as such represents a valuable basis for the extension to dual sourcing. Moreover, a newly developed integrated multi-echelon framework with single sourcing also represents an additional candidate for a potential dual-sourcing extension. However, this thesis focuses on the extension of the guaranteed-service approach only. The integration of dual sourcing into any other approaches is postponed to future research.

1.3 Structure and overview

This thesis is divided into 5 chapters. After the presentation of the research motivation, the specific research questions, and the overall structure of the thesis in this chapter, Chapter 2 outlines fundamentals that are required for a thorough understanding of the thesis and reviews the relevant literature. Section 2.1 starts by briefly discussing the relevant demand distributions in Section 2.1.1. Demand is regarded as the primary source of uncertainty in the inventory models studied in this thesis. In Section 2.1.2, the basic inventory control terminology is introduced followed by a description of several performance measures for the evaluation of an inventory control policy. Next, a basic infinite-horizon inventory model is presented, i.e. the single-echelon periodic-review order-up-to level model with single sourcing. Section 2.1.3 is concerned with the basics of multi-echelon inventory control. The notion of a process, stockpoint, and stage is explained together with the basic supply network structures that can be encountered in practice.

The literature review in Section 2.2 follows the general structure of the thesis. First, the body of literature on single-echelon inventory models with multiple sourcing is discussed in Section 2.2.1. The contributions are characterized into works that are concerned with the derivation of the optimal policy and works that deal with the parameter optimization for a given (not necessarily optimal) policy. This classification also reflects the historical development quite well. Whereas early contributions have focused on the determination of the optimal policy, more recent contributions rather center around policies that are applicable in practice. Second, Section 2.2.2 reviews multi-echelon inventory models with single sourcing. This field of literature can broadly be classified into models following the stochastic-service framework and those following the guaranteed-service framework. In each field, contributions that concentrate on the derivation of the optimal policy structure or the computation of the optimal policy parameters for different network structures are summarized. Moreover, the review includes works that compare or combine both modeling frameworks, since these are also aspects addressed in this thesis. Third, Section 2.2.3 gives an overview over multi-echelon inventory models with dual or multiple sourcing. Here, the available literature is rather limited, which demonstrates that there is room (and also need) for further model developments as provided in this thesis.

In Chapter 3, which is concerned with the first major research question, a single-echelon periodic-review inventory model with two suppliers is considered. Following an introduction of the main assumptions and notations in Section 3.2, Section 3.3 presents several dual-sourcing inventory control policies. First, the computation of the optimal policy is addressed in Section 3.3.2. While for the special case of consecutive lead times, i.e. a lead-time difference of one period between the two suppliers, the optimal policy is known to be the single-index policy, it is shown for offsetting lead times how the optimal policy can be found by using a Markov Decision Process (MDP) formulation. As is directly apparent from the MDP model and also reported in the literature by Veeraraghavan and Scheller-Wolf (2008), for instance, the optimal policy can only be computed for limited problem sizes in a reasonable amount of time. That is why in Section 3.3.3 several simpler and, in general, non-optimal policies are outlined. These include the single-index (SIP), constant-order (COP), dual-index (DIP), and the order-splitting policy (OSP). These policies have already been studied in the literature by different authors; however, mostly in isolation. In order to foster the understanding of the policy differences in view of the policy comparison in Section 3.4, their mode of operation and the major available results are reiterated in Section 3.3.3 using a unified notational framework. It is indicated at certain points where known results are complemented by new aspects. Each policy is studied in a backorder-cost model. Extensions to model formulations with different types of service-level constraints are presented, as well. It is shown how the policy parameters for these policies can be optimized. Except for special cases, which are addressed in a subsection of the respective policy section, the optimization can be performed by a one-dimensional search procedure over the relevant policy parameter region. In case of the COP and DIP a so-called stationary overshoot distribution needs to be derived. Exact and approximate approaches of how this can be done are outlined for these policies.

Section 3.4 is based on Klosterhalfen et al. (2010a) and is concerned with the comparison of the (non-optimal) dual-sourcing policies. The comparison focuses on the COP and DIP only. While the SIP and DIP have already been compared in Scheller-Wolf et al. (2007), the OSP is arguably inferior to the other policies in the deterministic lead-time setting analyzed in this chapter. The theoretical findings, which can be derived from the extreme strategies of the COP and DIP, suggest that the cost difference between the policies decreases as the lead-time difference increases. For a sufficiently large lead-time difference it can be presumed that the COP outperforms the DIP (Section 3.4.2). In order to support this presumption, a numerical study is conducted in Section 3.4.3. The numerical results confirm the finding. In settings with a significant lead-time difference and small expediting premium the COP is identified as an effective dual-sourcing policy alternative to the DIP. In general, however, the DIP shows a superior performance, but is also the more complex policy to manage. In situations with a small lead-time difference and large expediting premium, single sourcing is found to be a reasonable alternative to the DIP. From a practitioner's point of view, the outcome of the COP-DIP comparison is particularly interesting for two reasons. First, the COP is the more easily implementable and controllable policy in practice. Second, the guarantee of a constant order for one of the suppliers is helpful for supply negotiations. It prevents the supplier from any demand fluctuations or even the bullwhip effect, which facilitates the production planning.

Chapter 4 shifts the focus to multi-echelon inventory models and addresses the second major research question. First, the two main modeling frameworks in this body of literature are outlined in Section 4.2, i.e. the stochastic-service (SS) and guaranteed-service (GS) approach. In addition to the presentation of the existing models and results, the major point of criticism of the GS approach is addressed. This approach assumes that further countermeasures besides safety stock exist to cope with demand variability, if it exceeds a certain normal or reasonable level. This 'operating flexibility' is not explicitly modeled in the framework, however, and the inventory optimization only takes into account normal demand variability. In this section, which is partly based on Klosterhalfen and Minner (2010), the standard GS model is extended to explicitly consider the effect that operating flexibility measures have on the material flow in the system. One possible modeling option is analyzed in detail and its reasonability is tested in a simulation study. This option assumes that missing items are expedited from the pipeline inventory of a stage. Due to the model extension, the optimization problem minimizes the on-hand stock cost in the entire supply network in its objective function, in contrast to the safety stock cost in the standard GS formulation. Based on different arguments, Minner (1997) and Minner (2000) also suggest the use of the on-hand stock expression in the objective function rather than the safety stock one. Moreover, the extended GS model developed in this section permits a cost-based derivation of the maximum reasonable demand level, up to which demand variability is covered by safety stock only. This level can be expressed as an internal service level and easily determined by a closed-form expression provided that a cost parameter for the use of operating flexibility is available. The specification of such a cost might often be easier for management than the direct setting of a service level.

In Section 4.3, which is also partly based on Klosterhalfen and Minner (2010), both multi-echelon inventory optimization approaches are compared. For each approach, an individual benefit is identified on the basis of theoretical considerations. The SS approach possesses the allocation benefit, whereas the GS approach takes advantage of the decoupling benefit. The SS approach can make its stock allocation decision according to the holding-cost relationships between the stages, taking into account the final-stage service level(s) only. The GS approach, on the other hand, has to comply with the service level (internal or external) of each stage that holds stock. On the upside, however, the operating flexibility measures allow for a decoupling of the stages, i.e. no stochastic delays occur in case of supply shortages, which otherwise would increase the stock requirement at downstream stages. In order to gain further insights into the relative performance of both approaches a numerical study is conducted for serial and divergent systems. Internal and external service levels of the α-service level type are considered with internal service levels ranging from 17% to 75% and external service levels of 85%, 95%, and 99%. From this study, the following three important drivers of the advantage of one approach over the other are derived: processing-time pattern, final-stage service level(s), and internal service level (or operating flexibility cost). In compliance with the individual benefits, the GS approach is found to be superior to the SS model in settings with a degressive processing-time pattern, high final-stage service level(s), and a low internal service level. For the SS model, the opposite is true. Among the three drivers, the internal service-level parameter has the biggest influence on the approach superiority. Most importantly, the results of the numerical study show that none of the approaches is superior to the other, in general. Both approaches have their advantages and disadvantages in certain settings. Hence, the integration of dual sourcing into any of the two approaches represents a valuable extension. One such extension, namely the integration into the GS approach, is the focus of Section 4.5.

Before the dual-sourcing extension is addressed, the joint exploitation of both individual benefits is at the core of Section 4.4, where the SS and GS approaches are combined in the so-called hybrid-service (HS) approach. This approach itself represents a candidate for a potential extension to incorporate dual sourcing, which is postponed to future research, however. The section is based on Dittmar et al. (2009). The HS approach allows the entire network to consist of both SS and GS subnetworks. This makes the appropriate modeling of the subnetwork interfaces an important issue of this section. Moreover, for serial systems the optimization problem is formulated and a pseudo-polynomial time dynamic programming algorithm for the determination of the optimal network partitioning and stock sizing is developed. Extensions to divergent and convergent systems are also discussed. The major contribution of the HS approach is that a practitioner does not have to choose one of the two multi-echelon inventory optimization approaches exclusively for the entire supply network, but (s)he can make a stage-wise choice. Thus, at least the better of the two pure approaches is selected and in some instances even additional cost-savings can be realized through a hybrid-service structure. This finding is also supported by the results of a numerical study conducted for serial systems with up to five stages. The largest additional cost-saving amounts to 10.5% and the average to 1.9%. The best HS performance is observed in settings with relatively low internal service levels, a broad internal service-level range, degressive lead-time structure, and progressive holding-cost pattern.

Section 4.5 focuses on the integration of dual sourcing into the standard GS model and is based on Klosterhalfen et al. (2010b). The dual-sourcing extension of the other two multi-echelon frameworks (SS, HS) is postponed to future research. Note that even the GS approach with only a single supplier for each item can be interpreted as a kind of dual-sourcing model. Due to the assumed operating flexibility, items can be speeded up in case of an imminent stockout, which can be regarded as a second supply option with a shorter processing time. This option, however, is not regarded as dual sourcing in the way this term is understood in this thesis. In order to keep the dual-sourcing model analysis analytically tractable, the order-splitting policy (OSP) is selected as inventory control policy. Moreover, one of the policy parameters, namely the sourcing fraction, is assumed to be exogenous to the model. The model objective is to determine the optimal safety stock allocation and sizing. Extensions of this model to more than two suppliers or the simultaneous optimization of the sourcing fractions and safety stocks are discussed, as well. Moreover, the integration of other inventory control policies like the SIP, COP, or DIP is addressed. In the model development it is shown that certain changes to the standard GS approach with single sourcing are required. In the single-sourcing situation, each stockpoint is preceded by a single process, which allows for an aggregation of the process and the stockpoint into a stage with a single index. In the dual-sourcing setting, several processes can precede a stockpoint depending on the number of suppliers. This prohibits an aggregation into a stage with a single processing time. Instead of the stockpoint and its index, the process needs to be assigned to the arc connecting two stockpoints. Otherwise, differing processing times of the two supply processes cannot be accurately reflected and an exact computation of the safety stock at the dual-sourced stockpoint is not possible. A dynamic programming algorithm is developed for the optimization of the safety stocks in serial and convergent systems. It is shown that this approach represents an improvement of the only approximate modeling idea outlined in the final section of Graves and Willems (2005), which is one of very few contributions available in the literature that address a similar problem.

Chapter 5 concludes the thesis. It summarizes the major findings and discusses implications and extensions for future research.

2 Fundamentals and literature review

The goal of this chapter is to provide the reader with the basic terminology of single- and multi-echelon inventory control theory as well as an understanding of an elementary inventory control model that forms the basis of the upcoming chapters. Furthermore, it reviews the relevant literature related to this thesis.

2.1 Fundamentals

2.1.1 Demand

Companies operate in uncertain environments. Apart from uncertainties on the supply side resulting from possible vehicle or machine break downs or production rescheduling, for instance, a major difficulty arises on the demand side, because future customer orders cannot be predicted exactly (see, e.g., Simchi-Levi et al. (2008)). The latter is the source of uncertainty considered in this thesis. One way to still enable smooth operation and provide a high level of service/product availability is the introduction of inventory buffers, called safety stocks. Other measures include safety lead times or additional capacities, for instance. Yet another way is the use of additional suppliers, who offer fast service in emergency situations, which in turn also reduces the stock requirement, but causes higher procurement costs. Safety stocks and dual sourcing as countermeasures are analyzed in this thesis.

With regard to the incorporation of the demand into an analytical inventory control model, two approaches can be distinguished. Given a sample of demand data, the empirical demand distribution based on these data can be directly used in the model. Alternatively, the parameters of a theoretical distribution can be estimated from the available data. The latter approach is widely used in the inventory control literature, because it smoothes the tasks of analysis and calculation and thus enables the derivation of solution properties. This approach is also pursued in this thesis. The theoretical demand distributions used are briefly outlined in this section. The description is mainly based on Chapter 6 in Law and Kelton (2000). The probability density (pdf) or mass function (pmf) is represented by f for continuous or discrete random variables, respectively. The cumulative distribution function (cdf) is indicated by F. f_L and F_L denote their L-fold convolutions, respectively. Different superscripts are used to refer to the different distribution types.

2.1.1.1 Continuous distributions

Normal distribution

The probably most commonly used demand distribution in inventory theory is the normal distribution. It is characterized by two parameters, the demand expectation μ and standard deviation σ. The probability density function is given as

$$f^{norm}(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad -\infty \leq x \leq \infty \quad . \tag{2.1}$$

The corresponding cumulative distribution function is

$$F^{norm}(x) = \int_{-\infty}^{x} f^{norm}(u)\; du \quad . \tag{2.2}$$

For numerical computations any normal distribution can be transformed into the standard normal distribution with $\mu = 0$ and $\sigma = 1$ by substituting $z := \frac{x-\mu}{\sigma}$. The standard normal probability density, $\phi(z)$, and cumulative distribution function, $\Phi(z)$, are given as

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}} \tag{2.3}$$

$$\Phi(z) = \int_{-\infty}^{z} \phi(u)\; du \quad . \tag{2.4}$$

The standard normal distribution values are tabulated such that the computation of integral expressions is possible. Moreover, rational approximations exist for $\Phi(z)$ (see, e.g., Abramowitz and Stegun (1970), p. 932).

Note that, when using a normal distribution to model customer demand, negative demand values have a positive probability. If the probability mass for these negative values is small, however, this still represents a reasonable approximation. This is usually the case, if the coefficient of variation, $CV = \sigma/\mu$, is smaller than 0.5 (see, e.g., Schneider (1981)). In various settings for fast moving items, where the demand per period is relatively large, a reasonable goodness of fit of the normal distribution is reported (see, e.g., Tijms and Groenevelt (1984)).

Since ordered items are not received immediately, but after a lead time, during which further demands need to be satisfied, the demand distribution over the lead-time is relevant for inventory control. For a deterministic lead time L, the distribution is the L-fold convolution of the single period demand random variable, if the demand process is assumed to be stationary and the single period demands are identically and independently distributed ($i.i.d.$). Consequently, in case the single period demand has a normal distribution, the lead-time demand is also normally distributed with an expected value of $\mu \cdot L$ and standard deviation $\sigma \cdot \sqrt{L}$. For the lead-time demand computation in case of stochastic lead times see, e.g., Tijms and Groenevelt (1984) and Eppen and Martin (1988).

Gamma distribution

The normal distribution disadvantage of possible negative values, which is especially critical for large coefficients of variation, induces Burgin (1975) to propose the gamma distribution for inventory control. Other contributions like Tyworth et al. (1996) also assume gamma distributed demand. The gamma distribution is only defined for non-negative values. The probability density and cumulative distribution function are given as

$$f^{gam}(x) = \begin{cases} \frac{\theta^{\kappa} x^{\kappa-1} e^{-\theta x}}{\Gamma(\kappa,\theta)} & x \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.5}$$

$$F^{gam}(x) = \int_0^x f^{gam}(u)\ du \tag{2.6}$$

where $\theta > 0$ is a scale parameter and $\kappa > 0$ the shape parameter (modulus). The term $\Gamma(\kappa, \theta)$ is the complete gamma function

$$\Gamma(\kappa, \theta) = \int_0^\infty \theta^\kappa x^{\kappa-1} e^{-\kappa x} \, dx \tag{2.7}$$

ensuring that $\int_0^\infty f^{gam}(x) \, dx = 1$. If κ is an integer, then $\Gamma(\kappa, \theta) = (\kappa - 1)!$. The mean and variance are

$$\mu = \frac{\kappa}{\theta} \tag{2.8}$$

$$\sigma^2 = \frac{\kappa}{\theta^2} \quad . \tag{2.9}$$

Based on the first two moments of some observed data, the scale and shape parameter can be determined as

$$\kappa = \frac{\mu^2}{\sigma^2} \tag{2.10}$$

$$\theta = \frac{\mu}{\sigma^2} \quad . \tag{2.11}$$

Given that the period demand is *i.i.d.* gamma distributed with κ and θ, the lead-time demand for a deterministic lead time of L periods has shape parameter $L \cdot \kappa$ and scale parameter θ, i.e.

$$f_L^{gam}(x) = \frac{\theta^{L\kappa} x^{L\kappa-1} e^{-\theta x}}{\Gamma(L\kappa, \theta)} \quad . \tag{2.12}$$

For the computation of integral expressions the use of tables or rational approximations is required again (see, e.g., Abramowitz and Stegun (1970), p. 257).

2.1.1.2 Discrete distributions

Discrete demand is well modeled by the following three distribution types, which provide a range of shapes that satisfy a variety of demand patterns encountered in practice (see Banks et al. (2009), p. 182).

Poisson distribution

In situations where item demand is rather low, the use of the Poisson distribution is very common in inventory theory for several reasons (see, e.g., Zipkin (2000), p. 179). The distribution is easy to specify, because it has only one parameter λ. Further, in many situations the model is shown to be fairly accurate. Finally, its mathematical simplicity facilitates analytical calculations. The probability mass and cumulative distribution function are defined as

$$f^{Pois}(x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!} & x = 0, 1, ... \\ 0 & \text{otherwise} \end{cases} \tag{2.13}$$

$$F^{Pois}(x) = \begin{cases} e^{-\lambda} \cdot \sum_{i=0}^{x} \frac{\lambda^i}{i!} & x = 0, 1, ... \\ 0 & \text{otherwise} \end{cases} \quad . \tag{2.14}$$

The mean and variance are

$$\mu = \sigma^2 = \lambda \quad . \tag{2.15}$$

The sum of $i = 1, 2..., m$ independent Poisson random variables with parameters λ_i is Poisson distributed with parameter $\lambda = \sum_{i=1}^{m} \lambda_i$. For the lead-time demand random variable it follows that, if period demands are identically and independently distributed according to a Poisson distribution with parameter λ_1, the lead-time random variable for a deterministic lead time of L periods has a Poisson distribution with parameter $L \cdot \lambda_1$.

Geometric distribution

Besides the Poisson distribution, the geometric distribution is used in inventory theory to model demand (see, e.g., Beckmann (1964)). Due to its recursive probability structure (see (2.20)) it lends itself to a more thorough analytical analysis in some cases. The probability mass and cumulative distribution function of the geometric distribution are defined as

$$f^{geom}(x) = \begin{cases} p(1-p)^x & x = 0, 1, ... \\ 0 & \text{otherwise} \end{cases} \tag{2.16}$$

$$F^{geom}(x) = \begin{cases} 1-(1-p)^{x+1} & x = 0, 1, ... \\ 0 & \text{otherwise} \end{cases} \tag{2.17}$$

with $0 < p \leq 1$. The mean and variance are given as

$$\mu = \frac{1-p}{p} \tag{2.18}$$

$$\sigma^2 = \frac{1-p}{p^2} \quad . \tag{2.19}$$

In addition, the following recursion holds

$$f^{geom}(x) = (1-p) \cdot f^{geom}(x-1) \quad . \tag{2.20}$$

The sum of L independent geometrically distributed random variables with parameter p follows a negative binomial distribution with parameters L and p (cf. Law and Kelton (2000)). (The geometric distribution can be viewed as a special case of a negative binomial distribution with $L = 1$ and the same value for p.) In terms of the demand, this means that, if the single-period demand has a geometric distribution, the lead-time demand random variable has a negative binomial distribution.

Negative binomial distribution

In situations with low but highly variable item demand, e.g., for certain service parts (see Muckstadt (2005)) or in retailing (see Agrawal and Smith (1996)), the Poisson distribution might not fit well, because its fixed variance to mean ratio of one is too small. Here, the use of the negative binomial distribution for modeling demand is appropriate. The negative binomial distribution has two parameters r (a positive integer) and $0 < p < 1$. Its probability mass and cumulative distribution function are defined as

$$f^{nbin}(x) = \begin{cases} \binom{r+x-1}{x} p^r (1-p)^x & x = 0, 1, ... \\ 0 & \text{otherwise} \end{cases} \tag{2.21}$$

$$F^{nbin}(x) = \begin{cases} \sum_{i=0}^{x} \binom{r+i-1}{i} p^r (1-p)^i & x = 0, 1, ... \\ 0 & \text{otherwise} \end{cases} \quad . \tag{2.22}$$

The mean and variance are

$$\mu = r \cdot \frac{1-p}{p} \tag{2.23}$$

$$\sigma^2 = r \cdot \frac{1-p}{p^2} \quad . \tag{2.24}$$

For the computation of the lead-time demand random variable it is important to note that the sum of independent negative-binomially distributed random variables with the same value of the parameter p but the 'r-values' r_1 and r_2 is negative-binomially distributed with the same p but with 'r-value' $r_1 + r_2$ (cf. Law and Kelton (2000)). Consequently, if the single-period demands are identically and independently distributed according to a negative binomial distribution with parameters p and r, the lead-time demand random variable for a deterministic lead time of L periods has negative binomial distribution with parameters p and $L \cdot r$.

2.1.1.3 Discretized (continuous) distribution

Although the fit of a continuous distribution to the empirical data might be good, some mathematical models require a discrete demand distribution for computational reasons. Whenever a discretized distribution is referred to in this thesis, the following kind of distribution is meant. Let F denote the single-period cumulative distribution function of a demand random variable D with positive but unlimited support, e.g., a gamma distribution. (In case of the normal distribution, negative values would be neglected and their probability mass cumulated at zero.) Further, specify $\bar{D}$ as a very large number such that

$$Pr\left\{D > \bar{D}\right\} \leq \epsilon \tag{2.25}$$

where ϵ is a very small number, e.g., $\epsilon = 0.00001$. Then, the mass probability function is given as

$$f(x) = \begin{cases} F(0.5) & x = 0 \\ F(x+0.5) - F(x-0.5) & x = 1, 2, ...\bar{D}-1 \\ 1 - F\left(\bar{D} - 0.5\right) & x = \bar{D} \end{cases} . \tag{2.26}$$

2.1.2 Inventory control

2.1.2.1 Inventory level classifications

In probabilistic demand settings, the following terms are used for conceptually classifying inventories. The explanations are based on Silver et al. (1998).

On-hand stock, OH

This term describes the stock quantity that is physically on the shelf and is available for directly satisfying customer demand. The on-hand stock can never be negative. In connection with a period index t, OH_t denotes the on-hand stock at the end of period t before deliveries. Consequently, the quantity of available items at the beginning of period t (before any orders are received) corresponds to OH_{t-1}.

Backorders, BO

In case demand in a period exceeds the available stock, a shortage occurs. Provided that customers are willing to wait for their products, the backorders represent the quantity of items that have already been requested, but are still to be delivered (*backorder case*). If customers do not wait, the shortage quantity is lost (*lost-sales case*). If a period index t is introduced, BO_t refers to the backorders at the end of period t.

Net stock, NS

The net stock is defined as the difference between the on-hand stock and the backorders. Obviously, it can become negative. For a given period t, the net stock at the end of this period is

$$NS_t = OH_t - BO_t \quad . \tag{2.27}$$

Pipeline inventory, PI

The pipeline inventory denotes the outstanding orders, i.e. the quantity of items, for which an order has already been placed, but has not yet been received. Together with the period index t, PI_t specifies the outstanding orders at the beginning of a period, before any of these items arrive in stock at a location.

$$PI_t = \sum_{i=1}^{L} Q_{t-i} \tag{2.28}$$

where L denotes the replenishment lead time and Q_t the order placed in period t.

Inventory position, IP

The inventory position at the beginning of a period t before ordering and receipt of any order is calculated as follows:

$$IP_t = OH_{t-1} + PI_t - BO_{t-1} \quad . \tag{2.29}$$

The inventory position comprises of the relevant information to trigger an order, because it also includes the stock on order. Compared to a situation where net stock is used as a trigger, this avoids the ordering of materials today, for which an order has already been placed and which are due in tomorrow.

Safety stock, SST

'The safety (or buffer) stock is defined as the average level of the net stock just before a replenishment arrives.' (Silver et al. (1998), p. 235)

$$SST = \mathbb{E}[NS] \tag{2.30}$$

If demand during the replenishment lead time is larger than average, a positive safety stock provides coverage against possible stockouts resulting from this fact. The numerical value of the safety stock depends on whether customer demands that occur during a stockout period are backordered or lost. If demands are lost, the net stock remains zero throughout the stockout period. In the backordering case, however, net stock is negative just before the next replenishment arrives. Since safety stock represents the average net stock just before a replenishment arrives, its value also depends on which kind of assumption applies.

2.1.2.2 Performance measures for inventory control

An inventory control system or policy manages the inventory level at a location by providing answers to the following three questions (see Silver et al. (1998), p. 235):

1. How often should the inventory status be determined?
2. When should a replenishment order be placed?
3. How large should the replenishment order be?

The performance of an inventory control system can be measured either in terms of cost or service. Under a *cost performance measure*, the objective is to find control parameters that minimize the sum of ordering, holding, and stockout penalty/backorder costs (see Minner (2000), p. 30). However, in many practical situations backorder costs are generally hard to quantify. To overcome this difficulty, a *service performance measure* can be introduced such that the objective of the inventory control system is to achieve a predefined service level with minimal holding costs. van Houtum and Zijm (2000) show that for a variety of models a one-to-one relationship between cost models and service models exists. Three common measures of service are the α-, β-, and γ-service levels (also known as P_1-, P_2-, P_3-service measures or the non-stockout probability, fill rate, and ready rate or modified fill rate, see Silver et al. (1998)). These service measures can be related to different time intervals, e.g., average period, average replenishment cycle, or lead time. Since customers are usually only interested in the quality of demand satisfaction in every period and do not care about the order cycle, solely period-based service definitions are presented in the following paragraphs.

α-service level

In situations where only the occurrence of a stockout is important and not the quantity and duration of the shortage, the α-service level should be used. This service level is defined as the 'probability of satisfying demand in an arbitrary period' (Klemm (1973), p. 170). In the spirit of the β-service level description of Chen et al. (2003) and Thomas (2005), which follows in the next paragraph, the random variable defining the non-stockout probability for T periods is

$$\alpha_T \equiv \frac{\sum_{t=1}^{T} 1 \cdot \mathbf{I}\left\{FD_t = D_t\right\}}{T} \tag{2.31}$$

where D_t denotes the demand random variable of period t (assuming non-negativity) and FD_t the filled demand, i.e. the number of units of the demand in period t that can be satisfied from stock. Furthermore, $\mathbf{I}\{x\}$ denotes the indicator function of event x. Equivalently, in a backorder setting, α_T can be seen as the random variable indicating the probability that the net stock at the end of a period is non-negative, i.e.

$$\alpha_T \equiv \frac{\sum_{t=1}^{T} 1 \cdot \mathbf{I}\left\{NS_t \geq 0\right\}}{T} \quad . \tag{2.32}$$

For the infinite-horizon case $T \to \infty$ under backordering, it follows that

$$\alpha = \lim_{T \to \infty} \mathbb{E}[\alpha_T] \equiv Pr\left\{NS \geq 0\right\} \tag{2.33}$$

where NS denotes the net stock random variable.

β-service level

The β-service level is defined as the fraction of demand satisfied directly from stock (see, e.g., Silver et al. (1998), p. 245). Using D_t and FD_t as defined in the α-service level case, the fill-rate random variable for T periods is (see, e.g., Thomas (2005))

$$\beta_T \equiv \frac{FD_1 + \cdots + FD_T}{D_1 + \cdots + D_T} \quad . \tag{2.34}$$

In the infinite-horizon case $T \to \infty$ and assuming *i.i.d.* period demands,

$$\beta = \lim_{T\to\infty} \mathbb{E}[\beta_T] \equiv \frac{\mathbb{E}[FD]}{\mu} \tag{2.35}$$

where FD denotes the filled demand random variable. Assuming a backorder situation, this expression can be rewritten using the 'expected units short' (see, e.g., Silver and Bischak (2010)):

$$\beta = 1 - \frac{\text{expected units short per period}}{\text{expected demand per period}} = 1 - \frac{\mathbb{E}\left[BO\right] - \mathbb{E}\left[BO^{beg}\right]}{\mu} \quad . \tag{2.36}$$

In (2.36), $\mathbb{E}[BO]$ indicates the expected backorders at the end of a period, whereas $\mathbb{E}\left[BO^{beg}\right]$ the ones at the beginning of a period (after outstanding orders have been received and existing backorders have been satisfied as far as being feasible).

Chen et al. (2003) establish the following interesting result for a periodic-review order-up-to S model with a deterministic lead time and backordering (see Section 2.1.2.3 for details on this kind of inventory model):

$$\mathbb{E}\left[\beta_1\right](S) \geq \mathbb{E}\left[\beta_T\right](S) \geq \lim_{t\to\infty} \mathbb{E}\left[\beta_T\right](S) \quad , \tag{2.37}$$

i.e. the expected finite-horizon fill rate is greater than the infinite-horizon fill rate and less than the single-period expected fill rate.

γ-service level

Whereas the β-service level only takes into account new shortages in a period, the γ-service level considers the entire backorders (or cumulative shortages) at the end of a period (see Schneider (1981), p. 617). Therefore, this service measure is only relevant in the backorder case, where it provides a lower bound for the β-service level. Due to this relation and its simpler way of computing, the γ-service level is often used as an approximation for the β-service level. For the infinite-horizon case, it is defined as:

$$\gamma = 1 - \frac{\text{expected cumulative units short per period}}{\text{expected demand per period}} = 1 - \frac{\mathbb{E}[BO]}{\mu} \quad . \tag{2.38}$$

For the upcoming exposition it is sometimes convenient to express the service-level constraint in terms of a constraint on the expected backorders. For a given γ-service level target, γ^{target}, (2.38) can be reformulated into

$$BO^{\gamma}_{target} = \mathbb{E}[BO] \tag{2.39}$$

where BO^{γ}_{target} denotes the maximum permissible value given as

$$BO^{\gamma}_{target} = \left(1 - \gamma^{target}\right)\mu \quad . \tag{2.40}$$

In case of a finite horizon of T periods, it follows from (2.38) that the modified fill-rate random variable γ_T is

$$\gamma_T \equiv 1 - \frac{BO_1 + \cdots + BO_T}{D_1 + \cdots + D_T} \quad . \tag{2.41}$$

Note that whenever there is reference to a service level in this thesis, it applies to the infinite-horizon definition.

2.1.2.3 Single-echelon order-up-to level model with single sourcing

As a preliminary for the upcoming analyses that refer to dual-sourcing and multi-echelon inventory models, an elementary single-sourcing single-echelon stochastic inventory control model with periodic review and backordering is presented in this section. In settings with linear holding (h) and backorder costs (b) per unit and period, where fixed ordering costs are zero or negligible, an order-up-to level policy represents the optimal inventory control strategy (see, e.g., Veinott (1966)). Under such a policy the inventory position is checked at each review instant and, if necessary, an order Q_t is placed to raise it up to B, the order-up-to level. The inventory position is defined as the net stock at the end of the previous period plus all outstanding orders. Under the assumption that the review period is equal to one and it takes L periods for an order to arrive, the inventory position at the beginning of period t before ordering is

$$IP_t = NS_{t-1} + \sum_{i=1}^{L} Q_{t-i} \tag{2.42}$$

and the inventory position recursion is

$$IP_t = IP_{t-1} + Q_{t-1} - d_{t-1} = B - d_{t-1} \tag{2.43}$$

where d_{t-1} denotes the demand realization in period $t-1$. Due to the order-up-to structure, the inventory position after ordering in each period, $IP^+_{t-1} = IP_{t-1} + Q_{t-1}$, is equal to B. Hence, the order quantity in period t, $Q_t = (B - IP_t)^+ = d_{t-1}$, i.e. it corresponds to the demand of the previous period. Consequently, the net stock at the end of period t is

$$NS_t = IP^+_{t-L} - \sum_{i=0}^{L} d_{t-i} = B - \sum_{i=0}^{L} d_{t-i} \quad . \tag{2.44}$$

Under stationary conditions $t \to \infty$, the net stock is a random variable given as

$$NS = B - D(L+1) \tag{2.45}$$

where $D(L+1)$ denotes the demand random variable over $L+1$ periods. The expected on-hand stock and backorders are

continuous demand:

$$\mathbb{E}\left[OH(B)\right] = \int_{x=0}^{B} (B - x) \cdot f_{L+1}(x)\; dx \tag{2.46}$$

$$\mathbb{E}\left[BO(B)\right] = \int_{x=B}^{\infty} (x - B) \cdot f_{L+1}(x)\; dx \tag{2.47}$$

$$\textit{discrete demand:} \qquad \mathbb{E}\left[OH(B)\right] = \sum_{x=0}^{B} (B-x) \cdot f_{L+1}(x) \tag{2.48}$$

$$\mathbb{E}\left[BO(B)\right] = \sum_{x=B+1}^{\infty} (x-B) \cdot f_{L+1}(x) \quad , \tag{2.49}$$

respectively. In case of continuous (discrete) demand and the infinite horizon average cost criterion, the optimal B is the (smallest) B that satisfies the critical-fractile (in)equality:

$$F_{L+1}(B) \geq \frac{b}{b+h} \quad . \tag{2.50}$$

If instead of a backorder cost b per unit and period a service-level constraint is used, the optimal order-up-to level can be determined as follows. For the *α-service level* it is found from (2.33) and (2.45) that

$$\alpha = Pr\left\{B - D(L+1) \geq 0\right\} = Pr\left\{D(L+1) \leq B\right\} = F_{L+1}(B) \quad . \tag{2.51}$$

Comparing (2.50) and (2.51) shows that the following equivalence relation exists between the backorder cost and α-service-level approach, i.e.

$$\alpha = \frac{b}{b+h} \quad . \tag{2.52}$$

For the *β-service level* computation, the expected backorders at the beginning and the end of an arbitrary period are required. The latter ones are given by (2.47) or (2.49). The former ones can simply be calculated as the expected backorders over L periods (instead of $L+1$). Consequently,

$$\textit{continuous demand:} \quad \beta = 1 - \frac{\int_B^\infty (x-B) f_{L+1}(x)\ dx - \int_B^\infty (x-B) f_L(x)\ dx}{\mu} \tag{2.53}$$

$$\textit{discrete demand:} \quad \beta = 1 - \frac{\sum_{x=B+1}^\infty (x-B) f_{L+1}(x) - \sum_{x=B+1}^\infty (x-B) f_L(x)}{\mu} \quad . \tag{2.54}$$

Similarly, the γ-*service level* equations are

$$\textit{continuous demand:} \quad \gamma = 1 - \frac{\int_B^\infty (x-B) f_{L+1}(x)\ dx}{\mu} \tag{2.55}$$

$$\textit{discrete demand:} \quad \gamma = 1 - \frac{\sum_{x=B+1}^\infty (x-B) f_{L+1}(x)}{\mu} \quad . \tag{2.56}$$

2.1.3 Supply network modeling

2.1.3.1 Stockpoint, process, and stage

In this thesis, a (potential) location for holding inventory of an item is referred to as a stockpoint and graphically represented by an upside-down triangle. At each stockpoint only one specific item can be stocked, e.g., a raw material, component, or finished product. Before an item enters a stockpoint, it has to pass through a process, e.g., the manufacture of a subassembly or the transportation of an item from another stockpoint to this stockpoint. Such a process (together with the item availability at preceding stockpoints in a multi-echelon setting) determines the replenishment lead time of a stockpoint and is visualized by a circle. In a single-sourcing setting, a stockpoint and its preceding process can be combined and jointly represented by a stage (see Figure 2.1). Since each item is only sourced from a single supplier, there is only a single process preceding a stockpoint, which can then be associated directly with the stockpoint and its index (see, e.g., Graves and Willems (2000)).

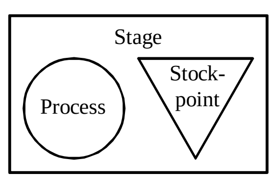


Figure 2.1: Stage structure

In a dual- or multiple-sourcing setting, an item can be delivered by several suppliers. Depending on the geographic distance, the transportation processes might take up different amounts of time. That means, a stockpoint can have several preceding processes depending on the number of suppliers. Consequently, an integration into a stage with a single processing time is not possible any more. Instead of associating the processing time with the stockpoint (and its index) it needs to be associated with the arc between the specific supplier and the stockpoint. This distinction becomes relevant in Section 4.5 where further details are provided.

2.1.3.2 Network structures

If the inventory model not only refers to a single stage (or stockpoint), but to multiple stages, which are linked with each other through supply-demand relationships, it is called a multi-stage or multi-echelon inventory model. The stages form a supply network, i.e. a directed graph where the nodes depict the stages and the arcs represent the supply-demand relationships (see Zipkin (2000), p. 108). Chapter 4 deals with multi-echelon inventory optimization under centralized control and single- and dual-sourcing aspects for different types of supply networks. The basic network structures are the following.

Serial system

The simplest way of linking several stages represents a serial system. Such a system consists of n stages where each stage supplies the next downstream one with its item. Only, the first (most upstream) stage is supplied by an external supplier and the most downstream stage faces external customer demand for the finished product (see Zipkin (2000), p. 108). In a serial system, each stage has a single direct predecessor and successor. For the upcoming exposition it is useful to assign a level code to each stage. Whereas this is less relevant in the serial system case, since there is only one stage on each level, it is of great importance for the other more complex structures. A practical example of this type of system can be found in the chemical industry, for instance, where a product passes through several consecutive chemical reaction processes. In other industries this system is of importance, if some level of aggregation is applied, i.e. if not each assembled part is modeled in detail. From an academic point of view the analysis of this system structure is a good starting point before investigating more complex ones.

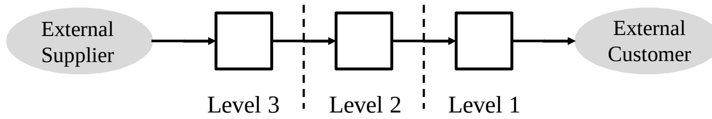


Figure 2.2: Serial system

Divergent system

Similar to a serial system, a divergent or distribution system is a network, in which there is a single most upstream stage that receives external supply. However, now several stages exist that supply external customers. The stages in such a system can be interpreted as warehouses where, e.g., a central warehouse supplies regional warehouses which, in turn, feed retail outlets. In terms of a production network, one can think of a raw material that is specialized into several products as it moves through the system (see Zipkin (2000), p. 109). The distinguishing feature of a divergent network is that each stage has only one direct predecessor, but can have several direct successors (see Muckstadt and Roundy (1993), p. 81).

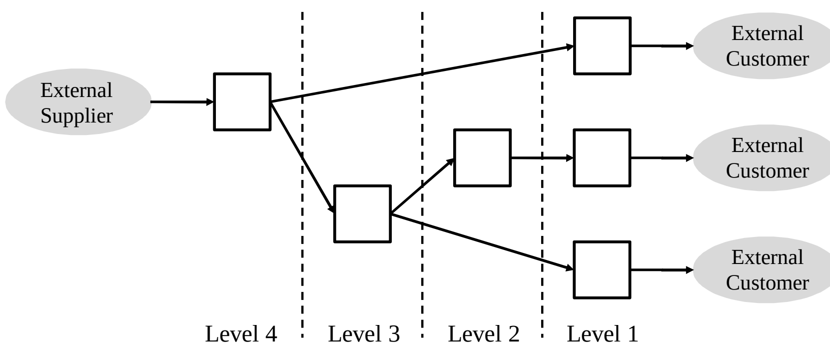


Figure 2.3: Divergent system

Convergent system

In a convergent or assembly system a single finished product is assembled from several components. These components, in turn, may be manufactured using several raw materials. Hence, a convergent network is characterized by the fact that each stage has at most one direct successor, but may have more than one direct predecessor (see Federgruen (1993), p. 144.). As before, all stages on the most upstream level receive items from external suppliers and the stage on the most downstream level meets external customer demand (see Zipkin (2000), p. 109).

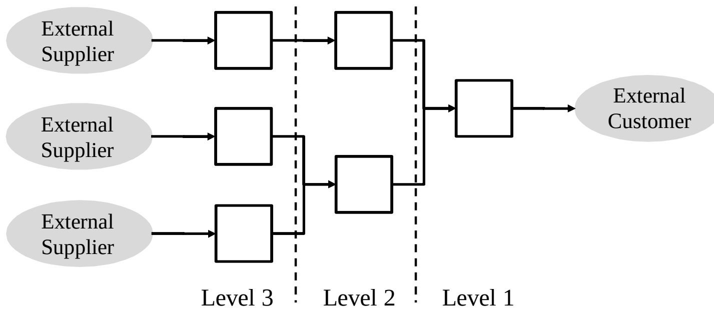


Figure 2.4: Convergent system

(Spanning-)Tree system

A tree system contains features of a convergent system and a divergent system. The first part of the network is roughly characterized by an assembly structure ending in one or more stages. From these stages, distribution structures may continue (see Zipkin (2000), p. 109). Any two stages must not be connected by more than one arc, however. Depending on the level of aggregation, many supply networks in reality exhibit either this kind of structure or a general (acyclic) one (see, e.g., Willems (2008) for real-world instances).

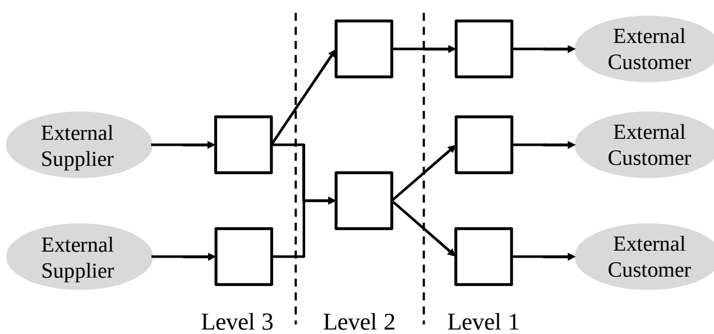


Figure 2.5: Tree system

General (acyclic) system

Finally, a general (acyclic) system relaxes the constraint of the tree system. Any possible links between stages are permitted except for links pointing in a backward direction (see Zipkin (2000), p. 109-110).

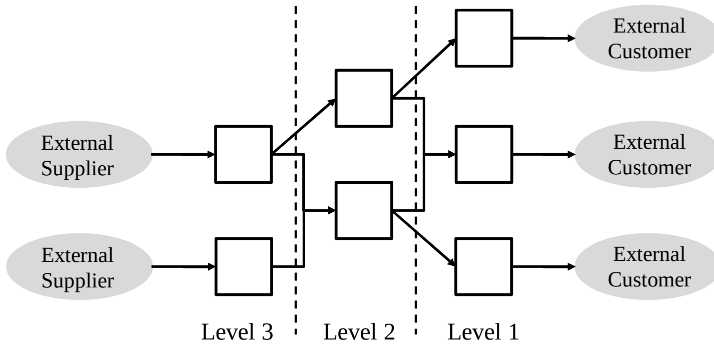


Figure 2.6: General (acyclic) system

2.2 Literature review

The study of multiple sourcing inventory models dates back about 50 years. Those early contributions by Barankin (1961), Daniel (1962), Fukuda (1964), and Whittemore and Saunders (1977), amongst others, focus on the derivation of the optimal policy for dual-sourcing problems in a single-echelon setting. In the subsequent decades, rather few contributions can be found addressing dual-sourcing models. It was not until the turn of the century that there has been a renewed interest in studying this problem induced by a need for replenishment decision support of companies such as Hewlett Packard (Beyer and Ward (2000)) or Caterpillar (Rao et al. (2000)), which rely on dual- or multiple supply modes. Analytical/theoretical results have been used to guide the development of heuristics like the constant-order, single-index, or dual-index policy. Section 2.2.1 presents a survey of the key contributions.

Dual or multiple sourcing in multi-echelon inventory models has rarely been studied. The yet considerable complexity of multi-echelon inventory models stemming from the objective of optimally allocating stocks across the supply network has made researchers focus almost exclusively on single-sourcing settings. The relevant works in this area are reviewed in Section 2.2.2.

Nevertheless some contributions are available in the literature that incorporate multiple sourcing in a multi-echelon context. An overview is provided in Section 2.2.3.

2.2.1 Single-echelon inventory models with multiple sourcing options

According to Minner (2003) the literature on inventory models with multiple sourcing options can be distinguished into two strands. The first is characterized by deterministic lead-time models where the use of multiple suppliers results from the motive of emergency ordering to prevent stockouts. The second is comprised of stochastic lead-time models where the motive for using multiple suppliers is the reduction of the effective lead time by order splitting. An overview over the body of literature on order-splitting models is available in Thomas and Tyworth (2006). This thesis focuses mainly on the first body of literature. Since a comprehensive review of these models up to around 2001 is provided in Minner (2003), only key results before that time are discussed in the following subsections and the main focus lies on relevant research since the turn of the century.

2.2.1.1 Derivation of the optimal policy

Early contributions focus on the structure of the optimal policy for periodic-review inventory systems with dual sourcing. Barankin (1961) studies a single-period problem with a lead time of the fast supplier of 0 and the slow supplier of 1 period. Daniel (1962), Bulinskaya (1964), and Neuts (1964) extend this model to the n-period and infinite horizon case. An extension to the case of an arbitrary (non-negative) deterministic lead time of the fast supplier and a lead-time difference of exactly one period, i.e. consecutive lead times, is presented in Fukuda (1964). The optimal policy in this situation is an order-up-to policy with one inventory position as the order trigger and two order-up-to levels, one for each supplier. This policy is called the single-index policy (SIP) in the upcoming sections.

For several extensions of the consecutive lead-time model, optimal policies can be derived. Yazlali and Erhun (2009) introduce minimum and maximum capacity limits on the orders and show that (what they term) a ‘two-level modified base stock policy’ is optimal without any restrictions on the ordering costs. This policy basically corresponds to a single-index policy, which takes into account the capacity constraints on the orders.

Fixed ordering costs as well as demand forecast updates are included by Sethi et al. (2003). They show that the optimal policies for both fast and slow orders are of an (s, S)-type. Fast orders are based on an inventory position, which includes the slow order issued in the previous period (*fast inventory position*). The slow order is triggered by an inventory position that takes the fast order into account, which is released during the period (*slow inventory position*).

Other contributions that derive the optimal policy in settings with fixed ordering costs but with zero lead times or positive, but identical lead times are Fox et al. (2006) and Yi and Scheller-Wolf (2003). The former paper shows that a reduced form of generalized (s, S)-policy is optimal for both finite and (discounted) infinite-horizon problems, provided that the demand density is strongly unimodal. A density $f(x)$ is strongly unimodal, if f is unimodal and if the convolution of f with any unimodal g is unimodal. The exposition focuses on the lost-sales case, but required modifications for the backorder case are discussed, too. Whereas given deterministic prices are usually assumed, the latter contribution by Yi and Scheller-Wolf (2003) allows for a stochastic price, which is formed at a spot market. In addition, a fixed cost is incurred when something is ordered at the spot market. The authors show that the optimal policy has a structure similar to the (s, S)-policy.

For a general lead-time difference or more than two suppliers with consecutive lead times Whittemore and Saunders (1977) and later Feng et al. (2006a) and Feng et al. (2006b) show that the optimal policy has a highly complex structure. Ordering decisions need to be based not only on a single state variable like the inventory position, but the system needs to keep track of all orders of the lead-time difference horizon individually. Due to this complexity, several simpler policies have been proposed in the literature for such cases, which will be reviewed next.

2.2.1.2 Parameter determination for a given (non-optimal) policy

The complexity of the optimal policy induced researchers to study the optimal parameter determination for given (non-optimal) policies. These policies can be classified by various dimensions. The following two are used here: (i) single- vs. dual-index policies depending on the number of inventory positions that are tracked and (within the former distinction) (ii) single vs. dual base-stock policies depending on the number of order-up-to levels employed by the policy.

The *constant-order policy*, which is a single-index policy with one order-up-to level (for the fast supplier) and one fixed order (for the slow supplier), is studied by Zhang and Hausman (1994) and Janssen and de Kok (1999). While both consider linear inventory costs, the latter one is slightly more general by also allowing a fixed order cost for both suppliers. In each of the two contributions, approximations are used to derive the optimal parameters for this periodic-review policy. A continuous-review version called the Tailored Base Surge policy is analyzed in Allon and van Mieghem (2010), who use a Brownian approximation for the parameter computation.

A *standing-order policy* is similar to the constant-order policy. However, in addition to the order-up-to level, also a dispose down-to-level is specified. This type of policy was first studied by Rosenshine and Obee (1976), who find the optimal parameters by modeling the system as a Markov Chain. While Rosenshine and Obee (1976) only determine the parameters after having predefined the policy structure as mentioned above, Chiang (2007) actually shows that for a predetermined standing-order size, which is larger than the expected period demand, the optimal policy exhibits this structure.

Another single-index policy with two order-up-to levels, one for each supplier, is investigated in Scheller-Wolf et al. (2007). They call it the *single-index policy* for short. For the periodically reviewed dual-sourcing inventory problem with a lead-time difference of one, this policy is optimal (see above). For larger differences they find that it provides reasonably good results. Furthermore, it is shown that the optimal policy parameters can be computed easily when demand distributions are mixtures of Erlang distributions. A continuous-review variant of this policy is considered in Bradley (2004). By using a Brownian approximation, a closed-form expression for one base stock (the slow supplier) and an analytical expression for the other are derived.

A dual-index policy with periodic review and two order-up-to levels referred to as the *dual-index policy* for brevity reasons by Veeraraghavan and Scheller-Wolf (2008) is analyzed in various works. Kiesmüller (2003) proposes the use of such a policy structure in the context of a remanufacturing system, which can be regarded as a special kind of a dual-sourcing model. The key idea in this contribution is to base the order decision with any source only on the information about orders that will arrive no later than the order, which is to be determined. This is exactly mirrored by the dual-index policy. Veeraraghavan and Scheller-Wolf (2008) study this policy in a dual-sourcing context. They provide a separability result, which allows to separate the originally two-dimensional optimization problem into two one-dimensional ones. Optimal parameters are found by a simulation-based optimization procedure due to the difficulty of deriving the stationary overshoot distribution. The overshoot is the quantity, by which the fast inventory position might exceed the order-up-to level of the fast supplier. Instead of using simulation, Arts et al. (2009) show that the stationary overshoot distribution can be efficiently approximated by using a one-dimensional Markov Chain based on limiting results. This approach also extends to a special case of stochastic lead times. Song and Zipkin (2009) study the dual-index policy in a continuous-time framework. They show that the system can be viewed as a network of queues with a state-dependent routing mechanism called an overflow bypass. Closed-form expressions for the policy evaluation and optimization are obtained. Furthermore, they present extensions to stochastic lead times, batch-ordering policies, non-Poisson demand processes, and multiple demand classes.

Recently, due to the relationship between the periodic review dual-sourcing problem and the lost-sales inventory problem established by Sheopuri et al. (2010), policies that show a good performance for the latter problem are transferred to the dual-sourcing problem. These include policies, which use a single index, namely the inventory position that takes into account all outstanding orders, and one order-up-to level plus, as a second decision variable, an allocation parameter that determines in each period which fraction of the total order quantity is sourced from each of the two suppliers. Consequently, these policies resemble an order-splitting policy, where the splitting decision, which is usually made once for all periods, is adjusted each period, however.

2.2.2 Multi-echelon inventory models with single sourcing

Due to the already increased complexity of multi-echelon inventory models (compared to single-echelon ones) caused by the task of the cost-optimal deployment of inventories across the various stages of the supply network, large parts of the literature only consider single sourcing in a multi-echelon context. Based on the two seminal papers by Simpson (1958) and Clark and Scarf (1960) on base-stock/order-up-to level models without lot-sizing, two competing research strands have developed over the years. The difference refers to the assumption made with regard to the role of safety stock and the resulting consequences for the material flow in the system, which coined the terms full-delay and no-delay (van Houtum et al. (1996)) or *stochastic-service* (SS) and *guaranteed-service* (GS) approaches (Graves and Willems (2003)). The SS approach assumes that safety stock is the only buffer against demand variability and thus explicitly takes into account that occasional material shortages at a supplying stage cause delays in the delivery of the material request to the ordering stage. Consequently, the service of a stage is stochastic. The GS framework assumes deterministic or 'guaranteed' service. That means, orders of any size can be met by the supplying stage after the committed service time. This 100% service is achieved by a combination of safety stock and so-called operating flexibility (i.e. some sort of emergency measure) in case of material shortages. How this additional means of flexibility influences the material flow in the system is not explicitly modeled in most of the GS contributions. This problem will be further discussed and resolved in Chapter 4.2.3 of this thesis. In the remainder of the thesis the classification of Graves and Willems (2003) is employed. Over the decades, both approaches have been treated mainly in isolation. Only few contributions have been concerned with the comparison or combination of the two approaches, which will be a major aspect addressed in Chapter 4 of this thesis. In the SS model strand, the focus has first been put on finding the optimal inventory control policy for different network structures. Then, efficient numerical methods to compute the optimal inventory control parameters have been developed. In the GS model strand, a base-stock policy is assumed and the primary focus has always been on the computation of the optimal parameters for different network structures.

2.2.2.1 Stochastic-service approach

The stochastic-service (SS) approach dates back to the seminal work by Clark and Scarf (1960). They establish the optimality of an echelon base-stock policy and derive a basic decomposition result for uncapacitated periodically reviewed serial systems with a finite horizon and without lot-sizing. The echelon stock consists of the stock at a given location plus all stock in transit to or on hand at locations located downstream in the supply chain minus backorders at the most downstream location. Thus, an echelon policy bases its ordering decision on the echelon inventory position that comprises of the echelon stock plus all units in transit to the given location. Federgruen and Zipkin (1984) extend the results to the stationary infinite-horizon setting. Rosling (1989) and Langenhoff and Zijm (1990) consider uncapacitated assembly systems and independently find that every convergent system can be transformed into an equivalent serial one. An extension of the echelon stock concept to divergent systems is already discussed by Clark and Scarf (1960), who recognize that due to the problem of imbalance, base-stock policies are not optimal in general. Diks and de Kok (1998) extend the exact two-level system analyses of Federgruen and Zipkin (1984) and Langenhoff and Zijm (1990) to the general N-level case and show that the decomposition result holds under the so-called balance assumption. This assumptions states that the echelon inventory positions with respect to all stockpoints that satisfy external customer demand are balanced after allocation in all periods. This implies, however, that it is allowed that some echelon inventory positions are decreased by the allocation, which corresponds to a negative shipment quantity to a stockpoint.

Numerical procedures for the inventory control parameter calculation in periodically reviewed uncapacitated serial systems are discussed, amongst others, in Federgruen and Zipkin (1984), van Houtum and Zijm (1991), and van Houtum and Zijm (1997). The latter two derive both approximate and exact algorithms based on incomplete convolutions of mixtures of Erlang distributions. For serial systems with Markov-modulated (integer) demand and Markov-modulated stochastic lead times Muharremoglu and Tsitsiklis (2008) provide an efficient algorithm for the calculation of the optimal base-stock levels based on a decomposition of the problem into single unit-customer pairs. They also show the optimality of state-dependent echelon base-stock policies. Shang and Song (2006) develop closed-form approximations for the calculation of the base-stock levels for serial systems with Poisson demand. Simulation-based heuristics for the base-stock level calculation are discussed in Daniel and Rajendran (2005), Kwon et al. (2006), and Daniel and Rajendran (2006), amongst others. An algorithm for the parameter calculation in distribution systems is presented in Diks and de Kok (1999). de Kok and Visschers (1999) show

how to decompose spanning-tree networks (also called general assembly networks) into pure serial or divergent ones, for which the inventory control parameters calculation is known again. They denote this concept as synchronized base-stock policies (see also de Kok and Fransoo (2003) for further details). The synchronization refers to the fact that common components are coordinated according to the insights from convergent systems and are allocated before they actually arrive in stock.

For spanning-tree structures, Lee and Billington (1993) develop a decentralized model. For each stage in the supply network, they develop an approximation for the base-stock calculation that takes into account random delays induced by shortages at upstream stages. Ettl et al. (2000) consider a similar setting, but under a continuous-time base-stock policy. They characterize the delays in the material delivery of a stage due to shortages by using a queueing model approximation. For the optimization of the safety factors or service levels at the stages they use conjugate gradient methods. Simchi-Levi and Zhao (2005) consider spanning tree network structures with stochastic lead times and Poisson demand and develop approximations and algorithms to coordinate the base-stock levels in these systems. In a subsequent paper, Zhao (2008a) extends the previous findings to compound Poisson demand and more general network structures with at most one directed path between two stages. This network class comprises of assembly, distribution, spanning tree, and two-level general networks as special cases. Zhao (2008b) analyzes general acyclic supply networks and derives the structural result that a dedicated stocking strategy (i.e. dedicated stock for each path) always outperforms the best shared stocking strategy (i.e. shared stock for all paths).

Extensions of the SS model are manifold. The optimal policy in a capacitated two-stage serial system is shown to be a modified echelon base-stock policy by Parker and Kapuscinski (2004). Whereas they use a dynamic programming approach to derive their result, Janakiraman and Muckstadt (2009) use a decomposition approach, i.e. an extension of the 'single-unit, single-customer' approach introduced by Axsäter (1990), to prove this. They also discuss the structure of the optimal policy in larger serial supply chains. Glasserman and Tayur (1995) use infinitesimal perturbation analysis to find the optimal policy parameters for capacitated spanning-tree systems under a base-stock policy. Simple approximations for the base-stock level determination are developed in Glasserman and Tayur (1996).

The first SS extension to incorporate lot-sizing has been made by Clark and Scarf (1962). They allow a fixed ordering cost at each stage and consider a periodic-review (s, S)-policy. Otherwise, the setting is identical to the one in Clark and Scarf (1960). A solution method is derived that successively computes the optimal (s, S) policy for each stage. Although this procedure does not guarantee the optimal multi-echelon policy, it gives upper and lower bounds on its cost. Many later extensions have been made for the continuous-review case, see, e.g., Chen (2000) for serial and assembly systems with batch ordering and various works by Axsäter for divergent systems like Axsäter (1993), Axsäter (1998), Axsäter (2000). Shang and Song (2007) develop optimal policy parameter bounds and approximations for serial supply chains with economies of scale based on single-stage considerations. Shang (2008) provides a simple heuristic for a serial system with fixed order costs that is controlled by an echelon-stock (R, nQ) policy. For a periodically reviewed serial system with batch ordering and fixed replenishment intervals Chao and Zhou (2009) derive the optimal ordering policy for given batch sizes, which is an echelon-stock (R, nQ) policy. Thus, they generalize the work of Chen (2000) and van Houtum et al. (2007). Moreover, they develop an efficient algorithm for the computation of the optimal reorder points. Shang and Zhou (2009a) consider a periodic-review serial system with echelon (R, nQ, T) policies and two types of fixed costs: one is incurred for each order batch and the other one for each inventory review. Under an echelon (R, nQ, T) policy an inventory location checks its echelon inventory position every T periods. If the inventory position is at or below R, the smallest multiple of batch size Q is ordered, which brings the inventory position above R again. The authors show how to compute the optimal parameters and also develop a near-optimal heuristic. In Shang and Zhou (2009b) the authors propose a simpler heuristic than the one of Shang and Zhou (2009a), which generates a solution by sequentially solving a deterministic demand problem, a subproblem with fixed reorder intervals, and a subproblem with fixed batch sizes. They find that this heuristic even outperforms the one of Shang and Zhou (2009a). Cachon (2001) analyzes a periodically reviewed two-level divergent system with one warehouse and N identical retailers. The author shows how to evaluate the average inventory, backorders, and fill rates at the locations exactly. While the safety stocks at the retailers are evaluated exactly, a good approximation is given for the safety stock at the warehouse.

Further extensions of the original Clark and Scarf framework focus on the consideration of time-correlated demand and systems with returns. Dong and Lee (2003) show that the structure of the optimal stocking policy of Clark and Scarf (1960) holds under time-correlated demand processes using a Martingale model of forecast evolution. Furthermore, they present an approximation, which gives a lower bound to the optimal order-up-to levels. Levi et al. (2006) provide computationally efficient approximations using a new marginal holding cost accounting approach (based on the findings in Levi et al. (2007) and Levi et al. (2008)) to determine provably good ordering policies. The policies are shown to be near-optimal in many instances with a worst-case performance of 2.

DeCroix (2006) analyzes a serial multi-echelon system with returns in addition to the traditional forward material flows. It is shown that, if remanufactured items enter the most upstream stage, the system can be optimized by decomposition into a sequence of single-stage systems. Each downstream stage follows an echelon base-stock policy and the most upstream stage follows a three-parameter policy with a simple structure. If remanufactured items enter a downstream stage, similar structural results are derived, but the definition of the echelon inventory needs to be adjusted for all stages upstream of the remanufacturing stage. DeCroix et al. (2005) study the steady-state behavior of a serial system with possibly negative stochastic demand, which basically represents returns. They develop exact and approximate methods for the evaluation of any echelon base-stock policy and an optimization procedure that returns a good policy. Extensions to a base-stock policy with local information and the occurrence of returns at several stages are discussed, too.

2.2.2.2 Guaranteed-service approach

The guaranteed-service (GS) framework makes use of the base-stock concept by Kimball (1988), i.e. each stage of the network operates a periodically reviewed base-stock policy. In his fundamental work, Simpson (1958) shows for uncapacitated serial systems that an all-or-nothing policy is optimal for this stock allocation problem, i.e. each stage either holds sufficient stock to completely decouple it from its successor or no stock at all. Based on this so-called extreme point property, Graves (1988) notes that the optimization problem can be solved by dynamic programming. In subsequent years, this approach has been extended to other network structures. Extensions to assembly and distribution systems, spanning trees or even general acyclic network structures can be found in Inderfurth (1991), Inderfurth and Minner (1998), Graves and Willems (2000), Minner (2000), Humair and Willems (2006), and Humair and Willems (2010).

Basically, all of the afore-mentioned contributions make use of dynamic programming as optimization technique. For general acyclic networks, for which Lesnaia (2004) shows that the optimization problem is NP-hard, Humair and Willems (2010) imbed the dynamic program developed for spanning trees into an overall branch-and-bound algorithm. Minner (2000) presents several heuristic approaches for this network type. Magnanti et al. (2006) approximate the concave objective function with piecewise linear functions and make use of powerful Linear Programming solvers.

Over the last two decades, the GS framework has been extended in several ways. Whereas the original (standard) GS model assumes a common review period at all stages, Bossert and Willems (2007) allow for an arbitrary, integer review period at each stage. Three different inventory control policies are analyzed, i.e. the constant base stock, constant safety stock, and adaptive base stock policy, and a solution to the inventory optimization problem is obtained by a modified version of the dynamic programming procedure of Graves and Willems (2000).

For products with short life cycles, Graves and Willems (2008) present an extension of the GS framework to non-stationary demand. For such situations a dynamic service-time policy is optimal. Since this policy is difficult to implement in practice, they consider a simpler so-called constant service-time policy and show that the optimization algorithm for the stationary demand case can be used for the safety stock determination.

Schoenmeyr and Graves (2009b) study the placement of safety stocks in supply chains, for which an evolving demand forecast exists. They show that under specific assumptions, the optimization problem is equivalent to the one for stationary demand and base-stock policies. Consequently, the optimal solution can be found by the already existing algorithms.

A generalization of the GS model to include capacity constraints is presented in Schoenmeyr and Graves (2009a). In their model, the order quantity of each stage is restricted by some capacity limit. It is shown that, if the objective function is concave and a base-stock policy is used, the all-or-nothing property holds as in the original contribution without capacities by Simpson (1958). Moreover, the authors establish that the optimization algorithm for uncapacitated systems can be employed to compute the optimal safety stock placement. In addition to the standard base-stock policy, they study a 'censored' ordering policy, where each stage orders the minimum of its capacity and the order it receives (plus extra quantities to 'catch up', as necessary). An optimal solution can be obtained by a slightly modified version of the dynamic programming algorithm of Graves and Willems (2000). Interestingly, Schoenmeyr and Graves (2009a) find that the inventory holding costs for the latter policy are less than for the standard base-stock policy. Sometimes, the costs are even smaller than those for the uncapacitated system. The authors argue that this effect results from a smoothing of the original demand process through the censored orders.

2.2.2.3 Comparison and combination of the stochastic- and guaranteed-service approach

Only few contributions in the literature can be found that compare, contrast, or try to combine both modeling strands. One such comparison is presented in Graves and Willems (2003). They apply both approaches to an assembly system and a spanning-tree network and find that (under their assumptions) the GS model outperforms the SS model. For two-level distribution systems, Klosterhalfen and Minner (2010) provide a model comparison and show that the superiority of any of the two approaches heavily depends on the specific parameter setting and cannot be established in general. Moreover, they present a method to derive appropriate internal service levels, which define the operating flexibility usage in the GS model, based on cost considerations.

Lawson and Porteus (2000) and Minner et al. (2003) combine aspects of both frameworks by including the possibility of expediting into the SS model. One way to think about the implicitly assumed operating flexibility in the GS approach is that missing items are speeded up from pipeline inventory. Minner et al. (2003) derive insights into the appropriate use of operating flexibility. Lawson and Porteus (2000) show for their extended SS model that a 'top-down base-stock policy' is optimal in each period. Muharremoglu and Tsitsiklis (2003) extend this work by allowing a more general cost structure for expediting, i.e. supermodular instead of additive. In a single-echelon continuous-time setting Gallego et al. (2007) also study the possibility of expediting existing orders and derive the optimal policy for Poisson demand, which is a threshold policy. Neither of these contributions allows the entire supply network to consist of subnetworks of *both* approaches, however.

2.2.3 Multi-echelon inventory models with multiple sourcing options

The majority of literature on multi-echelon inventory models with multiple sourcing options is limited to rather small supply chain settings with at most two echelons and mainly continuous review. Ganeshan (1999) considers a continuous review one-warehouse, multiple retailer system where the warehouse order is split across several identical suppliers, thus incorporating the multiple-supplier aspect. A near-optimal (s, Q) policy is presented and verified by means of a simulation study. A similar two-echelon system is analyzed by Muckstadt and Thomas (1980), where each location operates a $(S-1, S)$ policy. Instead of splitting the warehouse order across multiple suppliers, emergency orders with a shorter lead time than that of the regular orders are placed (both by the warehouse and the retailers) whenever the on-hand inventory drops to zero and another item is demanded. Aggarwal and Moinzadeh (1994) consider a two-echelon production/distribution system, where the retailers use a $(S-1, S)$ policy and the production facility does not hold any inventory, but produces to order. Retailer orders can either be of regular or emergency type depending on whether the number of outstanding orders exceeds a certain level or not. Emergency orders are processed first at the facility, which represents their advantage over regular orders. The objective is to determine the order-up-to levels and threshold values. Moinzadeh and Aggarwal (1997) study a one warehouse, multiple retailer system with a $(S-1, S)$ policy in place at each location, where each location can decide to either place a regular or emergency order similar to the single-stage model by Moinzadeh and Schmidt (1991). An optimization procedure for the policy parameters is developed and the benefit over a single resupply mode is illustrated in a numerical study.

Multi-echelon inventory models with lateral transshipments can also be regarded as a kind of dual-sourcing setting. Besides the regular replenishment from the warehouse, a retailer can also place an emergency order with another nearby located retailer for a lateral transshipment. Thus, the retailer has two supply options with differing lead times. A recent literature review on lateral transshipment models is available in Paterson et al. (2009). In some of these models, e.g., Alfredsson and Verrijdt (1999), there exist even more than two supply options. In addition to lateral transshipments, direct shipments from the warehouse or even direct shipments from an external supplier are incorporated.

For larger multi-echelon systems Graves and Willems (2005) present two ideas of how to incorporate dual sourcing in the guaranteed-service approach in the final section of their paper. Both approaches are only approximate, however. Assume that the demand split to the two sources is decided a priori. Then, in the first approach, the original stage is replaced by two stages. Each of these stages receives its respective fraction of demand and has a processing time equal to its transportation time. In this case, the safety stock computation is only approximate, because both stages are treated as if they provide different items and therefore hold separate stocks. In the second approach, a joint stage is created with a cost and replenishment time that represents a mixture of both supply options. Here, the difficulty lies in deriving these joint parameters. Since these parameters influence the safety stock calculation, this approach is approximate, too. An exact way of how the incorporation of dual sourcing can be achieved is presented in Section 4.5 of this thesis.

Note that the GS approach in its standard version, i.e. with a single supply source for each item, can be regarded as a kind of dual-sourcing model, too. The assumed operating flexibility enables a shortening of the processing time, if required. Thus, there are two supply modes available, a normal and an expedited one. However, in this thesis this special kind of sourcing flexibility is not understood or referred to as dual sourcing.

3 Single-echelon inventory model with dual sourcing

3.1 Introduction

In this chapter, a periodic-review single-echelon inventory model with two supply options and deterministic lead times is considered. The supply options can refer to either two different suppliers, e.g., one situated overseas and the other one close by, or two different modes of supply, e.g., transportation by air and ship. For clarity's sake, only the term supplier will be used in the remainder of the thesis to refer to both such options.

Dual-sourcing inventory models play a key role in practice. Many companies rely on two (or more) suppliers for their material procurement. Such sourcing strategies enable them to serve demand at low costs without compromising on service. Having two suppliers available, the majority of materials can be replenished from the cheaper one, which usually has a longer procurement lead time (*slow supplier*). In case of a surge in demand that leaves the inventory low with most outstanding orders far away, a replenishment order can be placed with the more expensive, but *faster supplier* in order to avoid future stockouts. Examples for such dual-sourcing practices are reported in the literature for Hewlett Packard (Beyer and Ward (2000)) or Caterpillar (Rao et al. (2000)), amongst others.

In contrast to a single-sourcing model, difficulties arise in the dual-sourcing context due to potential order crossing, which can occur if more than one supplier is used. This problem makes the analysis and derivation of the optimal policy or close-to-optimal heuristics a challenging task.

Consequently, although companies already employ dual sourcing, they are still asking for simple, yet effective policies to support their replenishment decisions. Hence, the main research question that is addressed in this chapter is: *What are effective dual-sourcing inventory control policies in a single-echelon setting?* 'Effective' in this context means easily implementable and performing close to optimal. In order to answer this broad question, it is broken down into several smaller and more specific research questions, which are addressed in turn. These include:

1. *How can the optimal inventory control policy for the studied dual-sourcing problem be determined? (Section 3.3.2)*
2. *What are simple non-optimal policy alternatives and how can their optimal parameters be computed? (Section 3.3.3)*
3. *How do these (non-optimal) inventory control policies perform? (Section 3.4)*

The outline of this chapter is as follows. In Section 3.2 the basic dual-sourcing inventory model and assumptions are set forth. Section 3.3 explains how the optimal policy can be derived and introduces several non-optimal policies, which are easier to compute and manage than the optimal one. Special cases and extensions of these policies are addressed in the respective policy subsections. A comparison of the dual-sourcing policies is presented in Section 3.4 on a theoretical and numerical basis.

3.2 Assumptions and notations

Throughout this chapter, a periodically reviewed single-item inventory model is considered with the following characteristics. Customer demand D per period is stochastic, with the demands of different periods being *i.i.d.* non-negative random variables from a stationary distribution. Without loss of generality, D is assumed to be discrete. Let F denote the cumulative distribution function of D with mean $\mathbb{E}[D] \stackrel{\text{def}}{=} \mu < \infty$ and standard deviation $\sqrt{\text{VAR}[D]} \stackrel{\text{def}}{=} \sigma < \infty$. F_L represents the L-period cumulative distribution function. Materials are replenished each period by placing orders with two potential suppliers. Via the slow supplier, s, it takes L^s

periods for the order to be delivered. The faster order arrives after L^f periods with $0 \leq L^f < L^s$. Let $L^\Delta = L^s - L^f$. (Alternatively, the terms regular and expedited supplier are used in the literature to refer to the slow and fast supplier.) Both lead times are assumed to be deterministic and an integer multiple of the base (review) period. Let $t \in \{0, 1, 2, ...\}$ denote the period index. The shorter lead time comes at a higher procurement cost per unit, $c^f > c^s$. Otherwise, since there are no fixed cost per order present in the model, one would only use the expedited supplier. The difference, $c = c^f - c^s$, is denoted as the expediting premium. For each unit of on-hand stock at the end of a period, an inventory holding cost h is incurred, which is independent of the procurement cost that has been paid for this unit. Unsatisfied customer demand is backordered at a cost of b per unit and period. The performance measure used is the infinite horizon expected average cost. (Extensions to service-level criteria are presented in the respective policy sections.) The total expected average cost, TC, consists of (i) inventory holding costs for the on-hand stock, OH, (ii) backorder costs for the backordered quantity, BO, and (iii) procurement costs for the quantities ordered with both suppliers, Q^s and Q^f:

$$TC = h \cdot \mathbb{E}\left[OH\right] + b \cdot \mathbb{E}\left[BO\right] + c^s \cdot \mathbb{E}\left[Q^s\right] + c^f \cdot \mathbb{E}\left[Q^f\right] \quad . \tag{3.1}$$

Note that the pipeline stock costs do not have to be taken into account explicitly in the cost function. If it is assumed that the holding cost h is also paid for each unit in the two pipelines, the actual procurement cost of the slow supplier c^s can simply be increased by $h \cdot (L^s - L^f)$ to account for the additional holding costs due to the longer lead time. Alternatively, in some situations it might be reasonable to assume that all materials are paid for after receipt. On average, the sum of both orders equals the period demand, i.e.

$$\mu = \mathbb{E}\left[Q^s\right] + \mathbb{E}\left[Q^f\right] \quad . \tag{3.2}$$

By inserting (3.2) into (3.1) and exploiting the fact that $c^s \cdot \mu$ cannot be influenced by the inventory control decisions, (3.1) reduces to the total relevant cost

$$TRC = h \cdot \mathbb{E}\left[OH\right] + b \cdot \mathbb{E}\left[BO\right] + c \cdot \mathbb{E}\left[Q^f\right] \quad , \tag{3.3}$$

which shows that only the expediting premium, c, is relevant and not the specific values of c^s and c^f.

In order to avoid trivial solutions, it is required that $c < b \cdot L^{\Delta}$. If $c \geq b \cdot L^{\Delta}$, it is cheaper to wait for the slow order to arrive and incur backorder costs than to use the faster supplier. Consequently, the problem would reduce to a standard single-sourcing inventory problem (see Section 2.1.2.3).

It is assumed that in each period the sequence of events is as follows:

- placement and arrival of orders (at the beginning of a period),
- satisfaction of backorders (at the beginning of a period),
- occurrence and satisfaction of demand (sometime during the period),
- assessment of costs (at the end of a period).

All terms in equation (3.3) depend on the inventory control policy that is in place, because this influences not only the on-hand stock and backorder quantity but also the orders placed with each supplier and therefore the procurement costs.

3.3 Inventory control policies

3.3.1 Definitions

Since the terms, by which dual-sourcing inventory control policies are described in the literature, are not always used consistently, it is shortly explained at this point what is meant by specific terms in this thesis. The most common distinctions are between single- and dual-index policies, depending on whether one or two inventory positions are tracked and, somewhat independently of the previous distinction, between single and dual base-stock policies, depending on the number of order-up-to levels used by the policy. For the purpose of this thesis, the following inventory control policies are defined.

Definition 3.3.1.1 *A single-index policy with two order-up-to levels, one for each supplier, is called a single-index policy (SIP) for short.*

Definition 3.3.1.2 *A single-index policy with a fixed order quantity for the slow supplier and an order-up-to level for the fast order determination is called a constant-order policy (COP).*

Definition 3.3.1.3 *A dual-index policy with two order-up-to levels, one for each supplier, is called a dual-index policy (DIP) for short.*

A slightly different dual-sourcing strand in the literature are *order-splitting* models. They are concerned with the optimal allocation of the demand to the different suppliers and mostly studied under deterministic demand and stochastic lead times. In stochastic demand, deterministic lead-time settings as analyzed in a single-echelon context in this chapter, an order-splitting policy is not a reasonable candidate, because demand is always allocated across the suppliers according to fixed portions (see Section 3.3.3.4 for more details). Nevertheless, such a policy is very appealing from the point of view of analytical tractability, which becomes especially relevant in the multi-echelon context in Chapter 4. Therefore, this policy is also defined and addressed here.

Definition 3.3.1.4 *A policy that at each review instant splits the total replenishment order according to fixed fractions among the suppliers is called an order-splitting policy (OSP).*

The indices that are used by these inventory control policies are defined as follows:

- **Inventory position used for replenishments with the fast supplier** (*Fast inventory position*)

  The fast inventory position at the beginning of period t before ordering comprises of the net stock at the end of the previous period plus all outstanding orders with any of the two suppliers that will arrive prior to or together with the fast order that is to be determined.

$$IP_t^f = NS_{t-1} + \sum_{i=L^\Delta}^{L^s} Q_{t-i}^s + \sum_{j=1}^{L^f} Q_{t-j}^f \tag{3.4}$$

- **Inventory position used for replenishments with the slow supplier** (*Slow inventory position*)

  The slow inventory position consists of the net stock at the end of the previous period plus all outstanding orders in the system excluding the potential fast order that is placed in period t.

$$IP_t^s = NS_{t-1} + \sum_{i=1}^{L^s} Q_{t-i}^s + \sum_{j=1}^{L^f} Q_{t-j}^f = IP_t^f + \sum_{i=1}^{L^\Delta - 1} Q_{t-i}^s \qquad (3.5)$$

3.3.2 Optimal policy

For a single-sourcing problem with the above specified cost structure a stationary order-up-to policy is optimal (see, e.g., Veinott (1966)). For the dual-sourcing problem, policies with an order-up-to structure are only optimal for the special case of consecutive lead times ($L^\Delta = 1$). Barankin (1961), Daniel (1962), Bulinskaya (1964), and Neuts (1964) show for the case with $L^f = 0$ and $L^s = 1$ and Fukuda (1964) for the more general case with $L^f = m$ and $L^s = m + 1$, $m \in \{0, 1, ...\}$ that the optimal policy is a SIP (see Section 3.3.3.1 for details).

For offsetting lead times ($L^\Delta > 1$), order-up-to policies are no longer optimal, as shown by Whittemore and Saunders (1977). This also directly follows from the recently established analogy of the dual-sourcing problem to the lost-sales inventory problem by Sheopuri et al. (2010). For the lost-sales problem the non-optimality of order-up-to policies is well-known (see Karlin and Scarf (1958)). Consequently, the optimal policy needs to be computed by dynamic programming, i.e. formulating the problem as a discrete-time *Markov Decision Process* (MDP). Although Veeraraghavan and Scheller-Wolf (2008) mention that they compute the optimal policy in this way, they do not provide any modeling details, as is done here.

3.3.2.1 General Markov Decision Process formulation

The following general description of a Markov Decision Process (MDP) is based on Puterman (1994). An MDP model consists of five elements:

1. Decision epochs (or periods),

2. States,

3. Actions (or decisions),

4. Transition probabilities, and

5. Rewards (or cost).

Decision epochs characterize the points in time, at which a decision is made. The number of epochs can be either finite or infinite and decisions can be made at discrete points in time or continuously. For the purpose of this study, the focus is on discrete-time infinite horizon MDPs, i.e. time is divided into periods or stages, $t = 0, 1, \dots$.

At each decision epoch the system is in a certain *state*, $\mathbf{i}$. Since a state can comprise of one or multiple state variables, the vector notation is used. The set of all possible states, also known as the state space, is denoted as $\mathcal{SSP}$.

Given a certain state $\mathbf{i}$, the decision maker can choose an *action* or decision $a(\mathbf{i})$ from the set of all possible decisions in state $\mathbf{i}$, $\mathcal{DSP}(\mathbf{i})$, i.e. $a \in \mathcal{DSP}(\mathbf{i})$. (For ease of presentation, the dependency of a on state $\mathbf{i}$ is not explicitly indicated, if it is already indicated in the decision space $\mathcal{DSP}(\mathbf{i})$.) It is assumed that $\mathcal{SSP}$ and $\mathcal{DSP}$ do not vary in t. Usually, these sets are arbitrary finite or countably infinite sets.

The choice of decision $a \in \mathcal{DSP}(\mathbf{i})$ in state $\mathbf{i}$ at decision epoch t results in a *reward,* $r(\mathbf{i}, a)$ for the decision maker. If the reward is negative, it is also referred to as cost. In the application of the MDP later on only costs are considered. That is why in the remainder only the term cost is used. The cost might consist of the

- cost of being in state $\mathbf{i}$,

- cost associated with decision a, and

- cost associated with the transition from state $\mathbf{i}$ to state $\mathbf{j}$.

A transition, as mentioned above as the last cost aspect, is induced by the chosen decision. The decision moves the system from state $\mathbf{i}$ at decision epoch t to another state $\mathbf{j}$ at decision epoch $t + 1$ with a certain *transition probability*, $p_{\mathbf{ij}}(a)$.

A *decision rule* prescribes a procedure for action selection in each state at a specified decision epoch. Only a deterministic decision rule is considered here, i.e. this rule chooses exactly one action/decision (with certainty) that is to be taken in a specific state. A *policy* specifies the decision rule to be used at all decision epochs. Let $a^*(\mathbf{i})$ denote the optimal decision in state $\mathbf{i}$ and $(a^*)_\infty$ the optimal policy.

The average cost, AC, of a given policy, $(a)_\infty$, is

$$AC((a)_\infty) = \sum_{\mathbf{i} \in \mathcal{SSP}} \pi_\mathbf{i} \; r(\mathbf{i}, a) \tag{3.6}$$

with $a \in (a)_\infty$. $\pi_\mathbf{i}$ denotes the steady-state probability of being in state $\mathbf{i}$ under a given policy and results from solving the following system of linear equations:

$$\pi_\mathbf{j} = \sum_{\mathbf{i} \in \mathcal{SSP}} \pi_\mathbf{i} \; p_{\mathbf{ij}}(a) \qquad \mathbf{j} \in \mathcal{SSP} \tag{3.7}$$

$$\sum_{\mathbf{i} \in \mathcal{SSP}} \pi_\mathbf{i} = 1 \quad . \tag{3.8}$$

An optimal policy $(a^*)_\infty$ is one, for which

$$AC\left((a^*)_\infty\right) \leq AC\left((a)_\infty\right) \quad , \tag{3.9}$$

i.e. a policy, for which the average cost of this policy is smaller than or equal to that of all other feasible policies. An optimal policy can be found by using a value iteration algorithm, policy iteration algorithm, or linear programming (see, e.g., Puterman (1994)).

3.3.2.2 Application to the dual-sourcing inventory problem

Decision epoch, state, and decision definition

In the dual-sourcing inventory problem, the decision variables are the fast and slow order quantities in period t, Q_t^f and Q_t^s, respectively. That means a decision is given by $a = \left(Q_t^s, Q_t^f\right)$. Since we assume a periodically reviewed system, a decision about these quantities can be made at the beginning of each period, which defines the decision epoch.

In order to find an appropriate state definition, it is first analyzed which information is required to make optimal ordering decisions in this system. At the beginning of period t before ordering the following information is available:

- The net stock, which is equal to the net stock at the end of the previous period, NS_{t-1}.
- The orders placed with the fast supplier that are still outstanding, i.e. $Q^f_{t-L^f}, ..., Q^f_{t-1}$.
- The orders placed with the slow supplier that are still outstanding, i.e. $Q^s_{t-L^s}, ..., Q^s_{t-1}$.

The earliest point in time that can be influenced by the ordering decision made in period t is the period, in which the fast order, Q^f_t, will arrive, i.e. period $t + L^f$. The costs of all previous periods $\{t, t+1, ..., t+L^f-1\}$ are sunk. Consequently, all information about the (fast and slow) orders that will arrive prior to or together with this order can be compressed into one number. Together with the current stock situation this corresponds to the fast inventory position at the beginning of period t before ordering, IP^f_t (see (3.4)). Only the remaining slow orders that have already been determined and are outstanding need to be tracked individually. Hence, the state of the system can be described by a $L^s - L^f = L^\Delta$ dimensional vector $\left(IP^f_t, Q^s_{t-1}, Q^s_{t-2}, ..., Q^s_{t-L^\Delta+1}\right)$.

Let $\mathbf{Y}_t = \left(y_t, q_{t,1}, q_{t,2}, ..., q_{t,L^\Delta-1}\right)$ describe a generic state vector of period t with

$$y_t = IP^f_t \tag{3.10}$$

$$q_{t,i} = Q^s_{t-i} \qquad i = 1, 2, ..., L^\Delta - 1. \tag{3.11}$$

The state space $\mathcal{SSP}$ consists of the permutation of all individual state variables and their respective admissible values. In order to reduce the complexity, lower and upper bounds for all variables are helpful. For tractability reasons, demand is assumed to be limited by some upper bound $\bar{D}$, i.e. $d \in \left\{0, 1, ..., \bar{D}\right\}$. Then, the following upper and lower bounds can be derived:

- **Order quantities**

  Veeraraghavan and Scheller-Wolf (2008) find that the optimal orders never exceed the maximum possible demand $\bar{D}$, which is to be expected due to the stationary cost structure and uncapacitated orders. Hence, the upper bound on the order quantities can be set to $\bar{D}+1$. Moreover, the minimum order quantity is zero since negative quantities are not permitted. Consequently,

$$Q^s, Q^f, q \in \left\{0, 1, ..., \bar{D}+1\right\} \quad . \tag{3.12}$$

- **On-hand stock**

  Since $h > 0$, the maximum on-hand stock is restricted to $(L^s+1)\cdot(\bar{D}+1)$. This can be explained as follows. In order to avoid *any* backorders in future periods, the system would place the largest feasible order each period, i.e. $\bar{D}+1$. This means that $L^s \cdot (\bar{D}+1)$ units are outstanding in total plus one order of size $\bar{D}+1$ that arrives in the current period. If there is a series of zero demands, all of these orders would arrive and increase the on-hand stock to a maximum of $(L^s+1)\cdot(\bar{D}+1)$. Since the risk of backorders is 0 and $h > 0$, no additional orders would be placed.

- **Backorders**

  Since $b > 0$, the backorders are restricted to $(L^s+1)\cdot\bar{D}$. The reasoning is similar to the on-hand stock line of thought. In the worst case, only orders of zero are placed and the maximum demand, $\bar{D}$, occurs several times in a row, resulting in maximum backorders of $(L^s+1)\cdot\bar{D}$.

- **Net stock**

  From the on-hand stock and backorder bounds, it follows that the admissible values of the net stock are

$$NS \in \left\{-(L^s+1)\cdot\bar{D}, ..., (L^s+1)\cdot(\bar{D}+1)\right\} \quad . \tag{3.13}$$

- **Slow inventory position after ordering**

  The slow inventory position after ordering, IP^{s^+}, corresponds to the sum of the net stock and *all* outstanding orders in the system (including the current orders). The maximum value of this expression is

$$IP^{s^+}_{max} = (L^s + 1) \cdot (\bar{D} + 1) \tag{3.14}$$

because this already ensures a service of 100%, i.e. no backorders. So, there is no need for any further orders.

Let i denote an element of state vector $\mathbf{i}$. Based on the above-derived bounds, the set of feasible decisions in state $\mathbf{i}$ is given by

$$\mathcal{DSP}(\mathbf{i}) = \left\{ \left(Q^s, Q^f\right) \middle| \sum_{\forall i \in \mathbf{i}} i + Q^s + Q^f \leq IP^{s^+}_{max} \right\} \quad . \tag{3.15}$$

Transition probabilities

In period $t+1$, the state vector is $\mathbf{Y}_{t+1} = \left(y_{t+1}, q_{t+1,1}, q_{t+1,2}, ..., q_{t+1,L^\Delta - 1}\right)$ with

$$y_{t+1} = y_t + Q^f_t + q_{t,L^\Delta-1} - d_t \tag{3.16}$$

$$q_{t+1,1} = Q^s_t \tag{3.17}$$

$$q_{t+1,k} = q_{t,k-1} \qquad k = 2, ..., L^\Delta - 1. \tag{3.18}$$

Let $p_{\mathbf{ij}}(a) = Pr\left\{\mathbf{Y}_{t+1} = \mathbf{j} \mid \mathbf{Y}_t = \mathbf{i}\right\}$ denote the transition probability from state $\mathbf{i}$ in t to $\mathbf{j}$ in $t+1$ under decision $a = \left(Q^s_t, Q^f_t\right)$. Under stationary conditions $t \to \infty$, $\mathbf{Y} = (y, q_1, q_2, ..., q_{L^\Delta-1})$. Then,

$$p_{\mathbf{ij}}(a) = \begin{cases} Pr\left\{D = y(\mathbf{i}) - y(\mathbf{j}) + Q^f + q_{L^\Delta-1}(\mathbf{i})\right\} & \text{if } Q^s = q_1(\mathbf{j}), ..., q_k(\mathbf{j}) = q_{k-1}(\mathbf{i}), \\ & k = 2, ..., L^\Delta - 1 \\ 0 & \text{otherwise} \end{cases} . \tag{3.19}$$

These probabilities result from the fact that a transition only has positive probability, if the demand and the fast order transfer the fast inventory position from $y(\mathbf{i})$ to $y(\mathbf{j})$ and the remaining outstanding order vector has the same entries (shifted by one period and the latest order with the slow supplier inserted).

Costs

The total cost in a state $\mathbf{i}$, if decision a is chosen, consists of two components. First, the direct cost, DC, of the decision a is calculated as

$$DC(a) = c^s \cdot Q^s + c^f \cdot Q^f \quad . \tag{3.20}$$

Second, there are the expected holding and backorder costs in the first future period that can be influenced by the current decision, whose function is given as

$$HB(x) = h \cdot \sum_{d=0}^{x} (x-d) f_{L^f+1}(d) + b \cdot \sum_{d=x+1}^{(L^f+1)\cdot \bar{D}} (d-x) f_{L^f+1}(d) \quad . \tag{3.21}$$

Consequently, the total cost in state $\mathbf{i}$, if decision a is chosen, is

$$r(\mathbf{i}, a) = DC(a) + HB\left(y(\mathbf{i}) + Q^f\right) \quad . \tag{3.22}$$

3.3.3 Non-optimal policies

For large lead-time differences and demand distributions with a large number of possible realizations, the MDP suffers from the curse of dimensionality, because the state and decision space grow considerably. That means the optimal policy can no longer be computed in a reasonable amount of time. Due to the computational complexity of the optimal policy, several simpler policies have been proposed in the literature for this problem. The most prominent ones are the *single-index*, *constant-order*, and *dual-index policy*. The exposition of these policies in the upcoming subsections contains many results that are already known from previous contributions. As the policies have been studied mainly in isolation in those works, however, the reiteration of the major results in this section using a unified notational framework is intended to foster a better understanding of the policy differences, which is particularly relevant for the policy comparison in Section 3.4. It is indicated explicitly, whenever known results are complemented by new aspects.

Recently, further policies have received increased attention because of the analogy of the periodic-review dual-sourcing problem to the lost-sales inventory problem established by Sheopuri et al. (2010). These policies exhibit features of an *order-splitting policy*. They only consider the total order quantity in a period and decide about the allocation of this quantity across the two suppliers, which is in contrast to the other policies that determine the two order quantities by using an order-up-to structure for the fast order and a separate control mechanism for the slow one. That is why a simple version of an order-splitting policy is addressed at the end of this section, too. This policy also plays a key role when it comes to the integration of dual sourcing into a multi-echelon setting (see Section 4.5).

3.3.3.1 Single-index policy

Policy description

The single-index policy (SIP) is analyzed in Scheller-Wolf et al. (2007), which forms the basis of this section. The presented lemmata are basically a modification of the ones derived by Scheller-Wolf et al. (2007) for continuous demand distributions to the discrete demand case, which is studied in this chapter. As already mentioned above, the SIP is optimal for consecutive lead times. The SIP specifies two order-up-to levels, one for the replenishments with the fast supplier (B^f) and one for the replenishments with the slow supplier (B^s). Let $\Delta = B^s - B^f$. Due to this relation, either $\left(B^f, B^s\right)$, $\left(B^f, \Delta\right)$, or (Δ, B^s) can be chosen as decision variables. Since the latter combination simplifies the upcoming analysis, this one is used. The system keeps track of a single inventory position, i.e. the slow inventory position, IP^s, which comprises of all units on-hand, on order, and owed to the customer (see (3.5)). In each period, the inventory position is checked against the fast order-up-to level first, and, if it is below, an order is placed to make up for the difference. Then, a slow order is made to bring the inventory position up to the slow order-up-to level. The inventory position recursion under this policy is

$$IP_t^s = IP_{t-1}^s + Q_{t-1}^f + Q_{t-1}^s - d_{t-1} \quad , \tag{3.23}$$

which reduces to

$$IP_t^s = B^s - d_{t-1} = B^f + \Delta - d_{t-1} \quad . \tag{3.24}$$

The order quantities are given as

$$Q_t^f = \left(B^f - IP_t^s\right)^+ = \left(B^f - \left(B^f + \Delta - d_{t-1}\right)\right)^+ = \left(d_{t-1} - \Delta\right)^+ \tag{3.25}$$

$$Q_t^s = \left(B^s - \left(IP_t^s + Q_t^f\right)\right)^+ = \left(B^s - \left(B^s - d_{t-1} + \left(d_{t-1} - \Delta\right)^+\right)\right)^+ = \min\left\{\Delta, d_{t-1}\right\}. \tag{3.26}$$

From (3.25) and (3.26) it can be seen that in any period the fast order corresponds to the portion of demand that exceeds Δ and the rest is ordered from the slow supplier with a maximum quantity of Δ. Furthermore, it holds that the total order quantity in a period corresponds to the demand of the previous period

$$Q_t^f + Q_t^s = d_{t-1} \quad . \tag{3.27}$$

In the extreme cases the SIP can mimic both fast and slow single sourcing with an order-up-to policy by setting $\Delta = 0$, i.e. $B^f = B^s$, where B^f is the solution to (2.50) with $L = L^f$ or $\Delta = \infty$ and B^s as the solution to (2.50) with $L = L^s$, respectively.

Policy evaluation

In order to evaluate a parameter combination, (Δ, B^s), expressions for the expected on-hand stock, backorders, and fast order quantity are required for the computation of TRC according to (3.3). From (3.25) it directly follows that

$$\mathbb{E}\left[Q^f\right] = \mathbb{E}\left[(D - \Delta)^+\right] \quad , \tag{3.28}$$

which results in the following lemma that is stated without proof, since it is obvious from (3.28):

Lemma 3.3.3.1 *Under the SIP, the expected fast order quantity, $\mathbb{E}\left[Q^f\right]$, is solely determined by Δ, independent of B^s or B^f.*

In the process of deriving expressions for the on-hand stock and backorders at the end of a period, first an expression for the net stock at the end of a period is developed. The net stock is related to the other two quantities via (2.27), i.e.

$$OH_t = (NS_t)^+ \tag{3.29}$$

$$BO_t = (NS_t)^- \tag{3.30}$$

where $(x)^+ = \max\{0, x\}$ and $(x)^- = \max\{0, -x\}$. Using (3.25), (3.26), and (3.27), the net stock at the end of period t (assuming w.l.o.g. that $NS_0 = B^s$) is

$$\begin{aligned} NS_t &= B^s - \sum_{i=0}^{t} d_{t-i} + \sum_{i=L^s}^{t} Q^s_{t-i} + \sum_{i=L^f}^{t} Q^f_{t-i} \\ &= B^s - \sum_{i=0}^{L^s} d_{t-i} + \sum_{i=L^f}^{L^s+1} Q^f_{t-i} = B^s - \sum_{i=0}^{L^s} d_{t-i} + \sum_{i=L^f}^{L^s+1} (d_{t-i} - \Delta)^+ \\ &= B^s - \sum_{i=0}^{L^f} d_{t-i} - \sum_{i=L^f+1}^{L^s} \min\{d_{t-i}, \Delta\} \quad . \end{aligned} \tag{3.31}$$

Let the random variable $\hat{D}(\Delta)$ be defined as:

$$\hat{D}(\Delta) = \sum_{i=0}^{L^f} D_i + \sum_{i=L^f+1}^{L^s} \min\{D_i, \Delta\} \quad . \tag{3.32}$$

Then, under stationary conditions $t \to \infty$ the net stock is given as

$$NS = B^s - \hat{D}(\Delta) \tag{3.33}$$

with

$$\mathbb{E}\left[\hat{D}(\Delta)\right] = (L^s + 1) \cdot \mu - L^\Delta \cdot \mathbb{E}\left[(D - \Delta)^+\right] \quad . \tag{3.34}$$

From (3.33) it is obvious that for a given Δ the net stock at the end of a period and consequently also the on-hand stock and the backorders are only determined by B^s. The optimal B^s results according to the following lemma.

Lemma 3.3.3.2 *Under the SIP, for a given* Δ*, the optimal* B^s*,* $B^{s^*}(\Delta)$*, is the smallest value that satisfies*

$$Pr\left\{\hat{D}(\Delta) \leq B^s\right\} \geq \frac{b}{b+h} \quad . \tag{3.35}$$

Proof:
See Appendix B.1. □

Given $B^{s^*}(\Delta)$, the expected on-hand stock and backorders are computed according to (2.48) and (2.49) as

$$\mathbb{E}\left[OH(B^{s^*}(\Delta))\right] = \sum_{x=0}^{B^{s^*}} \left(B^{s^*} - x\right) \cdot Pr\left\{\hat{D}(\Delta) = x\right\} \tag{3.36}$$

$$\mathbb{E}\left[BO(B^{s^*}(\Delta))\right] = \sum_{x=B^{s^*}+1}^{\infty} \left(x - B^{s^*}\right) \cdot Pr\left\{\hat{D}(\Delta) = x\right\} \quad . \tag{3.37}$$

Remark. The above analysis can also be done slightly different from the one in Scheller-Wolf et al. (2007), namely in terms of B^f instead of B^s. Substituting $B^s = B^f + \Delta$ in the net stock computation (3.31) yields

$$NS_t = B^f - \sum_{i=0}^{L^f} d_{t-i} + \Delta - \sum_{i=L^s}^{L^s+L^\Delta-1} \min\left\{d_{t-i}, \Delta\right\} \quad . \tag{3.38}$$

The first two terms on the right-hand side correspond to the net stock computation in the single-sourcing order-up-to level model, if only the fast supplier was available (see (2.44)). Due to the additional slow supply option, the maximum inventory value in period $t - L^s$ is not equal to B^f, but equal to

$$B^f + \Delta - \sum_{i=L^s}^{L^s+L^\Delta-1} \min\left\{d_{t-i}, \Delta\right\} \quad . \tag{3.39}$$

The additional quantity can be seen as a so-called overshoot in period $t - L^s$:

$$O_{t-L^s} = \Delta - \sum_{i=L^s}^{L^s+L^\Delta-1} \min\{d_{t-i}, \Delta\} \quad . \tag{3.40}$$

Note that the stationary distribution of such an overshoot is of relevance in the analysis of the other dual-sourcing policies, too. That is why this alternative analysis is presented here. Redefining the random variable $\hat{D}(\Delta)$ as

$$\hat{D}(\Delta) = \sum_{i=0}^{L^f} D_i - \left(\Delta - \sum_{i=L^f+1}^{L^s} \min\{D_i, \Delta\}\right) \tag{3.41}$$

yields the net stock as

$$NS = B^f - \hat{D}(\Delta) \tag{3.42}$$

with

$$\mathbb{E}\left[\hat{D}(\Delta)\right] = \left(L^f + 1\right) \cdot \mu - \left[\Delta - L^\Delta \cdot \left(\mu - \mathbb{E}\left[(D-\Delta)^+\right]\right)\right] \quad . \tag{3.43}$$

The expected on-hand stock and backorders follow from (3.36) and (3.37) with B^s replaced by B^f.

Policy optimization

Due to Lemma 3.3.3.2 the optimal parameter combination (Δ^*, B^{s^*}) can be computed by a one-dimensional search over Δ. The relevant region is $0 \leq \Delta \leq \infty$ (or $\bar{D}$ in case there exists a maximum demand value). The stepsize of Δ corresponds to the difference between two adjacent demand realizations. Since the SIP can mimic both fast and slow single-sourcing order-up-to policies, it holds that $B^s_{min} \leq B^s(\Delta) \leq B^s_{max}$, where B^s_{min} and B^s_{max} are the optimal order-up-to levels to the fast and slow single-sourcing problem, respectively. The following procedure yields the optimal SIP parameters:

1. *For each* Δ*, the optimal* B^s *is computed via (3.35) where* $\hat{D}(\Delta)$ *is given by (3.32).*

2. *The total relevant cost of each parameter combination,* $TRC(\Delta, B^{s^*}(\Delta))$*, is calculated according to (3.3) using (3.36), (3.37), and (3.28).*

3. *The optimal combination* (Δ^*, B^{s^*}) *is found as*

$$(\Delta^*, B^{s^*}) = \underset{(\Delta, B^{s^*}(\Delta))}{\operatorname{argmin}} \; TRC\big(\Delta, B^{s^*}(\Delta)\big) \quad . \tag{3.44}$$

Unfortunately, it cannot be shown that the TRC function is unimodal in Δ. However, all numerical tests conducted so far confirm this (cf. Scheller-Wolf et al. (2007)). Therefore, instead of searching over the entire space, Δ could be increased gradually until the total relevant cost increases for the first time.

Special case: Consecutive lead times

For the case of consecutive lead times, $L^f = m$ and $L^s = m + 1$, $m \in \{0, 1, 2, ...\}$, Whittemore and Saunders (1977) show that for continuous demand the optimal B^f is given by

$$B^{f^*}_{L^\Delta=1} = F^{-1}_{L^f+1}\left(\frac{b-c}{b+h}\right) . \tag{3.45}$$

The optimal B^s or Δ results as the solution to

$$-c + (c-b)F\left(B^s - B^f\right) + (h+b)\int_0^{B^s-B^f} F_{L^f+1}\left(B^s - x\right) f(x)\, dx = 0 \quad , \tag{3.46}$$

which becomes (by using the definition $B^s - B^f = \Delta$)

$$-c + (c-b)F(\Delta) + (h+b)\int_0^{\Delta} F_{L^f+1}\left(B^f + \Delta - x\right) f(x)\, dx = 0 \quad . \tag{3.47}$$

For $m = 0$, this has previously been shown by Bulinskaya (1964) and later also by Zhang and Hausman (1994), and Zhang (1996).

In case of discrete demand, the optimal B^f is the smallest value that satisfies

$$F_{L^f+1}\left(B^f\right) \geq \frac{b-c}{b+h} \quad . \tag{3.48}$$

Similarly, (3.47) can be rewritten and the optimal Δ found as the smallest value that satisfies

$$-c + (c-b)F\left(\Delta\right) + (h+b)\sum_{x=0}^{\Delta} F_{L^f+1}\left(B^f + \Delta - x\right) f(x) \geq 0 \quad , \tag{3.49}$$

which can be rearranged as

$$\sum_{x=0}^{\Delta} F_{L^f+1}\left(B^f + \Delta - x\right) f(x) \geq \frac{c \cdot \bar{F}\left(\Delta\right) + b \cdot F\left(\Delta\right)}{b+h} \tag{3.50}$$

where $\bar{F}(x) = 1 - F(x) = Pr\left\{D > x\right\}$. This reformulation allows for an intuitive interpretation. The left-hand side shows that due to the second supply option, the inventory position (after ordering from the fast supplier) at the beginning of a period is not equal to B^f, but can be higher. Depending on this value, which characterizes the stock availability, holding and backorder costs are incurred, which are represented by $b+h$ in the denominator on the right-hand side as in the standard single-sourcing order-up-to level model. In contrast to the standard model, a distinction has to be made in the nominator, however. If an order is placed with the fast supplier, the expediting premium is incurred. This happens with probability $\bar{F}(\Delta)$. Otherwise, only the standard backorder cost accrues. This finding about the 'critical ratio' might be helpful in the development of simple heuristics.

Extensions

As extensions, the optimization model under various service-level constraints is considered first. Afterwards, further aspects are addressed including the incorporation of capacities on the orders and the sourcing from more than two suppliers.

Service-level model. In this section, the dual-sourcing problem under the SIP is studied, if instead of a backorder cost per unit and period a service-level constraint is used. For didactical reasons the α-service level is addressed first, followed by the

γ one, and finally the β-service level.

In the single-sourcing order-up-to level model it holds that $\alpha = \frac{b}{b+h}$ (see (2.52)). Furthermore, it is shown in this section that the computation of the net stock under the SIP corresponds to the one in the single-sourcing order-up-to level model, if the lead-time demand $D(L+1)$ is replaced by $\hat{D}(\Delta)$. Consequently, the optimal slow order-up-to level under the SIP can be determined by (3.35). Following the same line of argument as in van Houtum and Zijm (2000), it can be shown that the relation $\alpha = \frac{b}{b+h}$ also holds under the SIP. Under this policy, each period orders are placed with both suppliers such that the slow inventory position is raised to B^s. The characterization of (3.35) shows that in the cost model the slow order-up-to level in period t should be chosen such that the probability that no stockout occurs at the end of period $t + L^s$ (i.e. $NS_{t+L^s} \geq 0$, cf. (3.31)) is larger than or equal to $\frac{b}{b+h}$. This probability is just the definition of the α-service level, which establishes the above-mentioned relation. Consequently, the parameter optimization for the α-service level case can be done as described above.

Under a γ-*service level* constraint the optimization problem can be formulated as follows (in general terms)

$$\min \; TRC^{\gamma}_{SIP}(\Delta, B^s) = c \cdot \mathbb{E}\left[Q^f\right] + h \cdot \mathbb{E}\left[OH\right] \tag{3.51}$$

$$s.t. \;\; 1 - \frac{\mathbb{E}\left[BO\right]}{\mu} \geq \gamma^{target} \tag{3.52}$$

$$\Delta, B^s \in \mathbb{N} \tag{3.53}$$

for $\gamma^{target} \in (0,1)$. The integrality constraint on Δ and B^s follows from the discrete nature of demand (and is not required for continuous demand). Using (3.28), (3.33), (3.29), (3.30), and (2.39) the optimization problem can be reformulated as

$$\min \; TRC^{\gamma}_{SIP}(\Delta, B^s) = c \cdot \mathbb{E}\left[(D-\Delta)^+\right] + h \cdot \mathbb{E}\left[\left(B^s - \hat{D}(\Delta)\right)^+\right] \tag{3.54}$$

$$s.t. \;\; \mathbb{E}\left[\left(\hat{D}(\Delta) - B^s\right)^+\right] \leq BO^{\gamma}_{target} \tag{3.55}$$

$$\Delta, B^s \in \mathbb{N} \quad . \tag{3.56}$$

Analyzing the properties of this formulation reveals:

Lemma 3.3.3.3 *For the γ-service level problem, given a fixed Δ:*

1. *The objective function is constant for $B^s \leq 0$ and non-decreasing for $B^s > 0$.*
2. *The average backorders $\mathbb{E}\left[\left(\hat{D}(\Delta) - B^s\right)^+\right]$ are non-increasing in B^s; $\mathbb{E}\left[\left(\hat{D}(\Delta) - B^s\right)^+\right] \uparrow \infty$ as $B^s \to -\infty$; $\mathbb{E}\left[\left(\hat{D}(\Delta) - B^s\right)^+\right] \downarrow 0$ as $B^s \to \infty$.*

Proof:
See Appendix B.2. □

Lemma 3.3.3.3 implies that the smallest B^s satisfying (3.55) is the optimal one for a given Δ. Consequently, the same one-dimensional search procedure over Δ as in the backorder cost approach can be conducted to find the optimal parameter combination.

Scheller-Wolf et al. (2007) show this lemma for *continuous demand* (with $0 < F(x) < 1$ for all $x \in (0, \infty)$ and $F(0) = 0$) together with an additional property that holds, namely: *For a given Δ, there is a unique finite positive value, $B^s(\Delta)$, for which (3.55) is satisfied at equality. At optimality this equality holds: $B^s = B^s(\Delta)$.* Based on this lemma and a lower bound on the optimal value of Δ, which they derive, they develop an optimization procedure for the continuous demand case. Furthermore, for the special case of mixed-Erlang distributed demand they derive exact closed-form expressions for $\mathbb{E}\left[\left(\hat{D}(\Delta) - B^s\right)^+\right]$ and $\mathbb{E}\left[(D - \Delta)^+\right]$ such that the average cost for a given $B^s(\Delta)$ can be easily determined.

In contrast to the γ-service level, the *β-service level* only takes into consideration the new backorders each period. The optimization problem can be formulated as

$$\min \ TRC^{\beta}_{SIP}(\Delta, B^s) = c \cdot \mathbb{E}\left[Q^f\right] + h \cdot \mathbb{E}\left[OH\right] \quad (3.57)$$

$$s.t. \ \ 1 - \frac{\mathbb{E}\left[BO\right] - \mathbb{E}\left[BO^{beg}\right]}{\mu} \geq \beta^{target} \quad (3.58)$$

$$\Delta, B^s \in \mathbb{N} \quad (3.59)$$

for $\beta^{target} \in (0, 1)$. In comparison to the γ-service level formulation an additional expression for $\mathbb{E}\left[BO^{beg}\right]$ is required. To this end, a new random variable is defined

$$\hat{D}^{beg}(\Delta) = \sum_{i=1}^{L^f} D_i + \sum_{i=L^f+1}^{L^s} \min\{D_i, \Delta\} \quad . \tag{3.60}$$

Then,

$$\mathbb{E}\left[BO^{beg}(B^s)\right] = \mathbb{E}\left[\left(\hat{D}^{beg}(\Delta) - B^s\right)^+\right] = \sum_{x=B^s+1}^{\infty} (x - B^s) \cdot Pr\left\{\hat{D}^{beg}(\Delta) = x\right\}. \tag{3.61}$$

Using the specific expressions for the SIP the optimization problem reads:

$$\min \ TRC^{\beta}_{SIP}(\Delta, B^s) = c \cdot \mathbb{E}\left[(D - \Delta)^+\right] + h \cdot \mathbb{E}\left[\left(B^s - \hat{D}(\Delta)\right)^+\right] \tag{3.62}$$

$$s.t. \ \ \mathbb{E}\left[\left(\hat{D}(\Delta) - B^s\right)^+\right] - \mathbb{E}\left[\left(\hat{D}^{beg}(\Delta) - B^s\right)^+\right] \leq BO^{\beta}_{target} \tag{3.63}$$

$$\Delta, B^s \in \mathbb{N} \quad . \tag{3.64}$$

with $BO^{\beta}_{target} = \left(1 - \beta^{target}\right) \mu$. For this optimization problem, an analog of Lemma 3.3.3.3 can be proven (see Scheller-Wolf et al. (2007)) such that the optimal B^s results as the smallest value that satisfies (3.63).

Further aspects. Scheller-Wolf et al. (2007) present extensions of the SIP model to capacitated suppliers and more than two supply modes. Both aspects can be incorporated into the SIP quite easily. If there is a *capacity* CAP^f on the fast orders, Lemma 3.3.3.2 holds with the modification that

$$\hat{D}(\Delta) = \sum_{i=0}^{L^f} D_i + \sum_{i=L^f+1}^{L^s} W_i \tag{3.65}$$

where

$$W_i = \begin{cases} D_i & D_i \leq \Delta, \\ \Delta & \Delta < D_j \leq CAP^f + \Delta, \\ D_i - CAP^f & CAP^f + \Delta < D_i. \end{cases} \quad (3.66)$$

If there is a capacity CAP^s on the slow orders all the results above hold, with the additional constraint that $\Delta \leq CAP^s$.

Also in the case of *more than two supply modes*, the SIP yields a considerable dimensional reduction compared to the optimal policy. For three suppliers, for instance, it follows that: *If there are three supply modes, for each* Δ_1 *and* Δ_2*, given* $L_1^\Delta < L_2^\Delta < L_3^\Delta$ *and* $B_3 = B_2 + \Delta_2, B_2 = B_1 + \Delta_1$*, Lemma 3.3.3.2 holds with the modification that:*

$$\hat{D}(\Delta_1, \Delta_2) = \sum_{i=0}^{L_1} D_i + \sum_{i=L_1+1}^{L_2} \min\{D_i, \Delta_1 + \Delta_2\} + \sum_{i=L_2+1}^{L_3} \min\{D_i, \Delta_2\} \quad . \quad (3.67)$$

3.3.3.2 Constant-order policy

Policy description

The constant-order policy (COP) is studied in Zhang and Hausman (1994), Janssen and de Kok (1999), and Klosterhalfen et al. (2010a). Under this policy, in each period a fixed quantity, $Q_t^s = Q$, is ordered from the slow supplier. The fast order, Q_t^f, is determined according to an order-up-to logic with order-up-to level, B^f. Thus, the decision variables are Q and B^f. The COP generates a constant inflow of Q units from the slow supplier each period, irrespective of L^s. Only L^f is relevant for the inventory position that is used as fast order trigger, IP_t^f. Since $Q_t^s = Q$, (3.4) becomes

$$IP_t^f = NS_{t-1} + (L^f + 1) \cdot Q + \sum_{j=1}^{L^f} Q_{t-j}^f \quad . \quad (3.68)$$

The slow constant order that arrives in period t might cause the inventory position to assume a value above B^f resulting in a so-called overshoot. This overshoot at the beginning of a period after ordering from the fast supplier is denoted as

$$O_t = IP_t^f + Q_t^f - B^f = \left(IP_t^f - B^f\right)^+ \tag{3.69}$$

where $(x)^+ = \max\{x, 0\}$. The order quantities are

$$Q_t^s = Q \tag{3.70}$$

$$Q_t^f = \left(B^f - IP_t^f\right)^+ \quad . \tag{3.71}$$

The COP is very appealing from a practical viewpoint. The constant-order property might facilitate the supply negotiations with the slow supplier. This supplier does not suffer from any demand variability or even the bullwhip effect, for instance, which makes the production planning much easier.

As an extreme case, the COP can mimic single sourcing from the fast supplier with an order-up-to policy, i.e. $Q = 0$. Single sourcing from the slow supplier requires $Q = \mu$, which is not of an order-up-to level type, however. For simplicity reasons, this less interesting boundary case is excluded from the upcoming analysis, because the optimal slow single-sourcing policy for the specified inventory model is known to be of an order-up-to level type (see Section 2.1.2.3), so that $Q = \mu$ cannot be better than that.

Policy evaluation

The evaluation of a parameter combination, (Q, B^f), requires expressions for the expected on-hand stock, backorders, and expedited order quantity. For a given Q, the latter quantity immediately results from (3.2) as

$$\mathbb{E}\left[Q^f\right] = \mu - Q \quad . \tag{3.72}$$

The other two quantities depend on the overshoot in the system, where the following can be noted (cf. Klosterhalfen et al. (2010a)).

Lemma 3.3.3.4 *The overshoot under a COP is a function of Q independent of B^f, $O_t(Q)$.*

Proof:
See Appendix B.3. □

Due to the possible overshoot, the inventory position after ordering in each period t is $IP_t^{f^+} = B^f + O_t(Q)$. Thus, the net stock calculation is

$$NS_t = IP_{t-L^f}^{f+} - \sum_{i=0}^{L^f} d_{t-i} = B^f - \left(\sum_{i=0}^{L^f} d_{t-i} - O_{t-L^f}(Q) \right) \quad . \tag{3.73}$$

Denote $\tilde{D}_t(Q) = \sum_{i=0}^{L^f} d_{t-i} - O_{t-L^f}(Q)$ as the net demand, i.e. the realized demands in periods $t - L^f$ to t (including periods $t - L^f$ and t) are convolved with the negative overshoot at the beginning of period $t - L^f$. Since they are independent, the stationary distribution of $\tilde{D}(Q)$ can be determined as the convolution of the demand over $L^f + 1$ periods with the overshoot random variable, $O(Q)$, which is independent of B^f (Lemma 3.3.3.4), but dependent on Q, i.e.

$$Pr\left\{\tilde{D}(Q) = x\right\} = \sum_{i=0}^{\infty} Pr\left\{D(L^f+1) = x+i\right\} \cdot Pr\left\{O(Q) = i\right\} \quad . \tag{3.74}$$

Using $\tilde{D}(Q)$, the net stock under stationary conditions $t \to \infty$ follows from (3.73) as

$$NS = B^f - \tilde{D}(Q) \quad . \tag{3.75}$$

Thus, as in the SIP case, (3.75) resembles the net stock calculation in an order-up-to level system with lead-time demand $\tilde{D}(Q)$ (see (2.45)). Therefore, the optimal B^f for a given Q, $B^{f^*}(Q)$, is the smallest B^f that satisfies

$$Pr\left\{\tilde{D}(Q) \leq B^f\right\} \geq \frac{b}{b+h} \quad . \tag{3.76}$$

Given $B^{f^*}(Q)$, the expected on-hand stock and backorders are computed according to (2.48) and (2.49) as

$$\mathbb{E}\left[OH(B^{f^*}(Q))\right] = \sum_{x=0}^{B^{f^*}} \left(B^{f^*} - x\right) \cdot Pr\left\{\tilde{D}(Q) = x\right\} \tag{3.77}$$

$$\mathbb{E}\left[BO(B^{f^*}(Q))\right] = \sum_{x=B^{f^*}+1}^{\infty} \left(x - B^{f^*}\right) \cdot Pr\left\{\tilde{D}(Q) = x\right\} \quad . \tag{3.78}$$

It remains to be shown how to derive the stationary overshoot distribution for a given Q, which enters (3.76), (3.77), and (3.78) through the net demand. Three ways to do this are explained here, out of which the first one is exact and the other two are approximations:

1. **One-dimensional Markov Chain**

   According to Klosterhalfen et al. (2010a), the overshoot process can be modeled as a Markov Chain. It is completely defined by recurrence relation (B.2), i.e.

   $$O_{t+1} = (O_t + Q - d_t)^+ \quad , \tag{3.79}$$

   because Q and the demand probability mass function are known. The state space is infinite, $O_t \in \mathcal{SSP} := \{0, 1, ...\}$. The transition probabilities $p_{ij} = Pr\{O_{t+1} = j \mid O_t = i\}$ of the transition matrix $\mathcal{P} = (p_{ij})$ can be obtained by distinguishing between the cases $j = 0$ and $j > 0$. For $j = 0$, it follows from (3.79) and $t \to \infty$ that

   $$p_{i0} = Pr\{i + Q - D \leq 0\} = Pr\{D \geq i + Q\} \quad . \tag{3.80}$$

   Similarly, for $j > 0$:

   $$p_{ij} = Pr\{j = i + Q - D\} = Pr\{D = i + Q - j\} \quad . \tag{3.81}$$

   Let $o_i = Pr\{O(Q) = i\}$, $i \in \mathcal{SSP}$, $\mathbf{o} = [o_0, o_1, ...]$, and $\mathbf{e} = [1, 1, ...]^T$. Then the stationary distribution $\mathbf{o}$ can be obtained as the eigenvector of $\mathcal{P}$ for the eigenvalue 1 (see, e.g., Steward (2009), Chapter 10)

$$\mathbf{o} = \mathcal{P}\mathbf{o} \text{ with } \mathbf{oe} = 1 \quad . \tag{3.82}$$

2. **Queueing model equivalence**

   Janssen and de Kok (1999) find that the overshoot recursion corresponds to the recursion of the waiting time in a $GI/D/1$ queue with the distribution of the inter-arrival times equal to the period demand distribution $F(.)$ and deterministic service Q. An overview of the available literature regarding the waiting time in a $GI/D/1$ queue is given in Chaudhry (1992), which also provides an exact method. However, Janssen and de Kok (1999) consider this exact method as relatively hard to implement. They rather suggest an approximate moment-iteration method developed for $GI/G/1$ queues (see de Kok (1989)). Based on the equivalence relation either the entire overshoot distribution or at least values for the first two moments of the stationary overshoot distribution and the probability of an overshoot can be computed. In the latter case, an appropriate distribution needs to be fitted to these moments again, e.g., a mixed-Erlang distribution.

3. **Simulation**

   Instead of using any of the two above-described methods, the system can be simulated. For a sufficiently large number of periods, the resulting overshoot distribution will be close to the optimal one. This approach is suggested by Veeraraghavan and Scheller-Wolf (2008) in their analysis of the DIP.

Policy optimization

The optimal policy parameters can be computed by a one-dimensional full enumerative search over Q (cf. Janssen and de Kok (1999)). The relevant region is $0 \leq Q < \mu$, which follows from (3.2) and the fact that both order quantities must be non-negative together with the remarks on the extreme cases at the end of the COP 'Policy description' section. In case of discrete integer demand the smallest stepsize of Q is 1. The procedure is as follows:

1. *For each* Q*, the optimal* $B^{f^*}(Q)$ *results from (3.76).*

2. *The total cost of each parameter combination,* $TRC(Q, B^{f^*}(Q))$*, is computed according to (3.3) using (3.77), (3.78), and (3.72).*

3. *The optimal combination,* (Q^*, B^{f^*})*, is found as*

$$(Q^*, B^{f^*}) = \underset{\left(Q, B^{f^*}(Q)\right)}{\operatorname{argmin}} \; TRC(Q, B^{f^*}(Q)) \quad . \tag{3.83}$$

Similar to the SIP case, it could not be shown so far that TRC is unimodal with respect to Q. However, in all numerical tests that have been conducted no counterexample to this assumption has been found, either. When assuming unimodality, not the entire feasible region of Q needs to be searched, but one could simply start with the largest feasible value for Q and decrease it gradually until the total relevant cost increases for the first time.

Special case: Geometric demand

If demand is distributed according to a geometric distribution (see Section 2.1.1), the stationary overshoot distribution is found to be computable by recursive expressions, which have not yet been derived in previous works.

Lemma 3.3.3.5 *If period demand follows a geometric distribution, the steady-state probabilities,* o_i*, of the overshoot distribution can be derived as*

$$\begin{aligned} o_1 = {} & \frac{1}{1 - \sum_{j=Q+1}^{\infty} \mathcal{H}(j)} \\ & \cdot \frac{\left[Pr\{D = Q-1\} - \sum_{j=Q+1}^{\infty} \mathcal{H}(j) Pr\{D = Q+j-1\}\right]}{1 - \left[\sum_{j=1}^{Q} (1-p)^{1-j} Pr\{D = Q+j-1\} + \sum_{j=Q+1}^{\infty} \mathcal{G}(j) Pr\{D = Q+j-1\}\right]} \\ & \cdot \frac{1}{1 + \frac{\sum_{j=1}^{Q}(1-p)^{1-j} + \sum_{j=Q+1}^{\infty}\mathcal{G}(j)}{1-\sum_{j=Q+1}^{\infty}\mathcal{H}(j)} \frac{\left[Pr\{D=Q-1\} - \sum_{j=Q+1}^{\infty}\mathcal{H}(j)Pr\{D=Q+j-1\}\right]}{1-\left[\sum_{j=1}^{Q}(1-p)^{1-j}Pr\{D=Q+j-1\} + \sum_{j=Q+1}^{\infty}\mathcal{G}(j)Pr\{D=Q+j-1\}\right]}} \end{aligned} \tag{3.84}$$

$$o_0 = \frac{1 - o_1 \left(\sum_{j=1}^{Q} (1-p)^{1-j} + \sum_{j=Q+1}^{\infty} \mathcal{G}(j)\right)}{1 - \sum_{j=Q+1}^{\infty} \mathcal{H}(j)} \tag{3.85}$$

$$o_i = o_1(1-p)^{1-i} \qquad i = 2, ..., Q \tag{3.86}$$

$$o_i = o_1\mathcal{G}(i) - o_0\mathcal{H}(i) \qquad i \geq Q+1 \tag{3.87}$$

with

$$\mathcal{G}(i) = \frac{1}{(1-p)^{x(i)Q+x(i)-2+j(i,x)}} + \sum_{n=1}^{x(i)-1}(-1)^n p^n \frac{\mathcal{X}^n_{x(i)}(j(i,x))}{(1-p)^{x(i)Q+x(i)+j(i,x)-2-nQ}} + (-1)^{x(i)}p^{x(i)}\frac{\mathcal{X}^{x(i)}_{x(i)}(j(i,x))}{(1-p)^{x(i)+j(i,x)-2}} \tag{3.88}$$

$$\mathcal{H}(i) = \frac{p}{(1-p)^{(x(i)-1)Q+x(i)-1+j(i,x)}} + \sum_{n=2}^{x(i)-1}(-1)^{n-1} p^n \frac{\mathcal{Y}^n_{x(i)}(j(i,x))}{(1-p)^{x(i)Q+x(i)+j(i,x)-1-nQ}} + (-1)^{x(i)-1}p^{x(i)}\frac{\mathcal{Y}^{x(i)}_{x(i)}(j(i,x))}{(1-p)^{x(i)+j(i,x)-1}} \tag{3.89}$$

and $x(i)$ and $j(i,x)$ computed as

$$x(i) = \left\lfloor \frac{i}{Q+1} \right\rfloor \tag{3.90}$$

$$j(i,x) = i - x(Q+1) + 1 \tag{3.91}$$

and the functions $\mathcal{X}^n_x(j)$ and $\mathcal{Y}^n_x(j)$ defined as follows for $x \geq 1$

$$\mathcal{X}^n_x(j) = \begin{cases} \sum_{k=1}^{j-1} \mathcal{X}^{n-1}_{x-1}(k+1) & x = n \\ \mathcal{X}^n_{x-1}(Q+1) + \sum_{k=1}^{j} \mathcal{X}^{n-1}_{x-1}(k) & x > n \end{cases} \qquad n \geq 1 \tag{3.92}$$

$$\mathcal{X}^0_x = 1 \tag{3.93}$$

$$\mathcal{Y}^n_x(j) = \begin{cases} \sum_{k=1}^{j} \mathcal{Y}^{n-1}_{x-1}(k) & x = n, n \geq 2 \\ \mathcal{Y}^n_{x-1}(Q+1) + \sum_{k=1}^{j} \mathcal{Y}^{n-1}_{x-1}(k) & x > n, n \geq 1 \end{cases} \tag{3.94}$$

$$\mathcal{Y}^1_1 = \begin{cases} 0 & \textit{if } x \textit{ from (3.90) for the first/largest } Q\textit{-cycle is equal to } 1 \\ 1 & \textit{otherwise} \end{cases} \tag{3.95}$$

$$\mathcal{Y}^0_x = 0 \quad . \tag{3.96}$$

Proof:
See Appendix B.4. □

Extensions

As in the SIP section, the extensions refer to service-level models and capacity constraints. Moreover, a short remark is made on fixed ordering costs as well as on a variant of the COP called a standing-order policy.

Service-level model. Based on the same arguments as in the SIP case, the optimal B^f for a given Q can be determined by (3.76) for the *α-service level* problem, too. The optimal Q is found by a full enumerative search.

Similar to the SIP, but adjusted to the COP expressions, the optimization problem under a *γ-service level* constraint can be stated as follows

$$\min \; TRC^{\gamma}_{COP}(Q, B^f) = c \cdot (\mu - Q) + h \cdot \mathbb{E}\left[\left(B^f - \tilde{D}(Q)\right)^+\right] \tag{3.97}$$

$$s.t. \;\; \mathbb{E}\left[\left(\tilde{D}(Q) - B^f\right)^+\right] \leq BO^{\gamma}_{target} \tag{3.98}$$

$$Q, B^f \in \mathbb{N} \quad . \tag{3.99}$$

The properties stated in Lemma 3.3.3.3 for the SIP also hold for the COP, which can be shown as follows. The expected backorders on the left-hand side can be rewritten using (3.74):

$$\mathbb{E}\left[\left(\tilde{D}(Q) - B^f\right)^+\right] = \sum_{i=0}^{\infty} \mathbb{E}\left[\left(D(L^f+1) - B^f - i\right)^+\right] Pr\left\{O(Q) = i\right\} \quad . \tag{3.100}$$

Similarly, the on-hand stock expression in the objective function is

$$\mathbb{E}\left[\left(B^f - \tilde{D}(Q)\right)^+\right] = \sum_{i=0}^{\infty} \mathbb{E}\left[\left(B^f + i - D(L^f+1)\right)^+\right] Pr\left\{O(Q) = i\right\} \quad . \tag{3.101}$$

Expression (3.101) is non-decreasing in B^f, which is easy to see by recalling that probabilities are non-negative and using finite differences. This implies that the smallest integer B^f that satisfies (3.98) is the optimal solution to the optimization problem. Since it can also be shown that (3.100) is non-increasing in B^f, the optimal B^f can be found by a simple numerical method like a bisection procedure. The optimal Q is found by the search procedure described in the backorder cost approach.

The optimization problem for the *β-service level* is:

$$\min \; TRC^{\beta}_{COP}(Q, B^f) = c \cdot (\mu - Q) + h \cdot \mathbb{E}\left[\left(B^f - \tilde{D}(Q)\right)^+\right] \tag{3.102}$$

$$s.t. \; \mathbb{E}\left[\left(\tilde{D}(Q) - B^f\right)^+\right] - \mathbb{E}\left[\left(\tilde{D}(Q)^{beg} - B^f\right)^+\right] \leq BO^{\beta}_{target} \tag{3.103}$$

$$Q, B^f \in \mathbb{N} \tag{3.104}$$

with $BO^{\beta}_{target} = (1 - \beta^{target})\,\mu$ for $\beta^{target} \in (0,1)$, where the random variable $\tilde{D}(Q)^{beg}$ is defined as

$$Pr\left\{\tilde{D}(Q)^{beg} = x\right\} = \sum_{i=0}^{\infty} Pr\left\{D(L^f) = x + i\right\} \cdot Pr\left\{O(Q) = i\right\} \quad . \tag{3.105}$$

Again, the properties of Lemma 3.3.3.3 adjusted to the COP case hold, which determines the optimal B^f as the smallest value that satisfies (3.103).

Further aspects. A *fixed ordering cost* for each supplier can also be incorporated in the COP. This is shown by Janssen and de Kok (1999), who consider a system with a linear holding cost and a β-service level constraint.

Capacity constraints on the orders have not yet been considered in the COP. The integration of a capacity CAP^s on the slow order is straightforward. Simply the constraint $Q \leq CAP^s$ needs to be added, which can be done without affecting any of the above results. A capacity CAP^f on the fast order requires further analysis. Due to such a constraint the fast order might not be able to bring the fast inventory position up to the fast order-up-to level, i.e. fast shortfall occurs. By using similar arguments as Veeraraghavan and Scheller-Wolf (2008) do in the DIP case (see their Lemma 5.1, p. 854) it can be shown that the shortfall as well as the overshoot are functions of Q, which are independent of B^f. The optimal B^f for a given Q can be computed as the solution to a critical fractile inequality similar to (3.76). However, instead of the net demand, a stationary distribution of the convolution of the fast lead-time demand, the fast shortfall, and the negative overshoot is required. This distribution can be determined easily by the use of simulation. The optimization of the inventory control parameters can be done by a full enumerative search over Q.

A similar, but slightly different variant of the COP is the *standing-order policy*. Such a policy not only specifies a fixed order quantity and an order-up-to level, but also a dispose-down-to level. It is studied in Rosenshine and Obee (1976) and Chiang (2007), who show how to compute the optimal values.

3.3.3.3 Dual-index policy

Policy description

The dual-index policy (DIP), which is analyzed by Veeraraghavan and Scheller-Wolf (2008), Arts et al. (2009), and Klosterhalfen et al. (2010a), specifies two order-up-to levels, one for the slow (B^s) and one for the fast supplier (B^f). As in the SIP case, the difference is denoted as $\Delta = B^s - B^f$. Thus, as decision variables either (B^f, B^s), (B^f, Δ), or (Δ, B^s) can be used. For ease of presentation of the upcoming exposition (B^f, Δ) is chosen. For the execution of replenishment orders, the DIP keeps track of two inventory positions, IP_t^f and IP_t^s, which are defined in (3.4) and (3.5), respectively. That means each inventory position at the beginning of a period t is given by the net stock at the end of the previous period plus all outstanding orders with any of the two suppliers that will arrive no later than the order of the specific supplier, which is to be determined in t. In each period, the inventory position of the fast supplier, IP_t^f, is checked first and a potential fast order is placed. (Note that the inventory position of the slow supplier, IP_t^s, does not include this fast order.) The order quantities are given as

$$Q_t^f = \left(B^f - IP_t^f\right)^+ \tag{3.106}$$

$$Q_t^s = B^s - \left(IP_t^s + Q_t^f\right) \quad . \tag{3.107}$$

$IP_t^s + Q_t^f$ is always equal to or larger than B^f. Thus, the maximum regular order quantity is Δ. For a lead-time difference of one, both inventory positions are identical. Hence, the DIP reduces to the SIP, which represents the optimal policy.

Moreover, the DIP can mimic both single-sourcing strategies in the form of order-up-to policies by either setting $B^f = B^s \Rightarrow \Delta = 0$ and B^f as the solution to (2.50) with $L = L^f$ (*fast single sourcing*) or B^s as the solution to (2.50) with $L = L^s$ and $B^f = -\infty \Rightarrow \Delta = \infty$ (*slow single sourcing*). Note that the DIP *cannot* mimic the ordering behavior of the COP. In the DIP both orders vary according to the demand variability and the outstanding orders in the system. In the COP only the fast order varies, but not the slow one.

Policy evaluation

As in the COP case, it is possible in the DIP that the fast inventory position in period t lies above the respective order-up-to level, B^f, representing an overshoot, because the slow order placed $t - L^\Delta$ periods ago enters IP_t^f. The same applies to periods $t+1, ..., t+L^\Delta$, because the slow orders that will enter the fast inventory position calculation in these periods are already known at time t, but not taken into account in IP_t^f. In the DIP, Veeraraghavan and Scheller-Wolf (2008) show in their Proposition 4.1 (p. 854) that the overshoot only depends on Δ, but is independent of B^f. This is denoted as $O(\Delta)$ and it is obvious that the maximum overshoot is Δ. Consequently, once the stationary distribution of $O(\Delta)$ is known, the net demand distribution can be computed as in the COP case via (3.74). The optimal B^f for a given Δ, $B^{f^*}(\Delta)$, follows from (3.76) and the expected on-hand stock and backorder quantities from (3.77) and (3.78). The fast order quantity, $\mathbb{E}\left[Q^f\right]$, which is also required for the computation of $TRC(\Delta, B^{f^*}(\Delta))$ according to (3.3), can be derived as follows for a given Δ. Using (3.4), (3.5) can be rewritten as

$$IP_t^s = IP_t^f + \sum_{i=1}^{L^\Delta - 1} Q_{t-i}^s \quad . \tag{3.108}$$

After placement of the slow order in t the relation becomes

$$B^s = B^f + O_t + \sum_{i=1}^{L^\Delta - 1} Q_{t-i}^s + Q_t^s \quad . \quad \Rightarrow \quad O_t = \Delta - \sum_{i=0}^{L^\Delta - 1} Q_{t-i}^s \tag{3.109}$$

$$\Rightarrow \mathbb{E}\left[O\right] = \Delta - L^\Delta \cdot \mathbb{E}\left[Q^s\right] \tag{3.110}$$

Once the stationary overshoot distribution is known, $\mathbb{E}\left[O\right]$ is known as well (or can be computed). From (3.110), $\mathbb{E}[Q^s]$ follows and through relation (3.2) $\mathbb{E}\left[Q^f\right]$ is found.

Note that the term $\sum_{i=0}^{L^\Delta-1} Q^s_{t-i}$ in (3.109) represents the pipeline stock after ordering in period t that will not arrive within the fast lead time. Denote this quantity as A_t. The following interesting relation holds (see Arts et al. (2009), Lemma 4.1):

$$\Delta = O_t + A_t \quad . \tag{3.111}$$

This relation shows that for a given Δ, any knowledge about A_t implies knowledge about O_t, which is useful for the derivation of the stationary overshoot distribution described next.

The stationary overshoot distribution can be computed by using different approaches. The first one presented is exact, whereas the other two are approximations.

1. **Multi-dimensional Markov Chain**

   According to Klosterhalfen et al. (2010a) the exact stationary overshoot distribution can be computed via a multi-dimensional Markov Chain. Similar to (3.79) and proven by Veeraraghavan and Scheller-Wolf (2008) the overshoot in the DIP satisfies

   $$O_{t+1} = \left(O_t + Q^s_{t-(L^\Delta-1)} - d_t\right)^+ \quad . \tag{3.112}$$

   As explained above the overshoot is caused by the outstanding regular orders that are already determined at time t and will arrive in future periods, but are not included in the fast inventory position, i.e. $Q^s_t, Q^s_{t-1}, ..., Q^s_{t-(L^\Delta-1)}$. Consequently, these orders need to be stored in the state information. Given these orders and the demand probability mass function, which is known, the Markov Chain is completely defined by (3.112). Hence, the state is described by a $(L^\Delta+1)$-dimensional vector

   $$\mathbf{Z}_t = \left(Q^s_t, Q^s_{t-1}, ..., Q^s_{t-(L^\Delta-2)}, Q^s_{t-(L^\Delta-1)}, O_t\right) \quad . \tag{3.113}$$

   Note that the state is defined after ordering, because the overshoot distribution and not the distribution of the inventory position before ordering is of interest.

From (3.109) it follows that the sum of all state variables equals Δ. Consequently, the state space is given by all possible state variable combinations that fulfill this condition and its size is

$$\sum_{k=0}^{\Delta} \sum_{x_1=0}^{k} \sum_{x_2=0}^{k-x_1} \cdots \sum_{x_{L^\Delta}=0}^{k-\sum_{i=1}^{L^\Delta-1} x_i} \binom{k}{x_1, x_2, ..., x_{L^\Delta}} \tag{3.114}$$

where

$$\binom{k}{x_1, x_2, ..., x_{L^\Delta}} = \frac{k!}{x_1! x_2! \cdots x_{L^\Delta}!} \quad \text{and} \quad \sum_{i=1}^{L^\Delta} x_i = k \quad . \tag{3.115}$$

In period $t+1$, the state is

$$\mathbf{Z}_{t+1} = \left(Q^s_{t+1}, Q^s_t, ..., Q^s_{t-(L^\Delta-3)}, Q^s_{t-(L^\Delta-2)}, O_{t+1} \right) \quad . \tag{3.116}$$

With regard to the determination of the transition probabilities $p_{\mathbf{ij}} = Pr\{\mathbf{Z}_{t+1} = \mathbf{j} \mid \mathbf{Z}_t = \mathbf{i}\}$, it is important to note that $p_{\mathbf{ij}} > 0$, if (and only if) the state vector elements of $\mathbf{j}$ at positions 2 to L^Δ correspond to the ones of $\mathbf{i}$ at positions 1 to $(L^\Delta - 1)$. Otherwise, $p_{\mathbf{ij}} = 0$. Under stationary conditions $t \to \infty$, $\mathbf{Z} = (Q^s_1, ..., Q^s_{L^\Delta}, O)$. Then, using $Q^s_{t+1} = O_t - O_{t+1} + Q^s_{t-(L^\Delta-1)}$, which follows from (3.113), (3.116), and (3.109), and $O_{t+1} = \left(O_t + Q^s_{t-(L^\Delta-1)} - d_t \right)^+$ yields

$$p_{\mathbf{ij}} = \begin{cases} Pr\left\{D = Q^s_1(\mathbf{j})\right\} & \text{if } O(\mathbf{j}) > 0, Q^s_k(\mathbf{j}) = Q^s_{k-1}(\mathbf{i}) \ \forall k = 2, ..., L^\Delta \\ Pr\left\{D \geq Q^s_1(\mathbf{j})\right\} & \text{if } O(\mathbf{j}) = 0, Q^s_k(\mathbf{j}) = Q^s_{k-1}(\mathbf{i}) \ \forall k = 2, ..., L^\Delta \\ 0 & \text{otherwise} \end{cases} \quad . \tag{3.117}$$

As in the COP case, the stationary overshoot distribution can be obtained by solving equations (3.82).

2. **One-dimensional Markov Chain approximation**

   Arts et al. (2009) provide a one-dimensional Markov Chain approximation for the overshoot with $\Delta + 1$ states. Instead of studying O_t, they study A_t, for which the following recurrence relation holds:

$$A_{t+1} = \Delta - O_{t+1} = \min\left\{\Delta, A_t - Q^s_{t+1-L^\Delta} + d_t\right\} \quad . \tag{3.118}$$

The Markov Chain for A_t is defined by this relation, if the probability mass function of D and $\left\{Q^s_{t+1-L^\Delta} \mid A_t\right\}$ are known. Then, the transition probabilities $p_{ij} = Pr\{A_{t+1} = j \mid A_t = i\}$ can be obtained by distinguishing between the cases $j < \Delta$ and $j = \Delta$ (see Arts et al. (2009)). For $j < \Delta$ it follows that

$$\begin{aligned} p_{ij} &= Pr\left\{A_{t+1} = j \mid A_t = i\right\} \\ &= \ldots \\ &= \sum_{k=0}^{j} Pr\left\{Q^s_{t+1-L^\Delta} = i + k - j \mid A_t = i\right\} \; Pr\left\{D = k\right\} \quad . \end{aligned} \tag{3.119}$$

In case $j = \Delta$:

$$\begin{aligned} p_{i\Delta} &= Pr\left\{A_{t+1} = \Delta \mid A_t = i\right\} \\ &= \ldots \\ &= \sum_{k=0}^{i} Pr\left\{Q^s_{t+1-L^\Delta} = k \mid A_t = i\right\} \; Pr\left\{D \geq \Delta + k - i\right\} \quad . \end{aligned} \tag{3.120}$$

The transition probabilities form the transition matrix $\mathcal{P} = (p_{ij})$

$$\mathcal{P} = \begin{pmatrix} p_{00} & \cdots & p_{0\Delta} \\ \vdots & \ddots & \vdots \\ p_{\Delta 0} & \cdots & p_{\Delta\Delta} \end{pmatrix} \quad . \tag{3.121}$$

The stationary distribution results as described above (see (3.82)). The difficulty is that the demand probability mass function is known, but the distribution of $\left\{Q^s_{t+1-L^\Delta} \mid A_t\right\}$ is not. The latter one can be approximated based on the following limiting results (see Proposition 4.5 in Arts et al. (2009)):

a) As $\Delta \to \infty$,
$Pr\left\{Q^s_{t+1} = x\right\} \to Pr\left\{D_t = x\right\}$.

b) As $\Delta \to \infty$,
$Pr\left\{Q^s_{t+1-L^\Delta} = x \mid A_t = y\right\} \to Pr\left\{D_{t+1-L^\Delta} = x \mid \sum_{i=t+1-L^\Delta}^{t} D_i = y\right\}$.

c) For $\Delta = 1$,
$Pr\left\{Q^s_{t+1-L^\Delta} = x \mid A_t = y\right\} \to Pr\left\{D_{t+1-L^\Delta} = x \mid \sum_{i=t+1-L^\Delta}^{t} D_i = y\right\}$.

The first two results are obvious, because $\Delta = \infty$ corresponds to slow single sourcing. In this case it holds that $Q^s_{t+1} = d_t$. Since the second and third results show that for very large and very small Δ $Pr\left\{Q^s_{t+1-L^\Delta} = x \mid A_t = y\right\}$ is given by $Pr\left\{D_{t+1-L^\Delta} = x \mid \sum_{i=t+1-L^\Delta}^{t} D_i = y\right\}$, this expression is used to approximate $Pr\left\{Q^s_{t+1-L^\Delta} = x \mid A_t = y\right\}$ in general:

$$Pr\left\{Q^s_{t+1-L^\Delta} = x \mid A_t = y\right\} \approx Pr\left\{D_{t+1-L^\Delta} = x \left| \sum_{i=t+1-L^\Delta}^{t} D_i = y\right.\right\}$$
$$= \frac{Pr\left\{D = x\right\} \; Pr\left\{D(L^\Delta - 1) = y - x\right\}}{Pr\left\{D(L^\Delta) = y\right\}}. \tag{3.122}$$

With the help of this approximation the stationary distribution of A can be computed, $Pr\{A = x\}$, and the stationary overshoot distribution results as $Pr\left\{O = x\right\} = Pr\left\{A = \Delta - x\right\}$ due to relation (3.111). Note that this approximation is exact for $L^\Delta = 1$, where the DIP reduces to the SIP, which is the optimal policy.

3. **Simulation**

   As in the COP case, the overshoot distribution can also be computed via simulation (cf. Veeraraghavan and Scheller-Wolf (2008)). It has to be ensured that a sufficiently large number of periods is simulated in order to obtain a resulting distribution close to the exact one.

Given the stationary overshoot distribution, the net demand distribution results from (3.74).

Policy optimization

Since the overshoot only depends on Δ (but not on B^f), the optimal policy parameters are found by a one-dimensional search over Δ as in the SIP case. Again, the smallest stepsize of Δ is given by the difference of two adjacent demand realizations in case of a discrete demand distribution. The procedure is as follows:

1. *For each Δ, the stationary overshoot distribution is determined.*
2. *$B^{f^*}(\Delta)$ then follows from (3.76) and $B^{s^*} = B^{f^*}(\Delta) + \Delta$.*
3. *The globally optimal parameters result as*

$$(\Delta^*, B^{f^*}) = \underset{\left(\Delta, B^{f^*}(\Delta)\right)}{\operatorname{argmin}} \; TRC(\Delta, B^{f^*}(\Delta)) \tag{3.123}$$

where the cost components of TRC are computed via (3.77), (3.78), and $\mathbb{E}\left[Q^f\right]$ via $\mathbb{E}\left[Q^s\right]$ from (3.110) and relation (3.2).

In contrast to the COP, where Q is bounded from below by 0 and from above by μ, only a (finite) *lower* bound, 0, exists for Δ in the DIP. The *upper* bound is $\Delta = \infty$ ($\widehat{=}$ slow single sourcing). Although no proof for TRC being unimodal in Δ is available yet, no counterexample has been observed, either (see Veeraraghavan and Scheller-Wolf (2008)). Consequently, assuming that TRC is unimodal in Δ, a simple numerical search procedure like a golden section search could be performed to find the optimal value of Δ as already outlined for the SIP.

Analogy of the DIP and an order-up-to policy in a lost-sales inventory model

As previously mentioned, Sheopuri et al. (2010) show that the dual-sourcing problem is a generalization of the lost-sales problem. The orders placed with the fast supplier can be interpreted as 'lost sales' for the slow supplier. Due to this analogy, the DIP can be connected to a lost-sales model with an order-up-to policy in the way shown in Table 3.1. Let B^{LS} denote the order-up-to level in the lost-sales case. The overshoot process corresponds to the evolution of the on-hand stock at the end of a period.

DIP model	Lost-sales model
Δ	B^{LS} (order-up-to level)
O (overshoot)	on-hand stock at the end of a period
$\mathbb{E}\left[Q^s\right]$	expected average demand satisfied per period
$\mathbb{E}\left[Q^f\right]$	expected average demand lost per period
$L^{\Delta}-1$	lead time

Table 3.1: Relationship between DIP model and lost-sales model

That means, any results derived for the lost-sales problem can be transferred to the dual-sourcing problem. This relation is exploited in the next section, for instance.

Special case: Geometric demand

For the special case of geometric period demand, Johansen and Thorstenson (2008) present a closed-form expression for the distribution of the on-hand stock at the end of a period in a lost-sales model with an order-up-to policy. Since the on-hand stock process corresponds to the overshoot process in the DIP model (see Table 3.1), a direct computation of the overshoot distribution is possible for geometric demand.

Recall from Section 2.1.1 that if the one period demand has a geometric distribution with parameter p, the L-period demand has a negative binomial distribution with parameters L and p and is defined as follows:

$$f_L^{geom}(x) = f^{nbin}(x) = \binom{x+L-1}{L-1} p^L (1-p)^x \qquad x = 0, 1, 2, \dots \quad , \tag{3.124}$$

which for $L = 1$ returns the geometric distribution. The L-period cumulative distribution function results accordingly. Define

$$c^{LS} = \frac{1}{F_{L^{\Delta}-1}^{geom}(\Delta)} \quad . \tag{3.125}$$

The stationary probabilities of the overshoot are computed as

$$o_i = \begin{cases} 1 - c^{LS} F_{L^{\Delta}}^{geom}(\Delta - 1) & i = 0 \\ c^{LS} f_{L^{\Delta}}^{geom}(\Delta - i) & 1 \leq i \leq \Delta \end{cases} \quad . \tag{3.126}$$

The long-run average is given as

$$\mathbb{E}[O] = c^{LS} \sum_{i=1}^{\Delta} F_{L^{\Delta}}^{geom}(i-1) \quad . \tag{3.127}$$

The expected slow order quantity can be computed as

$$\mathbb{E}[Q^s] = c^{LS} \left(\sum_{i=1}^{\Delta} F_{L^{\Delta}-1}^{geom}(i-1) - \sum_{i=1}^{\Delta} F_{L^{\Delta}}^{geom}(i-1) \right) \tag{3.128}$$

or via (3.110) using (3.127). Then, the expected fast order quantity, $\mathbb{E}\left[Q^f\right]$, results via (3.2).

Extensions

As with the other two policies, the extensions cover the analysis of the service-level model formulations followed by capacity constraints and further aspects.

Service-level model. If an *α-service level* constraint is used, the optimal B^f for a given Δ results from (3.76) (using the net demand distribution computed for this Δ) due to the equivalence of $\alpha = \frac{b}{b+h}$, which can be shown following the same line of argument as in the SIP. Consequently, the same optimization procedure as described for the backorder cost problem can be applied.

Under a *γ-service level* constraint, the optimization problem can be formulated similarly to the SIP and COP case as (see Arts et al. (2009)):

$$\min \; TRC_{DIP}^{\gamma}(\Delta, B^f) = c \cdot \mathbb{E}\left[Q^f\right] + h \cdot \mathbb{E}\left[\left(B^f - \tilde{D}(\Delta)\right)^+\right] \tag{3.129}$$

$$s.t. \; \mathbb{E}\left[\left(\tilde{D}(\Delta) - B^f\right)^+\right] \leq BO_{target}^{\gamma} \tag{3.130}$$

$$\Delta, B^f \in \mathbb{N} \quad . \tag{3.131}$$

Since the expected fast order quantity is fixed for a given Δ, $c \cdot \mathbb{E}\left[Q^f\right]$ becomes a fixed constant for a given Δ. Moreover, the properties of Lemma 3.3.3.3 also hold for the DIP under a γ-service level constraint. The same argumentation as in the COP case can be pursued. That means, the smallest integer B^f that satisfies (3.130) is the optimal solution to the optimization problem and it can be found by a simple numerical method like a bisection procedure.

The optimization problem for the *β-service level* is given as:

$$\min\ TRC^{\beta}_{DIP}(\Delta, B^f) = c \cdot \mathbb{E}\left[Q^f\right] + h \cdot \mathbb{E}\left[\left(B^f - \tilde{D}(\Delta)\right)^+\right] \tag{3.132}$$

$$s.t.\ \ \mathbb{E}\left[\left(\tilde{D}(\Delta) - B^f\right)^+\right] - \mathbb{E}\left[\left(\tilde{D}(\Delta)^{beg} - B^f\right)^+\right] \leq BO^{\beta}_{target} \tag{3.133}$$

$$\Delta, B^f \in \mathbb{N} \tag{3.134}$$

with $BO^{\beta}_{target} = (1 - \beta^{target})\,\mu$ for $\beta^{target} \in (0,1)$, where the random variable $\tilde{D}(\Delta)^{beg}$ is defined as

$$Pr\left\{\tilde{D}(\Delta)^{beg} = x\right\} = \sum_{i=0}^{\infty} Pr\left\{D(L^f) = x + i\right\} \cdot Pr\left\{O(\Delta) = i\right\} \quad . \tag{3.135}$$

It can be shown that the properties of Lemma 3.3.3.3 adjusted to the DIP case still hold. Consequently, the optimal B^f is the smallest value that satisfies (3.133).

Further aspects. Veeraraghavan and Scheller-Wolf (2008) extend the DIP to incorporate *capacity constraints* on the orders placed with the fast and slow supplier. They show that, also under these circumstances, the overshoot process is independent of B^f and only dependent on Δ. Furthermore, since they use simulation to determine the stationary overshoot distribution, their method can also be applied to settings with *non-stationary demand, random stoppages, random yield*, and certain types of *lead-time variability*.

Arts et al. (2009) also address the aspect of lead-time variability in the DIP. They allow for a *stochastic integer lead time* of the slow supplier. Provided that the lower bound of the support of this random variable is at least $L^f + 1$, it is shown that the separability result of the deterministic lead-time case still holds, i.e. the overshoot and order processes only depend on Δ and not on the concrete values of B^f and B^s. Furthermore, approximations of the transition probabilities for the one-dimensional Markov Chain approach (see above) based on limiting results are derived.

3.3.3.4 Order-splitting policy

Policy description

In the literature, order-splitting policies (OSPs) are usually studied in deterministic demand, stochastic lead-time settings as an effective means to pool lead-time risk. A recent overview is presented in Thomas and Tyworth (2006). In contrast to the previously described policies, which could also be used to this end, an OSP has a simpler policy structure and therefore lends itself to a more thorough analytical investigation. It only considers the total order quantity in a period. This quantity corresponds to the demand of the previous period, because (in this thesis) it is assumed that an order-up-to policy is in place, which uses the slow inventory position as order trigger. The decision, which remains, is how to allocate the total order quantity across the two suppliers. The proportion that is sourced from each supplier is called the allocation or sourcing fraction. (In the remainder of the thesis only the latter term will be used.)

Since most contributions consider deterministic demand models, the (optimal) sourcing fraction is determined ex ante according to the demand and cost parameter settings. No periodical adjustment of the fraction takes place. In situations with stochastic demand and deterministic lead times, such a 'fixed' OSP is rather unusual, because it is presumably ineffective due its inflexibility to adjust the sourcing fraction. Intuitively, this can be explained as follows. Imagine several successive periods with low demand. The OSP would place orders with the fast and more expensive supplier, even though no stockout is imminent. In such situations, the previously mentioned policies (SIP, COP, DIP) would not order anything from the fast supplier and thus incur a lower cost.

The reason why such a fixed OSP is presented here, anyway, is its analytical tractability. This aspect is less important in the single-echelon context studied in this chapter, but becomes highly relevant in the multi-echelon setting analyzed in Chapter 4.5. Moreover, an analysis of the OSP with deterministic lead times, as presented in this section, is not available in the literature.

Policy evaluation

Let δ^m where $m \in \{f, s\}$ denote the sourcing fraction with the fast and slow supplier. The following logical constraints need to hold:

$$\delta^m \geq 0 \qquad m \in \{f, s\} \tag{3.136}$$

$$\delta^f + \delta^s = 1 \quad . \tag{3.137}$$

In each period, the order quantities with both suppliers are given as

$$Q_t^f = \delta^f \cdot d_{t-1} = (1 - \delta^s) \cdot d_{t-1} \tag{3.138}$$

$$Q_t^s = \delta^s \cdot d_{t-1} \quad , \tag{3.139}$$

i.e. they only depend on the sourcing fractions. On average, the expected order quantities are

$$\mathbb{E}\left[Q^f\right] = \delta^f \cdot \mu = (1 - \delta^s) \cdot \mu \tag{3.140}$$

$$\mathbb{E}\left[Q^s\right] = \delta^s \cdot \mu \quad . \tag{3.141}$$

Next, an expression for the net stock is derived. W.l.o.g. it is assumed that $NS_0 = B^s$.

$$\begin{aligned} NS_t &= B^s - \sum_{i=0}^{t} d_{t-i} + \sum_{i=L^s}^{t} Q_{t-i}^s + \sum_{i=L^f}^{t} Q_{t-i}^f \\ &= B^s - \sum_{i=0}^{t} d_{t-i} + \delta^s \cdot \sum_{i=L^s+1}^{t} d_{t-i} + \delta^f \cdot \sum_{i=L^f+1}^{t} d_{t-i} \\ &= B^s - \sum_{i=0}^{L^s} d_{t-i} + \delta^f \cdot \sum_{i=L^f+1}^{L^s} d_{t-i} \\ &= B^s - \sum_{i=0}^{L^f} d_{t-i} - \delta^s \cdot \sum_{i=L^f+1}^{L^s} d_{t-i} \quad . \end{aligned} \tag{3.142}$$

Define

$$\check{D}(\delta^s) = D(L^f + 1) + \delta^s D(L^\Delta) \quad . \tag{3.143}$$

Then, under stationary conditions $t \to \infty$ the net stock results as

$$NS = B^s - \check{D}(\delta^s) \tag{3.144}$$

with

$$\mathbb{E}\left[\check{D}(\delta^s)\right] = \left((L^f + 1) + \delta^s L^\Delta\right)\mu \tag{3.145}$$
$$\text{VAR}\left[\check{D}(\delta^s)\right] = \left((L^f + 1) + [\delta^s]^2 L^\Delta\right)\sigma^2 \quad . \tag{3.146}$$

(3.144) shows that for a given δ^s, the net stock only depends on B^s. Consequently, given the stationary distribution of $\check{D}(\delta^s)$ the optimal B^s follows from (3.35) with $\hat{D}(\Delta)$ replaced by $\check{D}(\delta^s)$. The expected on-hand stock and backorders for $B^{s^*}(\delta^s)$ are computed according to (2.48) and (2.49) as

$$\mathbb{E}\left[OH(B^{s^*}(\delta^s))\right] = \sum_{x=0}^{B^{s^*}} \left(B^{s^*} - x\right) \cdot Pr\left\{\check{D}(\delta^s) = x\right\} \tag{3.147}$$
$$\mathbb{E}\left[BO(B^{s^*}(\delta^s))\right] = \sum_{x=B^{s^*}+1}^{\infty} \left(x - B^{s^*}\right) \cdot Pr\left\{\check{D}(\delta^s) = x\right\} \quad . \tag{3.148}$$

Policy optimization

According to (3.3),

$$TRC_{OSP}(\delta^s, B^{s^*}(\delta^s)) =$$
$$h\sum_{x=0}^{B^{s^*}} \left(B^{s^*} - x\right) Pr\left\{\check{D}(\delta^s) = x\right\} + b \sum_{x=B^{s^*}+1}^{\infty} \left(x - B^{s^*}\right) Pr\left\{\check{D}(\delta^s) = x\right\} + c\,(1 - \delta^s)\,\mu. \tag{3.149}$$

Lemma 3.3.3.6 *If the critical ratio assumes a value such that a positive safety stock is required, i.e.* $B^{s^*}(\delta^s) \geq \mathbb{E}\left[\check{D}(\delta^s)\right]$*, and the demand distribution is strongly unimodal, the* TRC_{OSP} *function is unimodal in* δ^s*.*

Proof:

See Appendix B.5. □

If TRC_{OSP} is unimodal in δ^s, the optimal δ^s can be found by a simple one-dimensional search procedure like a golden section search over the feasible region $0 \leq \delta^s \leq 1$. Unfortunately, this could only be shown so far for critical ratios that require a positive safety stock, i.e. $B^{s^*}(\delta^s) \geq \mathbb{E}\left[\check{D}(\delta^s)\right]$, and strongly unimodal demand distributions, i.e. a distribution that is still unimodal after convolution, which applies to all distributions analyzed in this thesis (Lemma 3.3.3.6).

If these prerequisites are not met, discretization needs to be applied to δ^s. Then, the optimal inventory control parameters can be found by a full-enumerative one-dimensional search over δ^s over the feasible region $0 \leq \delta^s \leq 1$. The procedure is as follows:

1. *For each δ^s, compute the respective B^{s^*} from (3.35) with $\hat{D}(\Delta)$ replaced by $\check{D}(\delta^s)$.*

2. *The globally optimal parameters are determined as*

$$(\delta^{s^*}, B^{s^*}) = \underset{\left(\delta^s, B^{s^*}(\delta^s)\right)}{\operatorname{argmin}} \; TRC_{OSP}\big(\delta^s, B^{s^*}(\delta^s)\big) \quad . \tag{3.150}$$

Special case: Normally-distributed demand

If period demand is normally distributed, the demand random variable $\check{D}$ can be standardized (see Section 2.1.1), which yields

$$k^{\check{D}} = \frac{x - \mathbb{E}\left[\check{D}(\delta^s)\right]}{\sqrt{\mathrm{VAR}\left[\check{D}(\delta^s)\right]}} = \frac{x - \left((L^f+1) + \delta^s L^{\Delta}\right)\mu}{\sigma\sqrt{(L^f+1) + [\delta^s]^2 L^{\Delta}}} \quad . \tag{3.151}$$

Applying standardization to B^s gives

$$\begin{aligned} k &= \frac{B^s - \left((L^f+1) + \delta^s L^{\Delta}\right)\mu}{\sigma\sqrt{(L^f+1) + [\delta^s]^2 L^{\Delta}}} \\ \Rightarrow \; B^s &= \left((L^f+1) + \delta^s L^{\Delta}\right)\mu + k \cdot \sigma\sqrt{(L^f+1) + [\delta^s]^2 L^{\Delta}} \quad . \end{aligned} \tag{3.152}$$

Similar to the other policies, the optimal order-up-to level for a given δ^s can be determined according to (3.35) with $\hat{D}(\Delta)$ replaced by $\check{D}(\delta^s)$. Due to the continuous nature of the demand, equality must hold in (3.35). Using k and the standard normal probability density and cumulative distribution function, $\phi(x)$ and $\Phi(x)$, instead of B^s and $\check{D}(\delta^s)$, the optimal k results as

$$F_{\check{D}(\delta^s)}(B^s) = \frac{b}{b+h} \;\Rightarrow\; \Phi\left(k\right) = \frac{b}{b+h} \;\Rightarrow\; k = \Phi^{-1}\left(\frac{b}{b+h}\right) \quad . \tag{3.153}$$

Applying standardization to the expected backorder integral yields

$$\mathbb{E}\left[BO\right] = \sigma\sqrt{(L^f+1)+[\delta^s]^2L^\Delta}\int_k^\infty (x-k)\phi(x)\;dx \tag{3.154}$$

$$= \sigma\sqrt{(L^f+1)+[\delta^s]^2L^\Delta}\left(\phi(k)-k(1-\Phi(x))\right) \quad . \tag{3.155}$$

Similarly, the expected on-hand stock integral becomes

$$\mathbb{E}\left[OH\right] = \sigma\sqrt{(L^f+1)+[\delta^s]^2L^\Delta}\int_{-\infty}^k (k-x)\phi(x)\;dx \tag{3.156}$$

$$= \sigma\sqrt{(L^f+1)+[\delta^s]^2L^\Delta}\left(k\Phi(k)+\phi(k)\right) \quad . \tag{3.157}$$

Consequently, the TRC function can be rewritten as follows

$$\begin{aligned} TRC_{OSP}^{norm} = {} & h\cdot\sigma\sqrt{(L^f+1)+[\delta^s]^2L^\Delta}\left(k\Phi(k)+\phi(k)\right) \\ & + b\cdot\sigma\sqrt{(L^f+1)+[\delta^s]^2L^\Delta}\left(\phi(k)-k(1-\Phi(x))\right) \\ & + c\cdot(1-\delta^s)\mu \quad . \end{aligned} \tag{3.158}$$

The optimal δ^s can be found by solving the following optimization problem:

$$\min \;\; TRC_{OSP}^{norm}\left(\delta^s\right) \tag{3.159}$$

$$\text{s.t.} \;\; 0 \le \delta^s \le 1 \tag{3.160}$$

and exploiting the following lemma:

Lemma 3.3.3.7 *The cost function TRC_{OSP}^{norm} is convex in δ^s.*

Proof:

See Appendix B.6. □

Convexity of the objective function (Lemma 3.3.3.7) implies that the first-order conditions are sufficient for optimality. For ease of presentation of (3.158), define

$$r_{OH} = h \cdot \sigma\left(k\Phi(k) + \phi(k)\right) \tag{3.161}$$

$$r_{BO} = b \cdot \sigma\left(\phi(k) - k(1 - \Phi(x))\right) \tag{3.162}$$

which are both non-negative for all feasible values of k. Then, the Lagrange function is

$$\begin{aligned} L := L(\delta^s, \omega_1, \omega_2) = {} & r_{OH}\sqrt{(L^f+1) + [\delta^s]^2 L^\Delta} + r_{BO}\sqrt{(L^f+1) + [\delta^s]^2 L^\Delta} \\ & + c \cdot (1-\delta^s)\mu + \omega_1 \delta^s + \omega_2(1-\delta^s) \end{aligned} \tag{3.163}$$

with the Lagrangian multipliers ω_1 and ω_2 associated with (3.160). The *Karush-Kuhn-Tucker conditions (KKT)* for an optimal solution can be stated as

$$\begin{aligned} \frac{\partial L}{\partial \delta^s} = {} & r_{OH}\left((L^f+1) + [\delta^s]^2 L^\Delta\right)^{-\frac{1}{2}} \delta^s L^\Delta + r_{BO}\left((L^f+1) + [\delta^s]^2 L^\Delta\right)^{-\frac{1}{2}} \delta^s L^\Delta - c \cdot \mu \\ & + \omega_1 - \omega_2 \stackrel{!}{=} 0 \end{aligned} \tag{3.164}$$

where ω_1 and ω_2 must satisfy the complementary slackness conditions

$$\omega_1 \geq 0 \quad \text{and} \quad \omega_1 \cdot \delta^s = 0 \tag{3.165}$$

$$\omega_2 \geq 0 \quad \text{and} \quad \omega_2 \cdot (1-\delta^s) = 0 \quad . \tag{3.166}$$

From (3.164), (3.165), and (3.166) it follows that the cases (i) $\omega_1^* > 0$ and $\omega_2^* = 0$, (ii) $\omega_1^* = 0$ and $\omega_2^* > 0$, and (iii) $\omega_1^* = \omega_2^* = 0$ lead to feasible solutions. For all three cases, the KKT are satisfied, which indicates that an optimal solution can either be an inner solution (case (iii)) or one of the extreme solutions (case (i), i.e. $\delta^s = 0$, or case (ii), i.e. $\delta^s = 1$). By solving (3.164) for δ^s the optimal values for the different cases are found as

$$\delta^{s^*} = \begin{cases} 0 & L^{\Delta}(r_{OH}+r_{BO})^2 \leq (c\cdot\mu)^2 \\ 1 & \frac{\left(L^{\Delta}(r_{OH}+r_{BO})\right)^2}{L^f+1+L^{\Delta}} < (c\cdot\mu)^2 < L^{\Delta}(r_{OH}+r_{BO})^2 \\ \sqrt{\frac{L^f+1}{\left(\frac{L^{\Delta}(r_{OH}+r_{BO})}{c\cdot\mu}\right)^2 - L^{\Delta}}} & \text{otherwise} \end{cases} \tag{3.167}$$

(see Appendix B.7).

Extensions

The aspects covered in this subsection extend to the service-level optimization models under the OSP. At the end, a brief remark on similar policies to the OSP is made, too.

Service-level model. Due to the backorder cost-service level relation $\alpha = \frac{b}{b+h}$, which is known from the single-sourcing order-up-to level model, the optimal B^s for a given δ^s follows from (3.35) with $\check{D}(\delta^s)$ (instead of $\hat{D}(\Delta)$) in the *α-service level* case, because the net stock computation in the OSP model can be transformed to resemble the one in the single-sourcing order-up-to level model. Then, the same arguments as in the SIP case can be applied. Hence, the optimization can be done as in the backorder cost model.

The *γ-service level* optimization model is

$$TRC^{\gamma}_{OSP}(\delta^s, B^s) = c\cdot(1-\delta^s)\mu + h\cdot\mathbb{E}\left[\left(B^s - \check{D}(\delta^s)\right)^+\right] \tag{3.168}$$

$$s.t.\ \ \mathbb{E}\left[\left(\check{D}(\delta^s) - B^s\right)^+\right] \leq BO^{\gamma}_{target} \tag{3.169}$$

$$0 \leq \delta^s \leq 1 \tag{3.170}$$

$$B^s \in \mathbb{N} \quad . \tag{3.171}$$

For a given δ^s the expected fast order quantity (first term in the objective function) is fixed. Since the properties of Lemma 3.3.3.3 also hold for the OSP (based on the same arguments), the smallest integer B^s that satisfies (3.169) is the optimal one. Thus, the optimization only needs to be done over δ^s.

In the *β-service level* problem, the random variable $\check{D}(\delta^s)^{beg}$ needs to be computed, which is

$$\check{D}(\delta^s)^{beg} = D(L^f) + \delta^s D(L^\Delta) \quad . \tag{3.172}$$

Given this random variable, the optimization problem is similar to the γ-service level one with the same properties.

$$TRC^{\beta}_{OSP}(\delta^s, B^s) = c \cdot (1 - \delta^s)\mu + h \cdot \mathbb{E}\left[\left(B^s - \check{D}(\delta^s)\right)^+\right] \tag{3.173}$$

$$s.t. \;\; \mathbb{E}\left[\left(\check{D}(\delta^s) - B^s\right)^+\right] - \mathbb{E}\left[\left(\check{D}(\delta^s)^{beg} - B^s\right)^+\right] \leq BO^{\beta}_{target} \tag{3.174}$$

$$0 \leq \delta^s \leq 1 \tag{3.175}$$

$$B^s \in \mathbb{N} \quad . \tag{3.176}$$

where $BO^{\beta}_{target} = (1 - \beta^{target})\,\mu$. Therefore, the optimal B^s is the smallest value that satisfies (3.174).

Further aspects. Recently, policies that adjust the sourcing fractions periodically according to the outstanding orders have been studied. The interested reader is being referred to Sheopuri et al. (2010).

3.3.4 Summary and implications

In this section, several inventory control policies for a single-echelon periodic-review inventory model with dual sourcing have been studied. It has been shown how to derive the optimal policy. While for the case of consecutive lead times, i.e. a lead-time difference of one period between the two suppliers, the optimal policy is known to be the single-index policy, in the case of offsetting lead times it needs to be derived by using a Markov Decision Process formulation. Due to the increased complexity and computational intractability of the optimal policy in larger settings, several simpler non-optimal policies have been discussed, i.e. the single-index (SIP), constant-order (COP), dual-index (DIP), and order-splitting policy (OSP). For these policies, relevant results from the literature have been reiterated using a unified notational framework and complemented at certain points by new aspects.

For each policy, the backorder-cost model has been addressed in detail and a procedure for the optimal policy parameter determination has been presented. Service-level models have also been dealt with in the 'Extension' subsection of each policy section. In case of the COP and DIP, the derivation of the stationary overshoot distribution has been identified as the biggest challenge in the optimization process. Exact and approximate methods for the computation of this distribution have been discussed. A simplification results for the special case of geometric demand, where a recursive procedure for the derivation of the overshoot distribution can be applied in the COP case and a closed-form expression is available in case of the DIP. The parameter optimization is also simplified in the setting with consecutive lead times, where one of the parameters of the SIP can be calculated directly via a critical fractile (in)equality and an analytical expression for the other parameter can be derived, from which the optimal value can be computed numerically. Moreover, in case of normally distributed period demand, a closed-form expression for the optimal sourcing fraction of the OSP has been derived. The other policy parameter can be computed as the solution to an (in)equality.

3.4 Comparison of the constant-order and dual-index policy

3.4.1 Introduction

In the previous section, several dual-sourcing policies have been described and procedures for the optimal inventory control parameter computation presented. In order to be able to choose the right policy for a specific setting, knowledge about their cost performance and about the drivers of a potential performance gap is required. Such an analysis is conducted in this section.

A comparison of the cost performance of the SIP and the DIP has already been done by Scheller-Wolf et al. (2007). They find that the best DIP is superior to the SIP in most of the investigated instances (even though often not by far). The maximum difference is about 3%. Furthermore, Veeraraghavan and Scheller-Wolf (2008) show that the DIP mimics the behavior of the optimal policy very closely in settings where the optimal policy can still be computed via the MDP described in Section 3.3.2. Based on these findings the DIP can be regarded as a quite effective dual-sourcing policy.

Moreover, as already explained in the respective policy section, the OSP with fixed sourcing fractions is not a reasonable dual-sourcing policy for the single-echelon setting. Consequently, this policy is excluded from the upcoming analysis. Hence, the only two policies, whose performance has not yet been compared, are the COP and the DIP (excluding the SIP due to its slightly inferior performance to the DIP.) This comparison is the main focus of this section, which is based on Klosterhalfen et al. (2010a).

Section 3.4.2 develops certain presumptions about the cost performance of the two policies based on theoretical considerations. In Section 3.4.3, a numerical study is conducted to gain further insights. Section 3.4.4 summarizes the main findings.

3.4.2 Theoretical considerations

The COP and DIP result in very different order processes. Due to the employment of two order-up-to levels the DIP can vary both order quantities. In periods with high demand, large replenishment orders can be placed. In case of low demand, a small order can be made. The COP, on the other hand, does not have such flexibility options. The minimum quantity that is delivered each period corresponds to the fixed order quantity from the slow supplier. Only an increase of the total order quantity through a fast order is possible. Thus, at first glance one would conjecture that this lack of flexibility puts the COP at a major disadvantage, i.e. causing higher cost and rendering the COP less favorable compared to the DIP.

However, a more detailed analysis based on the extreme strategies that both policies can prescribe, i.e. slow and fast single sourcing, produces slightly different insights into the policy costs and their relative performance. Since the DIP can mimic both single-sourcing order-up-to policies, the optimal DIP is at least as good as the best single-sourcing strategy. Whether slow or fast single sourcing is advantageous depends on the trade-off between the expediting premium, which has to be paid when sourcing from the faster supplier, and the higher inventory holding costs in case of slow single sourcing due to the longer lead time and therefore larger safety stocks. Thus, as the lead-time difference increases, slow single sourcing becomes less attractive and, at some point, is dominated by fast single sourcing. Figure 3.1 illustrates this relationship. Fast single sourcing is represented by a horizontal line, because its cost does not vary in the slow lead time. On the contrary, the slow single-sourcing cost increases, as the slow lead time increases. Even though it starts off at a lower cost, once the slow lead time exceeds a certain length, its cost lies above the fast single-sourcing cost.

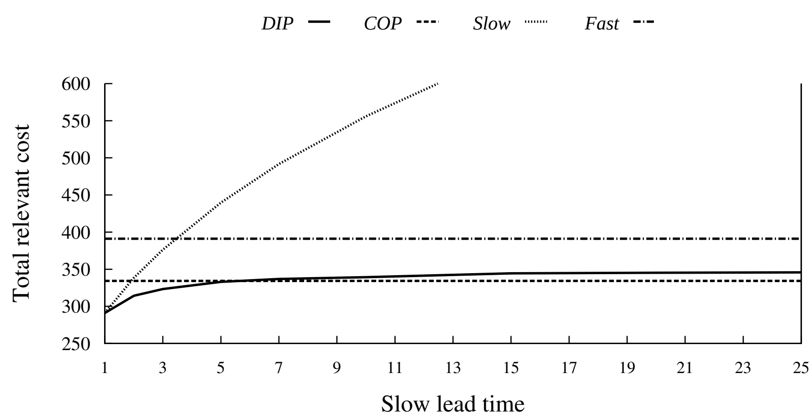


Figure 3.1: Relationship between single- and dual-sourcing policy costs for $L^f = 1$

The same reasoning also induces the DIP to source more items from the fast supplier, as the slow lead time increases, and therefore it approaches the fast single-sourcing cost. However, since the DIP can mimic both single-sourcing order-up-to policies, it is never worse than any of two. On the other hand, the COP cost (similar to the fast single-sourcing cost) is independent of the slow lead time, because it does not influence the choice of the control parameters. Consequently, its cost is also given by a horizontal line in Figure 3.1. Since the maximum possible COP cost corresponds to the fast single-sourcing cost, there must be some intersection (or at least cost equality) of the two dual-sourcing policies.

Due to the complexity of both dual-sourcing policies, an analytical derivation of an intersection is omitted. Instead, a numerical study is conducted to gain further insights.

3.4.3 Numerical study

3.4.3.1 Numerical design

Three different demand distributions are considered: two frequently used discrete ones, i.e. (1) the Poisson distribution to model small coefficients of variation (CVs) and (2) the negative binomial (nbin) distribution for larger CVs. Thirdly, a discretized Gamma distribution is used, which can reflect both small and large CVs (see Section 2.1.1 for details). Since only the expediting premium, c, is relevant for TRC, $c^s = 100$ is fixed and only $c^f \in \{102, 105, 110\}$ varies. This represents a relative premium, $\frac{c}{c^s}$, of 2%, 5%, and 10%. The holding cost per unit and period is $h \in \{0.1, 0.5, 1.0\}$, which corresponds to a yearly rate for interest and storage on c^s of 5%, 25%, and 50% under weekly ordering. For each h the corresponding backorder cost per unit and period is determined according to a $\frac{b}{b+h}$-ratio of 95% and 99%. As lead-time difference the smallest one, for which the optimal policy is unknown, i.e. $L^\Delta = 2$, is considered as well as larger differences of 5 and 10. All parameters are summarized in Table 3.2.

	Type	Parameter	Value	
Demand	Poisson	Mean, μ	10	100
		CV	0.32	0.1
	negative binomial	Mean, μ	10	100
		CV	0.49, 1.05	0.51, 1.0
	discretized Gamma	Mean, μ	10, 100	
		CV	0.5, 1.0, 1.5	
Lead times		L^f	1, 3	
		L^Δ	2, 5, 10	
Costs		c^f (given $c^s = 100$)	102, 105, 110	
		h	0.1, 0.5, 1.0	

Table 3.2: COP-DIP comparison – Parameter values

In total, 1224 instances are analyzed. (By simply permuting all factors and levels 1296 instances result. 72 instances, for which $c > b \cdot L^\Delta$, are excluded. For these instances it is cheaper to wait and incur backorder costs than to use the fast supplier.) Note that the numerical design considered by Veeraraghavan and Scheller-Wolf (2008) in their DIP analysis is not applied, because for many real-world settings their holding cost parameter seems to be fairly high.

3.4.3.2 Computational aspects

The complexity of the Markov Chain for the exact computation of the stationary overshoot distribution in the DIP increases considerably with an increase in L^{Δ} and μ, because the state space grows significantly (see Section 3.3.3.3, 'Policy evaluation'). That is why only instances with small values for these parameters can be efficiently solved with this approach. The theoretical analysis in Section 3.4.2 reveals, however, that especially instances with a large lead-time difference are of interest, because in such settings the COP might outperform the DIP. Therefore, not the Markov Chain approach, but the simulation-based optimization procedure proposed by Veeraraghavan and Scheller-Wolf (2008) is employed, which does not suffer from this dimensionality problem. The DIP parameter optimization procedure remains the same as described in Section 3.3.3.3 ('Policy optimization'). The only difference is that the stationary overshoot distribution is computed via simulation instead of using the Markov Chain. In order to ensure a fair comparison, the same is done in the COP optimization (see Section 3.3.3.2 'Policy optimization'). In the simulation models of both policies common random numbers are used. The optimal single-sourcing order-up-to levels are also determined based on the random number sequences used in the simulation instead of the theoretical distribution. For each instance, 10 simulation runs with 100,000 periods each are conducted.

The simulation-optimization results are validated by comparing them to the results of the Markov Chain for instances with $L^{\Delta} = 2$ and $\mu = 10$, for which the Markov Chain results can still be obtained in a reasonable amount of time. For computational reasons in the COP, a maximum overshoot state is determined such that the probability for larger states is negligible. The appropriate choice is checked by simulation (see, e.g., Tijms (1994), p. 119). Similarly, all considered demand distributions are restricted to a maximum value $\bar{D}$, which in itself unifies the remaining probability mass of 0.001%. Given $\bar{D}$, the upper bound on Δ in the DIP can be set to $(L^s + 1) \cdot \bar{D}$ instead of ∞. The results reveal that the cost difference is within about 1% for all instances, although the optimal parameter combination sometimes slightly differs. This is due to the flat shape of the cost curve around the optimum. Since all policies are optimized with respect to common random numbers, the relative policy performance is not affected, however.

In order to check the significance of the cost difference between the policies an ANOVA is conducted with a level of significance of 95%. The required normal distribution assumption of the 10 cost results is confirmed by the Kolmogorov-Smirnov test (significance level: 99%). All numbers in the upcoming tables refer to instances with a significant cost difference. The significance test of the cost difference is not performed for the comparison between the DIP and single sourcing, because the DIP will in fact prescribe single sourcing in some instances. Here, simply the average across the 10 simulation runs is reported.

3.4.3.3 Results

The analysis reveals that the tendency of the results for instances with $\mu = 10$ and $\mu = 100$ basically coincides. That is why for ease of presentation, the results analysis and interpretation in this section is based on instances with $\mu = 100$ only. The tables with the results for instances with $\mu = 10$ are presented in Appendix A.1. First, the advantage of dual sourcing over single sourcing is analyzed. The following observations can be made.

Observation 1 *If inventory holding is inexpensive, the COP performs worse than the best single-sourcing order-up-to policy.*

For $h = 0.1$, the COP performs worse than the best single-sourcing order-up-to policy in all instances with $L^{\Delta} = 2$ or 5 and in almost all instances with $L^{\Delta} = 10$ (see Table 3.3). By using a single supplier up to 69% of the cost could be saved. For higher holding cost values ($h = 0.5$ or 1.0), the COP performance improves. For $h = 1.0$ and $L^{\Delta} = 10$, the COP outperforms single sourcing in all instances.

In instances with a low h, slow single sourcing is the superior single-sourcing strategy (unless L^{Δ} is very large). Generally speaking, the COP's advantage over slow single sourcing lies in a lower on-hand stock, which is obtained by relying on the fast supplier. However, if h is low, little can be gained from an on-hand stock reduction. Furthermore, this advantage only materializes for large lead-time differences, where single sourcing requires a large safety/on-hand stock. In instances with a small L^{Δ}, the COP additionally suffers from the fact that the total order quantity can only be increased above Q, but not lowered. In order to allow for some variation of the total order quantity, which is reasonable in the case of stochastic demand, at least 1 unit, on average, needs to be ordered from the fast supplier. For this unit the expediting premium has to be paid. Since this premium is large compared to h, the COP is inferior.

	$TRC_{BS^a} \geq TRC_{DIP}$				$TRC_{BS} > TRC_{COP}$				$TRC_{BS} < TRC_{COP}$				
	No.	$\frac{TRC_{BS}-TRC_{DIP}}{TRC_{DIP}}$			No.	$\frac{TRC_{BS}-TRC_{COP}}{TRC_{COP}}$			No.	$\frac{TRC_{COP}-TRC_{BS}}{TRC_{COP}}$			n/s[b]
$h = 0.1$	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max	Inst.
$L^{\Delta} = 2$													
$(1, 3)$	24	1.40%	0.00%	6.01%	0	–	–	–	24	44.12%	20.87%	69.07%	0
$(3, 5)$	24	0.44%	0.00%	2.47%	0	–	–	–	24	41.37%	21.00%	65.16%	0
$L^{\Delta} = 5$													
$(1, 6)$	30	4.87%	0.00%	15.53%	0	–	–	–	30	32.75%	3.86%	59.86%	0
$(3, 8)$	30	2.30%	0.00%	8.53%	0	–	–	–	30	33.01%	8.47%	57.58%	0
$L^{\Delta} = 10$													
$(1, 11)$	36	10.46%	0.00%	29.30%	8	8.84%	1.18%	17.47%	28	29.07%	1.71%	56.60%	0
$(3, 13)$	36	6.32%	0.00%	18.36%	3	5.65%	4.11%	8.16%	30	28.96%	4.37%	55.70%	3
$h = 0.5$	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max	Inst.
$L^{\Delta} = 2$													
$(1, 3)$	36	5.06%	0.01%	12.66%	4	2.92%	1.49%	5.42%	32	22.29%	1.73%	49.80%	0
$(3, 5)$	36	2.06%	0.00%	6.45%	0	–	–	–	35	20.37%	2.21%	45.65%	1
$L^{\Delta} = 5$													
$(1, 6)$	36	16.05%	1.83%	29.52%	18	16.34%	2.95%	26.90%	18	13.64%	0.75%	34.15%	0
$(3, 8)$	36	8.96%	0.58%	17.50%	13	10.57%	3.52%	14.90%	21	14.50%	2.31%	33.81%	2
$L^{\Delta} = 10$													
$(1, 11)$	36	30.92%	9.52%	54.57%	32	28.77%	1.18%	58.10%	4	6.02%	1.71%	14.12%	0
$(3, 13)$	36	20.64%	5.51%	35.87%	27	20.92%	4.11%	37.80%	6	8.74%	4.37%	17.76%	3
$h = 1.0$	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max	Inst.
$L^{\Delta} = 2$													
$(1, 3)$	36	8.28%	0.59%	16.24%	12	7.61%	2.51%	11.63%	23	14.56%	0.59%	34.06%	1
$(3, 5)$	36	3.76%	0.06%	8.54%	8	3.35%	1.62%	5.83%	28	13.11%	1.02%	31.08%	0
$L^{\Delta} = 5$													
$(1, 6)$	36	21.08%	6.74%	37.40%	30	18.45%	2.95%	38.06%	6	5.55%	0.75%	13.49%	0
$(3, 8)$	36	12.71%	3.01%	22.34%	24	12.39%	3.45%	22.44%	10	6.82%	1.39%	16.07%	2
$L^{\Delta} = 10$													
$(1, 11)$	36	36.18%	6.87%	70.12%	36	36.44%	8.20%	77.74%	0	–	–	–	0
$(3, 13)$	36	25.09%	5.90%	44.48%	36	24.27%	4.30%	49.50%	0	–	–	–	0

[a]BS = Best single-sourcing policy
[b]n/s = not significant

Table 3.3: Single- vs. dual-sourcing cost for $\mu = 100$

Observation 2 *If the lead-time difference is small, single sourcing is a reasonable strategy.*

In instances with $L^{\Delta} = 2$, where the COP delivers poor results, the overall benefit that can be gained from dual sourcing using the DIP is not very large, either. It ranges between 0.4-8% on average and 16% at most (see Table 3.3). Contrasting these expected gains of the DIP with the additional policy complexity, single sourcing is found to be a reasonable strategy in these cases. Far greater benefits from dual sourcing of up to 78% can be realized in instances with a larger L^{Δ} of 5 or 10.

Due to Observations 1 and 2, instances with $h = 0.1$ and $L^{\Delta} = 2$ are excluded from the further analysis. A more detailed COP-DIP comparison is conducted for those instances, where dual sourcing is most valuable. All effects are mentioned first and the explanation presented thereafter. The comparison reveals the following issues.

Observation 3 *With an increase in the lead-time difference the performance gap between the COP and DIP closes.*

From Tables 3.4 and 3.5 it can be observed that as L^{Δ} increases from 5 to 10, the COP cost inferiority decreases, on average, from 16% to 5% for Poisson, 12% to 3% for negative binomial, and 11.5% to 3.5% for Gamma demand. The maximum difference also diminishes from 55% to 28%, 39% to 18%, and 39% to 19%, respectively.

Observation 4 *With an increase in the holding cost the performance gap between the COP and DIP closes.*

As h increases from 0.5 to 1.0, the average cost inferiority decreases to about 0.5% for $L^{\Delta} = 10$. This means that both policies perform almost equally well. For $L^{\Delta} = 5$, the COP still results in about 8-9% higher costs.

Observation 5 *With an increase in the expediting premium the performance gap between the COP and DIP increases.*

While for $L^{\Delta} = 5$ and $c^f = 102$, the average cost difference is only about 2%, this increases to 9-14% for $c^f = 105$ and 20-33% for $c^f = 110$ (for the different demand types). However, similar to the worse COP performance, the average DIP benefit over single sourcing diminishes from about 20% to 5-10%. Maximum cost savings decrease from about 35% to 13-22%. For $L^{\Delta} = 10$ and $c^f = 102$ even an average cost superiority of the COP over the DIP of about 2-3% can be observed. With a c^f-increase to 105 and 110, this turns into a cost inferiority of about 2-3% and 9-15%, respectively. Maximum values increase from about -1% to 18-29%. The striking observation that can be made here is the following:

Observation 6 *The COP can outperform the DIP.*

Already for $L^{\Delta} = 5$, instances can be found, where the COP outperforms the DIP. This is indicated by a negative number in the next-to-last 'Min'-column. The cost savings only range between 0.3-1%, however. For $L^{\Delta} = 10$, larger cost savings of 1.5-5% can be realized by using the COP instead of the DIP. For instances with $c^f = 102$ this becomes most obvious. Here, the COP delivers better results irrespective of the other parameter values. Surely, the cost savings are not very large, but the major finding is that the COP *can* outperform the DIP at all.

Observation 7 *With an increase in the $\frac{b}{b+h}$-ratio, the performance gap between the COP and DIP closes.*

As $\frac{b}{b+h}$ increases from 0.95 to 0.99 the cost difference decreases on average by about 4% to 14% (Poisson) and 10% (others) for $L^{\Delta} = 5$ and by about 3% to 4% (Poisson) and 3% (others) for $L^{\Delta} = 10$. Moreover, the dual-sourcing advantage over single sourcing becomes larger. The higher backorder cost forces the single sourcing policy to hold more stock, whereas the dual-sourcing policies can rely on the fast supplier.

	Best single (BS) vs. DIP				Best single (BS) vs. COP				COP vs. DIP			
	No.	$\frac{TRC_{BS}-TRC_{DIP}}{TRC_{DIP}}$			No.	$\frac{TRC_{BS}-TRC_{COP}}{TRC_{COP}}$			No.	$\frac{TRC_{COP}-TRC_{DIP}}{TRC_{DIP}}$		
Poisson	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max
(L^f, L^s)												
$(1, 6)$	12	15.80%	1.83%	35.05%	12	2.08%	-34.15%	36.42%	12	16.60%	-1.01%	54.64%
$(3, 8)$	12	8.40%	0.58%	20.20%	12	-4.59%	-33.81%	20.53%	12	16.03%	-0.28%	51.96%
h												
0.5	12	8.80%	0.58%	24.22%	12	-9.59%	-34.15%	20.04%	12	23.64%	3.48%	54.64%
1.0	12	15.41%	3.01%	35.05%	12	7.08%	-16.07%	36.42%	12	8.99%	-1.01%	23.39%
$\frac{b}{b+h}$												
0.95	12	9.82%	0.58%	29.64%	12	-5.04%	-34.15%	30.41%	12	18.93%	-0.59%	54.64%
0.99	12	14.38%	2.93%	35.05%	12	2.53%	-25.46%	36.42%	12	13.70%	-1.01%	41.22%
c^f												
102	8	20.82%	9.45%	35.05%	8	18.58%	3.52%	36.42%	8	2.07%	-1.01%	5.73%
105	8	10.42%	3.01%	21.05%	8	-2.32%	-16.07%	14.24%	8	13.67%	5.96%	23.39%
110	8	5.06%	0.58%	12.55%	8	-20.03%	-34.15%	-4.18%	8	33.21%	16.19%	54.64%
nbin	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max
(L^f, L^s)												
$(1, 6)$	24	19.16%	3.86%	37.17%	24	8.66%	-25.03%	37.93%	22	12.22%	-0.55%	38.53%
$(3, 8)$	24	11.22%	1.61%	22.24%	22	1.53%	-25.73%	22.09%	21	12.39%	0.54%	36.81%
CV												
0.51	24	15.05%	1.61%	37.17%	24	4.27%	-25.73%	37.93%	22	13.09%	-0.55%	38.53%
1.0	24	15.33%	2.93%	29.17%	22	6.32%	-21.32%	26.02%	21	11.48%	0.54%	32.39%
h												
0.5	24	12.75%	1.61%	29.17%	23	-1.18%	-25.73%	26.02%	24	15.73%	2.45%	38.53%
1.0	24	17.63%	5.08%	37.17%	23	11.68%	-10.63%	37.93%	19	7.98%	-0.55%	18.01%
$\frac{b}{b+h}$												
0.95	24	13.41%	1.61%	33.41%	24	2.06%	-25.73%	33.85%	22	14.00%	-0.32%	38.53%
0.99	24	16.96%	5.07%	37.17%	22	8.73%	-17.04%	37.93%	21	10.52%	-0.55%	28.69%
c^f												
102	16	21.97%	12.39%	37.17%	16	20.05%	7.60%	37.93%	11	2.33%	-0.55%	4.45%
105	16	14.81%	5.08%	26.64%	15	5.54%	-10.63%	21.91%	16	9.49%	3.65%	18.01%
110	16	8.79%	1.61%	19.01%	15	-10.83%	-25.73%	7.40%	16	21.97%	10.14%	38.53%
Gamma	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max
(L^f, L^s)												
$(1, 6)$	36	19.09%	3.66%	37.40%	36	9.34%	-25.34%	38.06%	32	11.45%	-0.47%	38.85%
$(3, 8)$	36	11.39%	1.56%	22.34%	34	2.46%	-25.95%	22.44%	31	11.44%	0.53%	37.15%
CV												
0.5	24	14.85%	1.56%	37.40%	24	4.26%	-25.95%	38.06%	22	12.93%	-0.47%	38.85%
1.0	24	15.45%	2.98%	29.52%	22	6.75%	-21.01%	26.90%	20	11.67%	2.06%	31.82%
1.5	24	15.44%	4.70%	29.03%	24	7.04%	-17.03%	25.43%	21	9.68%	0.53%	27.40%
h												
0.5	36	13.58%	1.56%	29.52%	35	0.68%	-25.95%	26.90%	36	14.34%	2.06%	38.85%
1.0	36	16.91%	4.90%	37.40%	35	11.31%	-10.70%	38.06%	27	7.59%	-0.47%	17.91%
$\frac{b}{b+h}$												
0.95	36	13.96%	1.56%	33.18%	36	3.34%	-25.95%	33.61%	32	13.20%	-0.32%	38.85%
0.99	36	16.52%	4.84%	37.40%	34	8.81%	-17.00%	38.06%	31	9.64%	-0.47%	28.70%
c^f												
102	24	19.77%	7.23%	37.40%	24	18.04%	6.70%	38.06%	15	2.28%	-0.47%	4.29%
105	24	16.01%	4.90%	27.22%	23	7.51%	-10.70%	22.83%	24	8.55%	3.26%	17.91%
110	24	9.95%	1.56%	21.69%	23	-8.09%	-25.95%	11.60%	24	20.07%	8.58%	38.85%

Table 3.4: Single- vs. dual-sourcing cost for $L^{\Delta} = 5$, $h = 0.5$ and 1.0, $\mu = 100$

	Best single (BS) vs. DIP				Best single (BS) vs. COP				COP vs. DIP			
	No.	$\frac{TRC_{BS}-TRC_{DIP}}{TRC_{DIP}}$			No.	$\frac{TRC_{BS}-TRC_{COP}}{TRC_{COP}}$			No.	$\frac{TRC_{COP}-TRC_{DIP}}{TRC_{DIP}}$		
Poisson	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max
(L^f, L^s)												
$(1, 11)$	12	36.59%	9.52%	70.12%	12	33.06%	-14.12%	77.74%	12	4.46%	-4.96%	27.53%
$(3, 13)$	12	23.15%	5.51%	44.48%	12	18.45%	-17.76%	49.50%	12	5.43%	-3.43%	28.29%
h												
0.5	12	23.51%	5.51%	51.67%	12	15.11%	-17.76%	56.40%	12	9.37%	-3.03%	28.29%
1.0	12	36.23%	12.83%	70.12%	12	36.40%	4.30%	77.74%	12	0.51%	-4.96%	8.18%
$\frac{b}{b+h}$												
0.95	12	25.97%	5.51%	61.66%	12	21.01%	-17.76%	70.09%	12	6.08%	-4.96%	28.29%
0.99	12	33.77%	11.21%	70.12%	12	30.50%	-7.54%	77.74%	12	3.81%	-4.29%	20.42%
c^f												
102	8	46.18%	26.93%	70.12%	8	51.09%	28.63%	77.74%	8	-3.15%	-4.96%	-1.32%
105	8	27.38%	12.83%	46.20%	8	24.38%	4.30%	48.84%	8	2.78%	-1.77%	8.18%
110	8	16.05%	5.51%	30.88%	8	1.80%	-17.76%	24.84%	8	15.21%	4.84%	28.29%
nbin	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max
(L^f, L^s)												
$(1, 11)$	24	34.49%	12.95%	57.73%	24	31.84%	-3.91%	64.08%	24	2.68%	-3.87%	17.55%
$(3, 13)$	24	23.77%	7.79%	42.28%	23	21.22%	-8.82%	46.01%	20	4.13%	-2.55%	18.22%
CV												
0.51	24	31.44%	7.79%	57.73%	23	29.78%	-8.82%	64.08%	22	3.40%	-3.87%	18.22%
1.0	24	26.82%	10.25%	45.29%	24	23.64%	-4.84%	47.36%	22	3.28%	-2.50%	15.86%
h												
0.5	24	26.83%	7.79%	54.57%	23	21.89%	-8.82%	58.10%	24	5.73%	-2.63%	18.22%
1.0	24	31.43%	12.33%	57.73%	24	31.21%	9.72%	64.08%	20	0.48%	-3.87%	5.85%
$\frac{b}{b+h}$												
0.95	24	27.60%	7.79%	57.73%	24	24.18%	-8.82%	64.08%	22	3.97%	-3.87%	18.22%
0.99	24	30.66%	12.33%	54.57%	23	29.22%	4.68%	58.10%	22	2.72%	-3.04%	13.04%
c^f												
102	16	34.46%	12.33%	57.73%	16	37.18%	13.83%	64.08%	16	-1.93%	-3.87%	-0.68%
105	16	31.87%	16.13%	49.63%	16	29.89%	9.72%	51.86%	12	2.38%	-1.56%	5.85%
110	16	21.05%	7.79%	37.13%	15	11.95%	-8.82%	33.12%	16	9.33%	3.02%	18.22%
Gamma	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max
(L^f, L^s)												
$(1, 11)$	36	31.91%	6.87%	59.42%	36	29.09%	-4.35%	65.77%	34	2.86%	-3.83%	18.27%
$(3, 13)$	36	22.16%	5.90%	42.25%	34	19.90%	-9.10%	46.11%	27	4.45%	-2.64%	18.95%
CV												
0.5	24	31.65%	8.12%	59.42%	23	29.86%	-9.10%	65.77%	23	3.39%	-3.83%	18.95%
1.0	24	26.93%	10.57%	46.23%	24	23.85%	-4.37%	47.78%	21	3.38%	-2.51%	15.63%
1.5	24	22.52%	5.90%	40.09%	23	20.20%	6.70%	36.36%	17	4.05%	-1.67%	13.68%
h												
0.5	36	25.84%	8.12%	54.26%	34	21.30%	-9.10%	58.08%	32	6.23%	-2.52%	18.95%
1.0	36	28.23%	5.90%	59.42%	36	27.77%	6.70%	65.77%	29	0.64%	-3.83%	6.09%
$\frac{b}{b+h}$												
0.95	36	26.27%	8.12%	59.42%	35	23.38%	-9.10%	65.77%	32	4.02%	-3.83%	18.95%
0.99	36	27.80%	5.90%	54.26%	35	25.87%	4.11%	58.08%	29	3.07%	-3.12%	13.60%
c^f												
102	24	27.67%	5.90%	59.42%	24	29.89%	6.70%	65.77%	19	-1.90%	-3.83%	-0.87%
105	24	30.68%	16.30%	49.39%	24	28.64%	9.62%	51.77%	18	2.32%	-1.57%	6.09%
110	24	22.76%	8.12%	40.09%	22	14.50%	-9.10%	36.36%	24	8.83%	2.73%	18.95%

Table 3.5: Single- vs. dual-sourcing cost for $L^{\Delta} = 10$, $h = 0.5$ and 1.0, $\mu = 100$

The only parameter effect not studied so far is the demand variability (CV). From Table 3.4 it can be observed that a CV-increase leads to a better COP performance. The results of Table 3.5 suggest the same for negative binomially demand, but the opposite for Gamma. A more detailed comparison for ($L^f = 1, L^s = 11$) reveals that the effect, which demand variability has on the two policies, seems to depend on which policy is superior (see Table 3.6). In instances with $c^f = 102$, the COP results in lower costs. A CV-increase reduces the COP cost advantage, i.e. the DIP performance improves. For $c^f = 105$ and $h = 0.5$, the DIP performs better. Now, a CV-increase narrows the relative cost gap, too, i.e. the COP performance improves.

		$c^f = 102$			$c^f = 105$		
h	CV	Poisson	nbin	Gamma	Poisson	nbin	Gamma
0.5	0.1	-2.90%	–	–	7.00%	–	–
	0.5 (0.51)	–	-2.63%	-2.52%	–	4.83%	5.09%
	1.0	–	-1.67%	-1.74%	–	4.23%	4.21%
	1.5	–	–	-1.07%	–	–	3.86%
1.0	0.1	-4.96%	–	–	-1.18%	–	–
	0.5 (0.51)	–	-3.87%	-3.83%	–	-1.56%	-1.41%
	1.0	–	-2.50%	-2.51%	–	-0.88%	-0.93%
	1.5	–	–	-1.67%	–	–	-0.45%

Table 3.6: CV-effect on $\frac{TRC_{COP} - TRC_{DIP}}{TRC_{DIP}}$ for ($L^f = 1, L^s = 11$), $\frac{b}{b+h} = 0.95$, $\mu = 100$

For the explanation of the effects, in particular the COP superiority in some instances, the on-hand stock, backorders, and order quantities of both policies are analyzed (see Table 3.7). Note that the non-integer numbers in the 'Slow order' COP column result from the fact that the average quantity across the 10 simulation runs is reported. The sum of the expected regular and expedited order does not correspond exactly to 100, because the actual mean per period of the random numbers used in the simulation deviates slightly. As expected, an increase in the expediting premium makes both policies order more units from the slow supplier. On the contrary, as the demand variability increases, both policies reduce the expected slow order. This reduction is usually larger in the COP than in the DIP case. (Only for $c^f = 102$ and $h = 1.0$ it is almost identical.) The reason is that in the COP the total order cannot be adjusted downwards below the slow (constant) order. Thus, there is an increased risk of excessive on-hand stock, if the slow order is set too large. Moreover, it can be observed that an increase in h causes an increase in the expected fast order of both policies to reduce the on-hand stock. This increase is larger in the DIP than in the COP, which becomes most obvious for $c^f = 102$. (For $c^f = 105$ and $c^f = 110$, this tendency can also be observed, but only for $CV = 0.5$ or 1.0 and an increase from $h = 0.5$ to 1.0.) Although the DIP starts off with a far lower fast order quantity for $h = 0.1$ and $c^f = 102$, it orders basically the same amount as the COP for $h = 1.0$, irrespective of the CV. While the backorders are basically identical, too, the COP produces a lower on-hand stock. This eventually leads to a lower total cost. Even in instances, where the slow order of the COP is slightly higher than that of the DIP, e.g., for $c^f = 102$ and $h = 0.5$, the cost savings from the on-hand stock reduction outweigh the higher expediting cost and render the COP superior.

			Slow order		Fast order		On-hand stock		Backorders		Superior
c^f	CV	h	COP	DIP	COP	DIP	COP	DIP	COP	DIP	policy
102	0.5	0.1	93.00	97.76	7.01	2.25	217.07	245.20	2.37	2.63	DIP
		0.5	86.20	88.70	13.81	11.31	157.08	172.85	2.29	2.33	COP
		1.0	82.10	81.48	17.91	18.53	146.23	153.44	2.25	2.27	COP
	1.0	0.1	87.00	95.00	13.03	5.03	436.63	484.38	6.04	6.31	DIP
		0.5	74.80	78.36	25.23	21.67	322.18	345.19	5.93	5.97	COP
		1.0	67.50	66.55	32.53	33.48	301.54	311.95	5.91	5.92	COP
	1.5	0.1	80.70	91.92	19.28	8.06	633.58	712.18	10.96	10.95	DIP
		0.5	65.00	69.13	34.98	30.85	484.12	510.47	11.01	10.97	COP
		1.0	56.90	55.64	43.08	44.34	461.38	471.28	10.97	11.00	COP
105	0.5	0.1	95.50	99.46	4.51	0.55	295.07	286.53	2.47	3.19	DIP
		0.5	90.70	95.02	9.31	4.99	184.94	210.51	2.34	2.45	DIP
		1.0	87.40	90.55	12.61	9.46	162.30	180.69	2.29	2.36	COP
	1.0	0.1	91.40	98.47	8.63	1.57	580.90	580.62	6.13	6.84	DIP
		0.5	82.30	89.63	17.73	10.41	369.33	413.30	5.99	6.13	DIP
		1.0	76.80	81.64	23.23	18.39	331.01	359.23	5.93	6.00	COP
	1.5	0.1	86.80	96.81	13.18	3.17	826.86	864.78	10.98	10.82	DIP
		0.5	74.50	84.50	25.48	15.47	545.56	608.45	10.97	10.95	DIP
		1.0	67.60	73.45	32.38	26.52	496.07	529.67	11.00	10.98	COP
110	0.5	0.1	96.80	99.99	3.21	0.02	387.66	302.10	2.58	4.24	DIP
		0.5	93.00	97.76	7.01	2.25	217.07	245.20	2.37	2.63	DIP
		1.0	90.70	95.02	9.31	4.99	184.94	210.51	2.34	2.45	DIP
	1.0	0.1	93.70	99.89	6.33	0.14	744.53	630.23	6.22	9.53	DIP
		0.5	87.00	95.00	13.03	5.03	436.63	484.38	6.04	6.31	DIP
		1.0	82.30	89.63	17.73	10.41	369.33	413.30	5.99	6.13	DIP
	1.5	0.1	90.20	99.50	9.78	0.48	1064.67	973.30	10.94	15.39	DIP
		0.5	80.70	91.92	19.28	8.06	633.58	712.18	10.96	10.95	DIP
		1.0	74.50	84.50	25.48	15.47	545.56	608.45	10.97	10.95	DIP

Table 3.7: Quantity details for $(L^f = 1, L^s = 11)$, $\frac{b}{b+h} = 0.95$, gamma, $\mu = 100$

From this analysis it follows that the major advantage of the COP lies in a reduction of the on-hand stock, which for high holding costs and low expediting premiums even offsets the sometimes higher expediting costs. Demand fluctuations have no influence on the slow (constant) order. For small lead-time differences, this is a disadvantage, because a quick reaction to demand peaks (or drops) is not possible. Consequently, the effects of these disruptions cannot be alleviated in a timely manner unless the fast supplier is used at a higher cost. In case of large lead-time differences, a quick rectifying action via the slow supplier is no longer possible. Nevertheless, the DIP would sometimes still place very large orders, which, at the time of arrival, are not required to such an extent and therefore are put on stock. The COP avoids these extreme cases and thus reduces the extent of the supply-demand mismatch causing less left-over stock.

3.4.4 Summary and implications

In this section, the constant-order (COP) and dual-index policy (DIP) have been compared. From the theoretical considerations, which have been based on the analysis of the extreme strategies that both policies can prescribe, it has been suggested that the cost difference between the COP and DIP closes as the lead-time difference increases. At some point, the COP might even outperform the DIP. Since the complexity of both policies has prohibited a further analytical investigation, a numerical study has been conducted.

Based on the findings from the numerical study a generally good performance of the DIP can be confirmed as already stated by Veeraraghavan and Scheller-Wolf (2008). Furthermore, the COP-DIP comparison allows for a supplementation of some of their assertions. They claim that

> '*the performance of the dual-index policy brings significant savings when the sourcing options differ significantly in lead times, as often is the case*' (p. 859) and '*dual sourcing is especially beneficial when [...] expediting costs are moderate, or when single sourcing via the expedited or regular channels have similar costs*' (p. 864).

In such settings, also the COP delivers a satisfactory performance, sometimes even outperforming the DIP, which renders the COP an effective policy here, too. The general tendency is that as the lead-time difference increases, the COP-DIP cost gap closes. If, in addition the expediting premium is small, a superior COP performance can be observed. This confirms the presumption derived from the theoretical considerations. However, in instances where inventory holding is inexpensive relative to the expediting premium, the DIP is clearly superior to the COP. Here, regular single sourcing is a reasonable and simple alternative to the DIP, if the lead-time difference is rather small.

The identification of settings with only a small cost performance difference between the DIP and COP is particularly of practical relevance for two reasons. First, the COP is the more easily implementable and controllable policy in practice. Secondly, such a policy is of particular benefit in supply negotiations. Being able to guarantee the supplier a constant order increases a company's bargaining power. The supplier will be more willing to make concessions, because a constant order facilitates his production planning significantly and avoids the bullwhip effect.

The latter aspect, namely the pattern of the order process is of particular importance in multi-echelon settings, because this corresponds to the demand process that the supplying stage faces and which needs to be taken into account in the inventory optimization there. Multi-echelon inventory optimization models are the subject of the next chapter.

4 Multi-echelon inventory model with dual sourcing

4.1 Introduction

In this chapter, the focus is shifted from a single-echelon inventory model to a multi-echelon one. The main research question that is addressed is: *How can dual sourcing be integrated into a multi-echelon setting?*

Since there is not only a single multi-echelon inventory modeling approach available in the literature, the different frameworks need to be analyzed first, before the integration of dual sourcing is discussed. Based on the two pioneering contributions to multi-echelon inventory research without lot-sizing by Simpson (1958) and Clark and Scarf (1960), two competing research strands have developed over the years. Although they solve the same inventory optimization problem in their core, they make a different assumption concerning the role of safety stock. The choice of the appropriate framework basically follows from answering the question whether safety stocks are supposed to protect supply chain performance against all variability or just against a maximum reasonable variability (see Graves (1988)). The resulting consequences for the material flow in the system coined the terms full-delay and no-delay (van Houtum et al. (1996)), or stochastic-service (SS) and guaranteed-service (GS) approaches (Graves and Willems (2003)).

The SS approach assumes the former and regards safety stock as the only means to deal with demand variability. Therefore, upstream material shortages cause stochastic delays in the material flow. The GS approach, on the other hand, makes the latter assumption. Here, it is assumed that, if demand variability exceeds a normal level, additional countermeasures like overtime or accelerated production are available. This 'operating flexibility' ensures the timely delivery of ordered items to the next downstream stage(s). Safety stock is sized to cope with variability up to this reasonable level only.

Based on each of these approaches, commercial software solutions have been developed by companies such as Optiant®, LogicTools (which is now a part of IBM), SmartOps®, etc., which are widely used by large multinationals like IBM, Intel, or Philips. Which approach forms the basis can be conjectured from taking a closer look at the affiliated scientists and their scientific contributions in this area. Thus, both assumptions seem to be justifiable in practice. Since no comprehensive comparative analysis of the two approaches in terms of their cost performance is available so far, the question arises, whether any of these approaches is generally superior to the other. If this was the case, it would be reasonable with respect to the main research question of this chapter to only extend the superior approach. That is why the main research question of this chapter is broken down into several smaller and more specific research questions, which are addressed in turn and eventually lead up to the answer of the main one:

1. *Given the characteristic assumptions and features of the two competing multi-echelon inventory optimization model strands in the literature, i.e. the stochastic- and guaranteed-service framework, is one of them generally superior to the other?*

   If this is not the case, in which settings does each approach perform well? (Sections 4.2 and 4.3)

2. *Depending on the outcome of the first question, is a mutually exclusive implementation of a single multi-echelon approach for the entire supply network reasonable?*

   Put differently, do situations exist where a combination of both approaches provides additional benefits and how can such an integrated approach be designed? (Section 4.4)

3. *Provided that none of the approaches is generally superior to the other, how can dual sourcing be accommodated in the guaranteed-service approach? (Section 4.5)*

Whereas the first specific research question aims at a comparison between the two existing single-sourcing multi-echelon frameworks, the second question goes one step further and targets a combination of both approaches. In view of the ultimate goal of this chapter, i.e. the integration of dual sourcing into a multi-echelon setting, the idea is to first exploit the gained knowledge about the two multi-echelon approaches in terms of the development of a new integrated framework, which then basically represents an additional candidate for a potential dual-sourcing extension. Due to the increased complexity of such an integrated approach, however, only the GS framework is extended to incorporate dual sourcing (third specific research question). The extension of any other approaches is postponed to future research.

The outline of this chapter is as follows. In Section 4.2 both the SS and GS approach and their underlying assumptions as well as the individual optimization models and procedures are presented. Section 4.3 addresses the first of the smaller research questions by providing a theoretical and numerical comparison of the two approaches. The combination of both approaches, which is the second smaller research question, is the main focus of Section 4.4, in which an integrated framework is developed together with a pseudo-polynomial time dynamic programming algorithm for the optimization of serial supply chains. Section 4.5 deals with the third smaller research question, namely the extension of the GS model (as one of the three multi-echelon frameworks) to accommodate dual sourcing.

4.2 Multi-echelon inventory optimization approaches

For ease of presentation, the two competing multi-echelon modeling frameworks without lot-sizing, i.e. the SS and GS models, are only described for a serial system and an α-service level constraint in detail (unless stated otherwise). For other network structures and service-level types relevant references are provided in the respective literature review in Section 2.2.2.

4.2.1 Common assumptions

A serial production/inventory system is considered consisting of n stages, which are numbered from $i = 1, ..., n$ starting with the most upstream stage (see Figure 4.1). All stages operate a periodic (echelon) order-up-to level policy with a common review period. Each stage performs a certain processing function, e.g., a step in a manufacturing process, and represents a potential location for holding stock after the process has finished. The processing time at each stage is given by T_i and is assumed to be deterministic and a multiple of the review period. It is assumed that T_n includes the review period. No capacity constraints exist at any of the stages or processes. Customer demand occurs at stage n and is assumed to be stationary and independent across non-overlapping intervals with mean μ and standard deviation σ. Let F_{T_i} denote the T_i-period demand cumulative distribution function. At any of the stages unsatisfied demands are backordered and for each unit of left-over stock at the end of a period a linear holding cost of h_i is incurred. For ease of presentation, it is assumed that an item at a downstream stage requires exactly one item of the upstream stage that is connected to it, i.e. production coefficients are set equal to 1. Note that the relaxation of this assumption is not difficult to include into the model. It would simply make the presentation more complicated and is therefore omitted. The objective is to determine an optimal order-up-to level for each stage such that the system cost is minimized subject to a service-level constraint at the final stage.

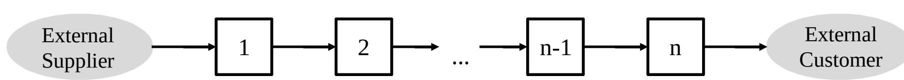


Figure 4.1: Serial system illustration

4.2.2 Stochastic-service approach

In the SS approach it is assumed that the only buffer against demand uncertainty is safety stock, i.e. the production system is inflexible and safety stock needs to account for all contingencies. If this stock quantity is chosen too small, delays in the material flow occur, because the delivery of the shortage quantity is delayed until new material becomes available. Consequently, the service of a stage, i.e. the ability to readily provide the requested materials, depends on its stock level and is therefore stochastic. Under the SS assumption, the system behavior can be fully described analytically as explained below.

4.2.2.1 Optimization model

For a serial system, the optimization problem in the SS approach, $\mathbf{P}^{SS}_{\langle 1,n\rangle}$, can be stated as

$$\begin{aligned} \mathbf{P}^{SS}_{\langle 1,n\rangle} \quad \min\ C^{SS}_{\langle 1,n\rangle}(\vec{B}_{\langle 1,n\rangle}) &= \sum_{i=1}^{n} h_i \cdot \mathbb{E}\left[OH_i(B_i)\right] \\ \text{s.t.} \quad \alpha_n(\vec{B}_{\langle 1,n\rangle}) &= \alpha_n^{target} \quad . \end{aligned} \tag{4.1}$$

The decision variables are the local order-up-to levels at all stages, B_i for $i = 1, ..., n$, which are summarized by vector $\vec{B}_{\langle 1,n\rangle}$. They are chosen such that the sum of the inventory holding costs in the entire system $\langle 1, n\rangle$ are minimized subject to the fact that the final-stage target service level, α_n^{target}, is achieved. Given a certain vector of local order-up-to levels, $\vec{B}_{\langle 1,n\rangle}$, the expected on-hand stock at stage i can be calculated as

$$\mathbb{E}\left[OH_i(B_i)\right] = (B_i - \mathbb{E}\left[BO_{i-1}(B_{i-1})\right]) - T_i \cdot \mu + \mathbb{E}\left[BO_i(B_i)\right] \quad i = 1, ..., n \tag{4.2}$$

with $\mathbb{E}\left[BO_0\right] = 0$, since it is assumed that the external supplier has ample stock. Due to potential shortages at the supplying stage ('stochastic service') given by the expected backorders, $\mathbb{E}\left[BO_{i-1}\right]$, the inventory position at stage i can only be raised to $B_i - \mathbb{E}\left[BO_{i-1}\right]$. Subtracting the expected demand during the replenishment time, $T_i \cdot \mu$, and adding the expected backorders of stage i itself, $\mathbb{E}\left[BO_i\right]$, results in the expected on-hand stock. $\mathbb{E}\left[BO_i\right]$ is given as

$$\mathbb{E}\left[BO_i(B_i)\right] = \mathbb{E}\left[(BO_{i-1}(B_{i-1}) + D(T_i) - B_i)^+\right] \qquad i = 1, ..., n \tag{4.3}$$

where BO_{i-1} is a random variable indicating the backorders (shortfall) at the preceding stage and $BO_0 \equiv 0$. $D(T_i)$ denotes the demand random variable over T_i periods and $(x)^+ = \max\{0, x\}$.

4.2.2.2 Optimization procedure

In a single-echelon model consisting of stage n only, the optimal solution to (4.1) is to set B_n as the solution to (cf. (2.51))

$$F_{T_n}(B_n) = \alpha_n^{target} \quad . \tag{4.4}$$

Since stage n orders from an external supplier, who by assumption has ample stock, the inventory position can be raised to B_n each period. Consequently, the probability that no stockout occurs is given by the probability that the demand over the processing time is smaller than or equal to the local order-up-to level B_n.

In the multi-echelon model, a material shortage at an upstream stage might occur. That means, except for the most upstream stage 1, the inventory positions of all other stages $i = 2, ..., n$ cannot necessarily be increased to B_i in every period, because the supplying stage $i-1$ might not have sufficient stock. The most upstream stage 1 experiences a material shortage, if the demand over the processing time, $D(T_1)$, exceeds the local order-up-to level B_1. Let $F_{T_1}^{B_1}$ denote the distribution function of the shortfall random variable $\big(D(T_1) - B_1\big)^+$. Due to the shortfall at stage 1, the inventory position at stage 2 can only be raised to $B_2 - \big(D(T_1) - B_1\big)^+$. Along the same lines (and rearranging terms), the shortfall random variable at stage 2 is given as $\big((D(T_1) - B_1)^+ + D(T_2) - B_2\big)^+$. The distribution of the expression $(D(T_1) - B_1)^+ + D(T_2)$ is called the two-fold incomplete convolution $F_{T_1}^{B_1} * F_{T_2}$, where $*$ denotes the convolution operator, because it can also be represented as

$$F_{T_1}^{B_1} * F_{T_2}(x) = \int_0^x F_{T_1}(x + B_1 - u)\, dF_{T_2}(u) \qquad x \geq 0 \quad . \tag{4.5}$$

Using this notation, the optimal local order-up-to levels $B_i, i = 1, ..., n$ are to be set such that the on-hand stock cost in the entire system is minimized subject to

$$\left(\left(\left(F_{T_1}^{B_1} * F_{T_2}\right)^{B_2} * ... * F_{T_{n-1}}\right)^{B_{n-1}} * F_{T_n}\right)(B_n) = \alpha_n^{target} \quad . \tag{4.6}$$

The optimal local order-up-to levels, B_i, can be derived from the echelon order-up-to levels, S_i, which have been shown by Clark and Scarf (1960) to constitute an optimal policy for such a multi-echelon system. The echelon stock of stage i

denotes all stock at that stage plus all materials in transit to or on hand at any downstream stages minus backorders at the most downstream stage. The echelon inventory position of stage i, which for the order determination is checked against the echelon order-up-to level S_i, is defined as its echelon stock plus all materials in transit to the stage. Given a certain set of echelon order-up-to levels, the local ones result as characterized in (4.9) and (4.10) below. Optimal echelon order-up-to levels can be obtained starting at the final stage n due to the decomposition result derived by Clark and Scarf (1960). For already determined optimal echelon order-up-to levels $S_{i+1}^*, ..., S_n^*$, the optimal S_i is chosen to satisfy

$$F^{[S_i^*,...,S_n^*]}\left(\tilde{S}_n\right) = \frac{p + \sum_{j=1}^{i-1} h_j^e}{p + h_n} = \alpha_i^{SS} \qquad i = 1, ..., n \tag{4.7}$$

where $\tilde{S}_n = \min\{S_i^*, ..., S_n^*\}$ and $h_j^e = h_j - h_{j-1}$ denotes the echelon holding cost of stage $j = 1, ..., n$ with $h_0 = 0$. For all $S_i, ..., S_n$, $F^{[S_i,...,S_n]}(x)$ is defined as

$$F^{[S_i,...,S_n]}(x) := \left(\left(\left(F_{T_i}^{B_i} * F_{T_{i+1}}\right)^{B_{i+1}} * ... * F_{T_{n-1}}\right)^{B_{n-1}} * F_{T_n}\right)(x) \qquad x \in \mathbb{R} \tag{4.8}$$

with

$$B_n := \tilde{S}_n \tag{4.9}$$

$$B_j := \tilde{S}_j - \tilde{S}_{j+1} \qquad j = i, ..., n-1 \tag{4.10}$$

$$\tilde{S}_j := \min\{S_i, ..., S_j\} \qquad j = i, ..., n \quad . \tag{4.11}$$

(For $i = n$, read $F^{[S_n]}(x) := F_{T_n}(x), x \in \mathbb{R}$.) For more details the reader is referred to, e.g., van Houtum and Zijm (1991) and van Houtum and Zijm (1997).

For a given α-service level constraint, the penalty cost, p, in (4.7) can be derived from the equivalence relationship between cost and service models (see, e.g., van Houtum et al. (1996))

$$\alpha_n^{target} = \frac{p}{p + h_n} \quad \Leftrightarrow \quad p = \frac{\alpha_n^{target}}{1 - \alpha_n^{target}} \cdot h_n \quad . \tag{4.12}$$

Note that the service level, α_n^{SS}, with regard to which the optimal final-stage order-up-to level is sized, is larger than the service level, α_n^{target}, which actually is to be achieved (see (4.7) and (4.12)). This is due to the potential shortfalls at the preceding stages that need to be taken into account.

In order to derive an estimate of the computational complexity of the optimal values, it is first noted that an order-up-to level needs to be calculated exactly once for each of the n stages due to the decomposition result. Each optimal order-up-to level is determined by using a bisection procedure, which contains (in)complete convolution computations. For an arbitrary but fixed targeted precision, e.g., $\epsilon = 10^{-6}$, and a predefined customer service level, e.g., $\alpha_n^{target} = 95\%$, the computational complexity of the bisection procedure is a logarithmic function in the optimal echelon order-up-to level of the most upstream stage, S_1^*, where $F_M(S_1^*) = \alpha_n^{target}$ and $M = \sum_{i=1}^{n} T_i$, i.e. the sum of all processing times. Convolution operations are done in $O(r\log(r))$ using Fast Fourier Transform (FFT) (see, e.g., Cooley and Tukey (1965)) with $r = \bar{D}$ where $F_M\left(\bar{D}\right) = 1 - \epsilon$ in this case. This yields a total complexity of $O\left(n\bar{D}\log(\bar{D})\log(S_1^*)\right)$.

4.2.3 Guaranteed-service approach

In the GS approach, the production system is regarded as being more flexible than in the SS framework. It is assumed that further countermeasures besides safety stock exist to cope with demand variability. These additional measures are summarized by the term 'operating flexibility' and comprise of, e.g., overtime or accelerated production. Thus, safety stock is only used to cover demand variability up to a certain level, the so-called maximum reasonable demand level (see, e.g., Graves (1988)). If demand exceeds this level, the company reverts to the operating flexibility measures in order to make the requested units available in time. Consequently, due to this combination of safety stock and operating flexibility there are no stochastic delays in the material flow. A stage can always guarantee 100% service to its successor(s) after the promised service time. The service time is the time it takes until the materials ordered by a stage are received and ready for processing.

A description of the system behavior under the GS assumption in an exact analytical way proves more difficult than in the SS case. The difficulty stems from the additionally assumed operating flexibility, which would have to be modeled explicitly in order to derive an exact analytical reflection of the real system. To this end the following two questions need to be answered:

1. How is the maximum reasonable demand level to be set, i.e. how can a specification be done of what is normal and what is not?
2. Given this level, how can the effect of the operating flexibility measure on the material flow in the system be modeled?

Most of the GS contributions only provide answers to the first question. These models are referred to as the standard GS models in the remainder of this thesis and described in Section 4.2.3.1. The second question is simply neglected in large parts of the GS literature, which has caused a lot of criticism of this framework over the years. In order to counteract this criticism, the second aspect is addressed in this thesis in greater detail and a so-called extended GS model formulation is presented in Section 4.2.3.2.

4.2.3.1 Standard optimization model and procedure

Optimization model

The standard GS models only address the question of how the maximum reasonable demand level can be set. Graves and Willems (2000) argue that for the end-item a demand bound can be established by management, for instance. It expresses how often a manager is willing to resort to other tactics to cover demand variability. For example, under the typical assumption of normally distributed demand with mean μ and standard deviation σ, the demand bounds for varying time horizons τ_n at the final stage n can be specified as:

$$D_n(\tau_n) = \tau_n \mu + k_n \sigma \sqrt{\tau_n} \tag{4.13}$$

where k_n indicates the percentage of time that the safety stock covers the demand variation indicating the manager's willingness to use other countermeasures. In a serial system, this demand bound directly defines the bounds of the upstream stages (through the production coefficients a_{ij}, which are assumed to be equal to 1 here), i.e.

$$D_i(\tau_i) = a_{ij} D_j(\tau_j) = D_j(\tau_j) \qquad i = 1, ..., n-1, \; j = i+1 \quad . \tag{4.14}$$

This bound linkage between adjacent stages can be viewed as an unnecessary restriction, however. A slightly different and more general interpretation can be found in van Houtum et al. (1996) or Minner (2000), for instance, and is as follows. The maximum reasonable demand level can be understood as an indicator of a stage's flexibility, which can be expressed by an *internal service-level requirement*, SL_i. If only the fact that an extraordinary operational action has to be taken matters, the flexibility can be expressed by an α*-service level*, which represents the target probability of this event. If, on the other hand, the quantity that needs to be made available is of relevance, the β- or, as a simpler approximation, γ*-service level* can be used. If a stage has a lot of slack capacity, which it can use at no additional cost in an emergency situation, the service level would be low. High service levels reflect a less flexible process. However, the flexibility of adjacent stages does not depend on one another. It can be specified for each stage individually, i.e. $k_i(SL_i), \; i = 1, ..., n$ in (4.13) can vary across the stages of a single supply chain.

In the following exposition, the optimization model and algorithm are described for the α-service level case only. The analysis also applies to the other service-level types with minor modifications (see, e.g., Inderfurth and Minner (1998)). Given a service-level target for each stage, α_i^{target}, which specifies this stage's flexibility (or the service that is to be guaranteed towards the external customer, if $i = n$), the optimal local order-up-to level at a stage can be computed by the well-known single-echelon formula

$$B_i(\tau_i) = F_{\tau_i}^{-1}\left(\alpha_i^{target}\right) \qquad i = 1, ..., n \tag{4.15}$$

where τ_i denotes the time span, for which safety stock has to be held. In the single-echelon case, this is the replenishment time (plus the review period). In the multi-echelon case, it is called the net replenishment time and is given as

$$\tau_i = ST_{i-1} + T_i - ST_i \qquad i = 1, ..., n \quad . \tag{4.16}$$

Here, ST_{i-1} and ST_i denote the incoming and outgoing service time of stage i. ST_{i-1} represents the time it takes until the materials ordered by stage i are received (from stage $i-1$) and ready for processing at stage i. Thus, the net replenishment time τ_i consists of the time it takes until any ordered units are received, processed, and put on stock at stage i, $ST_{i-1} + T_i$ (i.e. the replenishment time of stage i), minus any coverage requirements that are postponed to succeeding stages via ST_i. Through (4.15) there exists a one-to-one relationship between the net replenishment time defined by the adjacent service times and the order-up-to level of a stage. Consequently, instead of searching for the optimal order-up-to levels for the entire supply chain, one can also try to find the optimal service time or net replenishment time at each stage.

If it is assumed that any demands that exceed the available stock are dealt with by operating flexibility measures, no backorders occur at any of the intermediate stages $i = 1, .., n-1$. This means that in the long run the average expected (on-hand) stock quantity of a stage, for which inventory holding costs are incurred, corresponds to its safety stock, which for a certain τ_i is given as

$$SST_i(B_i(\tau_i)) = B_i(\tau_i) - \tau_i \mu \qquad i = 1, ..., n-1 \quad . \tag{4.17}$$

Only at the final stage n might backorders be permitted. Since the service-level at this stage is usually quite high, $\alpha_n^{target} \geq 90\%$, the backorder quantity is rather small. That means, for this stage the expected safety stock quantity is also a good approximation for the expected on-hand stock, the actual quantity, for which holding costs are incurred, due to relation (2.27) (see, e.g., Silver et al. (1998)).

Since the effect of operating flexibility on the material flow is not explicitly modeled in the standard GS approach, the pipeline inventory at each stage is simply assumed to be

$$\mathbb{E}[PI_i] = T_i \cdot \mu \quad . \tag{4.18}$$

This expression cannot be influenced by the service time or net replenishment time choice. Hence, it can be neglected in the optimization problem formulation (without affecting optimality). Using the common assumption that the external supplier has ample stock, i.e. $ST_0 = 0$, and the external customer requires immediate demand satisfaction, i.e. $ST_n = 0$, the optimization problem $\mathbf{P}^{GS^{std}}_{\langle 1,n\rangle}$ is (cf., e.g., Inderfurth and Minner (1998))

$$\mathbf{P}^{GS^{std}}_{\langle 1,n\rangle} \quad \min \; C^{GS^{std}}_{\langle 1,n\rangle}(\vec{\tau}_{\langle 1,n\rangle}) = \sum_{i=1}^{n} h_i \cdot SST_i(B_i(\tau_i)) \tag{4.19}$$

$$\text{s.t.} \quad \sum_{j=1}^{i} \tau_j \leq \sum_{j=1}^{i} T_j \qquad i = 1, ..., n-1$$

$$\sum_{j=1}^{n} \tau_j = \sum_{j=1}^{n} T_j$$

$$\tau_i \geq 0 \qquad i = 1, ..., n \quad .$$

$\vec{\tau}_{\langle 1,n\rangle}$ denotes the vector of net replenishment times in the network from stage 1 to n. The objective function minimizes the inventory holding costs in the entire system. The first constraint ensures that the cumulative net replenishment time until stage i does not exceed the cumulative processing time. The second constraint makes sure that the cumulative processing time in the entire system is covered. Finally, the net replenishment time must be non-negative.

Optimization procedure

The optimal net replenishment time combination for this optimization problem can be found by a dynamic program (DP) with backward recursion. The states, z_i, represent the time that is still to be covered at a specific stage of the supply chain. The decision variables, u_i, represent the net replenishment times, τ_i. The complexity of the state and decision space and consequently the DP complexity depends on whether the objective function is a concave function of τ_i or not. In the former case, $\mathbf{P}^{GS^{std}}_{\langle 1,n\rangle}$ is a concave minimization problem under linear constraints, for which an extreme point property holds (see, e.g., Horst and Tuy (1996)). That means, an optimal decision is found at a vertex of the decision space. Hence, a stage either holds sufficient stock to completely decouple itself from its successor, i.e. $\tau_i = ST_{i-1} + T_i$, or no stock at all, i.e. $\tau_i = 0$ (see Simpson (1958)). In the latter case, (assuming integrality of the processing times) all feasible integer net replenishment times need to be considered as decision candidates at a stage. Since the concavity of the objective function in τ_i does not hold for all demand distributions, the DP formulation of

Graves and Willems (2000), which does not depend on this assumption, is presented as a solution algorithm. The state space is given as

$$\mathcal{Z}_i = \{z \in \mathbb{N} \mid T_i \leq z \leq M_i\} \qquad i = 1, ..., n \tag{4.20}$$

where M_i denotes the maximum replenishment time for stage i, i.e. $M_i = \sum_{l=1}^{i} T_l$. The decision space is

$$\mathcal{U}_i(z_i) = \{u_i \in \mathbb{N} \mid 0 \leq u_i \leq z_i\} \qquad i = 1, ..., n-1 \tag{4.21}$$

$$\mathcal{U}_n(z_n) = \{z_n\} \quad . \tag{4.22}$$

The state transition equation looks as follows:

$$z_{i+1} = z_i + T_{i+1} - u_i \qquad i = 1, ..., n-1 \quad . \tag{4.23}$$

For each u_i the inventory holding cost can be calculated as explained above by substituting u_i for τ_i. The value function is given as

$$g_n(z_n) = h_n \cdot SST_n(B_n(u_n)) \qquad \forall z_n \in \mathcal{Z}_n \tag{4.24}$$

$$g_i(z_i) = \min_{u_i \in \mathcal{U}_i(z_i)} \{h_i \cdot SST_i(B_i(u_i)) + g_{i+1}(z_i + T_{i+1} - u_i)\} \qquad \forall z_i \in \mathcal{Z}_i \ , \quad i = 1, ..., n-1 \ . \tag{4.25}$$

Starting at stage n (down to 1), for each state the preliminary optimal decision is computed. Having reached stage 1 the overall optimal decisions can be found by a forward calculation. The complexity of the dynamic program is $O(nM^2)$ (where M is the maximum replenishment time, which is bounded by the sum of processing times $\sum_{l=1}^{n} T_l$, see Graves and Willems (2000)). In case of a concave objective function the complexity reduces to $O(n^2)$ (see, e.g., Minner (1997)).

4.2.3.2 Extended optimization model and procedure

The GS optimization model formulation of the previous section has been based on addressing only one of two crucial aspects, namely the specification of the maximum reasonable demand levels expressed as internal service levels. The second aspect, the effect of the operating flexibility measures on the material flow, has not been taken into account. Further, in the standard GS objective function, costs for using operating flexibility are not included explicitly, but only implicitly through setting the internal service level. These two facets are studied in this section, which is partly based on Klosterhalfen and Minner (2010).

Operating flexibility modeling options

In order to incorporate operating flexibility in the analysis, it needs to be clarified first what kind of measures operating flexibility can take. There are several possibilities for operating flexibility to achieve the guaranteed service, i.e. ensure that a material shortage at stage i does not affect the arrival of all ordered items in stock at stage $i+1$ after the replenishment time $S_i + T_{i+1}$. The three probably most obvious and relevant ones for practice are listed here. The first two modeling interpretations have already been outlined in Minner (2000).

1. The shortage quantity is directly speeded up from pipeline inventory (PI) of the stage. Thus, there is no shortage in fact. Due to the expediting, $\mathbb{E}[PI_i] < T_i \cdot \mu$. ('Production setting')

2. The shortage quantity is not speeded up from pipeline inventory of the stage, but the stage waits for the items to arrive after the regular processing. Once the missing quantity is available at the stage, it is sent to the next downstream stage via a faster transportation mode. Thus, although a shortage occurred at a stage, it does not affect the next downstream stage, because the timely arrival of all ordered items in stock at the downstream stage is still ensured. This option also results in $\mathbb{E}[PI_i] < T_i \cdot \mu$, since in some periods fewer items are in the pipeline than actually ordered. ('Transportation setting')

3. The shortage quantity is sourced from an external/outside supplier. Also in this case, the average replenishment order placed with the internal supplier would no longer be equal to μ and $\mathbb{E}[PI_i] < T_i \cdot \mu$ under the assumed base-stock policy. Note that this modeling option resembles a lost-sales situation. ('Outside supply setting')

In all three cases the average expected pipeline inventory is smaller than $T_i \cdot \mu$, which is in contrast to the standard GS model that simply assumes $\mathbb{E}[PI_i] = T_i \cdot \mu$. Since this thesis is concerned with production rather than transportation settings, the second modeling option is excluded from the further analysis. The third modeling aspect assumes a kind of second supply option. Although the incorporation of dual sourcing into a multi-echelon inventory model is the ultimate goal of this chapter, this option represents only a special kind of usage of the second supplier and as such is rather restrictive. Only the missing items are sourced from this outside supplier and in addition these items are delivered immediately, which is a rather unrealistic assumption. Moreover, the demand process at the upstream stage would be influenced, if a second supplier was used. This would further complicate the analysis. That is why the focus of the remainder of this section lies on the first operating flexibility interpretation.

A proper incorporation of the first operating flexibility option into the GS model requires knowledge about two things:

1. The quantity of items that is expedited from pipeline inventory, and
2. The timespan for which the expediting takes place.

The expediting quantity and, in turn, reduction in pipeline stock corresponds to the expected backorders of a stage, since this is the quantity that, in the absence of operating flexibility, would have to wait to be delivered until new material becomes available from incoming orders.

In order to get an idea about the expediting timespan, a simulation study of a single stage is conducted. Various parameter settings are tested (see Table 4.1). A full-factorial design is used. In total, 108 settings are analyzed. For each parameter setting, 10 simulation runs with 100,000 periods each are conducted.

The results of the simulation study are illustrated in Figure 4.2. The bars indicate the share of items that are expedited by a certain number of periods for different processing times and internal service levels given on the x-axis. The expediting timespan is represented by the different shades of the bars. The following results can be derived:

	Type	Parameter	Value
Demand	discretized Normal	Mean, μ CV	100 $0.1, 0.3, 0.5$
	discretized Gamma	Mean, μ CV	100 0.8, 1.0, 1.5
Processing time		T	2, 5, 10
Internal service level		α^{target}	16.67%, 25%, 50%, 75%, 85%, 95%

Table 4.1: Expediting analysis for the extended GS approach – Parameter values

- For $CV \leq 0.5$ and internal service levels of 50% or higher, more than half of the items are expedited by 1 period only. This means that the speeded-up items would have arrived in the next period. Segregating the analysis by the processing time yields:
  - For a short processing time ($T = 2$), the share of items, which are speeded-up by 1 period only, amounts to 90% or higher.
  - For a medium processing time ($T = 5$), this share is still above 70%.
  - Only for a long processing time ($T = 10$) is the minimum share 54%, but it increases as the internal service level increases or the coefficient of variation of period demand decreases.
- For $CV \leq 0.5$, the expediting timespan does not exceed the processing time of the stage. In a multi-echelon setting, this means that the expediting of items does not take place across upstream stages irrespective of the internal service level. (For $\alpha^{target} = 16.67\%$ and $\alpha^{target} = 25\%$, negligible shares of 0.24% and 0.06%, respectively, occurred.) Only items that are in the pipeline of the stage, which experiences the shortage, need to be speeded up. No items from upstream stages are required.
- If $CV = 0.8$, the internal service level needs to be at least 75% in order to ensure that the speeding up does not affect items from upstream stages in a significant way. For higher CVs, at least for the short processing time ($T = 2$) the share of items that requires expediting by more than 2 periods is not negligible any more.

- The setting with $CV = 0.8$ and $\alpha^{target} = 75\%$ can be viewed as the minimum combination, for which the share of items expedited by 1 period amounts to roughly 50% ($T = 10$) or higher ($T < 10$).

Given these findings, the following two simplifying assumptions can be made in situations with $CV \leq 0.5$ and an arbitrary internal service-level target or $CV = 0.8$ and $\alpha^{target} \geq 75\%$.

Assumption 4.2.3.1 *Expediting does not take place across various upstream stages, but only items in the pipeline of the stage, which experiences a shortage, are speeded up.*

Assumption 4.2.3.2 *If a cost per unit for the operating flexibility usage at a stage is to be specified, a single cost value, c_i^{OF}, is sufficient.*

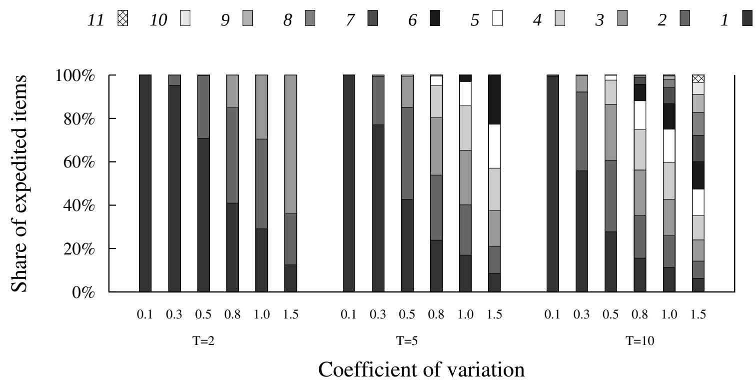
Since the majority of items is speeded up by 1 period only, a specification of several cost values depending on the expediting time span would unnecessarily complicate the analysis (without any significant additional benefit). Consequently, the standard GS cost function of a stage can be extended by a term that accounts for the operating flexibility costs in the following way. Since pipeline stock is reduced by the amount of expedited items, this term has to be included in the new cost function, as well.

$$\begin{aligned} C_i &= h_i \cdot (\mathbb{E}[PI_i] - \mathbb{E}[BO_i(B_i)] + \mathbb{E}[OH_i(B_i)]) + c_i^{OF} \cdot \mathbb{E}[BO_i(B_i)] \\ &= h_i \cdot (T_i \cdot \mu + \mathbb{E}[OH_i(B_i)]) + \left(c_i^{OF} - h_i\right) \cdot \mathbb{E}[BO_i(B_i)]. \end{aligned} \tag{4.26}$$

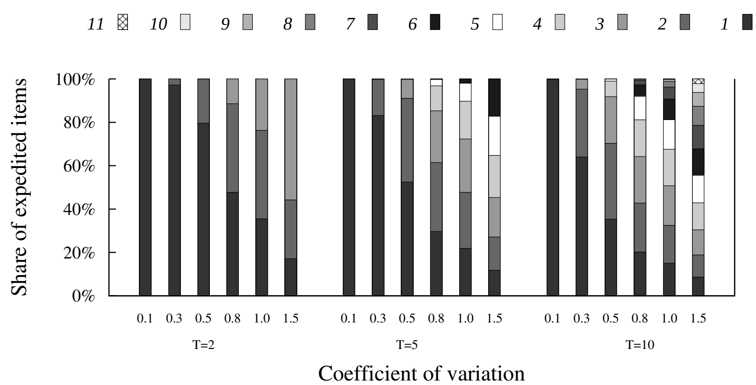
Note that it is assumed that the holding cost of stage i, h_i, is also paid for the units in transit to this stage. The quantity that needs to be made available by operating flexibility depends on the order-up-to level, which in turn depends on the internal service level (due to relation (4.15)), since this defines the expected backorders. In the following, two ways of how this service level can be specified at a stage are discussed.

Direct internal service-level specification

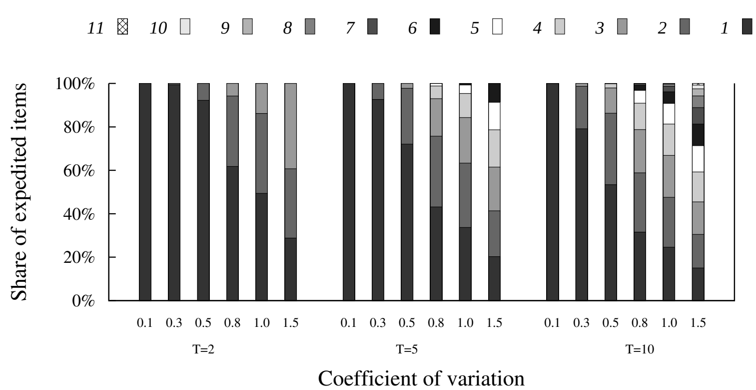
In some situations management might be able to directly specify an appropriate internal service-level target at a stage. By having good acumen of how often the



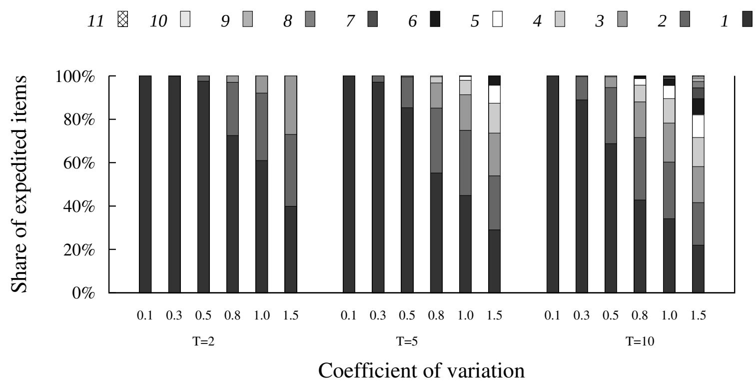
(a) Internal service level 16.67%



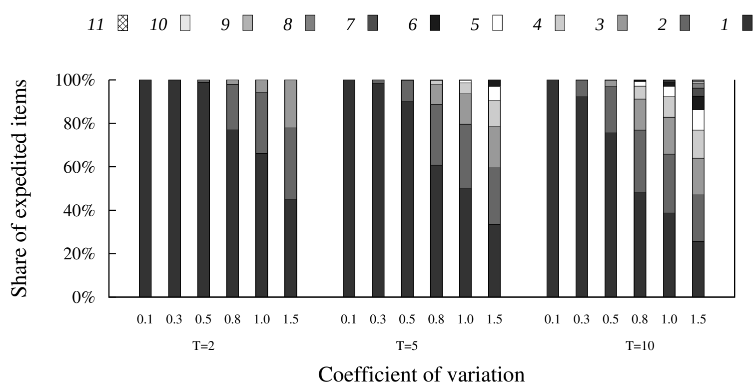
(b) Internal service level 25%



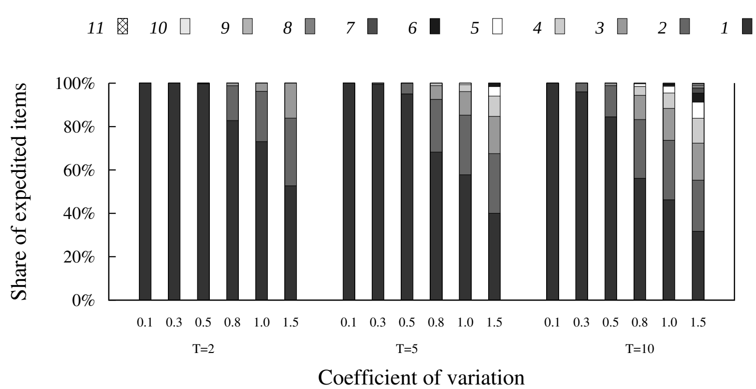
(c) Internal service level 50%



(d) Internal service level 75%



(e) Internal service level 85%



(f) Internal service level 95%

Figure 4.2: GS approach – Expedited items analysis

stage can deal with extraordinary situations or how many items can be provided in extreme situations without causing any additional cost, they can set a service level, which reflects the stage's flexibility. If the service level is specified in this way, the operating flexibility usage is free of any additional charge, which is a reasonable assumption at least in the short run (see van Houtum et al. (1996)). Since the current charge per item at a stage is the holding cost and the operating flexibility cost must not be higher, it must hold that $c_i^{OF} \leq h_i$. On the other hand, c_i^{OF} must not be smaller than h_i. Otherwise, it would be beneficial to always rely on operating flexibility and not stock any items at all. Consequently, $c_i^{OF} = h_i$ follows and (4.26) reduces to

$$C_i = h_i \cdot (T_i \cdot \mu + \mathbb{E}[OH_i]) \qquad i = 1, ..., n-1. \tag{4.27}$$

At the final stage n, the cost function is identical to (4.27). Here, the service-level target is given by the service that is to be guaranteed towards the external customer. At this stage no operating flexibility is required to achieve this service.

Cost-based internal service-level specification

In some situations it might be difficult to directly specify a service-level target for a stage representing its flexibility, i.e. the amount of extraordinary measures that can be used without any additional charge. The problem is basically comparable to the specification of the backorder cost per item in a single-echelon backorder cost model. In the latter model, it can be solved by deriving an implied backorder cost from the corresponding service-level model. Here, the difficulty points in the opposite direction. Instead of the service level, a cost per item for the operating flexibility usage might be available more easily. For instance, it might be quantifiable what the cost per additional worker is, which can be further broken down into a cost per item. From (4.26) a relation between the α-service level and the cost parameters c_i^{OF} and h_i can be derived provided that $c_i^{OF} \geq h_i$, which is a reasonable assumption, because otherwise it would be optimal not to stock anything at the stage. The following lemmata hold.

Lemma 4.2.3.3 *For a given τ_i, the extended cost function is convex in B_i and the unique optimum is given as*

$$B_i^*(\tau_i) = F_{\tau_i}^{-1}\left(1 - \frac{h_i}{c_i^{OF}}\right) \quad . \tag{4.28}$$

Proof:

See Appendix B.8. □

Lemma 4.2.3.4 *The internal α-service level can be determined independent of τ_i as*

$$\alpha_i^{target} = 1 - \frac{h_i}{c_i^{OF}} \quad . \tag{4.29}$$

Proof:

See Appendix B.9. □

Remark. (4.29) is comparable to the α-service level-backorder cost relation in a single-echelon order-up-to level model (see (2.52)). Setting $b_i = c_i^{OF} - h_i$, $\frac{b_i}{b_i+h_i}$ results.

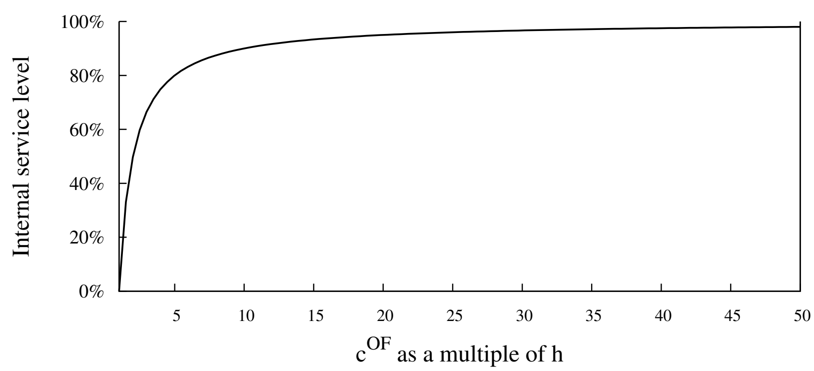


Figure 4.3: Internal service level as a function of the operating flexibility cost

Figure 4.3 shows the resulting internal service level as the operating flexibility cost (expressed as a multiple of the holding cost at the stage) is increased. The internal service level function is very sensitive to an operating flexibility cost of up to 10 times the holding cost. Then it flattens out. Once the internal service level has been determined for each stage, the optimal net replenishment times can be found by only considering the holding cost for the pipeline inventory and on-hand stock at a stage. The optimal trade-off between the holding cost and operating flexibility cost is taken into account by the internal service level. Thus, only cost function (4.27) is relevant for the optimization.

Special case: Normally-distributed demand. For the special case of normally-distributed demand, the following corollary holds due to the possible standardization of the order-up-to level, which results in the safety factor k_i (cf. Klosterhalfen and Minner (2010)).

Corollary 4.2.3.5 *Under an (internal) α-service level constraint and normally distributed period demand, the extended GS cost function of a stage i (4.26) is a convex function of the safety factor, k_i, and has a unique optimum, which is independent of the net replenishment time, τ_i.*

$$\Phi(k_i) = \alpha_i^{target} = 1 - \frac{h_i}{c_i^{OF}} \;\Leftrightarrow\; k_i^* = \Phi^{-1}\left(\alpha_i^{target}\right) \tag{4.30}$$

Lemma 4.2.3.6 *Under an (internal) α-service level constraint and normally distributed period demand, the extended GS cost function of a stage is a concave function of the net replenishment time, τ_i, for a given safety factor, k_i.*

Proof:
See Appendix B.10. □

Optimization model

Given both ways of how to specify the internal service level, the objective function of the standard GS model, which minimizes safety stock costs across the supply chain, is replaced by the sum of the on-hand stock costs across all stages. Note that the pipeline inventory expression in (4.27) can be neglected, because it cannot be influenced by the decision variables, τ_i. For a given τ_i, the expected on-hand stock at a stage is

$$\mathbb{E}\left[OH_i(B_i(\tau_i))\right] = B_i(\tau_i) - \tau_i \cdot \mu + \mathbb{E}\left[BO_i(B_i(\tau_i))\right] \qquad i = 1, ..., n \tag{4.31}$$

with $B_i(\tau_i)$ from (4.15) and

$$\mathbb{E}\left[BO_i(B_i(\tau_i))\right] = \mathbb{E}\left[(D(\tau_i) - B_i(\tau_i))^+\right] \qquad i = 1, ..., n. \tag{4.32}$$

Note that unlike in (4.2) and (4.3) no backorders from the previous stage have to be taken into account in the expected on-hand stock and backorder calculation due to the guaranteed service.

Using the same assumptions as in the standard GS model, namely that the external supplier has ample stock, i.e. $ST_0 = 0$, and the external customer requires immediate demand satisfaction, i.e. $ST_n = 0$, the optimization problem $\mathbf{P}^{GS}_{\langle 1,n\rangle}$ is

$$\begin{aligned}
\mathbf{P}^{GS}_{\langle 1,n\rangle} \quad \min\ C^{GS}_{\langle 1,n\rangle}(\vec{\tau}_{\langle 1,n\rangle}) &= \sum_{i=1}^{n} h_i \cdot \mathbb{E}[OH_i(B_i(\tau_i))] && (4.33)\\
\text{s.t.} \quad \sum_{j=1}^{i} \tau_j &\leq \sum_{j=1}^{i} T_j && i = 1, ..., n-1\\
\sum_{j=1}^{n} \tau_j &= \sum_{j=1}^{n} T_j &&\\
\tau_i &\geq 0 && i = 1, ..., n \quad .
\end{aligned}$$

$\vec{\tau}_{\langle 1,n\rangle}$ denotes the vector of net replenishment times in the network from stage 1 to n. The objective function minimizes the cost of the expected on-hand stock in the entire system. The first constraint ensures that the cumulative net replenishment time until stage i does not exceed the cumulative processing time. The second constraint makes sure that the cumulative processing time in the entire system is covered. Finally, the net replenishment time must be non-negative.

The main finding is that the explicit modeling of the operating flexibility effect on the material flow yields a different objective function than in the standard GS model. Whereas in the extended model an inventory holding cost for each unit of on-hand stock is incurred, the standard GS model calculates this cost on the safety stock units, which are usually less except for very high service levels.

Optimization procedure

If the internal service levels are specified directly, the optimal net replenishment time combination for optimization problem $\mathbf{P}^{GS}_{\langle 1,n \rangle}$ can be found directly by using the dynamic program (DP) with backward recursion of the standard GS model (see Section 4.2.3.1). Only the stage cost function needs to be adjusted to the on-hand stock cost.

If the internal service level is derived via the cost-based approach, Lemma 4.2.3.4 allows for a sequential solution procedure. First, the internal service level (which is independent of the net replenishment time) is determined for each stage based on the operating flexibility cost per unit. Second, optimal net replenishment times, τ_i, are computed given the internal service level via the standard DP again. Under normally distributed demand, the extended GS cost function of a stage is concave in the net replenishment time under an α-service level constraint (Lemma 4.2.3.6). Consequently, an extreme point property holds for this extended GS model. Optimal net replenishment time values can therefore be found by the simplified standard DP.

4.2.4 Summary and implications

In this section, the two main multi-echelon modeling frameworks that can be found in the literature have been outlined, i.e. the stochastic-service (SS) and guaranteed-service (GS) approach. Besides summarizing the existing models and results, one of the main criticisms of the GS approach has been addressed. In its standard form, this approach assumes that safety stock is only sized to cover demand variability up to a certain level, the maximum reasonable demand. All variability exceeding this threshold is dealt with by other countermeasures, which are simply referred to as operating flexibility. In the mathematical model, it is not explicitly detailed how these operating flexibility measures work and what the effect on the material flow is. Moreover, it is difficult to define what is normal variability and what is not.

In this section, this criticism has been counteracted by modeling the effect of operating flexibility on the material flow in the system. It has been assumed that missing items are made available in time through expediting from the pipeline inventory of a stage. The reasonability of this assumption has been tested in a simulation study and confirmed. By taking into account the way operating flexibility works, the objective function of the standard GS model has been modified. Instead of minimizing the safety stock cost in the entire system, the on-hand stock cost has become relevant in the extended model. Through the model extension a cost-based derivation of the internal service level at a stage, which specifies the maximum reasonable demand level, has become possible. Provided that a cost parameter for the operating flexibility usage can be specified per unit, a closed-form expression has been shown to exist, which gives the corresponding internal service level. In many situations, the specification of an operating flexibility cost might be easier for management than specifying a service level directly.

4.3 Comparison of the stochastic- and guaranteed-service approach

4.3.1 Introduction

After the description of both multi-echelon frameworks, the question arises: *Is one of them superior to the other, in general? And, if this is not the case, in which settings does each approach perform well?*

In the literature, very few contributions are available that focus on such a comparison. The most prominent one is probably the one by Graves and Willems (2003). They compare the two approaches in a convergent and spanning-tree system and find that (under their assumptions) the GS model performs better. Since they use a slightly different variant of the SS approach from the one presented in the previous section, their results are not fully conferrable to the SS framework analyzed in this thesis. Moreover, they apply the demand bound assumption and specify an identical bound for each stage in the supply network. As mentioned above, this might not fully reflect reality, where different levels of flexibility can be present at different stages in the network. That is why in this section a separate theoretical and numerical comparison of both approaches is conducted.

The structure of this section is as follows. Section 4.3.2 establishes an individual benefit of each approach based on theoretical considerations. In Section 4.3.3 numerical studies are conducted in order to derive further insights, first for serial systems (Section 4.3.3.1), then for divergent systems (Section 4.3.3.2). Some remarks on convergent systems are presented in Section 4.3.3.3. The section closes with a summary and implications in Section 4.3.4.

4.3.2 Theoretical considerations – Benefits of the approaches

From the framework description in Section 4.2, clear differences between the two approaches can be observed in terms of the allocation and sizing of safety stock in the supply chain. These differences result in an individual benefit of each approach: the *allocation benefit* of the SS approach and the *decoupling benefit* of the GS approach. Minner (2000) also points out these differences in his discussion of the two approaches. For illustrative purposes assume a two-stage serial system where stage 1 receives external supply and supplies stage 2, which in turn satisfies the external customer.

4.3.2.1 Allocation benefit of the stochastic-service approach

In the GS model, there is a direct relation between the allocation decision and the stock quantity at a stage. If a stage holds stock, i.e. $\tau_i > 0, i = 1, 2$, the exact size follows directly from the service-level requirement (internal or external).

$$\alpha_i^{target} = \int_0^{B_i} f_{\tau_i}(u)\; du \quad \Rightarrow \quad B_i = F_{\tau_i}^{-1}\left(\alpha_i^{target}\right) \qquad i = 1, 2 \qquad (4.34)$$

Due to the predefined internal service level and 100% service guarantee of the predecessor, it is not possible to substitute (safety) stock at the predecessor for additional (safety) stock at the successor.

In contrast, the SS model allows for such a substitution. Due to the echelon stock concept, only the total quantity of (safety) stock in the entire supply chain (resulting from the order-up-to levels of all stages) has to be sufficient to meet the external service-level requirement (see (4.7) for $i = 1$). The ultimate allocation of (safety) stocks to the individual stages depends on the holding-cost relationship between them. This is done in the following way. First, an appropriate penalty cost is derived, which ensures an α-service level for the entire supply chain equal to the external service-level requirement using relationship (4.12)

$$\alpha_2^{target} = \frac{p}{p+h_2} \quad \Rightarrow \quad p = h_2 \cdot \frac{\alpha_2^{target}}{1-\alpha_2^{target}} \quad . \tag{4.35}$$

Next, using this penalty cost, implied service levels, $\alpha_i^{SS}, i = 1,2$, are derived according to (4.7).

$$\alpha_2^{SS} = \frac{p+h_1}{p+h_2} \quad , \quad \alpha_1^{SS} = \frac{p}{p+h_2} = \alpha_2^{target} \tag{4.36}$$

Starting with the final stage, these implied service levels are used to determine the order-up-to levels of the stages in the system.

$$\int_0^{S_2} f_{T_2}(u)\, du = \alpha_2^{SS} \tag{4.37}$$

$$\int_0^{S_2} \int_0^{S_1-u} f_{T_1}(v)\; f_{T_2}(u)\, dv\, du = \alpha_1^{SS} = \alpha_2^{target} \tag{4.38}$$

(4.37) shows that, in general, the SS approach prescribes larger order-up-to levels than the GS approach. Assuming that $ST_1 = 0$ (i.e. $\tau_2 = T_2$) in the GS model, the final stage order-up-to level is sized with respect to α_2^{target} (see (4.34)). However, from (4.36) and (4.37) it is obvious that $\alpha_2^{target} < \alpha_2^{SS}$, the service level with respect to which the final stage order-up-to level in the SS model is dimensioned (except for the case where $h_1 = 0$). The larger order-up-to level is necessary to a certain extent, since the SS approach does not assume any operating flexibility measures and consequently delivery delays at the predecessor might occur. Given the final-stage order-up-to level (and safety stocks) in the SS framework, the order-up-to level of the predecessor is then set such that the external service-level requirement is met (see (4.38)). Thus, the SS approach benefits from the flexibility of either shifting more (safety) stock to upstream stages reducing the stockout risk or to the downstream ones allowing for a larger stockout probability of the upstream stages, whichever is cost-optimal. This is called the *allocation benefit*.

4.3.2.2 Decoupling benefit of the guaranteed-service approach

In the GS approach, on the other hand, this kind of flexibility with respect to the (safety) stock quantity is not available. There is no possibility of placing more (safety) stock at downstream stages than is actually required for the predefined (internal or external) service level. Nevertheless, there is another kind of flexibility available in the GS model, namely operating flexibility, which results in a decoupling effect between the stages. No shortfall and thus stochastic delay is propagated, which reduces the stock requirement of the next downstream stage. Each stage has a deterministic service time and can be viewed as an external supplier with sufficient stock to always fulfill its service-time guarantee. This is called the *decoupling benefit*.

In order to quantify the effects of these opposed flexibility types, a numerical study is conducted in the next section.

4.3.3 Numerical study

4.3.3.1 Serial systems

Numerical design

The simplest version of a serial supply chain is analyzed consisting of two stages, where stage 1 supplies stage 2 and the parameters given in Table 4.2. The parameters are chosen such that a large range of supply chain characteristics is captured. With regard to the processing time a short (2 periods), medium (5 periods), and long (10 periods) timespan is considered. Holding costs follow a value-adding structure with the upstream stage holding cost set to $h_1 = 10$ and an increase towards the downstream one by 20%, 50%, or 100%. The external/final-stage α-service level is varied between 85%, 95%, or 99%. In all instances, demand per period is assumed to follow a discretized normal distribution with the demands of different periods being *i.i.d.*. Mean demand per period is set to 100 and the coefficient of variation (CV) is either 0.1, 0.3, or 0.5. Thus, different levels of variability are captured. Moreover, the maximum CV of 0.5 ensures that the simplifying assumptions 4.2.3.1 and 4.2.3.2 with respect to the operating flexibility modeling are not violated irrespective of the height of the internal service level. Various internal service-level values are studied: 17%, 33%, 50%, 67%, and 75%. Using relation (4.29), these values correspond to an operating flexibility cost per unit of about 12, 15, 20, 30, 40, i.e. 1.2 times up to 4 times the holding cost of the stage 1.

Parameter	Description	Value
T_i	Processing time of stage $i \in \{1, 2\}$ including the review period at the final stage	2, 5, 10
h_1	Holding cost at stage 1	10
h_2	Holding cost at stage 2	12, 15, 20
μ	Mean period demand at stage 2	100
CV	Coefficient of variation of demand at stage 2	0.1, 0.3, 0.5
α_2^{target}	α-service level at stage 2 (final-stage)	85%, 95%, 99%
α_1^{target}	α-service level for GS at stage 1 (internal)	17%, 33%, 50%, 67%, 75%

Table 4.2: Serial system – Parameter values for simulation runs

For each internal service-level value, 243 instances are analyzed in a full-factorial design. For each instance, 10 simulation runs with 20,000 periods each are conducted.

Computational aspects

For both approaches a simulation model is built in ARENA v10. In case of the SS approach, optimal order-up-to levels are calculated by the algorithm described in Section 4.2.2.1 implemented in MAPLE v10. Numerical integration together with bisections for the echelon order-up-to levels are used to derive the optimal values.

For the GS approach, results are obtained by comparing the extreme points of all possible net replenishment time combinations and their respective on-hand stock costs. Computations are performed in MAPLE v10 for all different internal service levels, as well.

Since the comparison is based on simulation results, an ANOVA is conducted to establish the significance of the cost difference between the two approaches. The level of significance is 95%. The required normal distribution assumption of the 10 cost results is confirmed by the Kolmogorov-Smirnov test with a level of significance of 99%. All numbers reported in the upcoming tables and figures represent significant values.

Results

In the analysis, two performance measures are used. The *optimality share* indicates the fraction of instances where one or the other approach is optimal, $OS^m,\ m \in \{GS, SS\}$. The *relative benefit* of a respective approach quantifies the cost savings and is defined as

$$RB^{GS} = \left(1 - \frac{C^{GS}_{\langle 1,n \rangle}}{C^{SS}_{\langle 1,n \rangle}}\right)^+ , \ RB^{SS} = \left(1 - \frac{C^{SS}_{\langle 1,n \rangle}}{C^{GS}_{\langle 1,n \rangle}}\right)^+ \quad .$$

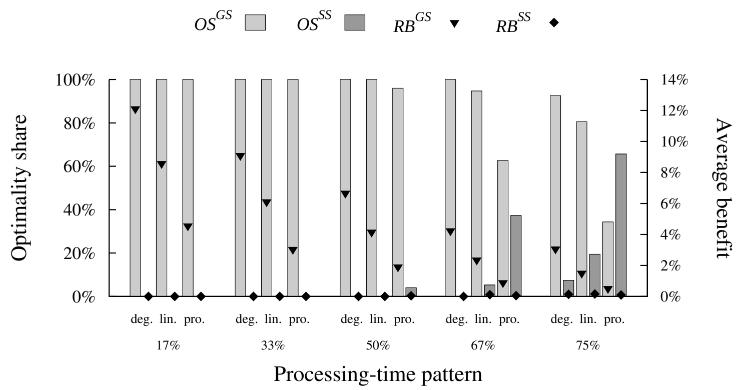
In the figures below, the shaded bars represent the optimality share and refer to the left-hand y-axis. The average relative benefit is illustrated by (upside-down) triangles and diamonds, which refer to the right-hand y-axis. On the x-axis the different levels of the specific factor are indicated for different internal service levels.

As discussed in Section 4.2.3.2, the internal service level can either be set directly by management or derived from the operating flexibility cost, c^{OF}. In the former case, no additional costs besides the on-hand stock cost would accrue. In the latter case, for each expedited item a cost of c^{OF} would have to be paid in addition to the on-hand stock cost. Results for both cases are reported.

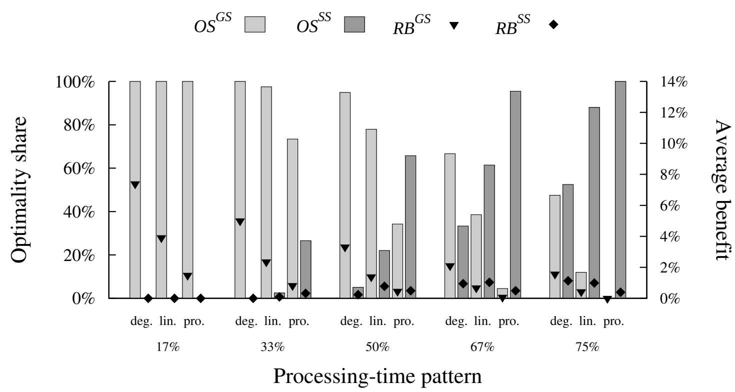
In order to identify drivers for the advantage of one approach over the other, three parameter dimensions are analyzed: processing time, final-stage service level, and coefficient of variation of period demand. Within each dimension, the effect of the varying internal service level is investigated, too. The following observations are made.

Processing time. Instead of analyzing all possible processing-time combinations individually, they are summarized into patterns: degressive (deg.), linear (lin.), and progressive (pro.). *Irrespective of the actual internal service level, it can be observed that the share of GS optima decreases from a degressive over linear to progressive processing-time pattern. Similarly, the average GS benefit decreases from a degressive over linear to progressive processing-time pattern.* (See Figures 4.4(a) and 4.4(b).)

In the situation without any additional operating flexibility costs, the GS optimality share for the degressive processing-time pattern is reduced only slightly from 100%



(a) without c^{OF}



(b) with c^{OF}

Figure 4.4: Serial system – Optimality share and relative benefit with respect to the processing time

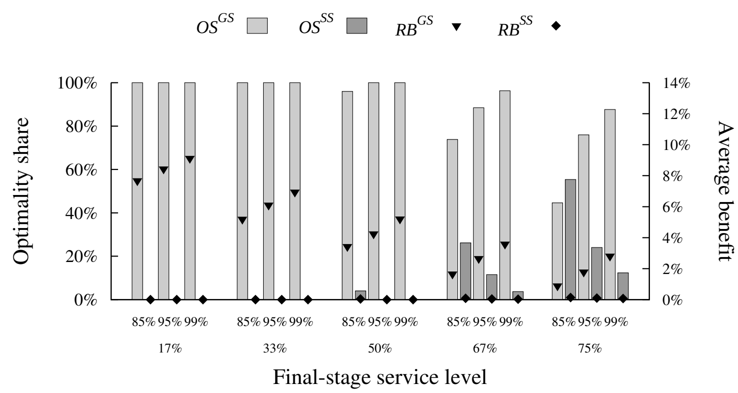
to 92.6% as the internal service level increases from 17% to 75%. For the linear and progressive pattern the effect is more pronounced with a decrease from 100% to 80.5% and 34.3%, respectively. In terms of the average GS benefit, for each internal service-level value the benefit shrinks to about a third (or even less) as the processing-time pattern changes from degressive over linear to progressive (see Figure 4.4(a)).

In the situation with additional operating flexibility costs, the GS optimality share is reduced by a larger extent. For a degressive, linear, and progressive processing-time pattern it shrinks from 100% to 47.5%, 12%, and 0%, respectively. As a matter of course, the SS optimality share increases accordingly. With regard to the average GS benefit, this benefit decreases to less than a fifth due to the change in the processing-time pattern for a given internal service level (see Figure 4.4(b)). For an internal service level of 75% the average relative GS benefit even completely disappears for a progressive processing-time pattern.

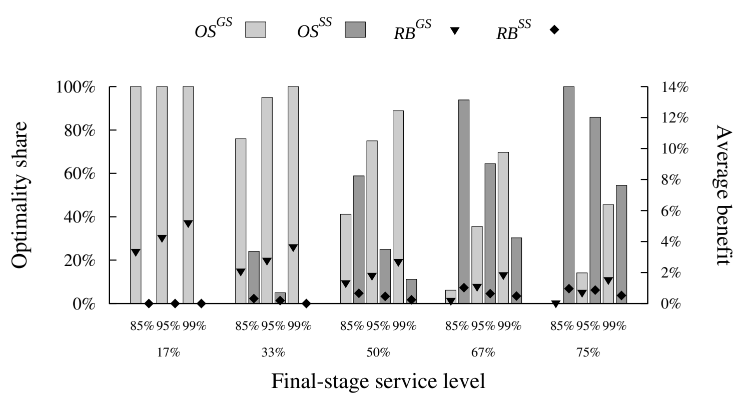
The reason for this effect lies in the different ways both approaches deal with a processing-time pattern change. The GS approach can use its decoupling benefit only in case of upstream coverage. Holding stock at upstream stages becomes more advantageous with larger processing times in the upstream part of the supply chain compared to the downstream part due to the square root effect of the processing time. This is reflected by a degressive pattern. Moreover, the GS approach does not change its internal service level according to the processing time. This is contrary to the SS approach, which raises the internal service level at the upstream stage as its processing time increases. This results in a higher stock quantity and thus higher holding costs compared to the GS model. Obviously, the quantity of expedited items in the GS approach increases with a longer processing time. However, even for an operating flexibility cost that is 4 times the holding cost, which translates into an internal service level of 75%, the optimality share of both approaches are almost equally balanced and the average GS benefit is still about 1.6% for a degressive processing-time pattern. Here, the average SS benefit amounts to about 1.1%.

Final-stage service level. From Figures 4.5(a) and 4.5(b) it can be observed that the GS optimality share increases as the final-stage service level rises. In the GS approach, a higher final-stage service level requirement simply leads to an increase in the safety stock at this stage. The internal service level and thus stock quantity at the upstream stage remains unchanged. This is possible due to the operating flexibility measure, which enables such a decoupling.

In the SS approach, the change in the final-stage service-level requirement results in an altered penalty cost and therefore different (safety) stock quantities at both stages. Both order-up-to levels and thus stock quantities are increased. Due to the square root effect in the safety stock formula, the safety stock quantity needed for an increase in the service level grows exponentially the larger the service level gets. Since the final-stage service level is already higher than the internal service level (cf. (4.36)), allocating more safety stock to the upstream stage and thus reducing the stockout probability there, is more efficient than only increasing the safety stocks at the downstream stage as is done in the GS approach. Nevertheless, the SS order-up-to level and stock quantity at the final stage is still higher than that in the GS approach, because the SS model has to take into account potential supply shortages. This results in high costs, if the final-stage target service level is high. For the same reasons, the average GS benefit increases with an increase in the final-stage service level. Therefore, *a higher final-stage service level favors the GS approach in terms of the optimality share and average relative benefit.*



(a) without c^{OF}



(b) with c^{OF}

Figure 4.5: Serial system – Optimality share and relative benefit with respect to the final-stage service level

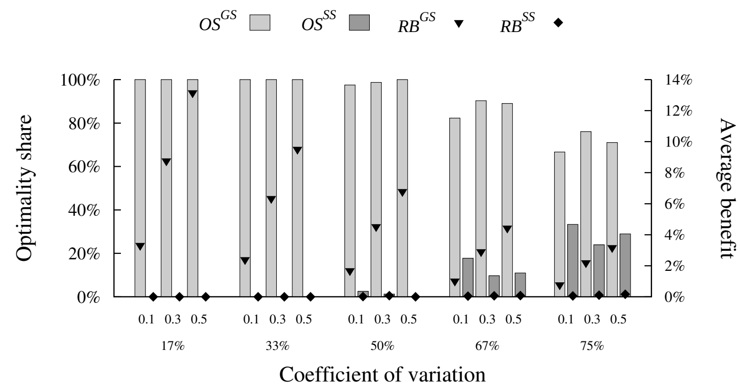
If no additional costs are incurred for the operating flexibility usage, the GS approach shows a very dominant performance over the SS approach in the considered settings (see Figure 4.5(a)). Only for an internal service level of 75% and a low final-stage service level of 85%, does the GS optimality share drop to 45%. In all other instances it varies above 75%. The average GS benefit increases by about 2% as the final-stage service level is increased from 85% to 99% for a given internal service level. Only for an internal service level of 67% or 75% can an average SS benefit be observed. This benefit is less than 0.2%, however.

If operating flexibility costs have to be paid, the SS optimality share increases considerably starting at an internal service level of 50% (see Figure 4.5(b)). In particular, for a rather low final-stage service level of 85% the SS optimality share rises from 58.8% to 100% as the internal service level increases from 50% to 75%. Similarly, the average SS benefit experiences a slight increase from 0.7% to 1%. In both situations the highest average relative GS benefit can be observed for a high final-stage and low internal service level with 9.1% and 5.2%, respectively.

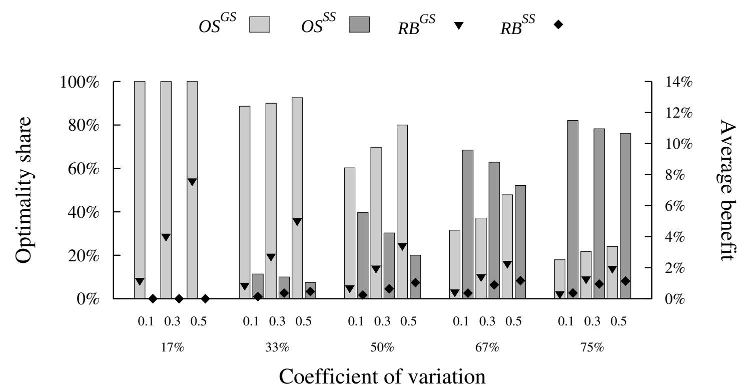
Coefficient of variation of period demand. *The GS approach performs better in terms of the optimality share and average relative benefit as demand becomes more variable.* This tendency becomes apparent in both situations, with and without operating flexibility costs (see Figures 4.6(a) and 4.6(b)). For each internal service level, the GS optimality share increases as the coefficient of variation (CV) changes from 0.1 to 0.3 to 0.5. (Note that the small anomaly in this tendency for an internal service level of 75% in the situation without additional operating flexibility costs results from the fact that for $CV = 0.3$ a relatively large number of instances with an average relative SS benefit exists, for which this benefit is not significant, however. Consequently, these instances are not taken into account in the performance measure calculation, which boosts the optimality share of the GS. If the significance test is neglected, the optimality shares for $CV = 0.1, 0.3$, and 0.5 amount to 64.2%, 66.7%, and 66.7%, i.e. the same tendency as for the other internal service levels can be observed.)

In the situation without any operating flexibility costs, the GS approach is clearly superior in the analyzed parameter settings (see Figure 4.6(a)). The average SS benefit only amounts to 0.2% at most, whereas the average GS benefit varies between 13.2% and 3.2% at most for the different internal service levels.

The clear GS superiority changes when additional operating flexibility costs are incurred (see Figure 4.6(b)). For internal service levels of 67% and 75% the SS optimality shares are larger than the GS ones. Also, the average SS benefit increases



(a) without c^{OF}



(b) with c^{OF}

Figure 4.6: Serial system – Optimality share and relative benefit with respect to the coefficient of variation of period demand

to about 1%. However, the average GS benefit lies above the SS one for almost all CVs and internal service levels.

The reasons for the better GS performance as the CV increases can be found in the operating flexibility. When the CV grows, the value of operating flexibility (if used) increases. Whereas the decoupling benefit of the GS becomes more important, the SS approach suffers from the CV-increase. This causes the decreasing optimality share. Nevertheless, in those instances where the SS model is superior to the GS one even with a low CV, the average SS benefit is more pronounced as the demand variability increases, i.e. *as demand becomes more variable, the SS optimality share decreases, but the average SS benefit increases.*

4.3.3.2 Divergent systems

Numerical design

In this section, which is based on Klosterhalfen and Minner (2010), the comparison of the two approaches is extended to the simplest divergent system, i.e. a network with one warehouse and two retailers (see Figure 4.7).

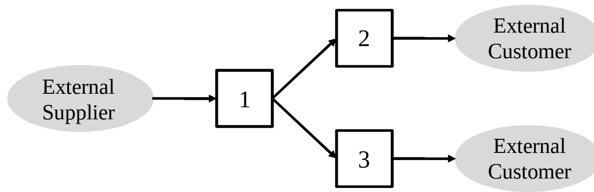


Figure 4.7: Divergent system illustration

The parameters are chosen similar to the ones of the numerical study for the serial system, but reduced to two possible parameter levels for each factor in order to avoid a too large increase in the number of total instances. All parameters are summarized in Table 4.3. With regard to the processing time, a time span of 2 periods (short) and 6 periods (medium-long) is considered. The warehouse holding cost is fixed at $h_1 = 10$ with a value adding towards the retailers by 50% or 100%. In all instances, demand is assumed to be *i.i.d.* normally distributed and discretized. There is no correlation between retailer demands, i.e. $\rho_{23} = 0$. The coefficient of variation is set to either 0.2 or 0.4. Thus, low as well as high demand variability is captured. Moreover, instances with rather low (85%) and high (95%) retailer service-level requirements are analyzed. The internal service-level range is identical to the serial system setting.

All possible parameter combinations are tested (except for cases where $\sigma_2 = 40$ and $\sigma_3 = 20$ for symmetry reasons with instances where $\sigma_2 = 20$ and $\sigma_3 = 40$), yielding 384 instances per internal service level in total. For each instance, 10 simulation runs with 20,000 periods each are conducted.

Computational aspects

As in the serial system case, the simulation is done in ARENA v10. In case of the SS approach, optimal order-up-to levels are calculated by the algorithm described

Parameter	Description	Value
T_i	Processing time of stage $i \in \{1, 2, 3\}$ including the review period for retailers	2, 6
h_1	Holding cost at the warehouse	10
h_i	Holding cost at retailer $i \in \{2, 3\}$	15, 20
μ_i	Mean period demand at retailer $i \in \{2, 3\}$	100
CV_i	Coefficient of variation of period demand at retailer $i \in \{2, 3\}$	0.2, 0.4
α_i^{target}	α-service level at retailer $i \in \{2, 3\}$	85%, 95%
α_1^{target}	α-service level for GS at warehouse	17%, 33%, 50%, 67%, 75%

Table 4.3: Divergent system – Parameter values for simulation runs

in Diks and de Kok (1999) implemented in MAPLE v10. Numerical integration together with bisections for the echelon order-up-to levels and other bisections for the allocation fractions using the consistent appropriate share (CAS) rationing policy have been used to derive the optimal values. (See van der Heijden et al. (1997) for details on this and other rationing policies.) For the GS approach, the optimal values are found by comparing the extreme points of all possible net replenishment time combinations and their respective costs.

In general, the balance assumption (see Section 2.2.2.1) is violated, because the physical stock at the warehouse is not always sufficient to ensure an allocation of non-negative quantities to all retailers. In the simulation, potential imbalance events are handled following a suggestion by Diks (1997), p. 29. If the inventory control policy prescribes to allocate a negative quantity to one of the retailers, this quantity is adjusted to zero and the other retailer gets all available items.

Results

The simulation results are analyzed in the same way as for the serial system. With respect to the different parameter dimensions, the same (or very similar) observations can be made for the same reason mentioned in Section 4.3.3.1. That is why only the observations together with the figures for the divergent system are presented, but the explanations omitted.

Processing time. In contrast to the serial system case, where the processing-time pattern characterization is straightforward, the divergent case requires a more specific explanation. The processing-time patterns are defined as follows:

- degressive:

$$T_1 \begin{cases} \geq \max\{T_2, T_3\} & , T_2 \neq T_3 \\ > \max\{T_2, T_3\} & , T_2 = T_3 \end{cases} \tag{4.39}$$

- linear:

$$T_1 = T_2 = T_3 \tag{4.40}$$

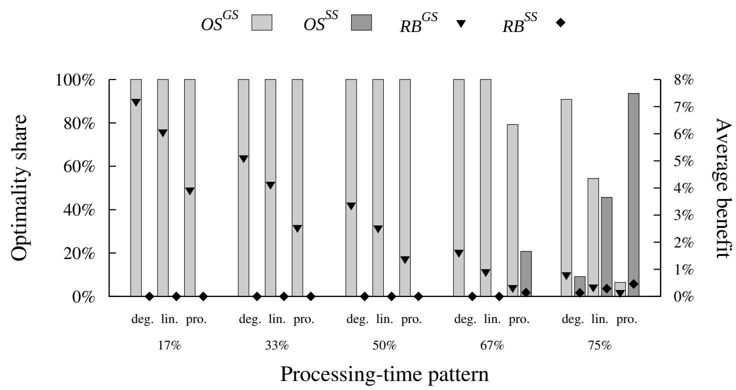
- progressive:

$$T_1 \begin{cases} \leq \min\{T_2, T_3\} & , T_2 \neq T_3 \\ < \min\{T_2, T_3\} & , T_2 = T_3 \end{cases} . \tag{4.41}$$

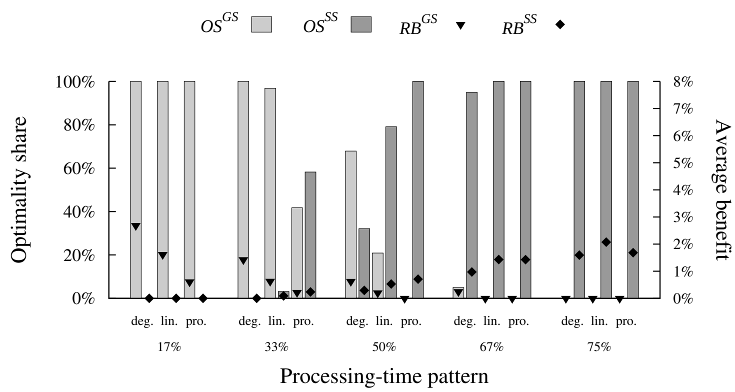
From Figures 4.8(a) and 4.8(b) the following observation can be made. *For a given internal service level, the GS optimality share as well as the average GS benefit decreases as the processing-time pattern switches from degressive over linear to progressive.*

In the situations without any additional operating flexibility costs, the first drop in the GS optimality share can be observed for an internal service level of 67% and a progressive processing-time pattern (see Figure 4.8(a)). The share decreases from 100% to 79.3%. As the internal service level is further increased to 75%, the GS optimality share decreases further to only 6.5%. The average GS benefit decreases even faster as the internal service level is raised from 17% to 75%. For the degressive, linear, and progressive processing-time pattern it drops from 7.2%, 6.1%, 3.9% to 0.8%, 0.3%, 0.1%, respectively.

In the situation with operating flexibility costs, the same tendency is apparent (see Figure 4.8(b)). However, the decrease of the GS optimality share and average benefit occurs for an internal service level of 33% already. Except for a small GS optimality share of 5% and a degressive processing-time pattern, the SS optimality share amounts to 100% for all patterns and internal service levels from 67%. Similarly, the average GS benefit drops from about 2.7% to 0% as the internal service level exceeds about 50%. The average SS benefit, on the other hand, increases from 0.1%



(a) without c^{OF}



(b) with c^{OF}

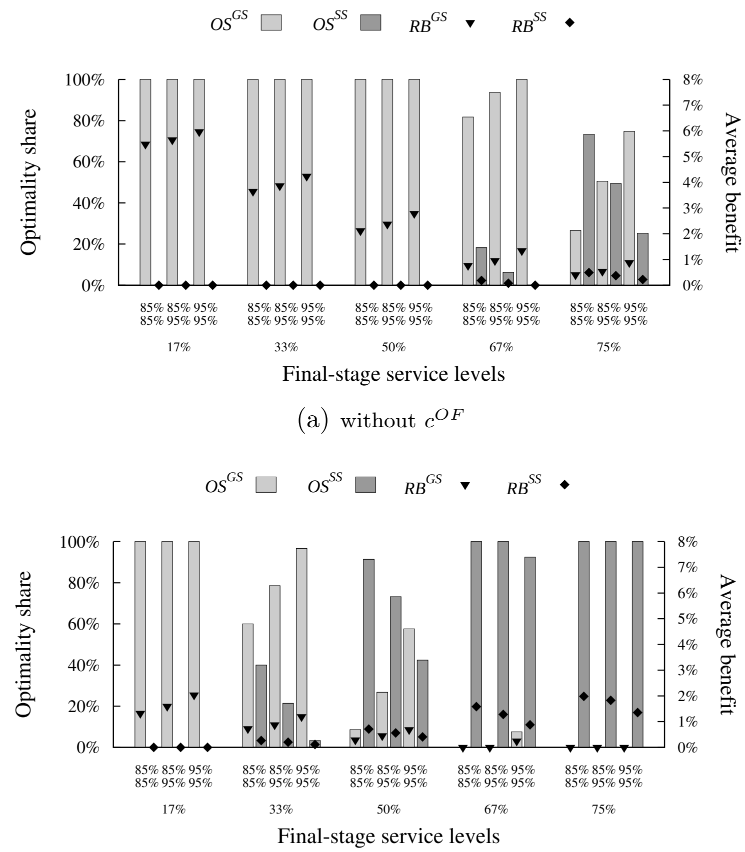
Figure 4.8: Divergent system – Optimality share and relative benefit with respect to the processing time

for an internal service level of 33% and a linear processing-time pattern to about 2% for an internal service level of 75% and a progressive pattern.

Final-stage service levels. *Higher final-stage service levels have a positive effect on the GS optimality share as well as the average relative GS benefit.* Figures 4.9(a) and 4.9(b) illustrate how the GS optimality share and average benefit increase for a given internal service level as one or both final-stage service levels are raised.

For each internal service level, the increase in the average GS benefit as the final-stage service levels are raised is rather small and amounts to about 0.5% for both situations, with and without any additional operating flexibility costs.

Whereas the GS approach shows a quite superior performance for the situation where



(a) without c^{OF}

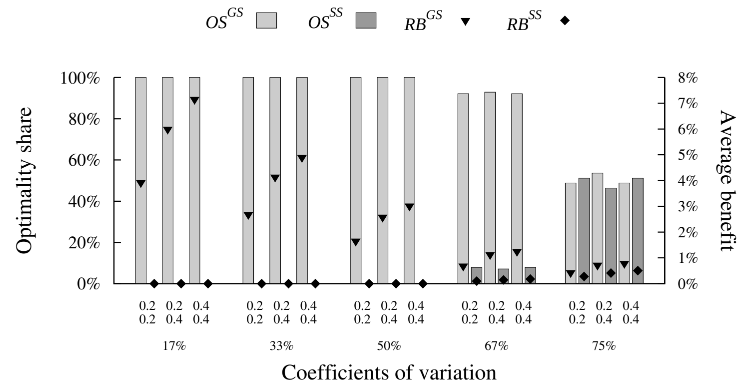
(b) with c^{OF}

Figure 4.9: Divergent system – Optimality share and relative benefit with respect to the final-stage service level

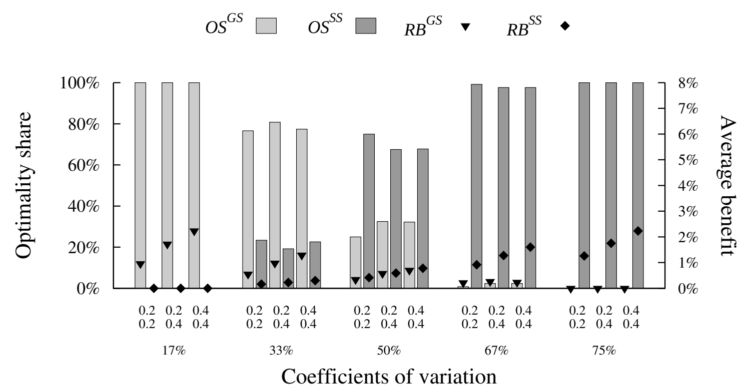
no operating flexibility costs are incurred, the SS approach is obviously dominant in the situation with operating flexibility costs and internal service levels larger than 50%. Nevertheless, an increase in the final-stage service level has a clear negative effect on the average SS benefit, which decreases from 1.6% to 0.9% and 2% to 1.4% for an internal service level of 67% and 75%, respectively.

Coefficient of variation of period demand. In the divergent system case an increase in the demand variability has only a minor effect on the optimality share of the two approaches (see Figures 4.10(a) and 4.10(b)). For a given internal service level the share remains fairly constant. The average benefit increases, however. For a low internal service level of up to 50%, this increase is larger in the GS than in the

SS case. Starting from an internal service level of 67% it is vice versa. Consequently, it can be stated that *the average benefit of the superior approach becomes larger as demand gets more variable.*



(a) without c^{OF}



(b) with c^{OF}

Figure 4.10: Divergent system – Optimality share and relative benefit with respect to the coefficient of variation of period demand

4.3.3.3 Convergent systems

In the SS approach, any convergent system can be transformed into an equivalent serial one (see Rosling (1989)). In the GS approach, this equivalence does not hold. However, the main feature and benefit of the GS approach, the decoupling, is also present in convergent systems. Therefore, it is expected that a numerical comparison would deliver results similar to the ones in the serial and divergent system case. Since the findings in both settings have not differed significantly, no separate numerical study is conducted for convergent systems.

4.3.4 Summary and implications

In this section, a comparison of two multi-echelon inventory optimization approaches, the stochastic- (SS) and guaranteed-service (GS) approach, has been provided. First, based on theoretical considerations an individual benefit of each approach has been established: the *allocation benefit* of the SS model and the *decoupling benefit* of the GS model. Next, a simulation study has been carried out in order to compare the cost performance of both approaches. In the GS framework both modeling options, with and without additional costs for using operating flexibility, have been considered.

From the numerical results, three important drivers of the advantage of one approach over the other can be identified: processing-time pattern, final-stage service level(s), and internal service level (or operating flexibility cost). The GS approach shows a superior performance for a degressive processing-time pattern, high final-stage service level(s), and a low internal service level. For the SS model, the opposite is true. Although the first two parameters have a significant effect, the superiority of one approach over the other mainly depends on the internal service-level parameter, which reflects a stage's level of flexibility.

The major finding from the comparison is that *none of the approaches is superior to the other, in general.* Both approaches have their advantages and disadvantage in certain settings. This outcome not only shows that the extension of *both* approaches to incorporate dual sourcing is valuable, it also raises another question, which is not directly related to dual sourcing at first glance: *Is it possible to combine the two approaches and thus benefit from both advantages?* Such an integrated approach would solve the dilemma of having to choose a single approach for the entire supply chain rather than for each stage individually. One can easily imagine that a single supply chain might comprise of stages with different levels of flexibility. Those with a high flexibility level would prefer the GS model, whereas the others would favor the SS model. However, also in terms of dual sourcing the development of a combined single-sourcing multi-echelon approach is of relevance. If an integration of both approaches is possible, the new approach will represent another candidate for a potential dual-sourcing extension. In the next section, such a combination of both approaches into an integrated framework is addressed.

4.4 Combination of the stochastic- and guaranteed-service approach

4.4.1 Introduction

The comparison of the SS and GS approach in Section 4.3 has shown that none of the two approaches is superior to the other, in general. Most of the multi-echelon literature as well as many commercial software solutions treat both approaches as mutually exclusive frameworks. Behind inventory optimization software tools offered by Optiant, LogicTools, SmartOps, etc., lie numerous extensions of the pioneering contributions to multi-echelon inventory research by Simpson (1958) and Clark and Scarf (1960) that form the basis of the GS and SS research strands. Consequently, the practitioner faces the dilemma of having to decide which approach is appropriate for the safety stock optimization of his entire supply chain knowing that both approaches might not fully exploit all cost-saving potentials due to the lack of a stage-wise choice.

Hence, before the focus is shifted to the integration of dual sourcing into one of the multi-echelon approaches in Section 4.5, the above-mentioned dilemma is resolved in this section by developing an integrated approach, called the hybrid-service (HS) approach. This newly developed framework represents yet another candidate for a potential dual-sourcing extension. The HS approach optimally and endogenously determines, which strategy is the best at each individual stage of the supply chain. The integration will implicitly provide the choice of the better of the two frameworks for a given system and additionally enable further cost savings by allowing for a stage-wise choice of a framework. A rough idea of a framework combination is given in Minner (2000). Whereas he only conceptionally outlines how to model the interfaces between the approaches and restricts the outline to local search methods, this section provides a detailed interface modeling description and presents a pseudo-polynomial time dynamic programming algorithm for the optimization of the integrated safety stocks in a serial system. Moreover, extensions to convergent and divergent systems are outlined. The benefit of the HS approach is tested in a numerical study for a serial system with up to five stages. From this study drivers that favor the use of hybrid-service structures are identified. The section is based on Dittmar et al. (2009).

The structure of this section is as follows. First, serial systems are addressed (Section 4.4.2). The interface modeling is described in Section 4.4.2.1. Section 4.4.2.2 presents the combination of both approaches into an integrated optimization model and the dynamic programming algorithm. Section 4.4.2.3 reports the results of a numerical comparison of the pure and hybrid approaches. In Sections 4.4.3 and 4.4.4 extensions to divergent and convergent systems are discussed. Section 4.4.5 provides a summary of the main findings.

4.4.2 Serial systems

In Section 4.3.2 the *allocation benefit* of the SS and the *decoupling benefit* of the GS approach have been established. The hybrid-service (HS) approach tries to jointly exploit the benefits of the two pure approaches. For each stage, the HS approach chooses the cost-optimal framework with regard to the entire supply chain. This leads to a partitioning of the supply chain into SS and GS subnetworks. Special care has to be taken at the interface of these subnetworks due to the differing underlying assumptions of the pure approaches. Each interface and the required adjustments for the order-up-to level calculation are addressed in turn.

4.4.2.1 Interface modeling

SS subnetwork with preceding GS subnetwork

Consider a situation where a GS subnetwork runs from stage l to $i-1$ and an SS subnetwork from i to j. Due to the operating flexibility of a GS stage, the succeeding SS subnetwork does not have to include any *stochastic* delays caused by its predecessor. However, the preceding GS stage might quote a positive service time to the SS stage, which can be viewed as a *deterministic* delay. Consequently, when the first SS stage determines its order-up-to level it has to do this with respect to the replenishment lead time $\tau_i^{SS} = ST_{i-1} + T_i$, and not just T_i. The term τ_i^{SS} is used to indicate the similarity to a GS stage through the incoming service time, ST_{i-1}. For each possible incoming service time, a different SS subnetwork needs to be solved. Note that due to the decomposition result, not all order-up-to levels need to be recalculated, but only the one of the first SS stage i. Given a certain incoming service time, the optimization algorithm of the pure SS approach can be applied.

Proposition 4.4.2.1 *Suppose an HS serial supply chain, where a GS subnetwork runs from l to $i-1$ and a SS subnetwork from i to j. Then, the net replenishment time candidate $\tau_{i-1} = 0$ can be excluded, i.e.* $1 \leq \tau_{i-1} \leq M_{i-1} \Leftrightarrow 0 \leq ST_{i-1} \leq M_{i-1} - 1$.

Proof:

See Appendix B.11. □

As already mentioned in Section 4.2.3.1, the computational complexity of GS models depends on the behavior of the objective function when the decision variables, namely the net replenishment times, are changed. A concave objective function leads to a complexity reduction due to the extreme point property.

Corollary 4.4.2.2 *If the GS objective function is a concave function of the net replenishment time, the extreme point property holds. Hence, in an optimal HS serial supply chain the final stage $i-1$ of a GS subnetwork that precedes an SS subnetwork (from stage i to j) holds sufficient stock to cover all uncovered processing times from its predecessors, i.e. $\tau_{i-1} = ST_{i-2} + T_{i-1}$. It completely decouples itself (and the upstream stages) from the downstream part of the supply chain by quoting a service time of 0 to the succeeding SS stage, i.e.* $ST_{i-1} = 0$.

Due to Corollary 4.4.2.2, there is only one possible replenishment time candidate for the first SS stage, i.e. $\tau_i^{SS} = T_i$. Consequently, only a single SS subnetwork needs to be optimized.

GS subnetwork with preceding SS subnetwork

The first GS stage might experience *stochastic* delays in the delivery of its order requests (shortfall). These delays have to be taken into account in the determination of the order-up-to level. Assume that the SS subnetwork runs from stage l to $i-1$

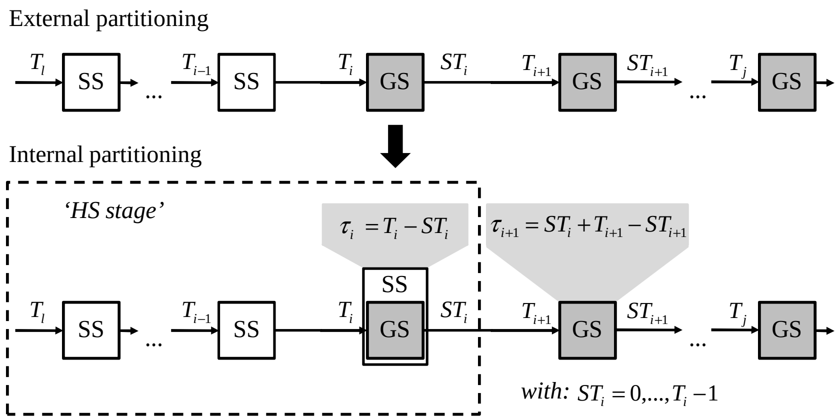


Figure 4.11: Interface modeling between an SS and a GS subnetwork

and the GS subnetwork from i to j (see Figure 4.11, 'external partitioning'). The order-up-to level of the first GS stage can be computed by treating stage i as if it was part of the preceding SS subnetwork ('internal partitioning'). Recall that in the SS subnetwork each stage has to cope with possible stochastic delays from its predecessor, too. The combination of the preceding SS subnetwork and the first GS stage is called a 'hybrid-service' (HS) stage. Consequently, the order-up-to level of GS stage i, B_i^{GS}, follows from optimizing the HS stage running from l to i. That means B_i^{GS} results from (see Section 4.2.2)

$$F_{\tau_i}\left(B_i^{GS}\right) = \frac{p + \sum_{k=l}^{i-1} h_k^e}{p + h_i} = \alpha_i^{SS} \quad \text{with} \quad p = h_i \cdot \frac{\alpha_i^{target}}{1 - \alpha_i^{target}} \quad . \tag{4.42}$$

Moreover, from the GS description in Section 4.2.3 it is known that a GS stage might quote a positive service time to its successor and thereby postpone some coverage requirement (except for the final stage). Hence, the net replenishment time of the GS stage, τ_i, that is required for the SS subnetwork optimization in (4.42) is not fixed to the processing time, T_i, a priori, but depends on the outgoing service time, ST_i. For each $\tau_i = T_i - ST_i$ a separate HS stage needs to be evaluated. The possible net replenishment times (or service times) of stage i can be reduced by the following proposition:

Proposition 4.4.2.3 *Suppose an HS serial supply chain, where an SS subnetwork runs from l to $i-1$ and a GS subnetwork from i to j. Then, the net replenishment time candidate $\tau_i = 0$ can be excluded, i.e.* $1 \leq \tau_i \leq T_i \Leftrightarrow 0 \leq ST_i \leq T_i - 1$.

Proof:
See Appendix B.12. □

Corollary 4.4.2.4 *If the GS objective function is a concave function of the net replenishment time, the extreme point property holds. Hence, in an optimal HS serial supply chain it holds for the first stage i of a GS subnetwork that succeeds an SS subnetwork (from stage l to $i-1$) that* $\tau_i = T_i \Leftrightarrow ST_i = 0$.

4.4.2.2 Combination of the approaches

Properties of HS systems

Given the interface modeling, an HS system can be transformed into a system of GS and HS stages. Figure 4.12 illustrates such an HS system where a GS subnetwork that ends at stage $i-1$ is succeeded by an SS subnetwork from i to j, which in turn is succeeded by another GS subnetwork starting at $j+1$. Through the introduction of HS stages, the HS system is similar to a pure GS system. Like a standard GS stage, the HS stage i faces an incoming service time, ST_{i-1}, and quotes an outgoing service time, ST_{j+1}. The incoming service time is quoted to the first SS stage, whereas the outgoing service time belongs to the first GS stage that directly succeeds the SS subnetwork. Once both service times of the HS stage are known, the optimal order-up-to levels and the resulting inventory holding cost of this stage can be calculated.

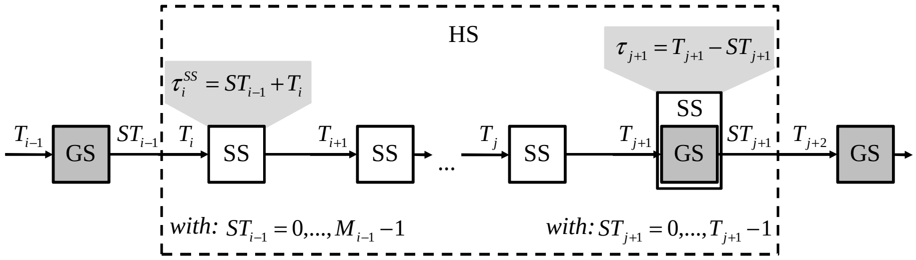


Figure 4.12: System illustration with a hybrid-service (HS) stage

Lemma 4.4.2.5 *In an HS system it holds that:*

1. *An HS stage consisting of a single stage comprises of a GS stage only.*
2. *An HS stage that runs until the most downstream stage n of the entire supply chain is a pure SS subnetwork.*
3. *Benefits of the HS approach over the pure approaches (SS, GS) start materializing in serial supply chains consisting of three stages or more.*

Proof:
See Appendix B.13. □

Optimization model

Although the overall objective is to find the optimal partitioning of the entire supply chain into SS and GS subnetworks, the optimization problem is formulated in terms of GS subnetworks and HS stages for simplicity reasons (see Figure 4.12). A back transformation is straightforward based on the explanations above. It is a two-stage optimization problem. At optimization stage 1, the objective is to find the cost-optimal partitioning of the entire serial supply chain into GS subnetworks and HS stages. The cost of each GS subnetwork/HS stage is the outcome of another optimization problem (optimization stage 2), namely the respective pure optimization problem $\mathbf{P}^{SS}$ or $\mathbf{P}^{GS}$ (see Sections 4.2.2 and 4.2.3). The GS subnetworks are indicated by superscript GS and the HS stages by superscript HS.

(Note that by only incorporating the on-hand stock cost as the total cost of a GS subnetwork it is implicitly assumed that the internal service levels are set directly by management such that no additional costs for the operating flexibility usage accrue. In the numerical study in Section 4.4.2.3 the internal service-level values will be chosen appropriately to make this a reasonable assumption.)

Stage 1: Partitioning into GS subnetworks and HS stages $\langle i,j\rangle$

$$\mathbf{P}^{HS} \quad \min\ C^{HS}_{\langle 1,n\rangle} = \sum_{\langle i,j\rangle \in \mathcal{J}} \left[c^{GS}_{\langle i,j\rangle} \cdot x^{GS}_{\langle i,j\rangle} + c^{HS}_{\langle i,j\rangle} \cdot x^{HS}_{\langle i,j\rangle}\right] \tag{4.43}$$

s.t.

$$\sum_{\langle i,j\rangle \in \mathcal{J}} a_{k,\langle i,j\rangle} \cdot \left(x^{GS}_{\langle i,j\rangle} + x^{HS}_{\langle i,j\rangle}\right) = 1 \qquad k = 1,...,n \tag{4.44}$$

$$x^{GS}_{\langle l,i-1\rangle} + x^{HS}_{\langle i,i\rangle} + x^{GS}_{\langle i,j\rangle} + x^{GS}_{\langle 1,n\rangle} + x^{HS}_{\langle 1,n\rangle} = 1 \qquad \forall \langle i,j\rangle \in \mathcal{J},\ 1 \le l < i \tag{4.45}$$

$$x^{GS}_{\langle i,j\rangle}\ ,\ x^{HS}_{\langle i,j\rangle} \in \{0,1\} \qquad \forall \langle i,j\rangle \in \mathcal{J} \tag{4.46}$$

$$a_{k,\langle i,j\rangle} \in \{0,1\} \qquad k = 1,...n,\ \forall \langle i,j\rangle \in \mathcal{J} \tag{4.47}$$

Stage 2: Determination of optimal cost for each GS subnetwork/HS stage

$$c^{type}_{\langle i,j\rangle} = \begin{cases} C^{GS}_{\langle i,j\rangle} & \text{, if } type = GS \\ C^{SS}_{\langle i,j\rangle} & \text{, if } type = HS \end{cases} \tag{4.48}$$

where

$\mathcal{J}$	$= \{\langle 1,1\rangle, ..., \langle 1,n\rangle, \langle 2,2\rangle, ..., \langle 2,n\rangle, ..., \langle n,n\rangle\}$, i.e. set of all feasible GS subnetworks/HS stages
$\langle i,j\rangle$	GS subnetwork/HS stage that runs from stage i to j (including i and j with $i \le j$)
$x^{type}_{\langle i,j\rangle}$	indicator variable that is 1 if the GS subnetwork ($type = GS$) or HS stage ($type = HS$) from i to j is chosen or 0 otherwise
$a_{k,\langle i,j\rangle}$	$k \times \langle i,j\rangle$-matrix that shows which stages are part of the GS subnetwork/HS stage $\langle i,j\rangle$
$c^{type}_{\langle i,j\rangle}$	cost of the on-hand stock in the GS subnetwork ($type = GS$) or HS stage ($type = HS$) from i to j

(4.44) ensures that stage k is part of either a GS subnetwork or an HS stage. (4.45) ensures that (i) a GS subnetwork does not succeed another GS subnetwork, (ii) an HS stage that only consists of a single stage does not succeed another GS subnetwork (Lemma 4.4.2.5 part (1)), and (iii) the entire network can be of a single type.

In order to compute the optimal cost of an HS stage in (4.48), which comprises of stages i to j, the on-hand stock calculation in optimization problem $\mathbf{P}^{SS}_{\langle i,j\rangle}$ needs to be adjusted slightly. Due to possible positive incoming and outgoing service times at the first and final stage of the HS stage, the time span that is used in the order-up-to level calculation is not equal to the processing time, but the net replenishment time of these stages. Therefore, (4.2) becomes

$$\mathbb{E}\left[OH_k(B_k)\right] = \left(B_k - \mathbb{E}\left[BO_{k-1}(B_{k-1})\right]\right) - \lambda_k \cdot \mu + \mathbb{E}\left[BO_k(B_k)\right] \qquad k = i, ..., j \tag{4.49}$$

$$\text{with} \qquad \lambda_k = \begin{cases} \tau_i^{SS} & \text{if } k = i \\ T_k & \text{if } i < k < j \\ \tau_j & \text{if } k = j \end{cases}$$

and τ_i^{SS} and τ_j as defined in Section 4.4.2.1.

Dynamic programming algorithm

The partitioning problem $\mathbf{P}^{HS}$ can be solved by dynamic programming (DP). At an HS stage, an SS optimization problem needs to be solved to obtain the optimal cost. Hence, in a pre-processing step the optimal order-up-to levels and resulting inventory holding costs of all possible HS stages as well as for all incoming and outgoing service-time combinations of these stages are calculated. Then, all possible net replenishment times at all stages can be enumerated to find the optimal ones.

State space. The state variable, z_k, represents the replenishment lead time of stage k, i.e. any uncovered processing times from preceding stages (including stage k).

$$z_k \in \mathcal{Z}_k = \{z \in \mathbb{N} \mid T_k \leq z \leq M_k\} \qquad k = 1, ..., n \tag{4.50}$$

where $M_k = \sum_{l=1}^{k} T_l$ denotes the maximum replenishment lead time for stage k.

Decision space. A three-dimensional decision variable is defined $\mathbf{u}_k = (u_k^1, u_k^2, u_k^3)$. u_k^1 indicates whether stage k is a GS stage or the first stage of an HS stage, i.e. $u_k^1 \in \{GS, HS\}$. u_k^2 indicates the next downstream GS stage that holds stock, if $u_k^1 = GS$, which implies $k \leq u_k^2 \leq n$, or the final stage of the HS stage, if $u_k^1 = HS$, which implies $k < u_k^2 \leq n$ due to Lemma 4.4.2.5 part (1). (Note that this final stage is also of type GS.) Finally, u_k^3 represents the net replenishment time of the stage u_k^2. If $u_k^1 = GS$,

$$u_k^3 \begin{cases} \in \left\{1, 2, ..., z_k + \sum_{j=k+1}^{u_k^2} T_j \right\} & , \text{if } u_k^2 < n \\ = z_k + \sum_{j=k+1}^{u_k^2} T_j & , \text{if } u_k^2 = n \end{cases} . \tag{4.51}$$

If $u_k^1 = HS$,

$$u_k^3 \begin{cases} \in \left\{1, 2, ..., T_{u_k^2} \right\} & , \text{if } u_k^2 < n \\ = T_{u_k^2} & , \text{if } u_k^2 = n \end{cases} . \tag{4.52}$$

Due to Propositions 4.4.2.1 and 4.4.2.3, $u_k^3 = 0$ can be excluded from (4.51) and (4.52), respectively, for $u_k^2 < n$. Further, since immediate customer service is assumed, the final stage n has to cover its replenishment time (including the review period). The entire decision space of stage k for a given state, $\mathcal{U}_k(z_k)$, is specified by all feasible combinations of the three decision variable elements.

State transition equation. The state transition equation denotes how the state of a succeeding stage of k depends on the state and the decision made by k. It needs to be distinguished whether stage k acts as a GS stage or starts an HS stage.

$$z_{u_k^2+1}(\mathbf{u}_k) = \begin{cases} z_k + \sum_{j=k+1}^{u_k^2+1} T_j - u_k^3 & , \text{if } u_k^1 = GS \\ \sum_{j=u_k^2}^{u_k^2+1} T_j - u_k^3 & , \text{if } u_k^1 = HS \end{cases} \tag{4.53}$$

Value function. In the value function the minimum on-hand stock cost of the current stage and the downstream part of the entire supply chain are calculated depending on the current state and the decision made at the current stage k. Note that if an HS stage starts at stage k, the entire cost of the HS stage is assigned to its final stage indicated by u_k^2. In the GS case, u_k^2 indicates the next downstream GS stage that holds stock. The direct cost assigned to stage u_k^2, $DC_{u_k^2}$, can be calculated as

$$DC_{u_k^2}(\mathbf{u}_k) = \begin{cases} h_{u_k^2} \cdot \mathbb{E}\left[OH_{u_k^2}^{GS}(\mathbf{u}_k)\right] & \text{, if } u_k^1 = GS \\ \sum_{m=k}^{u_k^2} h_m \cdot \mathbb{E}\left[OH_m^{SS}(\mathbf{u}_k)\right] & \text{, if } u_k^1 = HS \end{cases} \quad (4.54)$$

with

$$\mathbb{E}\left[OH_{u_k^2}^{GS}(\mathbf{u}_k)\right] = B_{u_k^2}(u_k^3) - u_k^3 \cdot \mu + \mathbb{E}\left[BO_{u_k^2}\left(B_{u_k^2}(u_k^3)\right)\right] \quad (4.55)$$

$$\text{where } B_{u_k^2}(u_k^3) = F_{u_k^3}^{-1}\left(\alpha_{u_k^2}^{target}\right)$$

$$\mathbb{E}\left[OH_m^{SS}(\mathbf{u}_k)\right] = \left(B_m(\mathbf{u}_k) - \mathbb{E}\left[BO_{m-1}\left(B_{m-1}(\mathbf{u}_k)\right)\right]\right) - \lambda_m \cdot \mu + \mathbb{E}\left[BO_m\left(B_m(\mathbf{u}_k)\right)\right] \quad (4.56)$$

$$\text{where} \quad \lambda_m = \begin{cases} z_k & \text{if } m = k \\ T_m & \text{if } k < m < u_k^2 \\ u_k^3 & \text{if } m = u_k^2 \end{cases}$$

$$\text{and} \quad B_m(\mathbf{u}_k) = \left\{ B_m \in \vec{B}_{\langle k,u_k^2\rangle} \;\middle|\; \vec{B}_{\langle k,u_k^2\rangle} = \underset{\vec{B}_{\langle k,u_k^2\rangle}}{\operatorname{argmin}} \sum_{i=k}^{u_k^2} h_i \cdot \mathbb{E}\left[OH_i^{SS}(B_i)\right] \quad \text{s.t.} \quad \alpha_{u_k^2}\left(\vec{B}_{\langle k,u_k^2\rangle}\right) = \alpha_{u_k^2}^{target} \right\}$$

$$\text{and} \quad BO_{k-1} = 0 \quad .$$

Given $DC_{u_k^2}$, the value function is

$$g_n(z_n) = DC_n\left((GS, n, z_n)\right) \qquad \forall z_n \in \mathcal{Z}_n \quad (4.57)$$

$$g_k(z_k) = \min_{\mathbf{u}_k \in \mathcal{U}_k(z_k)} \left\{ DC_{u_k^2}(\mathbf{u}_k) + g_{u_k^2+1}\left(z_{u_k^2+1}(\mathbf{u}_k)\right) \right\} \quad \forall z_k \in \mathcal{Z}_k \ , \ k = 1, ..., n-1. \quad (4.58)$$

Computational complexity. The computational complexity of the algorithm is $O\left(n^3 M^2 \bar{D} \log(\bar{D}) \log(S_1^*)\right)$. During the pre-evaluation phase all possible HS stages are optimized. Taking into account all possible incoming and outgoing service times (at the first and final stage, respectively), the total number of HS stages is found to be of complexity class $O(n^2 M^2)$. The optimization procedure of an HS stage corresponds to that of a pure SS subnetwork, whose complexity is $O\left(n\bar{D} \log(\bar{D}) \log(S_1^*)\right)$ (see Section 4.2.2). Hence, the total pre-evaluation complexity is $O\left(n^3 M^2 \bar{D} \log(\bar{D}) \log(S_1^*)\right)$. The dynamic program for solving the HS approach does have complexity $O(n^3 M^2)$ resulting from a stage's state space lying in $O(M)$ and its decision space lying in $O(n^2 M)$. Combining these findings results in the total DP algorithm complexity.

Remark. Most of the computation time is required for the evaluation of the HS stages due to the (in)complete convolution computation. A major reduction can be achieved, if the moment-iteration approximation is used as a heuristic (see van Houtum and Zijm (1991)).

Special case: Concave objective function

If the on-hand stock is a concave function of the net replenishment time, the state and decision space of the dynamic program can be reduced considerably:

1. **State space.** Due to the extreme point property, the state space is restricted to:

$$\mathcal{Z}_k = \left\{ \sum_{l=j}^{k} T_l \ , \ j = 1, ..., k \right\} \qquad k = 1, ..., n \tag{4.59}$$

2. **Decision space.** From Corollary 4.4.2.2 and the new state space definition it follows that if $u_k^1 = GS$,

$$u_k^3 = \sum_{j=k}^{u_k^2} T_j \quad . \tag{4.60}$$

Due to Corollary 4.4.2.4, the net replenishment time of the GS stage within an HS stage corresponds to its processing time. Consequently, if $u_k^1 = HS$,

$$u_k^3 = T_{u_k^2} \quad . \tag{4.61}$$

The restrictions on k and u_k^2 remain unchanged.

A further simplification concerns equation (4.56), which becomes

$$\mathbb{E}\left[OH_m^{SS}(\mathbf{u}_k)\right] = (B_m(\mathbf{u}_k) - \mathbb{E}\left[BO_{m-1}(B_{m-1}(\mathbf{u}_k))\right]) - T_m \cdot \mu + \mathbb{E}\left[BO_m(B_m(\mathbf{u}_k))\right] \tag{4.62}$$

i.e. no case differentiation has to be made with respect to the timespan that is relevant for the order-up-to level computation within the HS stage. At all stages it is simply the processing time, T_m, that needs to be considered. Keeping these aspects in mind, the complexity reduces to $O\left(n^3 \bar{D} \log(\bar{D}) \log(S_1^*)\right)$.

4.4.2.3 Numerical study

Numerical design

In order to gain further insights into the performance of the HS approach a numerical study for serial supply chains of length $n \in \{3, 4, 5\}$ is conducted. Parameters are chosen such that a large range of supply chain characteristics is captured. External customer demand per period is characterized by a discretized Gamma distribution with a mean of 100 and coefficients of variation (CV) of 0.2, 0.4, and 0.8 to reflect different levels of variability. Although the model and DP algorithm itself are not limited to discrete demand distributions, a discretized version of the Gamma distribution is chosen such that the mathematical expressions for the SS subnetwork optimization, which contain multiple (in)complete convolutions, can be evaluated by exact numerical computations. For serial supply chains of more than three stages, the evaluation of the multiple-integral expressions, which would result for continuous demand distributions in these systems, can no longer be computed in an exact way, except for special cases. That means, approximate methods would have to be used, which might affect the results. The choice of a discrete distribution avoids these problems. Based on the assumption that operating flexibility might vary along the supply chain, (internal) service-level targets at the intermediate stages correspond to either 75%, 85%, 95%, or 99% and may differ between the stages of a single supply chain. (Note that in the model formulation it is assumed exclusively that no additional costs for using operating flexibility are incurred, which distinguishes this setting from the comparison of the SS and GS approach in Section 4.3. That is why the internal service levels are set rather high. In the most flexible production system considered at a stage, the probability that operating flexibility is used must not exceed 25%, which corresponds to an internal service level of 75%. This parameter choice still ensures that the simplifying assumptions 4.2.3.1 and 4.2.3.2 hold in view of the different CVs.) Final-stage service levels are set to 90%, 95%, and 99%, which reflect common values in practice. Inspired by the numerical design in Graves and Willems (2005), progressive, linear, and degressive holding cost and lead-time patterns are analyzed, which are characterized in Table 4.4 and by which a large variety of structures that exist in reality can be captured. The constant $\mathcal{C}$ in

Pattern	Holding cost ($i = 1, ..., n$)	Processing time ($i = 1, ..., n$)
progressive	$h_i^e = i$	$T_i = i$
linear	$h_i^e = \mathcal{C}$	$T_i = \lceil \frac{n+1}{2} \rceil$
degressive	$h_i^e = n - i + 1$	$T_i = n - i + 1$

Table 4.4: Holding-cost and processing-time patterns

the linear cost pattern can be set arbitrarily. In the upcoming calculations $\mathcal{C} = 2$ is used. In total, 1296 ($n = 3$), 5184 ($n = 4$), 20736 ($n = 5$) instances are tested. The DP algorithm is implemented in Java.

Results

Overview. In the analysis, the same performance measures as in Section 4.3, where the two pure approaches have been compared, are used, but in a slightly extended way. The optimality share indicates the fraction of instances where an approach is optimal (i) within only the pure approaches, OS_{pure}^m, $m \in \{GS, SS\}$, and (ii) within all approaches, OS_{all}^m, $m \in \{GS, SS, HS\}$. The relative benefit of a respective approach quantifies the cost savings and is defined as

$$RB^{GS} = \left(1 - \frac{C^{GS}_{\langle 1,n \rangle}}{C^{SS}_{\langle 1,n \rangle}}\right)^+, \; RB^{SS} = \left(1 - \frac{C^{SS}_{\langle 1,n \rangle}}{C^{GS}_{\langle 1,n \rangle}}\right)^+, \; RB^{HS} = \left(1 - \frac{C^{HS}_{\langle 1,n \rangle}}{\min\left\{C^{GS}_{\langle 1,n \rangle}, C^{SS}_{\langle 1,n \rangle}\right\}}\right)^+.$$

Table 4.5 presents the measures for the entire numerical study. The optimality share of the pure approaches is almost equally-balanced, if they are the only options available, with a slight increasing trend towards the GS approach as the supply chain length increases. If hybrid structures are allowed, the HS approach improves the best pure approach in about half of the cases or more. In many instances the HS model can exploit the *allocation benefit* by introducing SS subnetworks into an originally pure GS system. In fewer cases it can profit from the *decoupling benefit* within originally pure SS systems. In terms of the relative benefit, the superiority of the GS (SS) approach over the SS (GS) approach can be quite large with 32.6% (23.4%) at most and 7.6-9.8% (3.9-4.1%) on average. Besides finding the pure optimum, the HS approach obtains additional benefits of 10.5% at most and 1.4-1.9% on average.

(a) $n = 3$

	$OS^m_{\dots}$		RB^m	
m	*pure*	*all*	Max	Avg.
GS	48.69%	10.88%	24.56%	7.58%
SS	51.31%	41.28%	17.70%	4.06%
HS		47.84%	7.07%	1.40%

(b) $n = 4$

	$OS^m_{\dots}$		RB^m	
m	*pure*	*all*	Max	Avg.
GS	51.60%	3.36%	29.19%	8.77%
SS	48.40%	32.72%	21.03%	3.98%
HS		63.93%	8.63%	1.61%

(c) $n = 5$

	$OS^m_{\dots}$		RB^m	
m	*pure*	*all*	Max	Avg.
GS	55.50%	3.44%	32.59%	9.79%
SS	44.50%	27.41%	23.43%	3.85%
HS		69.15%	10.54%	1.88%

Table 4.5: HS result overview

In the remainder of this section a more detailed analysis of each parameter dimension for $n = 3$ is conducted. For the instances with $n = 4$ and $n = 5$ the same effects are observed and therefore these results are not reported in detail. (The figures for these instances can be found in Appendix A.2.)

Internal service levels. The internal service levels have a major influence on the preferability of a particular approach. For the sake of clarity, service-level ranges are used to analyze the results. By combining the four possible internal service levels as min-max pairs, ten ranges result that cluster all test instances into disjoint subsets. Ordering by min values first (decreasing) and by maximum values second (increasing), a sorted sequence of service-level ranges is obtained as shown on the x-axis in Figure 4.13. For each range, the shaded bars refer to the left-hand y-axis and indicate the optimality shares. The dashed-lined white bars show the original shares of the pure approaches that are now outperformed by the HS. The average relative HS benefit is illustrated by triangles, which refer to the right-hand y-axis.

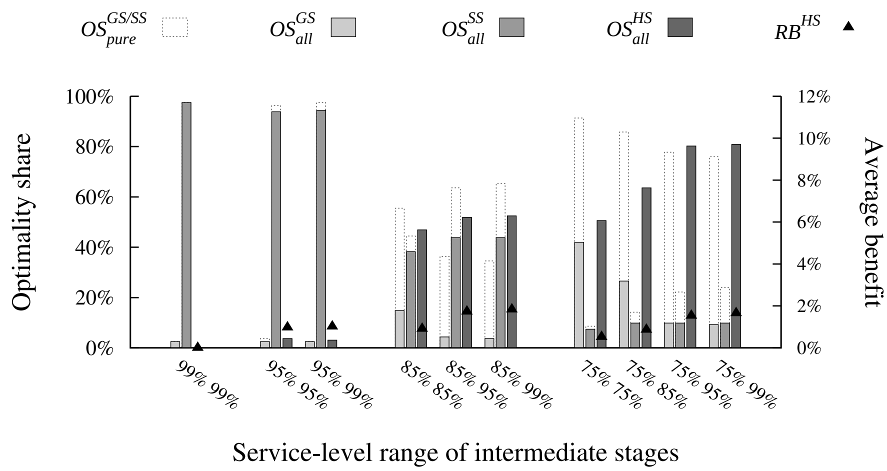


Figure 4.13: Optimality share and additional average HS benefit with respect to the internal service-level ranges

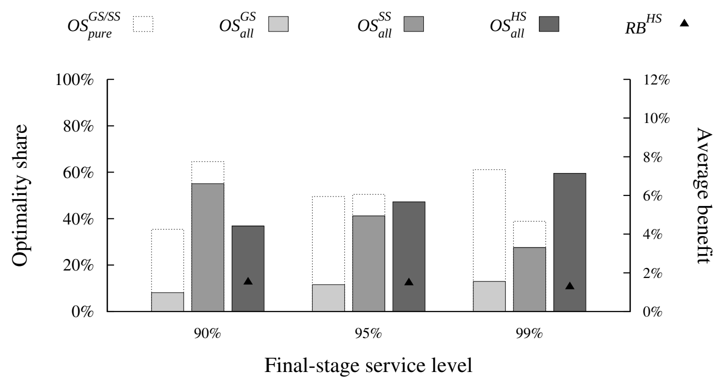
A real HS optimum becomes more likely with a low minimum internal service level and a broad internal service-level range. No HS optimality share can be observed in settings with all internal service levels set to 99% and only a small one for ranges starting at 95%. In these settings, the *decoupling benefit* is negligible. The high internal service levels indicate that there is only little operating flexibility at these stages and most of the demand variability has to be dealt with by using safety stock. Due to the *allocation benefit* pure SS solutions are dominant. In ranges starting at 85%, more operating flexibility is available and thus the extent of the *decoupling benefit* increases. Now, the HS approach can exploit both individual benefits and outperform the pure approaches. The HS superiority gets even more pronounced for service-level ranges starting at 75%.

Within all ranges starting at 95%, 85%, and 75%, respectively, the average relative HS benefit increases as illustrated by the solid black triangles. Thus, *the average relative benefit of a real HS optimum increases with the breadth of the internal service-level range.*

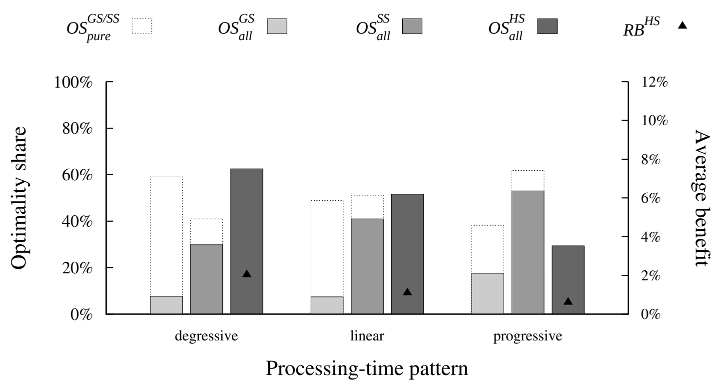
Final-stage service levels. When comparing the two pure approaches only, it can be observed that the optimality share of the SS approach decreases as the final-stage service level gets higher (see Figure 4.14(a)). An optimal SS solution has to ensure that the service level is achieved solely by safety stock taking into account potential supply shortages. This results in high costs, if the final-stage target service level is high. In contrast, the GS approach can exploit its *decoupling benefit* and carry more stock at the upstream stages. Although the HS optimality share increases, the relative HS benefit remains fairly constant. *A higher final-stage service level increases the HS optimality share, but has a negligible effect on the average relative HS benefit.*

Processing time. From degressive over linear to progressive, the share of HS optima decreases, the share of GS optima remains fairly constant (it decreases within the pure approach comparison, though), and the share of SS optima increases (see Figure 4.14(b)). The GS and HS approach can use their *decoupling benefit* only in case of upstream coverage. Holding stock at upstream stages becomes more advantageous with larger processing times in the upstream part of the supply chain compared to the downstream part due to the square root effect of the processing time. Similarly to the optimality share, the average relative HS benefit also decreases for the same reason. *Both the optimality share and the average relative benefit of the HS approach decrease from degressive over linear to progressive processing-time patterns.*

Holding cost. *The optimality share as well as the average relative benefit of the HS approach increase from degressive over linear to progressive holding-cost patterns.* (Figure 4.15(a).) Generally speaking, if value-adding is high, it is preferrable to hold more stock upstream. If operating flexibility of the upstream stages is high, the GS approach can do this at a lower cost than the SS model. In case of low operating flexibility, the SS approach's *allocation benefit* prevails. This causes the relatively



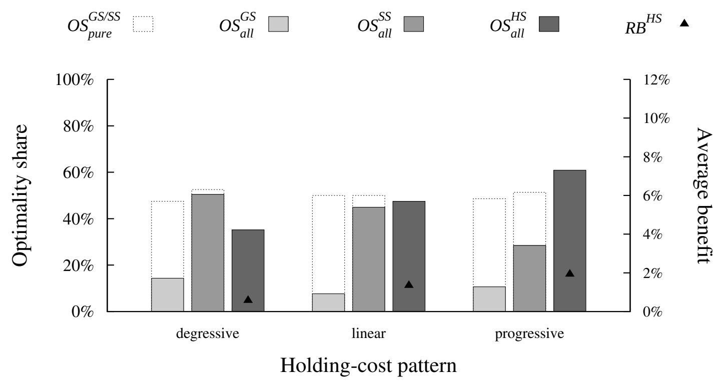
(a) Final-stage service level



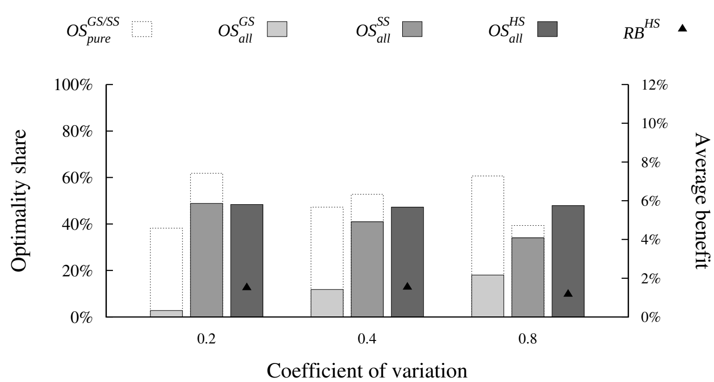
(b) Processing-time pattern

Figure 4.14: Optimality share and additional average HS benefit with respect to (a) and (b)

equally-balanced performance of both pure approaches across the different holding-cost patterns. Due to the joint exploitation of both individual benefits, the HS approach produces additional benefits. In particular, for a progressive holding-cost structure it is advantageous to have a GS subnetwork in the upstream part of the supply chain making use of the *decoupling benefit* and an SS one in the downstream part, where the *allocation benefit* can be realized.



(a) Holding-cost pattern



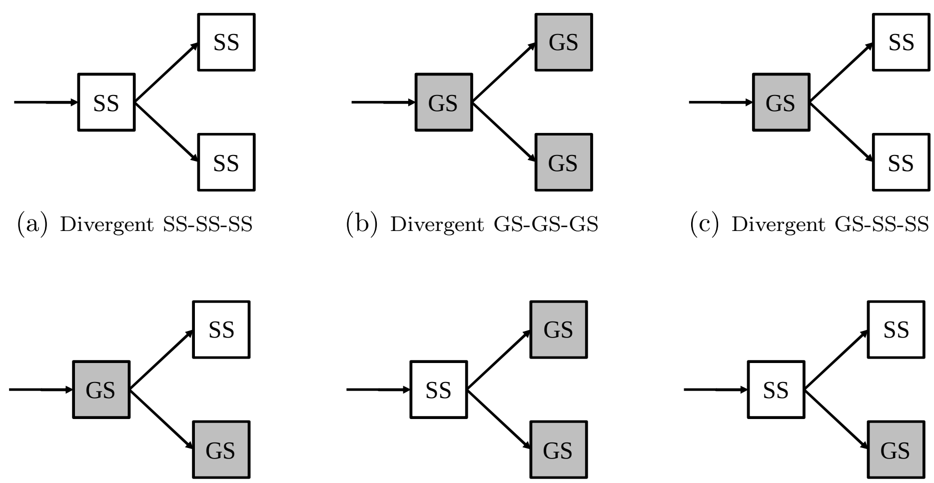
(b) Coefficient of variation of period demand

Figure 4.15: Optimality share and additional average HS benefit with respect to (a) and (b)

Coefficient of variation of period demand. When the coefficient of variation (CV) grows, the value of operating flexibility (if used) increases. Whereas the *decoupling benefit* of the GS approach becomes more important, the SS approach suffers from the CV-increase. Accordingly, the HS approach cannot make use of the *allocation benefit* to a larger extent. *The coefficient of variation has no significant* *effect on the HS optimality share and average relative benefit.* (Figure 4.15(b).)

4.4.3 Divergent systems

The HS approach directly extends to divergent systems. However, the computational complexity increases due to the larger number of HS stages that need to be evaluated, because a stage can have multiple successors now. Moreover, additional computational effort is required for the optimization of the order-up-to levels within an HS stage due to material rationing at a preceding SS stage. Not only optimal order-up-to levels, but also optimal allocation functions need to be determined (see, e.g., Diks and de Kok (1998)). In the GS subnetworks, due to the internal service of 100%, rationing problems do not exist. If it is further assumed that the service time, which a preceding GS stage quotes to its successors, is identical for all successors (which is quite common, see, e.g., Graves and Willems (2000)), the complexity of the dynamic program for pure divergent GS systems is the same as for serial ones.



(a) Divergent SS-SS-SS (b) Divergent GS-GS-GS (c) Divergent GS-SS-SS

(d) Divergent GS-SS-GS (e) Divergent SS-GS-GS (f) Divergent SS-SS-GS

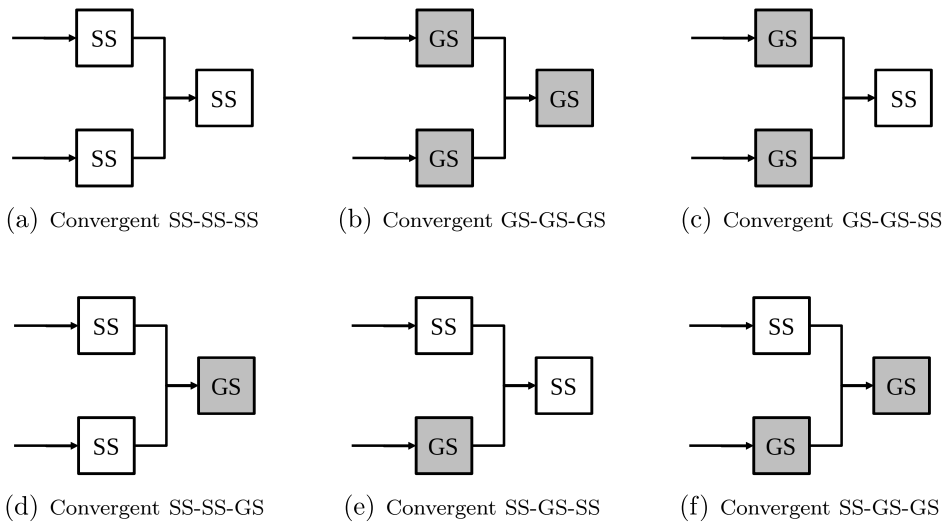
Figure 4.16: Divergent HS systems

In order to describe the interface modeling, the simplest version of an arborescent system is considered and displayed in Figure 4.16. Cases 4.16(a) and 4.16(b) are straightforward since they are pure systems. In Cases 4.16(c) and 4.16(d) the SS stages need to take into account a potential deterministic delay of the GS predecessor through a positive service time. The SS stages' order-up-to level calculation follows the same logic as in the serial system case. In Cases 4.16(e) and 4.16(f), the SS predecessor causes stochastic delays (shortfall). Hence, for the order-up-to level determination of the succeeding GS stages, these stages are simply added to the preceding SS subnetwork and form one HS stage, which is optimized with respect to different outgoing service times at the GS stages.

4.4.4 Convergent systems

In the SS approach, Rosling (1989) and Langenhoff and Zijm (1990) show that any convergent system can be transformed into an equivalent serial one. Then, optimal order-up-to levels are found by using the standard method for serial systems. In the GS approach, since all parts for an assembly need to be available before the process can start, the incoming service time of the common successor is the maximum of the predecessors' outgoing service times. This does not affect the complexity of the DP algorithm of Graves and Willems (2000), however, which makes pure convergent GS systems still easy to solve. The presence of multiple predecessors and thus enlarged number of possible SS systems/HS stages increases the computational effort required for optimally combining the two approaches. In addition, the HS stages need to be transformed into the corresponding serial ones first, before they can be evaluated.

The interface modeling is described for the simplest convergent systems illustrated in Figure 4.17. In the pure Cases 4.17(a) and 4.17(b) the standard solution methods can be applied. In Case 4.17(c) the order-up-to level of the SS successor needs to be optimized for all feasible incoming service times, which can be quoted by the GS stages, as in the serial and divergent case. Case 4.17(d) is solved by merging the GS stage and the preceding SS stages into a single HS stage. This HS stage is optimized according to Rosling (1989) with respect to all feasible outgoing service times that the comprised GS stage can quote. Cases 4.17(e) and 4.17(f) are more difficult to handle, because the succeeding stage faces a stochastic delay (shortfall) as well as a potential deterministic delay by the SS and GS predecessor, respectively. Roughly speaking, the joint shortfall (caused by these delays) needs to be taken into account at the succeeding stage, i.e. the distribution of a maximum expression of two random variables is to be computed, which makes the computation more complex.



(a) Convergent SS-SS-SS (b) Convergent GS-GS-GS (c) Convergent GS-GS-SS

(d) Convergent SS-SS-GS (e) Convergent SS-GS-SS (f) Convergent SS-GS-GS

Figure 4.17: Convergent HS systems

4.4.5 Summary and implications

In this section an approach has been presented to combine the stochastic-service (SS) and guaranteed-service (GS) framework. The integrated hybrid-service (HS) approach derives a cost-optimal partitioning of the supply chain into SS and GS subnetworks and calculates the optimal order-up-to levels. By the stage-wise choice of an appropriate approach instead of only a single decision for the entire supply chain, the HS approach not only resolves the practitioner's dilemma to find the better of the two pure approaches, but even realizes additional gains.

From a numerical study for serial systems with up to five stages it has been found that the cost superiority of the pure GS (SS) over the SS (GS) solution can be quite large amounting to 32.6% (23.4%) at maximum. The HS approach not only mitigates the risk of choosing the 'wrong' pure approach, but it has even achieved further cost-savings of up to 10.5% at most and 1.9% on average in the analyzed experimental design. The numerical results have shown that the largest additional HS benefits accrue in settings with relatively low internal service levels, a broad internal service-level range, degressive processing-time structure, and progressive holding-cost pattern.

Now that both multi-echelon modeling approaches have been outlined (Section 4.2), compared (Section 4.3), and even combined (in this section), the focus can be shifted to the integration of dual sourcing in a multi-echelon context. This aspect is dealt with in the next section.

4.5 Guaranteed-service approach with dual sourcing

4.5.1 Introduction

While the determination of optimal inventory levels in multi-echelon models already represents a challenging task, the incorporation of dual sourcing into such settings increases the complexity even further. That is why only few contributions can be found in the literature that address this aspect (see Section 2.2.3).

In order to keep the complexity still manageable even under dual sourcing, out of the three (single-sourcing) multi-echelon approaches that have been described and developed in the previous sections, the guaranteed-service (GS) approach is chosen as the basic framework for the dual-sourcing extension in this section for the following reasons. In contrast to the stochastic-service (SS) approach, where stochastic supply delays need to be taken into account in the order-up-to level determination of a stage, the GS approach only has to cope with deterministic delays through positive incoming service times. This facilitates the dual-sourcing analysis, which requires the coordination of orders in the presence of two replenishment lead times of different length. The hybrid-service (HS) approach, which one could argue is the most advanced approach, is also postponed to future research, because it not only suffers from the SS difficulty, but also from an already increased complexity in the single-sourcing setting resulting from the integration of the two pure multi-echelon approaches. Moreover, the GS approach has been shown to be (easily) extendable in various ways. Over the last two decades, it has been extended to incorporate, e.g., *stochastic processing times* (Minner (2000)) and *differing integer review periods* (Bossert and Willems (2007)) at the stages, *non-stationary demand* (Graves and Willems (2008)), *evolving forecasts* (Schoenmeyr and Graves (2009b)), and *capacity constraints* (Schoenmeyr and Graves (2009a)). Dual or multiple sourcing has not been addressed in this approach yet, except for a brief remark in the final section of Graves and Willems (2005). In contrast to the model developed in this section, their idea does not provide an accurate estimate of the safety stock, however. The difference between the two modeling ideas is outlined and an example for illustrating the cost difference is presented. In its *standard* model framework, the GS approach does not specify in detail how the operating flexibility measures work to provide the guaranteed service (recall Section 4.2.3.1). For tactical or strategic decision guidance, which most of the previously mentioned contributions focus on, this is of minor importance, anyway. Such a tactical/strategic perspective is also taken in this section.

Similar to the simplification regarding the underlying multi-echelon framework, not the most effective (and also most complex) dual-sourcing policy of the four policies investigated in Section 3.3.3 is chosen to be integrated into a multi-echelon model (i.e. the dual-index policy (DIP)), but a simple and thus analytically better manageable one is selected for the start, i.e. the order-splitting policy (OSP). Whereas the DIP (and also the single-index (SIP) and constant-order policy (COP)) determine the allocation of the demand to the different suppliers only indirectly via their policy parameters, the OSP does this directly as one of the policy parameters. This direct specification allows for an exact (and relatively simple) derivation of the demand process at the supplying (upstream) stages in the supply network in contrast to the other dual-sourcing policies. Moreover, the demand allocation to the suppliers can either be part of the optimization or treated as exogenous to the model. The latter is assumed here, i.e. the allocation is determined by company regulations, minimum production quantities for facilities, and/or supply chain security considerations, for instance. The primary focus is put on the optimization of safety stocks *given* a certain allocation.

Thus, the objective of the model developed in this section differs from the few existing dual-sourcing multi-echelon models (see Section 2.2.3) mainly in two respects:

1. This model does not view one of the suppliers as the regular (slow) one and the other one as an emergency option. Rather, it is assumed that a certain share of the demand shall be allocated to each supplier in every period.

2. The focus lies on larger supply chain settings than the previously studied ones. This also suggests the use of the GS approach, which has been shown to be applicable to supply networks of large sizes (see, e.g., Willems (2008)).

The section is based on Klosterhalfen et al. (2010b). The remainder of this section is organized as follows. In Section 4.5.2 dual sourcing is introduced in the GS approach for serial and convergent systems. In Section 4.5.2.1 the basic assumptions, notations, and changes to the standard GS approach are explained. Section 4.5.2.2 developes the single-echelon model, on which the multi-echelon one in Section 4.5.2.3 builds. A numerical example demonstrating the benefit of the model extension is presented in Section 4.5.2.4. Section 4.5.3 discusses the extension to other network structures. Section 4.5.4 deals with further extensions of this approach, i.e. more than two suppliers, the optimization of the sourcing fractions, and other inventory control policies. A summary is provided in Section 4.5.5.

4.5.2 Serial and convergent systems

In this section, the GS framework is modified to allow for two suppliers for a single item. Since a serial system represents a special case of a convergent system, the upcoming exposition focuses on convergent systems only. Other network structures are discussed in Section 4.5.3.

4.5.2.1 Assumptions, notations, and changes to the standard guaranteed-service approach

Most of the notations and assumptions referring to the GS model have already been explained in Section 4.2. Therefore, only some additional notations, which are due to the analysis of convergent systems instead of the previously studied serial ones, are briefly introduced at the end of this section. The focus lies on the changes to the standard GS framework that are required to accommodate dual sourcing. These changes include the following two aspects: (i) the replenishment policy has to allow for sourcing from more than one supplier and (ii) processing times have to be assigned to the arcs connecting two potential inventory locations (stockpoints) instead of the stages, which has been the case in the previous sections. These two aspects are addressed in more detail in turn.

Periodic-review order-up-to level replenishment policy with order splitting

In line with the standard GS approach it is assumed that the replenishment policy at each stockpoint follows a standard order-up-to level policy (with a common review period) where each period location i places an order that raises its inventory position to the desired order-up-to level, B_i. However, under the dual-sourcing regime, not the entire demand is ordered from a single supplier j, but only a fraction of $\delta_{j,i}$ is placed with each supplier. It is assumed that this fraction is exogenously determined and not part of the optimization model, which is developed in this section. Since the order-up-to level policy under steady-state conditions essentially replenishes current period demand at a stockpoint i, $d_{i,t}$, the replenishment orders are $Q_{(j,i),t} = \delta_{j,i} \cdot d_{i,t}$. Thus, the replenishment policy is basically an order-splitting policy (see Section 3.3.3.4).

In a multi-echelon model the replenishment order process is of particular importance, because it represents the demand process of the next upstream stockpoint(s). Under the assumed replenishment policy this process can be derived exactly and easily. In each period each supplier receives a certain fraction of the current period demand. This enables an exact inventory optimization at all stockpoints in the supply network. Under many other replenishment policies (including the ones analyzed in the single-echelon Chapter 3), the upstream demand process is more difficult and computationally much more challenging to derive exactly (see Section 4.5.4.3).

Processing times

As explained in Section 2.1.3, the supply network is modeled as a sequence of processes, graphically represented by circles, and potential stockpoints after each process visualized by triangles. Figure 4.18 illustrates a supply network consisting of two 'make' processes, that take T_1 and T_2 time periods, and two transportation processes, 'Ship 1,3' and 'Ship 2,3', which last $T_{1,3}$ and $T_{2,3}$ time periods with $T_{1,3} \neq T_{2,3}$.

In all previous sections the processing time, i.e. the time from when all of the inputs are available until production is completed and available to serve demand, has been assigned to a stockpoint (and its index) and both form a stage. Since each item has been sourced from a single supplier so far, there is only a single process preceding

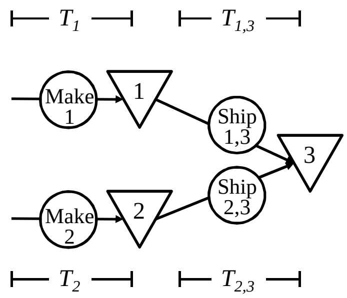
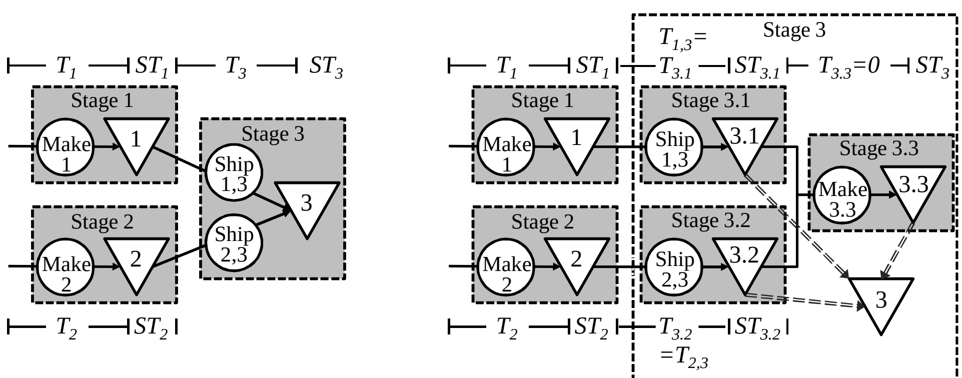


Figure 4.18: Dual-sourcing supply network represented by processes and stockpoints

a stockpoint. Hence, the respective processing time can be associated directly with this stockpoint and its index (see, e.g., Graves and Willems (2000)). In a dual- or multiple-sourcing setting, an item can be delivered by several suppliers. Depending on the geographic distance, the transportation processes might take up different amounts of time. That means, a stockpoint can have several preceding processes depending on the number of suppliers. Consequently, the aggregation into stages with only a single processing time is no longer possible, as is illustrated by stage 3 in Figure 4.19(a).



(a) With assignment to stages

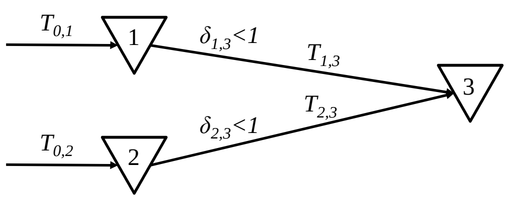
(b) Workaround in the standard GS approach

Figure 4.19: Supply network – Stage assignment and current workaround

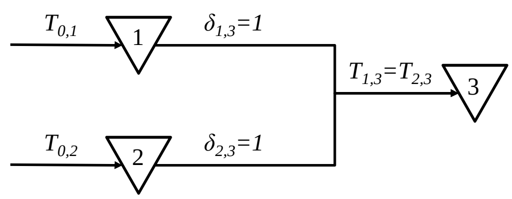
For such settings, one workaround in the standard GS model, which is outlined at the end of Graves and Willems (2005) and is currently the only one available in the literature, is to introduce additional 'dummy' stages that preserve the different processing times and also the property that each stage has only a single processing time (see Figure 4.19(b)). Stage 3 is basically replaced by three substages 3.1, 3.2, and 3.3. The first two stages are associated with the transportation processes and thus processing times $T_{3.1} = T_{1,3}$ and $T_{3.2} = T_{2,3}$. Stage 3.3 represents an assembly stage with a processing time of $T_{3.3} = 0$ since no actual process takes place there.

The clear drawback of this remodeling is that, through the breakdown of stage 3 into three substages, the possible inventory pooling at stockpoint 3 is not captured correctly. In fact, the remodeled system is only able to hold either two separate safety stocks after the two shipment processes, i.e. at stockpoints 3.1 and 3.2, or a common safety stock at stockpoint 3.3. Since the 'dummy' process preceding stockpoint 3.3 represents an assembly step, this common safety stock would have to buffer against the longer of the two transportation times, however (as is known from assembly systems with single sourcing). By accurately accounting for both processing times in the safety stock determination at stockpoint 3, a *lower* common safety stock could be held, because the risk interval could be reduced below the longer transportation time.

In order to be able to derive an exact computation of the safety stock quantity at stockpoint 3, the aggregation of the processes and stockpoints into stages is omitted. Instead of the stage, the *processing time is assigned to the arc* connecting two adjacent stockpoints. For stockpoints j and i, the respective processing time is denoted as $T_{j,i}$. Thus, the supply network of Figure 4.18 becomes the one depicted in Figure 4.20(a). (Note that index 0 denotes an external supplier.)



(a) Dual-sourcing network



(b) Assembly network

Figure 4.20: Dual-sourcing network vs. assembly network

The distinction between a dual-sourcing situation compared to an assembly situation is achieved by the use of different arc types. Whereas Figure 4.20(a) refers to the dual-sourcing case with $T_{1,3} \neq T_{2,3}$, Figure 4.20(b) illustrates an assembly situation. Stockpoints 1 and 2 provide different items for the assembly process, which takes $T_{1,3} = T_{2,3}$ periods. Hence, for the safety stock determination at stockpoint 3 either of the two processing times could be used. Apart from the arc types, the sourcing fractions $\delta_{j,i}$ are an indicator for dual or single sourcing, i.e. $\delta_{i,j} \in (0,1) \Rightarrow$ dual sourcing and $\delta_{i,j} = 1 \Rightarrow$ single sourcing.

In the model, it is further assumed that the processing time is not impacted by the size of the order, i.e. there are no capacity constraints on the production or transportation processes between stockpoints. Also, it is assumed that at a dual-sourced stockpoint no assembly process can take place in addition. Such an operational step would be modeled as a separate process after the dual-sourced stockpoint.

Additional assumptions and notations

The stockpoint facing external customer demand is denoted as n. Let $\mathcal{A}$ be the arc set for the network representation of the supply network and $(i,j) \in \mathcal{A}$ denote the arc between stockpoints i and j. In addition to the numbering of the stockpoints from 1 to n, a level code (LC) is assigned to each stockpoint, which is needed for the dynamic programming formulation later on. The demand stockpoint n receives level code 1, i.e. $LC(n) := 1$. All other stockpoints i have $LC(i) := 1 + LC(j)$ with $(i,j) \in \mathcal{A}$. Let N denote the highest level.

For ease of presentation, it is assumed that an item at a downstream stockpoint requires exactly one item of all the upstream stockpoints that are connected to it, i.e. the production coefficients $a_{i,j} = 1, \forall (i,j) \in \mathcal{A}$. Note that the relaxation of this assumption is not difficult to include into the model. It would simply make the presentation more complicated and is therefore omitted.

Further, it is assumed that external demand only occurs at stockpoint n. Due to the replenishment policy, the demand at an internal stockpoint i is given as

$$d_{i,t} = \delta_{i,j} \cdot d_{j,t} \qquad (i,j) \in \mathcal{A} \quad . \tag{4.63}$$

In case of single sourcing, $\delta_{i,j} = 1$. Demand per period is assumed to be *i.i.d.* with stationary mean μ_n and standard deviation σ_n. The coefficient of variation is denoted as $CV_n = \sigma_n / \mu_n$.

4.5.2.2 Single-echelon model

Inventory model

In this section, the single-echelon model is described, on which the multi-echelon one is built. Consider a supply network setting where stockpoint i is dual sourcing from j and k (see Figure 4.21).

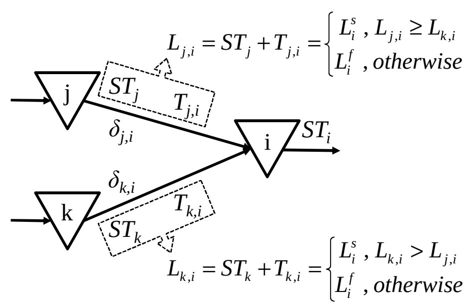


Figure 4.21: Supply network explaining notations

Given the GS assumptions, the items of stockpoint j with $(j, i) \in \mathcal{A}$ are available for processing after ST_j periods. Then, it takes another $T_{j,i}$ periods until the items are in stock at i. As a consequence, each replenishment placed with supplier j is received after a replenishment lead time $L_{j,i} = ST_j + T_{j,i}$. The same holds for k, respectively. Let the stockpoint index of the supplier of i that causes the longer (shorter) replenishment lead time be denoted as $\overline{pre}$ ($\underline{pre}$), i.e.

$$\overline{pre} = \underset{\forall l:(l,i)\in\mathcal{A}}{\operatorname{argmax}}\{ST_l + T_{l,i}\} \tag{4.64}$$

$$\underline{pre} = \underset{\forall l:(l,i)\in\mathcal{A}}{\operatorname{argmin}}\{ST_l + T_{l,i}\} \quad . \tag{4.65}$$

Note that at a dual-sourced stockpoint the maximization and minimization in (4.64) and (4.65) only needs to be done over two preceding stockpoints. The above characterization is more general, however, so that it can be used in the formulation of the multi-echelon optimization model in Section 4.5.2.3 without any modification. For an assembly process in the convergent network, multiple components (i.e. more than two), each delivered by a different stockpoint, might be required. In the optimization model formulation (as will be seen later), the largest outgoing service time that is quoted by any of the supplying stockpoints needs to be identified. Since the processing time in (4.64) is identical for all l, because it refers to the same assembly process, the stockpoint with the largest service time and the time itself are easily found by (4.64).

Given $\overline{pre}$ and $\underline{pre}$, the following parameters and variables can be defined for stockpoint i, where the superscript s denotes those of the 'slower' of the two suppliers and f the ones of the 'faster':

$$ST_i^s = ST_{\overline{pre}} \qquad ST_i^f = ST_{\underline{pre}} \tag{4.66}$$

$$T_i^s = T_{\overline{pre},i} \qquad T_i^f = T_{\underline{pre},i} \tag{4.67}$$

$$L_i^s = ST_i^s + T_i^s \qquad L_i^f = ST_i^f + T_i^f \tag{4.68}$$

$$\delta_i^s = \delta_{\overline{pre},i} \qquad \delta_i^f = \delta_{\underline{pre},i} = 1 - \delta_i^s \tag{4.69}$$

$$Q_{i,t}^s = \delta_i^s \cdot d_{i,t} \qquad Q_{i,t}^f = \delta_i^f \cdot d_{i,t} \quad . \tag{4.70}$$

Stockpoint i itself quotes a service time ST_i to its successor. In the calculation of the net stock (or inventory level) at the end of a period two cases need to be distinguished: (For ease of presentation the stockpoint index is dropped.)

1. The outgoing service time ST is shorter than (or equal to) both replenishment lead times, i.e. $L^s \geq L^f \geq ST$.
2. The outgoing service time ST is larger than (or equal to) the fast replenishment lead time, but shorter than (or equal to) the slow replenishment lead time, i.e. $L^s \geq ST \geq L^f$.

Note that in case $ST \geq L^s \geq L^f$, the stockpoint does not have to hold any stock at all, since the downstream stockpoint is willing to wait longer than the larger of the two replenishment lead times. If $ST > L^s$, the next upstream stockpoints could delay their deliveries even further by $ST - L^s$. In the other two above-mentioned cases, the stockpoint needs to hold inventory and the net stock calculation is as follows. W.l.o.g. it is assumed that the net stock at the end of period 0 is equal to the base-stock level, i.e. $NS_0 = B$. Let superscript m where $m \in \{1, 2\}$ indicate the specific case, which is considered.

Case 1: $L^s \geq L^f \geq ST$**.** The net stock at the end of period t can be calculated similarly to (3.142). Additionally, only the outgoing service time, ST, needs to be taken into account.

$$
\begin{aligned}
NS_t^1 &= B^1 \underbrace{- \sum_{i=0}^{t} d_{t-i}}_{(a)} \underbrace{+ \sum_{i=L^s}^{t} Q^s_{t-i}}_{(b)} \underbrace{+ \sum_{i=L^f}^{t} Q^f_{t-i}}_{(c)} \underbrace{+ \sum_{i=0}^{ST-1} d_{t-i}}_{(d)} \\
&= B^1 - \sum_{i=ST}^{t} d_{t-i} + \delta^s \cdot \sum_{i=L^s}^{t} d_{t-i} + \delta^f \cdot \sum_{i=L^f}^{t} d_{t-i} \\
&= B^1 - \sum_{i=ST}^{L^f-1} d_{t-i} - \delta^s \cdot \sum_{i=L^f}^{L^s-1} d_{t-i} \\
&= B^1 - \sum_{i=ST}^{ST^f+T^f-1} d_{t-i} - \delta^s \cdot \sum_{i=ST^f+T^f}^{ST^s+T^s-1} d_{t-i} \quad .
\end{aligned} \tag{4.71}
$$

In period t, all outstanding orders that have been placed with the slow supplier up to period $t-L^s$ (including the order of period $t-L^s$) have arrived at the stockpoint, i.e. (b) in (4.71). Similar, all outstanding orders that have been placed with the fast supplier up to period $t-L^f$ (including the order of period $t-L^f$) have arrived, i.e. (c) in (4.71). Moreover, demands have occurred in all periods up to t (including the demand in t). If the outgoing service time ST was zero, all of these demands would have depleted the stock level, i.e. (a) in (4.71). However, in case of a positive outgoing service time, the fulfillment of the demands of the most recent ST periods can be delayed, i.e. (d) in (4.71). That means, all demands that have occurred up to period $t-ST$ (including the demand in $t-ST$) have been filled by the stockpoint. Figure 4.22 illustrates the timeline for the calculation.

Consequently, the inventory shortfall is given as (cf. (4.71))

$$
\sum_{i=ST}^{ST^f+T^f-1} d_{t-i} + \delta^s \cdot \sum_{i=ST^f+T^f}^{ST^s+T^s-1} d_{t-i} \tag{4.72}
$$

and under stationary conditions $t \to \infty$ it can be represented by the following random variable

$$
\check{D}^1(ST^f, ST^s, ST, \delta^s) = D(ST^f + T^f - ST) + \delta^s D(ST^s + T^s - ST^f - T^f) \quad . \tag{4.73}
$$

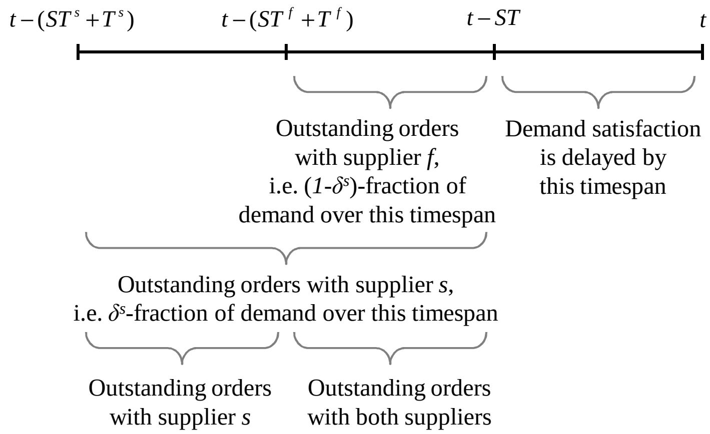


Figure 4.22: Case 1 – Timeline for net stock calculation

Instead of the replenishment lead times L^s and L^f, it is expressed in terms of the service times, which are the decision variables in the GS model. Then, the net stock results as

$$NS^1 = B^1 - \check{D}^1(ST^f, ST^s, ST, \delta^s) \tag{4.74}$$

with

$$\mathbb{E}\left[\check{D}^1(ST^f, ST^s, ST, \delta^s)\right] = \Big((ST^f + T^f - ST) + \delta^s(ST^s + T^s - ST^f - T^f)\Big)\,\mu \tag{4.75}$$

$$\mathrm{VAR}\left[\check{D}^1(ST^f, ST^s, ST, \delta^s)\right] = \Big((ST^f + T^f - ST) + [\delta^s]^2(ST^s + T^s - ST^f - T^f)\Big)\,\sigma^2. \tag{4.76}$$

Case 2: $L^s \geq ST \geq L^f$. In this case, the fraction of demand of period t that is sourced from the fast supplier $(1-\delta^s)\cdot d_t$ arrives after L^f periods and thus before (or just at the point in time) the demand of period d_t actually needs to be filled, which is after ST periods, because $ST \geq L^f$. Consequently, these items are put on stock. Figure 4.23 illustrates the timeline in this case.

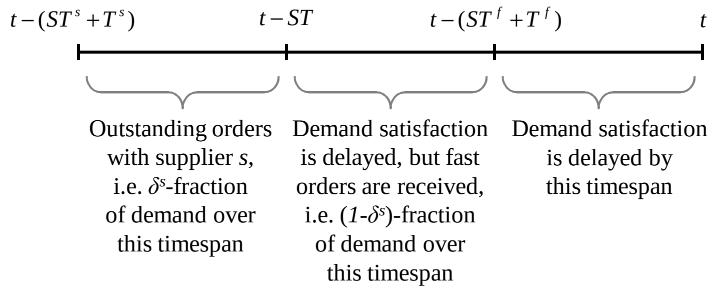


Figure 4.23: Case 2 – Timeline for net stock calculation

The net stock computation is similar to (4.71) and reads as follows:

$$\begin{aligned} NS_t^2 &= B^2 - \sum_{i=0}^{t} d_{t-i} + \sum_{i=L^s}^{t} Q^s_{t-i} + \sum_{i=L^f}^{t} Q^f_{t-i} + \sum_{i=0}^{ST-1} d_{t-i} \\ &= B^2 - \sum_{i=ST}^{t} d_{t-i} + \delta^s \cdot \sum_{i=L^s}^{t} d_{t-i} + \delta^f \cdot \sum_{i=L^f}^{t} d_{t-i} \\ &= B^2 - \delta^s \cdot \sum_{i=ST}^{L^s-1} d_{t-i} + \delta^f \cdot \sum_{i=L^f}^{ST-1} d_{t-i} \\ &= B^2 - \delta^s \cdot \sum_{i=ST}^{ST^s+T^s-1} d_{t-i} + (1-\delta^s) \cdot \sum_{i=ST^f+T^f}^{ST-1} d_{t-i} \quad . \end{aligned} \tag{4.77}$$

By defining $\check{D}^2$ as

$$\check{D}^2(ST^f, ST^s, ST, \delta^s) = \delta^s D(ST^s + T^s - ST) - (1-\delta^s) D(ST - ST^f - T^f) \tag{4.78}$$

the net stock under stationary conditions $t \to \infty$ follows from (4.77) as

$$NS^2 = B^2 - \check{D}^2(ST^f, ST^s, ST, \delta^s) \tag{4.79}$$

with

$$\mathbb{E}\left[\check{D}^2(ST^f, ST^s, ST, \delta^s)\right] = \\ \left(\delta^s(ST^s + T^s - ST) - (1-\delta^s)(ST - ST^f - T^f)\right)\mu \tag{4.80}$$

$$\text{VAR}\left[\check{D}^2(ST^f, ST^s, ST, \delta^s)\right] =$$
$$\left([\delta^s]^2(ST^s + T^s - ST) + (1-\delta^s)^2(ST - ST^f - T^f)\right)\sigma^2. \quad (4.81)$$

Based on the net stock calculations, the determination of the safety stock and order-up-to level for a specific service-level target can be addressed next.

Determination of safety stock and order-up-to level

For a given set of incoming and outgoing service times $\check{D}^m(ST^f, ST^s, ST, \delta^s)$ where $m \in \{1, 2\}$ is completely defined (keeping in mind that δ^s is exogenous to the model and thus given, too). As in the single-echelon case, the optimal order-up-to level for a given α-service level and a (discrete) continuous demand distribution is the (smallest) one, for which the following (in)equality holds

$$Pr\left\{\check{D}^m(ST^f, ST^s, ST, \delta^s) \le B^m\right\} \ge \alpha^{target} \qquad m \in \{1, 2\} \quad . \quad (4.82)$$

Denote this optimal value as $B^m(ST^f, ST^s, ST, \delta^s, \alpha^{target})$. Then, the safety stock is found as

$$SST^m(ST^f, ST^s, ST, \delta^s, \alpha^{target}) =$$
$$B^m(ST^f, ST^s, ST, \delta^s, \alpha^{target}) - \mathbb{E}\left[\check{D}^m(ST^f, ST^s, ST, \delta^s)\right] \quad m \in \{1, 2\}. \quad (4.83)$$

Special case: Normally-distributed demand. If period demand is assumed to be normally *i.i.d.* distributed, $\check{D}^m(ST^f, ST^s, ST, \delta^s)$ where $m \in \{1, 2\}$ also follows a normal distribution. Hence, the safety stock in **case 1**, which is required to achieve an α-service level, is given as

$$SST^1_{norm}(ST^f, ST^s, ST, \delta^s, \alpha^{target}) =$$
$$k(\alpha^{target})\sigma\sqrt{(ST^f + T^f - ST) + [\delta^s]^2\left(ST^s + T^s - ST^f - T^f\right)} \quad (4.84)$$

where k denotes the safety factor, which depends an the service-level target (cf. (3.153) and relation (3.35)). In **case 2**, the safety stock can be determined as

$$SST^2_{norm}(ST^f, ST^s, ST, \delta^s, \alpha^{target}) = \\ k(\alpha^{target})\sigma\sqrt{[\delta^s]^2\left(ST^s + T^s - ST\right) + (1-\delta^s)^2\left(ST - ST^f - T^f\right)}. \quad (4.85)$$

In both cases $m \in \{1, 2\}$, the optimal order-up-to level at a stockpoint is given as

$$B^m_{norm}(ST^f, ST^s, ST, \delta^s, \alpha^{target}) = \\ \mathbb{E}\left[\check{D}^m(ST^f, ST^s, ST, \delta^s)\right] + SST^m_{norm}(ST^f, ST^s, ST, \delta^s, \alpha^{target}). \quad (4.86)$$

Average order quantities and pipeline inventory

Due to the model assumptions, the average order quantities of stockpoint i with its suppliers are given as

$$\mathbb{E}\left[Q_{j,i}(\delta_{j,i})\right] = \delta_{j,i} \cdot \mu_i \qquad \forall j : (j,i) \in \mathcal{A} \quad . \quad (4.87)$$

The average pipeline or work-in-process inventory between supplier j and stockpoint i is given as the average quantity ordered from stockpoint j times the processing time from j to i, i.e.

$$\mathbb{E}\left[PI_{j,i}(\delta_{j,i})\right] = \delta_{j,i} \cdot \mu_i \cdot T_{j,i} \qquad \forall j : (j,i) \in \mathcal{A} \quad . \quad (4.88)$$

Since these two quantities are not influenced by the service times (decision variables), only by the predefined sourcing fraction, they do not need to be included in the optimization model.

Holding cost

Most of the models dealing with dual sourcing assume a holding cost per unit and period, which does not depend on the sourcing fraction. In Chapter 3 this assumption has also been made when the holding cost has been modeled as the foregone interest on the procurement cost of the cheap supplier (according to, e.g., Scheller-Wolf et al. (2007)). Since the inventory at the end of a period usually consists of items from both suppliers, this simplification is only reasonable in situations where the fraction of items sourced from the cheap supplier is close to 1. Otherwise, the holding cost is underestimated.

Under the assumed replenishment policy of the multi-echelon setting, the above assumption might be violated severely. That is why a more precise computation of the item holding cost is required, which takes the stock composition into account. Here, it is assumed that in the long run the stock at the end of a period is composed of exactly the same proportions of items, which are sourced from the two suppliers. There are two explanations to support this assumption. First, if the dispatch rule is such that, once the items are in stock, they are randomly picked to satisfy demand (i.e. there is no possibility of distinguishing between them), the share of items from supplier j in stock corresponds to $\delta_{j,i}$, i.e. the fraction of demand, which is sourced from j. Second, if a fraction of $\delta_{j,i}$ is sourced from each supplier, a reduction in (safety) stock by one unit corresponds to a reduction in tied-up capital by $\delta_{j,i}$ and $1-\delta_{j,i}$ times the cumulative cost and holding cost rate.

Let $c_{j,i}^{add}$ denote the cost added to an item when proceeding from stockpoint j to i and c_i^{cum} represent the cumulative cost of an item at stockpoint i. Then, the cumulative cost at a dual-sourced stockpoint i is given as

$$c_i^{cum} = \delta_{j,i}\left(c_j^{cum} + c_{j,i}^{add}\right) + (1-\delta_{j,i})\left(c_k^{cum} + c_{k,i}^{add}\right) \qquad j \neq k \text{ and } (j,i),(k,i) \in \mathcal{A} \tag{4.89}$$

and the holding cost per unit and period can be calculated as

$$h_i = \nu c_i^{cum} = \nu\left(\delta_{j,i}\left(c_j^{cum} + c_{j,i}^{add}\right) + (1-\delta_{j,i})\left(c_k^{cum} + c_{k,i}^{add}\right)\right) \quad \text{if } \delta_{j,i} \in (0,1) \tag{4.90}$$

where ν denotes the holding-cost/interest rate for the underlying base period, e.g., one day or one week. Since the demand at stockpoint i is also filled according to the δ-ratios, on average, the cumulative cost c_i^{cum} can simply be used in the cost calculation at the next downstream stockpoint.

If stockpoint i receives each of the required items for its (assembly) process from a single source, the holding cost is given as

$$h_i = \nu \sum_{\forall j:(j,i)\in\mathcal{A}} \left(c_j^{cum} + c_{j,i}^{add}\right) \qquad \text{if } \delta_{j,i} = 1 \quad . \tag{4.91}$$

4.5.2.3 Multi-echelon model

Optimization model

In this section, the optimization problem for finding the optimal service times in the network is formulated. For **case** $m \in \{1, 2\}$ using (4.83), the optimization problem can be stated as

$$\mathbf{P}^m \quad \min \sum_{i=1}^{n} h_i \left(B_i^m(ST_i^f, ST_i^s, ST_i, \delta_i^s, \alpha_i^{target}) - \mathbb{E}\left[\check{D}_i^m(ST_i^f, ST_i^s, ST_i, \delta_i^s) \right] \right) \tag{4.92}$$

s.t.

$$ST_i^s + T_i^s \geq \begin{cases} \left(ST_i^f + T_i^f\right) \cdot \mathbf{I}\{\delta_i^s \in (0,1)\} + ST_i \cdot \mathbf{I}\{\delta_i^s = 1\} \geq ST_i & m = 1 \\ ST_i \geq \left(ST_i^f + T_i^f\right) \cdot \mathbf{I}\{\delta_i^s \in (0,1)\} + ST_i \cdot \mathbf{I}\{\delta_i^s = 1\} & m = 2 \end{cases}$$

$$i = 1, 2, ..., n \tag{4.93}$$

$$ST_i \in \mathbb{N} \qquad i = 1, 2, ..., n \tag{4.94}$$

$$ST_0 = 0 \tag{4.95}$$

$$ST_n = 0 \tag{4.96}$$

where $\mathbf{I}\{x\}$ is the indicator function of the event x. For ease of presentation, the definition of the parameters and variables referring to the slow and fast supplier, which are given in (4.64)-(4.69), are not repeated. The objective of $\mathbf{P}^m$ is to minimize the inventory holding cost in the supply network. The constraints ensure that the service times are non-negative and integer (4.94), the external supplier delivers the items immediately (4.95), and the demand stockpoint satisfies the external customer demand immediately (4.96), which is commonly assumed. Constraint (4.93) ensures the relation between the replenishment lead times and the outgoing service time of stockpoint i, which is required by case $m \in \{1, 2\}$, if i is a dual-sourced stockpoint, or the relation between a single replenishment lead time and the outgoing service time of stockpoint i, if it is a single-sourced stockpoint.

Note that if there is only a single supplier for an item, all sourcing fractions between stockpoint i and its suppliers l are equal to 1, i.e. $\delta_{l,i} = 1$, $\forall l : (l, i) \in \mathcal{A}$, and thus

$\delta_i^s = 1$. Consequently, only ST_i^s influences the inventory holding cost at stockpoint i, which is obvious from the definitions of $\check{D}^m$, $m \in \{1, 2\}$ (cf. (4.73) and (4.78)). That means, a single incoming service time needs to be taken into account at stockpoint i, which corresponds to the maximum of the outgoing service times of all suppliers (cf. (4.64) and (4.66)). If this service time is denoted as SI_i instead of ST_i^s, (4.73) and (4.78) collapse to the demand random variable over the net replenishment time, $D(SI_i{+}T_i{-}ST_i)$, which is well-known from the single-sourcing model (cf. Graves and Willems (2000)), because the processing times between the supplying stockpoints and the receiving one are identical as they refer to the same process, i.e. $T_{l,i} = T_i$, $\forall l : (l, i) \in \mathcal{A}$ and thus $T_i^s = T_i^f = T_i$.

Lemma 4.5.2.1 *The objective function of* $\mathbf{P}^2$ *is decreasing in* ST_i^f *for given* ST_i^s *and* ST_i.

Proof:
See Appendix B.14. □

Due to Lemma 4.5.2.1 the optimal ST_i^f value under case 2 will always be the largest feasible, i.e. $ST_i^f = ST_i - T_i^f$, which comes from constraint (4.93). The cost function of case 2 then reduces to the one of case 1, which is also defined for $ST_i = ST_i^f + T_i^f$. Hence, the whole optimization problem can be expressed by using the problem formulation of case 1 only and it is sufficient to develop a single optimization algorithm.

Optimization procedure

In this section, it is shown how to solve $\mathbf{P}^1$ by dynamic programming. The dynamic program uses a forward recursion similar to the one presented in Minner (1997) for convergent systems with single sourcing, i.e. it proceeds from level N down to 1 and finds the solution to $\mathbf{P}^1$ for all stockpoints at the same level by evaluating a functional equation. Consequently, when proceeding to the next lower level the inventory holding costs for all possible outgoing service times of all supplying stockpoints have already been determined. For the specification of the functional equations, $\mathcal{N}_i$ is defined as the subset of stockpoints that are connected to i on the subgraph with stockpoints that have a higher level as i. $\mathcal{N}_i$ is determined by the following equation:

$$\mathcal{N}_i = \{i\} + \bigcup_{\forall j:(j,i)\in\mathcal{A}} \mathcal{N}_j \quad . \tag{4.97}$$

Functional equations. Two cost functions are defined, C_i^{SI} and C_i^{DS}, one for a single-sourced stockpoint and one for a dual-sourced stockpoint. The respective function calculates the minimum inventory holding cost for the subnetwork with stockpoint set $\mathcal{N}_i$.

1. Single-sourced stockpoint

   If each item is delivered by a different supplier, i.e. $\delta_{j,i} = 1, \forall j : (j,i) \in \mathcal{A}$ $\Rightarrow \delta_i^s = 1$, the minimum cost is a function of the incoming service time, $SI_i = ST_i^s = \max_{\forall j:(j,i)\in\mathcal{A}} \{ST_j\}$, and the outgoing service, ST_i (as specified in Graves and Willems (2000)). ST_i^f is irrelevant in the single-sourcing case, because it cancels out in $\check{D}_i^1$ (cf. (4.73)) and thus does not need to be further specified. Note that the function $w_j(.)$ is characterized after the introduction of the cost function for a dual-sourced stockpoint, since it is used in both formulations.

$$\begin{aligned} C_i^{SI}(SI_i, ST_i) = \\ & h_i \left(B_i^1(ST_i^f, ST_i^s, ST_i, \delta_i^s, \alpha_i^{target}) - \mathbb{E}\left[\check{D}_i^1(ST_i^f, ST_i^s, ST_i, \delta_i^s)\right]\right) \\ & + \sum_{\forall j:(j,i)\in\mathcal{A}} w_j(\min\{SI_i, M_j\}) \end{aligned} \tag{4.98}$$

$$\text{where} \qquad ST_i^s = SI_i = \max_{\forall j:(j,i)\in\mathcal{A}} \{ST_j\} \tag{4.99}$$

$$T_i^s = T_i^f = T_{j,i} \qquad \text{for any } j:(j,i)\in\mathcal{A} \tag{4.100}$$

M_i is the maximum replenishment lead time for stockpoint i, which is defined as

$$M_i = \max_{\forall j:(j,i)\in\mathcal{A}} \{M_j + T_{j,i}\} \tag{4.101}$$

with $M_0 = 0$, i.e. the maximum replenishment lead time of the external supplier(s) is zero by assumption.

2. Dual-sourced stockpoint

   At a stockpoint where the same item is sourced from two different suppliers j and k, i.e. $\delta_{j,i} \in (0,1)$ and $\delta_{k,i} = 1 - \delta_{j,i}$, the minimum cost is a function of both incoming service times, ST_j and ST_k, and the outgoing service time, ST_i.

$$\begin{aligned} C_i^{DS}(ST_j, ST_k, ST_i) = \\ & h_i\left(B_i^1(ST_i^f, ST_i^s, ST_i, \delta_i^s, \alpha_i^{target}) - \mathbb{E}\left[\check{D}_i^1(ST_i^f, ST_i^s, ST_i, \delta_i^s)\right]\right) \\ & + w_j(\min\{ST_j, M_j\}) + w_k(\min\{ST_k, M_k\}) \end{aligned} \tag{4.102}$$

   Note that $\delta_{j,i}$, ST_j, and ST_k are related to δ_i^s, ST_i^s, and ST_i^f as specified in (4.64)-(4.69).

In each cost function, the first term represents the inventory holding cost of stockpoint i resulting from the incoming and outgoing service times. The remaining term addresses the stockpoints in $\mathcal{N}_i$ that are upstream of i. If stockpoint i has a single supplier for each required item, for each stockpoint j that supplies stockpoint i with a different item the minimum inventory holding cost of the subnetwork with stockpoint set $\mathcal{N}_j$ is included as a function of the stockpoint's outgoing service time SI_i. If this outgoing service time is larger than the maximum replenishment lead time of the stockpoint, it is adjusted accordingly. This means that stockpoint i delays its orders from stockpoint j by $SI_i - M_j$ periods in order to avoid unnecessary inventory. If stockpoint i has two suppliers, j and k, the minimum inventory holding cost for the subnetworks with stockpoint sets $\mathcal{N}_j$ and $\mathcal{N}_k$ are included as a function of the stockpoints' outgoing service times ST_j and ST_k, respectively. Similarly, the service times are adjusted where necessary.

The following optimization is solved by enumeration to find the functional value $w_i(ST_i)$:

$$w_i(ST_i) = \min_{SI_i} \left\{ C_i^{SI}(SI_i, ST_i) \right\} \cdot \mathbf{I}\left\{\delta_{j,i} = 1\right\} + \min_{(ST_j, ST_k)} \left\{ C_i^{DS}(ST_j, ST_k, ST_i) \right\} \cdot \mathbf{I}\left\{\delta_{j,i} \in (0,1)\right\} \quad \text{for } (j,i) \text{ and/or } (k,i) \in \mathcal{A} \tag{4.103}$$

$$\text{s.t.} \quad \max\left\{0, ST_i - T_{j,i}\right\} \leq SI_i \leq M_i - T_{j,i} \quad \text{and } SI_i \text{ integer, for } (j,i) \in \mathcal{A} \tag{4.104}$$

$$\max\left\{0, ST_i - T_{j,i}\right\} \leq ST_j \leq M_i - T_{j,i} \quad \text{and } ST_j \text{ integer, for } (j,i) \in \mathcal{A} \tag{4.105}$$

$$\max\left\{0, ST_i - T_{k,i}\right\} \leq ST_k \leq M_i - T_{k,i} \quad \text{and } ST_k \text{ integer, for } (k,i) \in \mathcal{A} \tag{4.106}$$

The lower bound on ST_j, ST_k, and SI_i comes from $\mathbf{P}^1$, while the definition of M_i provides the upper bound.

Dynamic programming algorithm. The algorithm is as follows:

1. *For all stockpoints i with $LC(i) := N$ down to 1, evaluate $w_i(ST_i)$ for $ST_i = 0, 1, ..., M_i$.*
2. *For $i := n$ evaluate $w_i(ST_i)$ for $ST_i = 0$ (assuming immediate demand satisfaction).*
3. *Minimize $w_n(ST_n)$ for $ST_n = 0$ to obtain the optimal objective function value.*

An optimal set of service times is found by the standard backtracking procedure for a dynamic program.

Computational complexity. As one can observe from (4.103), in the GS approach with two suppliers it is not sufficient to only consider one incoming and outgoing service time at a time. Rather, at each dual-sourced stockpoint, the stockpoint's outgoing service time and the incoming service times of all suppliers need to be evaluated together in one step. That means, at each dual-sourced stockpoint, M^2 incoming service-time combinations have to be evaluated together with all feasible outgoing service times, M. At each stockpoint with only a single supplier for an item, M incoming service times need to be considered. Hence, the complexity of the algorithm is $P^{SI}M^2 + P^{DS}M^3$, where P^{SI} denotes the number of stockpoints with single sourcing, P^{DS} the number of stockpoints with dual sourcing and M the maximum replenishment lead time, which is also the maximum outgoing service time.

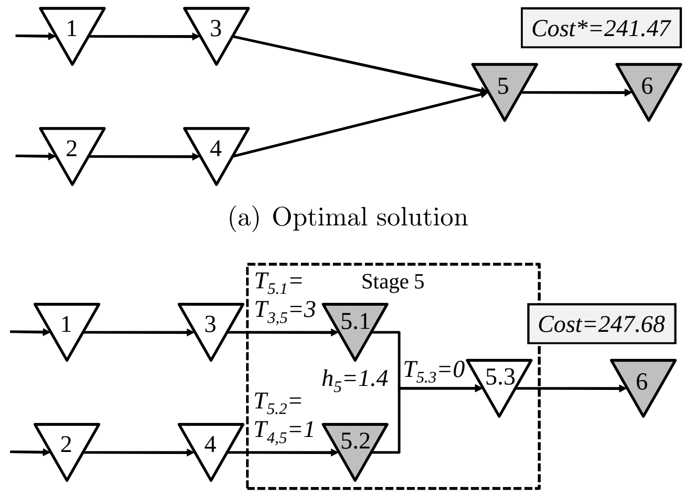
4.5.2.4 Numerical example

Even though the superiority of the above-described approach over the GS modeling with dummy stages (see Section 4.5.2.1) is quite obvious, which currently represents the only way to include dual sourcing in the GS model that is reported in the literature, the benefit is illustrated by a small numerical example in this section. Consider the sample supply network depicted in Figure 4.24(a) with the parameters given in Table 4.6. The system consists of 6 stockpoints. Stockpoint 5 has two suppliers for the same item. The sourcing fraction is given with $\delta_{3,5} = 0.7$. The processing times between all stockpoints are 1 except for the processing time between stockpoints 3 and 5, which is 3. For ease of computation, a holding-cost rate of $\nu = 1 = 100\%$ is assumed, i.e. the cumulative cost per unit at a stockpoint corresponds to its holding cost. Period demand is normally distributed with a mean of 100 and standard deviation 30.

		Stockpoint						
	i	0	0	1	2	3	4	5
Parameter	j	1	2	3	4	5	5	6
$c_{i,j}^{add}$		0.6	1.2	0.4	0.8	0.1	0.1	0.7
$T_{i,j}$		1	1	1	1	3	1	1
$\delta_{i,j}$		1	1	1	1	0.7	0.3	1
α_j^{target}		95%	95%	95%	95%	95%	95%	95%

Table 4.6: GS approach with dual sourcing – Parameter values

The optimal stock allocation pattern prescribes safety stock at stockpoints 5 and 6 (see Figure 4.24(a)) and results in a cost of 241.47. The remodeled supply network using dummy stages is illustrated in Figure 4.24(b). Stockpoint 5 is replaced by stockpoints 5.1, 5.2, and 5.3 with processing times $T_{5.1} = T_{3,5} = 3$, $T_{5.2} = T_{4,5} = 1$, and $T_{5.3} = 0$, respectively. All stockpoints have a holding cost of $h_{5.1} = h_{5.2} = h_{5.3} =$



(a) Optimal solution

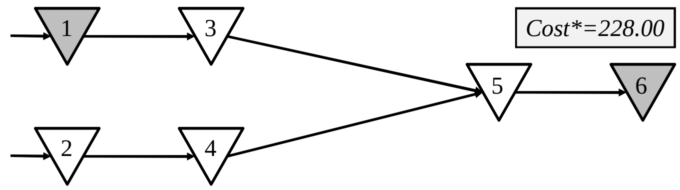
(b) Remodeled solution

Figure 4.24: Supply network with dual sourcing – Example 1

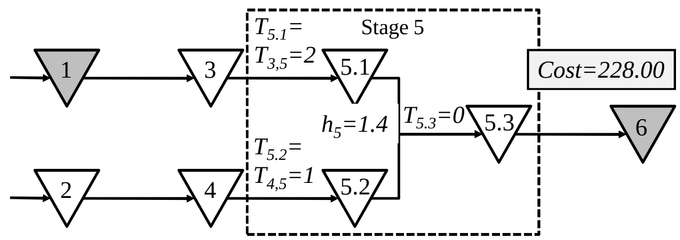
$h_5 = 1.4$. Here, the total cost amounts to 247.68, a 2.57% increase. Safety stock is held at the dummy stockpoints 5.1 and 5.2, which in fact means that two separate safety stocks are held at stockpoint 5. In addition, stockpoint 6 as final stockpoint holds stock.

The benefit of the new modeling approach cannot be realized in all settings, but only in those where the optimal allocation pattern prescribes safety stock at a dual-sourced stockpoint. At these stockpoints the remodeling with dummy stages is only approximate. In settings where in an optimal solution safety stocks are not held at a dual-sourced stockpoint, both approaches deliver the same result. As an example, the following situation is considered. If the processing time between stockpoint 3 and 5 is reduced to 2 periods, safety stocks are held at stockpoint 1 and at the final stockpoint in both approaches (see Figure 4.25). Consequently, the total safety stock cost is identical.

Nevertheless, it remains that if safety stock is prescribed at a dual-sourced stockpoint, the new approach computes the safety stock quantity exactly, whereas the other one does this only approximately.



(a) Optimal solution



(b) Remodeled solution

Figure 4.25: Supply network with dual sourcing – Example 2

4.5.3 Other network structures

The extension of the above-described approach to divergent network structures, spanning trees, or even more general networks results in an increase in the computational complexity. To illustrate this, consider the supply network setting in Figure 4.26. Stockpoints 4 and 5 are both dual-sourced stockpoints with one common supplying stockpoint, namely stockpoint 2. The solution algorithm for convergent systems would proceed as follows:

1. Evaluate the cost of stockpoints 1, 2, and 3 for all feasible outgoing service times.

2. Evaluate the cost of stockpoint 4 and the connected upstream subset of stockpoints for all incoming service-time combinations (ST_1, ST_2) and an outgoing service time of 0 (final stockpoint). Choose the cost-minimal set of service times as the optimal one.

3. Evaluate the cost of stockpoint 5 and the connected upstream subset of stockpoints for all incoming service-time combinations (ST_2, ST_3) and an outgoing service time of 0 (final stockpoint). Choose the cost-minimal set of service times as the optimal one.

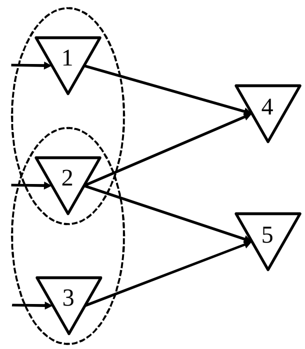


Figure 4.26: Sample network with dual sourcing and divergent substructure

In step 2 and 3, a problem arises. The service time ST_2 influences the cost calculation and therefore the decision of both stockpoints 4 and 5. Hence, a sequential solution procedure does not guarantee optimality. The service time ST_2 that is optimal for the cost minimization with regard to stockpoint 4 might not correspond to the optimal ST_2-value in the cost minimization with regard to stockpoint 5. Therefore, the service times of all five stockpoints have to be considered simultaneously in order to derive the overall optimal set of service times.

Generally speaking, the outgoing service times of all stockpoints, which are situated on adjacent levels of the supply network and connected with each other through some path, need to be considered simultaneously. In Figure 4.26, stockpoints 1 and 3 are connected through the path 1-4-2-5-3. Consequently, the outgoing service times of all five stockpoints need to be considered simultaneously. This makes the solution of this problem computationally much more complex. Nevertheless, a solution might still be obtained for moderate network sizes by a slightly modified version of the above-described approach. For larger systems, other solution methods or heuristics need to be developed.

The problem of simultaneously evaluating multiple service times is not specific to the dual-sourcing situation only. It also occurs in general (acyclic) network structures with single sourcing. Consider the above-mentioned setting, if the processes between the stockpoints are assemblies (see Figure 4.27). In this situation, too, the outgoing service time of stockpoint 2 influences the safety stock computation at stockpoints 4 and 5. Therefore, these stockpoints need to be considered together in the optimization of their incoming service times $SI_4 = \max\{ST_1, ST_2\}$ and

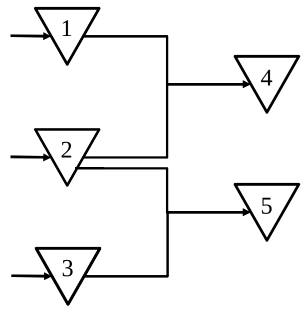


Figure 4.27: Sample network with single sourcing and general structure

$SI_5 = \max\{ST_2, ST_3\}$ with respect to an outgoing service time of $ST_4 = ST_5 = 0$. For networks with so-called clusters of commonality, which is a special case of general networks that contain commonality only between adjacent levels, Humair and Willems (2006) note that all incoming and outgoing service times in such a cluster are coupled and therefore the minimization needs to be done over all feasible combinations of these times. Nevertheless, the authors show that the safety stocks can still be optimized with a dynamic programming algorithm. However, for general networks beyond the clusters of commonality type, the combinatorial complexity induces Humair and Willems (2010) to develop a branch and bound algorithm as a solution method.

4.5.4 Extensions

4.5.4.1 Multiple sourcing

The dual-sourcing model for convergent systems can be easily extended to an arbitrary number of suppliers. Consider a supply network setting where stockpoint i has K suppliers for the same item (see Figure 4.28). As in the dual-sourcing case, the replenishment lead time from supplier k is $L_{k,i} = ST_k + T_{k,i}$.

The net stock at the end of period t can be calculated as follows. First, arrange the suppliers in increasing order according to their replenishment lead time, $L_{k,i}$. Denote the shortest lead time as L_i^1, the next one L_i^2, and so on until L_i^K. Moreover, denote the corresponding sourcing fractions as $\delta_i^1, \delta_i^2, ..., \delta_i^K$. In case two or more lead times are identical, represent them by a single L_i^k, sum up the sourcing fractions,

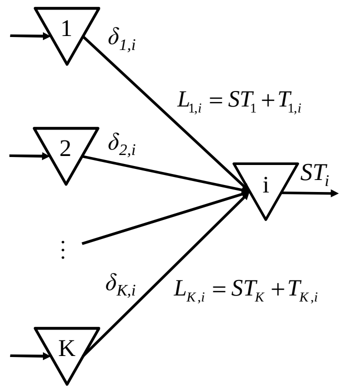


Figure 4.28: Supply network with multiple suppliers

and denote the sum as δ_i^k. For ease of presentation, assume that $ST_i = 0$ for the moment. Then, by using the same logic as in the dual-sourcing case, the net stock at time t at stockpoint i is given as

$$\begin{aligned} NS_{i,t} = B_i - &\left[\sum_{j=0}^{L_i^1-1} d_{i,t-j} + \left(1-\delta_i^1\right) \sum_{j=L_i^1}^{L_i^2-1} d_{i,t-j} + \left(1-\left(\delta_i^1+\delta_i^2\right)\right) \sum_{j=L_i^2}^{L_i^3-1} d_{i,t-j} \right. \\ &\left. + \cdots + \left(1-\sum_{l=1}^{K-1}\delta_i^l\right) \sum_{j=L_i^{K-1}}^{L_i^K-1} d_{i,t-j} \right] \\ = B_i - &\left[\sum_{j=0}^{L_i^1-1} d_{i,t-j} + \sum_{k=1}^{K-1} \left(1-\sum_{l=1}^{k}\delta_i^l\right) \sum_{j=L_i^k}^{L_i^{k+1}-1} d_{i,t-j} \right] \quad . \end{aligned} \tag{4.107}$$

Assume that $\sum_{j=a}^{b} x^j = 0$ for $b < a$. The term in parenthesis represents the inventory shortfall. If stockpoint i quotes a positive service time, ST_i, to its successor, the inventory shortfall is reduced by $\sum_{j=0}^{ST_i-1} d_{i,t-j}$, i.e.

$$NS_{i,t} = B_i - \left[\sum_{j=0}^{L_i^1-1} d_{i,t-j} + \sum_{k=1}^{K-1} \left(1-\sum_{l=1}^{k}\delta_i^l\right) \sum_{j=L_i^k}^{L_i^{k+1}-1} d_{i,t-j} - \sum_{j=0}^{ST_i-1} d_{i,t-j} \right] \ . \tag{4.108}$$

If $ST_i \leq L_i^1$, which is a valid condition as has been shown in the dual-sourcing case, this expression reduces to

$$NS_{i,t} = B_i - \left[\sum_{j=ST_i}^{L_i^1-1} d_{i,t-j} + \sum_{k=1}^{K-1} \left(1 - \sum_{l=1}^{k} \delta_i^l \right) \sum_{j=L_i^k}^{L_i^{k+1}-1} d_{i,t-j} \right] \quad . \tag{4.109}$$

For a stockpoint with three suppliers, Figure 4.29 illustrates the timeline for the net stock and inventory shortfall calculation, which represents a straightforward extension of the dual-sourcing case. From (4.109) it follows that the inventory

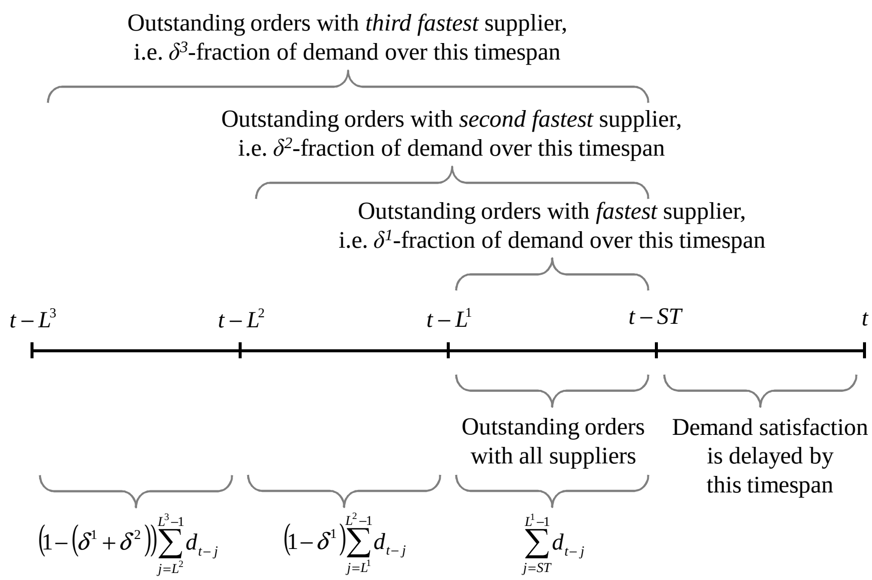


Figure 4.29: Timeline for the net stock calculation for a stockpoint with three suppliers

shortfall can be represented by the following random variable

$$\check{D}_i^{MS}(\vec{SI}_i, ST_i, \vec{\delta}_i) = D_i(L_i^1 - ST_i) + \sum_{k=1}^{K-1} \left(1 - \sum_{l=1}^{k} \delta_i^l \right) D_i(L_i^{k+1} - L_i^k) \tag{4.110}$$

with

$$\mathbb{E}\left[\check{D}_i^{MS}(\vec{SI}_i, ST_i, \vec{\delta}_i)\right] = \left(\left(L_i^1 - ST_i\right) + \sum_{k=1}^{K-1} \left(1 - \sum_{l=1}^{k} \delta_i^l \right) \left(L_i^{k+1} - L_i^k\right) \right) \mu_i \tag{4.111}$$

$$\text{VAR}\left[\check{D}_i^{MS}(\vec{SI}_i, ST_i, \vec{\delta}_i)\right] = \left(\left(L_i^1 - ST_i\right) + \sum_{k=1}^{K-1}\left(1 - \sum_{l=1}^{k}\delta_i^l\right)^2\left(L_i^{k+1} - L_i^k\right)\right)\sigma_i^2, \tag{4.112}$$

which depends on the outgoing service times of the suppliers represented by the vector $\vec{SI}_i$ (which influence the lead times L_i^k) and the outgoing service time, ST_i. The sourcing fractions represented by the vector $\vec{\delta}_i$ also influence the distribution of this random variable, but are predetermined. Under stationary conditions $t \to \infty$ the net stock follows from (4.109) as

$$NS_i = B_i - \check{D}_i^{MS}(\vec{SI}_i, ST_i, \vec{\delta}_i) \quad . \tag{4.113}$$

The optimal order-up-to level for a predefined α-service level target is given by the (smallest) value that satisfies the following (in)equality for continuous (discrete) period demand

$$Pr\left\{\check{D}_i^{MS}(\vec{SI}_i, ST_i, \vec{\delta}_i) \leq B_i\right\} \geq \alpha_i^{target} \quad . \tag{4.114}$$

The safety stock is found as

$$SST_i(\vec{SI}_i, ST_i, \vec{\delta}_i, \alpha_i^{traget}) = B_i(\vec{SI}_i, ST_i, \vec{\delta}_i, \alpha_i^{traget}) - \mathbb{E}[\check{D}_i^{MS}(\vec{SI}_i, ST_i, \vec{\delta}_i)] \quad . \tag{4.115}$$

In case of normally $i.i.d.$ period demand, the safety stock can be calculated as

$$SST_{i_{norm}}(\vec{SI}_i, ST_i, \vec{\delta}_i, \alpha_i^{traget}) = k_i(\alpha_i^{target})\sigma_i\sqrt{\left(L_i^1 + \sum_{k=1}^{K-1}\left(1 - \sum_{l=1}^{k}\delta_i^l\right)^2\left(L_i^{k+1} - L_i^k\right) - ST_i\right)}. \tag{4.116}$$

The complexity of the dynamic program increases considerably the more suppliers a stockpoint has, because all of these outgoing service times need to be considered simultaneously. Therefore, the computational complexity is given as $P^{SI}M^2 + \sum_{k=2}^{\bar{K}} P_K^{MS} M^{K+1}$, where P^{SI} denotes the number of stockpoints with single sourcing,

P_K^{MS} the number of multiple-sourced stockpoints with K supplier, $\bar{K}$ the maximum number of suppliers, and M the maximum replenishment lead time, which corresponds to the maximum outgoing service time.

4.5.4.2 Optimization of sourcing fractions

Optimization model

If the analysis is restricted to convergent systems with dual sourcing again, a simultaneous optimization of the safety stocks and the sourcing fractions is possible, too. In the upcoming exposition it is assumed that period demand is normally *i.i.d.*. Once the sourcing fractions become decision variables, the total relevant cost in the system needs to be expanded by two terms: (i) the cost of goods sold (COGS) and (ii) the pipeline inventory cost. Both terms are influenced by the sourcing fractions and therefore need to be included in the objective function of the optimization problem. The COGS at stockpoint i are given by

$$\eta \sum_{\forall j:(j,i)\in\mathcal{A}} c_{j,i}^{add} \delta_{j,i} \mu_i \qquad i = 1, ..., n \quad . \tag{4.117}$$

They represent the total cost of all units that are delivered to customers during a company-defined interval of time. η is a scalar that is used to express the COGS in the same time unit as the pipeline and safety stock cost (cf. Graves and Willems (2005)).

With regard to the cost valuation of the units in the pipeline various approaches can be taken:

1. The units can be costed using the holding cost of the next upstream stockpoint. Thus, it would be implicitly assumed that the value-adding only takes place once the units enter the next downstream stockpoint.
2. On the other extreme, the holding cost of the next downstream stockpoint can be used for all units in the pipeline, which assumes that the entire value is added right at the start of a process.
3. As a compromise, one can argue that the value-adding occurs during the time, which the units spent in the pipeline or process. Following Graves and Willems

(2005), the units in the pipeline between stockpoints j and i could then be valued at a cost per unit and period of

$$\nu\left(\frac{c_j^{cum}+c_i^{cum}}{2}\right)=\nu\left(\frac{c_i^{cum}-c_{j,i}^{add}+c_i^{cum}}{2}\right)=\nu\left(c_i^{cum}-\frac{c_{j,i}^{add}}{2}\right) \quad , \tag{4.118}$$

i.e. the cost for one unit of pipeline inventory between stockpoints j and i is the product of the holding cost rate ν and the average of the unit cost at stockpoints j and i.

Using the second approach, for instance, the entire optimization problem can be stated as

$$\begin{aligned}\mathbf{P}^3 \quad \min \sum_{i=1}^{n} \Bigg(& \underbrace{\eta \sum_{\forall j:(j,i)\in\mathcal{A}} c_{j,i}^{add}\delta_{j,i}\mu_i}_{\text{COGS}} + \underbrace{\nu \sum_{\forall j:(j,i)\in\mathcal{A}} \left(c_j^{cum}+c_{j,i}^{add}\right) T_{j,i}\delta_{j,i}\mu_i}_{\text{pipeline inventory cost}} \\ & + \underbrace{h_i k_i(\alpha_i^{target})\sigma_i\sqrt{ST_i^f+T_i^f+[\delta_i^s]^2\cdot\left(ST_i^s+T_i^s-ST_i^f-T_i^f\right)-ST_i}}_{\text{safety stock cost}}\Bigg)\end{aligned} \tag{4.119}$$

s.t.

$$ST_i^s+T_i^s \geq \left(ST_i^f+T_i^f\right)\cdot\mathbf{I}\{\delta_i^s\in(0,1)\}+ST_i\cdot\mathbf{I}\{\delta_i^s=1\}\geq ST_i \qquad i=1,2,...,n \tag{4.120}$$

$$ST_i \in \mathbb{N} \qquad i=1,2,...,n \tag{4.121}$$

$$ST_0 = 0 \tag{4.122}$$

$$ST_n = 0 \tag{4.123}$$

$$\delta_i^s+\delta_i^f=1 \qquad i=1,2,...,n \tag{4.124}$$

$$0\leq\delta_i^s\leq 1 \qquad i=1,2,...,n \tag{4.125}$$

$\mathbf{P}^3$ resembles $\mathbf{P}^1$, only the objective function contains additional terms and two constraints on the sourcing fractions are added. It can be solved by a dynamic programming algorithm. The procedure is outlined below.

Optimization procedure

Additional notations and labeling. In addition to the notations and labeling of stockpoints introduced in Section 4.5.2, the set of stockpoints in the subgraph from one dual-sourced stockpoint i to the next upstream one (or external supplier) l (including i, but excluding l) is defined as dual-sourcing subnetwork $\mathcal{DS}(i)$ (see Figure 4.30).

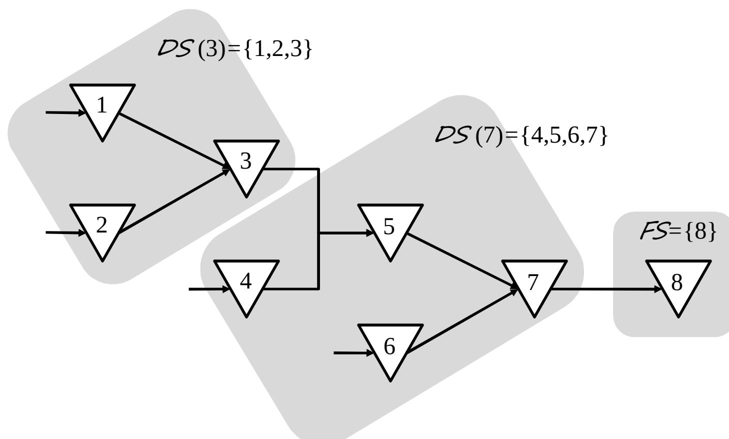


Figure 4.30: Dual-sourcing subnetworks

Furthermore, let $\mathcal{DS}^j(i)$ denote the set of stockpoints in $\mathcal{DS}(i)$, which are predecessors of stockpoint i's supplier j (including j). If no dual-sourcing subnetwork $\mathcal{DS}(n)$ exists, $\mathcal{FS}$ is defined as the final-stockpoint subnetwork, which comprises of all stockpoints that are not an element of any of the dual-sourcing subnetworks. For each stockpoint j the following function is defined:

$$Index(j) := \begin{cases} i & \text{if } j \in \mathcal{DS}(i) \\ n & \text{if } j \in \mathcal{FS} \end{cases}, \tag{4.126}$$

which returns the dual-sourced stockpoint of the dual-sourcing subnetwork, to which stockpoint j belongs, or the final stockpoint, if $\mathcal{DS}(n)$ does not exist. Moreover, the following stockpoint set is introduced

$$\mathcal{N}_j^{\mathcal{DS}(i)} = \{j\} + \bigcup_{\forall l:(l,j)\in\mathcal{A}} \mathcal{N}_l \qquad \forall l, j \in \mathcal{DS}(i) \quad , \tag{4.127}$$

which comprises of all stockpoints upstream of j in the dual-sourcing subnetwork $\mathcal{DS}(i)$.

For ease of presentation of the cost functions later on, the following expressions are defined

$$COPI_m = \sum_{\substack{\forall l:(l,m)\in\mathcal{A},\\ l\in\mathcal{DS}^j(i)}} \left(\nu(c_l^{cum} + c_{l,m}^{add})T_{l,m} + COPI_l\right) \qquad \forall m \in \mathcal{DS}^j(i) \tag{4.128}$$

$$COGS_m = \sum_{\substack{\forall l:(l,m)\in\mathcal{A},\\ l\in\mathcal{DS}^j(i)}} \left(\eta c_{l,m}^{add} + COGS_l\right) \qquad \forall m \in \mathcal{DS}^j(i) \quad , \tag{4.129}$$

which summarize the pipeline inventory cost and cost of goods sold up to stockpoint m in a dual-sourcing subnetwork without considering the mean demand at these stockpoints.

The dynamic program proceeds from the most upstream dual-sourcing subnetwork to the most downstream one (or final-stockpoint subnetwork) and finds the solution to $\mathbf{P}^3$ for each of them by evaluating a functional equation for each stockpoint in the subnetworks.

Functional equations. Two cost functions are defined, one for single-sourced stockpoints and one for dual-sourced stockpoints.

1. Singe-sourced stockpoint

   At each stockpoint $m \in \mathcal{DS}^j(i)$ (or $\mathcal{FS}$) the sourcing fraction is given and equal to 1, because only at a dual-sourced stockpoint a decision about the fraction can be made. Consequently, the pipeline inventory cost and the cost of goods sold are fixed, too. Moreover, the holding cost h_m is not influenced by the sourcing fraction. The only decision that has to be made at these stockpoints concerns the safety stock allocation. Hence, the following function is formulated, which gives the minimum of the sum of the safety stock cost in the subnetwork with stockpoint set $\mathcal{N}_m^{\mathcal{DS}(i)}$ (or $\mathcal{N}_m^{\mathcal{FS}}$) and the total cost of the preceding dual-sourcing subnetwork:

$$C_m^{SI^{opt}}(SI_m, ST_m) = h_m k_m(\alpha_m^{target})\sigma_m \sqrt{SI_m + T_m - ST_m} + \sum_{\forall l:(l,m)\in\mathcal{A}} v_l(\min\{SI_m, M_l\}, m) \quad . \tag{4.130}$$

The first term corresponds to the well-known GS safety stock cost expression. There is a single incoming service time SI_m, which is the maximum of the outgoing service times of all suppliers. Furthermore, the processing times between the suppliers and stockpoint m are identical and therefore only indicated by T_m. The second term summarizes the safety stock cost of the preceding stockpoints in the current dual-sourcing subnetwork and the total cost of the preceding dual-sourcing subnetworks. The function $v_l(.)$ is defined after the cost function of a dual-sourced stockpoint, because it applies to both.

2. Dual-sourced stockpoint

   At a dual-sourced stockpoint a decision about the sourcing fraction as well as the safety stock allocation is made. Therefore, all costs need to be considered. Without loss of generality, it is assumed that the following relation holds between the suppliers j and k of stockpoint i: $ST_j + T_j \geq ST_k + T_k \geq ST_i$. Then, for i from $\mathcal{DS}(i)$, the minimum total cost up to this stockpoint is given as

$$\begin{aligned} C_i^{DS^{opt}}&(ST_j, ST_k, ST_i, \delta_{j,i}) = \\ &h_i(\delta_{j,i})k_i(\alpha_i^{target})\sigma_i\sqrt{ST_k + T_k + [\delta_{j,i}]^2\left(ST_j + T_j - ST_k - T_k\right) - ST_i} \\ &+ \delta_{j,i}\mu_i\left(\nu(c_j^{cum} + c_{j,i}^{add})T_{j,i} + COPI_j + \eta c_{j,i}^{add} + COGS_j\right) \\ &+ (1-\delta_{j,i})\mu_i\left(\nu(c_k^{cum} + c_{k,i}^{add})T_{k,i} + COPI_k + \eta c_{k,i}^{add} + COGS_k\right) \\ &+ v_j(\min\{ST_j, M_j\}, i) + v_k(\min\{ST_k, M_k\}, i) \quad . \end{aligned} \tag{4.131}$$

The first term expresses the safety stock cost at stockpoint i. The second and third terms represent the cost for pipeline inventory and the cost of goods sold in the dual-sourcing subnetwork $\mathcal{DS}(i)$, which result from sourcing from supplier j and k, respectively. The fourth and fifth terms account for the safety stock cost in the current dual-sourcing subnetwork up to supplier j and k, respectively, as well as the total cost of all preceding dual-sourcing subnetworks.

In both cost functions,

$$\begin{aligned} v_l(ST_l, m) = &\min_{SI_l} \left\{ C_l^{SI^{opt}}(SI_l, ST_l) \right\} \cdot \mathbf{I}\{Index(l) = Index(m)\} \\ &+ \min_{(ST_j, ST_k)} \left\{ C_l^{DS^{opt}} \left(ST_j, ST_k, ST_l, \delta^*_{j,l}(ST_j, ST_k, ST_l) \right) \right\} \\ &\cdot \mathbf{I}\{Index(l) \neq Index(m)\} \qquad j \neq k \text{ and } (j,l),(k,l) \in \mathcal{A}. \end{aligned} \tag{4.132}$$

In contrast to the situation with predetermined sourcing fractions, where the sourcing fraction itself has been used as an indicator for whether the cost minimization is to be conducted for a single- or dual-sourced stockpoint, the distinction here is made by using the affiliation of a stockpoint to a specific dual-sourcing subnetwork, because the sourcing fraction itself is a decision variable. Due to the definition of a dual-sourcing subnetwork, each such network only contains one dual-sourced stockpoint, say i. All other stockpoints of this subnetwork are single-sourced ones. The $Index$-function of all these stockpoints returns i. Since it holds that stockpoint l is a direct predecessor of stockpoint $m \in \mathcal{DS}(i)$, l can only be a dual-sourced stockpoint, if it belongs to a different dual-sourcing subnetwork, i.e. $Index(l) \neq Index(m)$.

In both summands of (4.132) the optimization (for a given ST_l) can be done by enumeration. In the first summand, this can be done as in the standard GS approach over a single incoming service time, SI_l. In the second one, the enumeration has to be done over all incoming service-time combinations (ST_j, ST_k), which define an optimal supply fraction $\delta^*_{j,l}(ST_j, ST_k, ST_l)$.

The cost functions (4.130) and (4.131) still contain unknown parameters besides the decision variables. The mean demand, μ_i, and standard deviation, σ_i, of a stockpoint in an upstream dual-sourcing subnetwork are influenced by the sourcing fractions of the downstream subnetworks. Thus, the dynamic programming algorithm cannot start with the most upstream subnetwork unless some further modifications are made. To this end, the following lemmata are of help, which are stated without proofs since they are straightforward.

Lemma 4.5.4.1 *In pure convergent networks, the coefficient of variation is the same at each stockpoint and equal to the one of the final stockpoint, i.e.* $CV_i =$

$\sigma_i/\mu_i = CV_n$. *This holds irrespective of the sourcing fractions.*

Lemma 4.5.4.2 *The mean demand and demand variability at all stockpoints* $m \in \mathcal{DS}^j(i)$ *is identical, namely* $\mu_m = \delta_{j,i}\mu_i$ *and* $\sigma_m = \delta_{j,i}\sigma_i$.

Due to Lemma 4.5.4.1, μ_i can be replaced by σ_i/CV_n in (4.131), which gives

$$\begin{aligned} &C_i^{DS^{opt'}}(ST_j, ST_k, ST_i, \delta_{j,i}) = \\ &\quad h_i(\delta_{j,i})k_i(\alpha_i^{target})\sigma_i\sqrt{ST_k + T_k + [\delta_{j,i}]^2\left(ST_j + T_j - ST_k - T_k\right) - ST_i} \\ &\quad + \delta_{j,i}\frac{\sigma_i}{CV_n}\left(\nu(c_j^{cum} + c_{j,i}^{add})T_{j,i} + COPI_j + \eta c_{j,i}^{add} + COGS_j\right) \\ &\quad + (1-\delta_{j,i})\frac{\sigma_i}{CV_n}\left(\nu(c_k^{cum} + c_{k,i}^{add})T_{k,i} + COPI_k + \eta c_{k,i}^{add} + COGS_k\right) \\ &\quad + v_j(\min\{ST_j, M_j\}, i) + v_k(\min\{ST_k, M_k\}, i) \quad . \end{aligned} \tag{4.133}$$

Exploitation of Lemma 4.5.4.2 means that σ_m in (4.130) can be replaced by $\delta_{j,i}\sigma_i$, which yields

$$\begin{aligned} C_m^{SI^{opt'}}(SI_m, ST_m) = {} & h_m k_m(\alpha_m^{target})\delta_{j,i}\sigma_i\sqrt{SI_m + T_m - ST_m} \\ & + \sum_{\forall l:(l,m)\in\mathcal{A}} v_l(\min\{SI_m, M_l\}, m) \ . \end{aligned} \tag{4.134}$$

Now, it can be observed that in all terms of the cost functions (4.133) as well as (4.134) the factor σ_i appears. Consequently, it can be neglected without changing the ultimate outcome. The final modification is based on the following lemma, which is straightforward and therefore stated without proof.

Lemma 4.5.4.3 *For all stockpoints* $m \in \mathcal{DS}^j(i)$, $\delta_{j,i}$ *represents a constant factor in the cost function (4.134). Therefore, the optimal safety stock allocation is not influenced by it.*

Due to Lemma 4.5.4.3, the sourcing fraction $\delta_{j,i}$ can be included in the cost function of the dual-sourced stockpoint rather than the single-sourced stockpoints. As a result, it follows that

$$C_m^{SI^{opt''}}(SI_m, ST_m) = h_m k_m(\alpha_m^{target})\sqrt{SI_m + T_m - ST_m} + \sum_{\forall l:(l,m)\in\mathcal{A}} v_l(\min\{SI_m, M_l\}, m) \qquad m \in \mathcal{DS}^j(i) \text{ or } m \in \mathcal{DS}^k(i) \quad (4.135)$$

$$\begin{aligned} C_i^{DS^{opt''}}(ST_j, ST_k, ST_i, \delta_{j,i}) = \\ & h_i(\delta_{j,i})k_i(\alpha_i^{target})\sqrt{ST_k + T_k + [\delta_{j,i}]^2(ST_j + T_j - ST_k - T_k) - ST_i} \\ & + \delta_{j,i}\frac{1}{CV_n}\left(\nu(c_j^{cum} + c_{j,i}^{add})T_{j,i} + COPI_j + \eta c_{j,i}^{add} + COGS_j\right) \\ & + (1-\delta_{j,i})\frac{1}{CV_n}\left(\nu(c_k^{cum} + c_{k,i}^{add})T_{k,i} + COPI_k + \eta c_{k,i}^{add} + COGS_k\right) \\ & + \delta_{j,i}v_j(\min\{ST_j, M_j\}, i) + (1-\delta_{j,i})v_k(\min\{ST_k, M_k\}, i) \qquad i \in \mathcal{DS}(i)\ . \end{aligned} \quad (4.136)$$

These two cost functions contain no unknown parameters any more, only the decision variables, i.e. the service times and the sourcing fractions. Consequently, for the computations in the dynamic programming algorithm, $C_i^{SI^{opt}}$ and $C_i^{DS^{opt}}$ in (4.132) need to be replaced by $C_i^{SI^{opt''}}$ and $C_i^{DS^{opt''}}$.

Dynamic programming algorithm. The procedure for the simultaneous optimization of the service times and the sourcing fractions is as follows. For all feasible outgoing service times of a dual-sourced stockpoint i, ST_i, an enumeration over all feasible incoming service-time combinations (ST_j, ST_k) is conducted. For each surrounding service-time constellation, the corresponding optimal sourcing fraction, $\delta^*_{j,i}(ST_j, ST_k, ST_i)$, is determined. Unfortunately, the cost function is not convex in the sourcing fraction. Therefore, the optimal fraction needs to be found by use of some numerical method.

The dynamic programming algorithm can be stated as:

1. *For all $\mathcal{DS}(i)$ with $LC(i) := N$ down to 1 (or $\mathcal{FS}$)*
2. *For all stockpoints $m \in \mathcal{DS}(i)$ with $LC(m) := N$ down to i (or down to 1 in*

$\mathcal{FS}$)

3. *Evaluate $v_m(ST_m, i)$ for $ST_m = 0, 1, .., M_m$.*

4. *For $m := n$ evaluate $v_m(ST_m, n)$ for $ST_m = 0$ (assuming immediate demand satisfaction).*

5. *Minimize $v_n(ST_n, n)$ for $ST_n = 0$ and multiply the resulting expression by the standard deviation of the customer demand to obtain the optimal objective function value.*

An optimal set of service times is found by the standard backtracking procedure for a dynamic program.

Benefit of the sourcing fraction optimization

In order to assess the benefit of optimizing the sourcing fractions, the following analysis is conducted. Due to the decoupling effect in the GS model it is sufficient to consider a single dual-sourced stockpoint and compare the resulting total relevant cost of three scenarios:

1. Single sourcing from the slow supplier (*slow single sourcing*),

2. Single sourcing from the fast supplier (*fast single sourcing*), and

3. Dual sourcing.

For simplicity reasons, it is assumed that the incoming service times as well as the outgoing service time are 0. Furthermore, $T^s = xT^f$, i.e. the slow processing time is a multiple of the fast one. The total relevant cost for slow single sourcing is given as

$$TRC^s = \eta c^{add^s}\mu + \nu c^{add^s} xT^f \mu + \nu c^{add^s} k\sigma\sqrt{xT^f} \quad . \tag{4.137}$$

Similarly, for fast single sourcing

$$TRC^f = \eta c^{add^f}\mu + \nu c^{add^f} T^f \mu + \nu c^{add^f} k\sigma\sqrt{T^f} \tag{4.138}$$

and for dual sourcing

$$\begin{aligned} TRC^{DS} = &\eta\left(\delta^s c^{add^s} + (1-\delta^s) c^{add^f}\right)\mu \\ &+ \nu\left(\delta^s c^{add^s} x T^f + (1-\delta^s) c^{add^f} T^f\right)\mu \\ &+ \nu\left(\delta^s c^{add^s} + (1-\delta^s) c^{add^f}\right) k\sigma\sqrt{T^f + [\delta^s]^2 (xT^f - T^f)} \quad . \end{aligned} \tag{4.139}$$

Dual sourcing is better than slow single sourcing, if the following relation holds:

$$TRC^s > TRC^{DS}$$

$$\frac{c^{add^f}}{c^{add^s}} < \underbrace{\frac{\nu(1-\delta^s) x T^f \mu + \nu k\sigma\left(T^f + \sqrt{xT^f} - \delta^s\sqrt{[\delta^s]^2 (xT^f - T^f)}\right) + \eta(1-\delta^s)\mu}{\nu(1-\delta^s) T^f \mu + \nu k\sigma(1-\delta^s)\sqrt{T^f + [\delta^s]^2 (xT^f - T^f)} + \eta(1-\delta^s)\mu}}_{c^{add^f}_{upper}} . \tag{4.140}$$

The left-hand side of (4.140) expresses c^{add^f} as a multiple of c^{add^s}. That means, assuming c^{add^s} is given, the right-hand side of (4.140) defines an upper bound on c^{add^f} that must not be exceeded, if dual sourcing is to be better than slow single sourcing. For dual sourcing to be better than fast single sourcing, it has to hold that

$$TRC^f > TRC^{DS}$$

$$\frac{c^{add^f}}{c^{add^s}} > \underbrace{\frac{\nu\delta^s x T^f \mu + \nu k\sigma\delta^s\sqrt{T^f + [\delta^s]^2 (xT^f - T^f)} + \eta\delta^s\mu}{\nu\delta^s T^f \mu + \nu k\sigma\left(\sqrt{T^f} - (1-\delta^s)\sqrt{T^f + [\delta^s]^2 (xT^f - T^f)}\right) + \eta\delta^s\mu}}_{c^{add^f}_{lower}} . \tag{4.141}$$

Similarly to (4.140), (4.141) defines a lower bound on c^{add^f}. Consequently, the range, within which c^{add^f} (expressed as a multiple of c^{add^s}) can vary such that dual sourcing is advantageous, is defined by

$$c^{add^f}_{upper} - c^{add^f}_{lower} \quad . \tag{4.142}$$

This range is rather small for reasonable parameter settings as the following example illustrates (see Table 4.7 and Figure 4.31). If c^{add^f} is expressed as a percentage of c^{add^s}, the range, within which c^{add^f} has to lie in order to make dual sourcing preferable, is about 2% or less for a coefficient of variation of up to 1. For the specific example of $c^{add^s} = 100$, this means that fast single sourcing is better, if $c^{add^f} < 110$, and slow single sourcing is preferable, if $c^{add^f} > 112.5$ (see Figure 4.31).

Surely, one can benefit to a larger extent from dual sourcing, if demand is more variable. However, even for a coefficient of variation of period demand of 3 the c^{add^f}-range is just about 6% wide (see Table 4.7). It is very unlikely that the procurement cost of the fast supplier falls exactly within that range.

Consequently, in many instances the optimal sourcing fraction assumes one of the extreme values of the feasible region, 0 or 1, i.e. single sourcing is the optimal strategy. Presumably, this result will not change in larger supply network settings, because the major difference there is a potential increase in the processing times, which could also be reflected in this single-stockpoint model. Graves and Willems (2005) analyze a related model, where at each stage there is a choice between several processing/sourcing options differing in terms of the process length and added cost. One option has to be chosen exclusively, however. Thus, the model described in this section can be viewed as a generalization of their model. It confirms nevertheless that an extreme strategy is often a reasonable choice. This finding is further supported by the fact that the cost advantage of dual sourcing, if it is chosen at all, is not very large, either. It amounts to not even 1% (see Table 4.8).

		δ^s									Max:
CV		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	$c^{add^f}_{upper} - c^{add^s}_{lower}$
0.3	$c^{add^f}_{upper}$	110.53%	110.57%	110.59%	110.60%	110.61%	110.61%	110.61%	110.60%	110.60%	0.58%
	$c^{add^f}_{lower}$	109.95%	110.08%	110.19%	110.27%	110.33%	110.38%	110.41%	110.44%	110.45%	
0.5	$c^{add^f}_{upper}$	111.00%	111.07%	111.10%	111.12%	111.13%	111.13%	111.13%	111.13%	111.12%	0.97%
	$c^{add^f}_{lower}$	110.04%	110.27%	110.45%	110.58%	110.68%	110.75%	110.81%	110.85%	110.88%	
0.8	$c^{add^f}_{upper}$	111.71%	111.81%	111.87%	111.90%	111.91%	111.91%	111.91%	111.90%	111.88%	1.54%
	$c^{add^f}_{lower}$	110.17%	110.54%	110.83%	111.04%	111.20%	111.32%	111.41%	111.47%	111.52%	
1.0	$c^{add^f}_{upper}$	112.18%	112.30%	112.38%	112.41%	112.43%	112.43%	112.42%	112.41%	112.39%	1.92%
	$c^{add^f}_{lower}$	110.26%	110.72%	111.08%	111.35%	111.55%	111.69%	111.80%	111.88%	111.95%	
1.5	$c^{add^f}_{upper}$	113.35%	113.53%	113.63%	113.68%	113.70%	113.70%	113.68%	113.65%	113.62%	2.87%
	$c^{add^f}_{lower}$	110.48%	111.17%	111.71%	112.12%	112.41%	112.63%	112.79%	112.91%	113.00%	
2.0	$c^{add^f}_{upper}$	114.51%	114.74%	114.87%	114.94%	114.95%	114.94%	114.91%	114.87%	114.82%	3.80%
	$c^{add^f}_{lower}$	110.70%	111.63%	112.35%	112.88%	113.28%	113.57%	113.78%	113.94%	114.05%	
2.5	$c^{add^f}_{upper}$	115.65%	115.94%	116.10%	116.17%	116.18%	116.16%	116.12%	116.06%	115.99%	4.73%
	$c^{add^f}_{lower}$	110.92%	112.08%	112.98%	113.65%	114.14%	114.50%	114.76%	114.95%	115.09%	
3.0	$c^{add^f}_{upper}$	116.79%	117.13%	117.31%	117.38%	117.39%	117.36%	117.29%	117.21%	117.12%	5.64%
	$c^{add^f}_{lower}$	111.14%	112.53%	113.61%	114.42%	115.01%	115.44%	115.75%	115.97%	116.12%	

Table 4.7: c^{add^f}-range for a dual-sourcing cost advantage ($T^f = 5$, $T^s = 12 \cdot T^f$, $\nu = 45\%$, $\eta = 250$, $k = 2.33$)

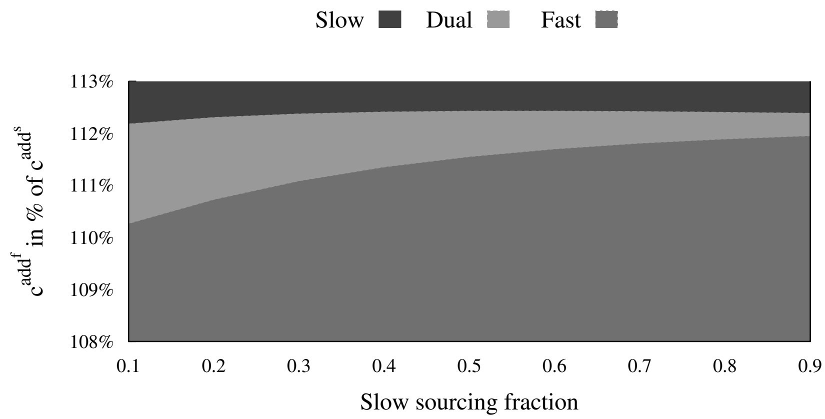


Figure 4.31: c^{add^f}-range for a dual-sourcing cost advantage ($T^f = 5$, $T^s = 12 \cdot T^f$, $\nu = 45\%$, $\eta = 250$, $k = 2.33$, $CV = 1$)

			Dual-sourcing cost advantage over	
CV	$c_{upper}^{add^f}$	δ^{s*}	slow single sourcing	fast single sourcing
0.3	110	0.1	0.43%	0.01%
0.5	111	0.4	0.07%	0.15%
0.8	111	0.2	0.59%	0.08%
1.0	112	0.3	0.24%	0.25%
1.5	113	0.3	0.40%	0.34%
2.0	114	0.3	0.54%	0.43%
2.5	115	0.3	0.68%	0.52%
3.0	116	0.3	0.80%	0.61%

Table 4.8: Dual-sourcing cost advantage ($T^f = 5$, $T^s = 12{\cdot}T^f$, $c^{add^s} = 100$, $\nu = 45\%$, $\eta = 250$, $k = 2.33$)

The major drivers for this outcome can been seen in the following two aspects:

1. Total relevant cost components

   The safety stock cost, which benefits most from dual sourcing, makes up only a very small share of the total relevant cost. In the COGS expression, the mean demand is multiplied by $\eta = 250$, whereas the safety stock quantity is only multiplied by $\nu = 45\%$, $k = 2.33$ and σ, which is considerably lower in total.

2. Replenishment policy

   The replenishment policy might not be well-chosen in view of deterministic processing times. In the single-echelon case, order splitting is used almost exclusively in stochastic lead-time scenarios. For deterministic lead times, there are often better ways than allocating the demand variability evenly to both suppliers, such as the SIP, COP, or DIP (see Chapter 3).

   However, the problem with these other policies is that the upstream demand process changes depending on the policy parameters. Even for given parameters it is quite cumbersome and complex to determine the upstream demand processes as will be shown in the next section.

   Despite the quite disappointing results, in practice there might still be companies that employ a kind of order-splitting policy to allocate production volumes to their factories. For these companies, the optimal safety stock can be determined with the optimization model for given sourcing fractions (Section 4.5.2) and, if possible for the given cost parameters, even the sourcing fractions can be optimized by the approach outlined in this section.

4.5.4.3 Other inventory control policies

If, instead of an order-splitting policy with a predefined sourcing fraction, a different dual-sourcing policy, such as the SIP, COP, or DIP, is to be used at a stockpoint, additional aspects need to be analyzed and others modified. These include:

- the holding cost per unit and period,
- the total relevant cost,
- the order processes, and
- the optimization of service times.

Holding cost per unit and period

The OSP specifies the sourcing fraction explicitly as one of the policy parameters. Consequently, the holding cost per unit and period at a dual-sourced stockpoint can be easily computed according to (4.90). In case of the other dual-sourcing policies the sourcing fractions need to be derived indirectly from the policy parameters. (For ease of presentation the stockpoint index is dropped.) The analysis for each policy yields:

SIP:

$$\delta^f(\Delta) = \frac{\mathbb{E}\left[Q^f(\Delta)\right]}{\mu} \overset{(3.28)}{=} \frac{\mathbb{E}\left[(D-\Delta)^+\right]}{\mu} \tag{4.143}$$

COP:

$$\delta^f(Q) = \frac{\mathbb{E}\left[Q^f(Q)\right]}{\mu} \overset{(3.72)}{=} \frac{\mu - Q}{\mu} = 1 - \frac{Q}{\mu} \tag{4.144}$$

DIP:

$$\delta^f(\Delta) = \frac{\mathbb{E}\left[Q^f(\Delta)\right]}{\mu} \overset{(3.2)}{=} \frac{\mu - \mathbb{E}\left[Q^s(\Delta)\right]}{\mu} = 1 - \frac{\mathbb{E}\left[Q^s(\Delta)\right]}{\mu} \overset{(3.110)}{=} 1 - \frac{\Delta - \mathbb{E}\left[O(\Delta)\right]}{L^{\Delta} \cdot \mu} \tag{4.145}$$

and $\delta^s = 1 - \delta^f$. Given the sourcing fractions, the holding cost per unit and period can be determined according to (4.90).

It is important to note that under all three policies the sourcing fraction only depends on one policy parameter, either Δ or Q. Consequently, given this parameter the holding cost is fixed and the optimal value of the second policy parameter can be derived as previously explained.

Total relevant cost

Since the sourcing fractions are now dependent on the policy parameters, the cost of goods sold (COGS) and the pipeline inventory cost enter the total relevant cost function. This resembles the setting analyzed in the previous section, where the service times and the sourcing fractions have been optimized simultaneously. The cost expressions are identical to the ones outlined there.

Order processes

The third aspect refers to the determination of the order processes that a dual-sourced stockpoint induces. These represent the demand processes of the suppliers. Under the OSP each supplier faces a fraction of the demand of the downstream stockpoint each period. That means, the upstream stockpoints see a scaled version of the demand distribution of the downstream situated dual-sourced stockpoint. Several demand distributions commonly used in inventory theory, like the Normal or Gamma, are closed under a scale transformation. Consequently, the demand distribution of the suppliers can be derived easily and exactly. Under the SIP, COP, or DIP the derivation proves more difficult. For each of the three policies, the main aspects are addressed by analyzing the simplest multi-echelon dual-sourcing system depicted in Figure 4.32. The black and grey lines represent the flow of goods and information, respectively.

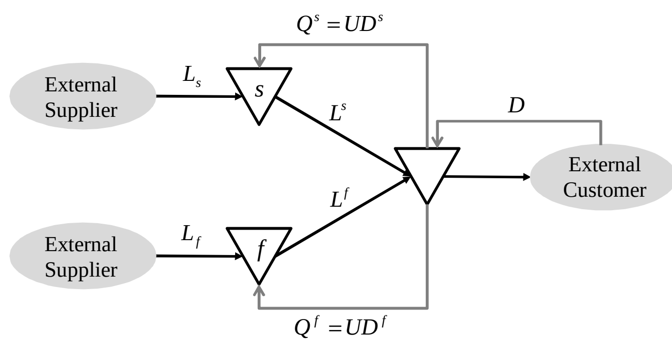


Figure 4.32: Simple supply network with dual sourcing

SIP. The analysis in Section 3.3.3 shows that the orders under the SIP are given as

$$Q^s = \min\{\Delta, D\} \tag{4.146}$$

$$Q^f = (D - \Delta)^+ \quad . \tag{4.147}$$

In case of a discrete (non-negative) demand distribution, the distributions of the upstream demand random variables for the slow and fast supplier, UD^s_{SIP} and UD^f_{SIP}, respectively, can be determined from (4.146) and (4.147) as

$$Pr\left\{UD^{s}_{SIP}=x\right\}=\begin{cases}Pr\left\{D=x\right\} & 0\leq x<\Delta\\ 1-\sum_{i=0}^{\Delta-1}Pr\left\{D=i\right\} & x=\Delta\\ 0 & x>\Delta\end{cases} \tag{4.148}$$

and

$$Pr\left\{UD^{f}_{SIP}=x\right\}=\begin{cases}\sum_{i=0}^{\Delta}Pr\left\{D=i\right\} & x=0\\ Pr\left\{D=x+\Delta\right\} & x>0\end{cases}. \tag{4.149}$$

In each period, the slow supplier sees a positive demand (unless the external demand is zero). The fast supplier does not necessarily face a positive demand in each period. Due to the order-up-to policy structure at the downstream stockpoint that governs its demand process, periods with zero demand can occur, if the inventory position is larger than the fast order-up-to level, B^f. The larger Δ, the more frequently this happens (see (4.149)).

Due to the differing nature of the demand processes at the two suppliers, different approaches for their order-up-to level determination need to be taken (assuming that both suppliers operate a periodic-review order-up-to policy). The order-up-to level of the slow supplier can be determined according to (2.50) as the smallest value that satisfies the following inequality

$$Pr\left\{[UD^{s}_{SIP}]_{L_s}\leq B_s\right\}\geq\frac{b_s}{b_s+h_s} \tag{4.150}$$

where $[UD^{s}_{SIP}]_{L_s}$ denotes the upstream-stage demand random variable over L_s periods and b_s and h_s the cost parameters at this stockpoint. (Note that the review period is not included in the characterization of (4.150), because it is assumed that the upstream stockpoints place their own order after they have observed the order of the downstream stockpoint.)

The sporadic nature of positive demand events at the fast supplier (intermittent demand) calls for a different approach. Instead of modeling lead-time demand by using some non-compound distribution (e.g. truncated normal) and applying (4.150) for the order-up-to level computation, the intermittent character is exploited, i.e. lead-time demand is modeled as a compound binomial process. This method, which is outlined in Teunter et al. (2010) for a spare parts application, is similar in nature to Croston's method of forecasting intermittent demand (see Croston (1972)), where the demand size and demand interval are updated separately.

The compound binomial (or Bernoulli) process is characterized by the fact that with a fixed probability there is a positive demand during a period, otherwise demand is zero. In case of a positive demand, the demand size follows another distribution. This process has been studied by Dunsmuir and Snyder (1989) for the continuous-review (s, Q) system and by Janssen et al. (1998) for the periodic-review (R, s, Q) inventory model. In contrast to the approach of Teunter et al. (2010), which is exact and applied here, Janssen et al. (1998) use approximations. Let p^f denote the probability that demand for the fast supplier is positive in a period, which is given as

$$p^f_{SIP} = Pr\left\{D > \Delta\right\} \quad . \tag{4.151}$$

Since the inventory position after ordering in each period equals the slow order-up-to level, p^f_{SIP} is identical for all periods and independent across periods. Consequently, the probability that the fast supplier faces exactly l positive demand events during the replenishment lead time L_f follows a binomial distribution, i.e.

$$Pr\left\{l, L_f\right\} = \binom{L_f}{l} \cdot [p^f_{SIP}]^l \cdot (1 - p^f_{SIP})^{L_f - l} \tag{4.152}$$

where

$$\binom{L_f}{l} = \frac{L_f!}{l!(L_f - l)!} \quad . \tag{4.153}$$

The distribution of the random variable $[UD^f_{SIP}]^{pos}$, which denotes the demand size for positive realizations only, is computed as

$$Pr\left\{[UD^f_{SIP}]^{pos} = x\right\} = \frac{1}{p^f_{SIP}} \cdot Pr\left\{UD^f_{SIP} = x\right\} \qquad x > 0 \quad . \tag{4.154}$$

Given these two distributions, the optimal order-up-to level can be computed as the smallest value that satisfies

$$\sum_{l=0}^{L_f} Pr\left\{l, L_f\right\} \cdot Pr\left\{[UD_{SIP}^f]_l^{pos} \leq B_f\right\} \geq \frac{b_f}{b_f + h_f} \tag{4.155}$$

where $[UD_{SIP}^f]_l^{pos}$ denotes the positive demand random variable over l periods.

Approximations. Since the computation of the exact discrete demand distributions and convolutions might be cumbersome (for distributions with many feasible realizations and long lead times), a discrete distribution function can be fitted using the first two moments of the respective random variable according to Adan et al. (1995).

Moreover, in case of continuous demand the following approximation can be performed. If external customer demand follows (or can be approximated by) a mixed-Erlang distribution, the moments of the order distributions can be computed exactly.

$$\mathbb{E}[(Q^s)^k] = \mathbb{E}\left[(UD_{SIP}^s)^k\right] \qquad k = 1, 2, ... \tag{4.156}$$

$$\mathbb{E}[(Q^f)^k] = \mathbb{E}\left[\left(UD_{SIP}^f\right)^k\right] \qquad k = 1, 2, ... \tag{4.157}$$

Given $p_{SIP}^f = Pr\left\{D > \Delta\right\}$,

$$\mathbb{E}\left[\left([UD_{SIP}^f]^{pos}\right)^k\right] = \frac{1}{p_{SIP}^f} \cdot \mathbb{E}\left[\left(UD_{SIP}^f\right)^k\right] \qquad k = 1, 2, ... \,. \tag{4.158}$$

Based on these first two moments, a new mixed-Erlang distribution can be fitted as described in Tijms (1994), which can then be used for the order-up-to level determination at the suppliers.

COP and DIP. Under the COP and DIP the probability that an order is placed with the fast supplier in a period depends on the realizations in previous periods, i.e. autocorrelation exists. Consequently, p^f is not independent across and identical for all periods. Thus, an exact computation is more complicated than in the SIP case and goes beyond the scope of this work. In order to keep the calculations simple, an idea would be to use the above-described approach for the SIP as an approximation for the COP and DIP. It needs to be tested in future research, however, how well this approximation works.

For the COP, the upstream demand distributions can be derived based on the slow and fast order expressions

$$Q^s = Q \tag{4.159}$$
$$Q^f = (D - (O(Q) + Q))^+ \tag{4.160}$$

as

$$Pr\left\{UD^s_{COP} = Q\right\} = 1 \tag{4.161}$$
$$Pr\left\{UD^f_{COP} = x\right\} = \begin{cases} \sum_{i=0}^{\infty} Pr\left\{D \leq i + Q\right\} \cdot Pr\left\{O(Q) = i\right\} & x = 0 \\ \sum_{i=0}^{\infty} Pr\left\{D = x + i + Q\right\} \cdot Pr\left\{O(Q) = i\right\} & x > 0 \end{cases} \tag{4.162}$$

given the overshoot distribution $O(Q)$ for a predefined Q. The slow supplier faces a constant demand of Q units each period. Since there is no variability attached, the supplier does not need to hold any safety stock and can operate in a just-in-time fashion. On the other hand, the demand for the fast supplier fluctuates with a potentially large probability mass at zero depending on Q (see (4.162)). The approximated probability for a positive demand at the fast supplier is

$$p^f_{COP} = 1 - Pr\left\{UD^f_{COP} = 0\right\} \quad . \tag{4.163}$$

Under the DIP, the upstream demand distributions need to be computed via a multi-dimensional Markov Chain (similar to the one for the determination of the overshoot distribution) or simply approximated by means of simulation.

Optimization of service times

In the analysis of the order processes, it has been implicitly assumed that the service times, which determine the replenishment lead times at the suppliers as well as the dual-sourced stockpoint, are fixed. However, in the GS optimization model, these service times represent decision variables, as well. That means, under any of the three discussed dual-sourcing policies, the entire optimization problem consists of two intertwined optimization problems. On the one hand, for the optimization of the service times in the entire supply network, the sourcing fractions or other policy parameters at all dual-sourced stockpoints need to be known, because they influence the upstream demand processes. On the other hand, the optimization of the policy parameters at a dual-sourced stockpoint requires knowledge about the surrounding service times, since they influence the replenishment lead times and thus coverage decision. Moreover, the sourcing decision at a dual-sourced stockpoint also affects the holding cost parameter computation for the pipeline inventory and safety stock held at downstream stockpoints.

This difficulty also occurred in Section 4.5.4.2, where the sourcing fractions for the OSP have been optimized. Due to the relatively simple ordering structure of this policy, the upstream demand processes could be easily derived, however. Furthermore, some substitutions finally allowed for a similar solution procedure of the problem as the one for predefined sourcing fractions starting at the most upstream stockpoints.

In case of the SIP, COP, or DIP the upstream demand processes are more difficult to derive as previously described. The solution procedure of the OSP case cannot be simply adjusted. The interdependencies between the dual-sourcing policy parameters and the service times call for the development of a more sophisticated solution algorithm, which goes beyond the scope of this work, however.

4.5.5 Summary and implications

In this section it has been shown how to integrate dual sourcing in the guaranteed-service framework. In the literature, this framework is reported to be widely used in practice emphasizing its relevance. The newly developed approach enables the optimization of the safety stock allocation and sizing even for large serial and convergent systems with dual sourcing. Although a rather simple dual-sourcing replenishment policy has been studied, i.e. an order-splitting policy, the model represents an improvement to the existing contributions in the literature. These either provide only an approximate approach for this problem or consider much smaller supply networks, mainly two-echelon systems.

Moreover, the presented approach has been shown to be easily extendable to more than two suppliers. Even the simultaneous optimization of the demand allocation to the suppliers and the safety stock allocation has been outlined for the dual-sourcing case. Preliminary results have indicated, however, that the benefit of such a simultaneous optimization is rather limited. The model has often prescribed single sourcing as an optimal strategy. Nevertheless, for companies, which intentionally decide to use two or more suppliers, the approach for a fixed demand allocation can be of considerable help for the appropriate sizing of the safety stocks in their supply networks. Finally, the incorporation of other dual-sourcing policies such as the single-index, constant-order, or dual-index policy has been discussed and the main difficulties of their integration have been identified, which deserve further investigation in future research.

5 Conclusions and outlook

This chapter summarizes the major findings of this thesis and discusses possible future research.

5.1 Conclusions

This thesis has contributed to the field of literature on dual-sourcing inventory models in a single- and multi-echelon setting. Since the integration of dual sourcing into a multi-echelon inventory model requires a thorough understanding of multi-echelon inventory models with single sourcing as a starting point, it has also provided a major contribution to this body of literature. The thesis has centered around two major research topics: (i) the detection of effective dual-sourcing inventory control policies in a single-echelon model (Chapter 3) and (ii) the integration of dual-sourcing into a multi-echelon inventory model (Chapter 4).

Chapter 3 has focused on a single-echelon periodic-review inventory model with two suppliers. Several dual-sourcing policies have been presented in Section 3.3. First, it has been shown how to compute the optimal policy in Section 3.3.2. For the special case of a lead-time difference of one period between the two suppliers (i.e. consecutive lead times), the optimal policy is known to be the single-index policy. For larger lead-time differences it has to be found via a Markov Decision Process (MDP) formulation, which has been outlined in this section.

From the MDP model it is apparent that the optimal policy can only be computed in a reasonable amount of time for limited problem sizes. The state and decision space increase considerably as the lead-time difference between the two suppliers grows or the mean demand and demand variability become larger.

That is why in Section 3.3.3 several simpler and, in general, non-optimal policies have been outlined. These include the single-index (SIP), constant-order (COP), dual-index (DIP), and the order-splitting policy (OSP).

In order to provide the reader with a thorough understanding of these policies, as a valuable basis for the policy comparison in Section 3.4, major results from the literature have been reiterated using a unified notational framework. At certain point, new aspects have been added. For each policy, the backorder-cost model as well as the service-level model has been addressed. Furthermore, it has been outlined how to compute the optimal policy parameters. In all four cases, the optimization can be performed by a one-dimensional search procedure over the relevant policy parameter region. Only in special cases are there easier ways for the optimization. In the case of consecutive lead times, one of the optimal parameters of the SIP, which is also the truly optimal policy in this setting, can be derived directly via a critical fractile (in)equality. For the other parameter, an analytical expression can be derived, from which the optimal value can be computed numerically. In case of the COP and DIP, it has been pointed out that the major difficulty in the parameter optimization lies in the derivation of the stationary overshoot distribution. An overshoot denotes the quantity by which the fast inventory position might exceed the order-up-to level that is used for determining replenishments with the fast supplier. Such an overshoot can occur, because the fast inventory position, which both policies use, only includes outstanding orders that arrive within the fast replenishment lead time, but not any other slow outstanding orders that have already been determined. Several ways (exact and approximate) of how this overshoot distribution can be derived have been summarized. For the special case of geometric demand a new recursive computation of the stationary overshoot distribution has been presented in case of the COP and a direct closed-form computation in case of the DIP. If period demand follows a normal distribution, a closed-form expression for the optimal sourcing fraction of the OSP has been derived. The other policy parameter can be computed as the solution to an (in)equality.

Section 3.4 has dealt with the comparison of the dual-sourcing policies. Since the SIP and DIP have already been compared in Scheller-Wolf et al. (2007) and the OSP is arguably inferior to the other policies in the studied deterministic lead-time setting, the section has focused on the remaining two policies, for which a comparison has not yet been available in the literature, the COP and DIP.

The main finding based on the theoretical considerations with respect to the extreme strategies of both policies has been that the cost gap between both policies closes as the lead-time difference increases. It suggests that at some point the COP might even outperform the DIP (Section 3.4.2). This presumption has been backed by the numerical study in Section 3.4.3. Whereas the generally good performance of the DIP has been confirmed, which has already been shown by Veeraraghavan and Scheller-Wolf (2008), the COP has been identified as an effective policy alternative in settings with significant lead-time differences and small expediting premiums. This finding is interesting for two reasons. First, the COP is the more easily implementable and controllable policy in practice. Second, the constant-order guarantee can be of particular importance in supply negotiations. This supplier does not face any demand fluctuations, which facilitates his production planning considerably. In situations with a small lead-time difference and large expediting premium, single sourcing has often been found to be a reasonable alternative to the more complex DIP.

The major contribution of Chapter 4 has been two-fold. First, the existing two main multi-echelon inventory modeling frameworks without lot-sizing, i.e. the stochastic-service (SS) and guaranteed-service (GS) approach, have been outlined (Section 4.2), compared (Section 4.3), and combined in the so-called hybrid-service (HS) approach (Section 4.4). Second, an extension of one of the three multi-echelon frameworks, namely the GS model, to accommodate dual sourcing has been presented (Section 4.5).

In Section 4.2, apart from summarizing the existing models and results, the problem of setting an appropriate internal service level at each stage of the supply network in the GS approach has been addressed. This level specifies the maximum reasonable demand, up to which all variability is to be covered by safety stock. Demand variability exceeding this threshold is assumed to be handled by other countermeasures, which are summarized by the term 'operating flexibility'. Operating flexibility measures have not been modeled explicitly in most of the GS contributions, but the analysis has focused exclusively on the 'normal' part of the demand variability (*standard GS approach*). This has caused a lot of criticism of this approach in the past. In Section 4.2.3.2, this criticism has been counteracted by taking into account the effect that operating flexibility has on the material flow in the system (*extended GS approach*). Although various ways of how operating flexibility can work have been described, only the most relevant one for the context of this thesis has been analyzed in detail, i.e. missing items are speeded up from the pipeline inventory of a stage. Interestingly, this new modeling has led to a change in the objective function of the GS model. Whereas in the standard GS approach the safety stock cost across the entire supply chain is minimized, it is the on-hand stock cost that is minimized in the extended GS approach. Furthermore, the explicit modeling of the operating flexibility has allowed for the derivation of a closed-form expression for the determination of the internal service level, provided that a cost parameter for the operating flexibility usage can be specified per unit. In many situations, this is probably easier to do for management than specifying a service level directly.

In Section 4.3 the SS and GS approaches have been compared. The main contribution here has been the identification of an individual benefit of each of the two approaches: the allocation benefit of the SS approach and the decoupling benefit of the GS approach. The SS approach is not restricted in any respect with regard to the stock allocation across the various stages of the supply network (apart from the final-stage service-level target). It can base its allocation decision completely on the holding-cost relationships between the stages. The GS approach does not have this kind of flexibility. It has to comply with the internal or external service-level requirement at each stage. However, it benefits from the operating flexibility, which allows for a decoupling of the stages, i.e. there are no stochastic delays that need to be taken into account at downstream stages in case of supply shortages.

A numerical study, which has been conducted for serial and divergent systems, has revealed the following three important drivers of the advantage of one approach over the other: processing-time pattern, final-stage service level(s), and internal service level (or operating flexibility cost). Due to the individual benefits, the GS approach has shown a superior performance for a degressive processing-time pattern, high final-stage service level(s), and a low internal service level, while the opposite has been true for the SS model. From the three drivers, the internal service-level parameter has been identified as the most important one. The major finding from the numerical comparison has been, however, that none of the approaches is superior to the other, in general. Both approaches have been found to have their advantages and disadvantages in certain settings.

In Section 4.4 both approaches have been combined in order to benefit from both individual advantages, before the dual-sourcing extension of the GS approach has been addressed in Section 4.5. The integrated approach has been called the hybrid-service (HS) approach. The HS approach allows the entire supply network to consist of subnetworks of both types. The interface modeling between the subnetworks has been outlined and a pseudo-polynomial time dynamic programming algorithm for the determination of the optimal network partitioning and stock sizing in serial systems has been developed. Furthermore, the extension to divergent and convergent systems has been discussed.

The main contribution and finding has been that the HS approach solves the practitioner's dilemma of having to choose one of the two multi-echelon inventory optimization approaches exclusively for the entire supply network. It enables a stage-wise choice. Moreover, the numerical study for serial systems with up to five stages has revealed that the cost difference between the two pure approaches can be quite significant. The HS approach prevents an erroneous choice and, in addition, has been shown to even achieve further cost-savings of up to 10.5% at most and 1.9% on average in the analyzed experimental design. It has performed best in settings with relatively low internal service levels, a broad internal service-level range, degressive processing-time structure, and progressive holding-cost pattern.

In Section 4.5, out of the three previously described and developed (single-sourcing) multi-echelon approaches (SS, GS, and HS), the standard GS approach has been chosen and extended to incorporate dual sourcing. Due to the already increased model complexity resulting from the shift from a single-echelon inventory model to a multi-echelon one (even under the GS framework), a rather simple dual-sourcing inventory control policy has been selected, i.e. the order-splitting policy. Moreover, it has been assumed that the determination of the sourcing fractions, which define the demand allocation to the suppliers, is exogenous to the model. Thus, the main focus has been put on the computation of the optimal safety stock allocation and sizing. Extensions to more than two suppliers, the simultaneous optimization of the sourcing fractions and safety stocks, as well as the integration of other inventory control policies have been briefly addressed in the 'Extensions' section.

It has been shown that in order to accurately incorporate dual sourcing into the GS model, a stockpoint and its preceding process can no longer be aggregated into a stage with a single processing time. In case of dual sourcing, a stockpoint is preceded by two processes of different length. A stage with a single processing time cannot capture the processing-time difference correctly. Consequently, a safety stock computation based on such an aggregation would only lead to an approximate quantity and cost. Therefore, instead of the stockpoint (and its index), the processing time needs to be assigned to the arc connecting two stockpoints.

For serial and convergent systems, a dynamic programming algorithm has been developed that optimally determines the safety stock allocation in the system as well as the exact stock size at each stockpoint. The approach represents an improvement of the only approximate modeling idea outlined in the final section of Graves and Willems (2005), which is one of very few contributions available in the literature that address a similar problem. Most of the other works only consider supply networks of smaller sizes.

5.2 Outlook

Within both bodies of literature, the single-echelon inventory models with dual sourcing as well as the multi-echelon inventory models with (and without) dual sourcing, various challenges remain.

Although procedures for the policy parameter optimization of the different dual-sourcing inventory control policies in a single-echelon setting have been presented in this thesis, the optimization of the COP and DIP parameters still requires knowledge about Markov Chain models or the development of a simulation tool, which might hamper their application in practice. Here, the development of heuristics that enable an approximate, but simple parameter computation in a spreadsheet model, for instance, would be of interest.

Once lead times become stochastic, the analysis gets more complicated, but the model is also brought closer to reality. Arts et al. (2009) consider a special type of stochastic integer lead times in the DIP. Only the slow lead time is stochastic with a lower bound of the support of this random variable that is larger than the fast lead time. A relaxation of this restriction might be an interesting extension. Moreover, the relative performance of the SIP, COP, DIP, and OSP might change in the presence of stochastic lead times. A comparison would now have to include all four dual-sourcing policies, because in such a setting the OSP cannot be excluded upfront.

With respect to the multi-echelon inventory models with single sourcing, the hybrid-service (HS) approach has been developed in this thesis. While the HS model idea directly extends to divergent and convergent systems (see Sections 4.4.3 and 4.4.4), the increase in the computational complexity calls for the development of good heuristics for networks of larger sizes. Furthermore, an extension of the HS approach to include stochastic processing times and lot-sizes are promising and relevant areas for future research.

The dual-sourcing integration into a multi-echelon model has been done for the guaranteed-service approach only. The integration into the stochastic-service or even hybrid-service framework also represents an interesting area for future research. Moreover, the developed dynamic programming algorithm can solely handle serial and convergent network structures. In the presence of divergent (sub)structures, the computational complexity of the current solution method increases considerably (see Section 4.5.3). Therefore, the development of other methods or heuristics for the extension to divergent systems might be worthwhile pursuing.

Furthermore, the integration of other ordering policies like the SIP, COP, or DIP in a multi-echelon setting requires further investigation. Whereas the major difficulties connected with these policies have been identified and discussed in this thesis in Section 4.5.4.3, the ultimate integration and development of an optimization procedure for the policy parameters has gone beyond the scope of this work.

Similar to the single-echelon case, the consideration of stochastic processing times might also represent a valuable extension in the multi-echelon dual-sourcing model. In such a setting the value of the OSP is presumably larger than in the deterministic processing-time setting as is known from single-echelon models.

For both the single- and multi-echelon model with dual sourcing, the inventory valuation problem might be an issue that deserves more investigation of its own. One approach to tackle this problem has been presented in this thesis. However, there are other ways to address or circumvent this problem, e.g., by using a discounted cash flow approach. It remains that the assumption of a single holding cost parameter that most dual-sourcing models make, which does not take into account the different procurement costs, is a simplification that is not justifiable in all situations.

Appendices

A Additional figures and tables

A.1 Comparison of the constant-order and dual-index policy for $\mu = 10$

	$TRC_{BS} \geq TRC_{DIP}$				$TRC_{BS} > TRC_{COP}$				$TRC_{BS} < TRC_{COP}$				
	No.	$\frac{TRC_{BS}-TRC_{DIP}}{TRC_{DIP}}$			No.	$\frac{TRC_{BS}-TRC_{COP}}{TRC_{COP}}$			No.	$\frac{TRC_{COP}-TRC_{BS}}{TRC_{COP}}$			n/s
$h = 0.1$	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max	Inst.
$L^{\Delta} = 2$													
$(1;3)$	24	1.40%	0.01%	6.22%	0	–	–	–	24	49.35%	20.50%	84.38%	0
$(3;5)$	24	0.45%	0.00%	2.59%	0	–	–	–	24	46.64%	20.76%	81.88%	0
$L^{\Delta} = 5$													
$(1;6)$	30	4.93%	0.02%	15.94%	0	–	–	–	30	39.56%	3.89%	79.68%	0
$(3;8)$	30	2.36%	0.01%	8.85%	0	–	–	–	30	39.58%	8.52%	77.95%	0
$L^{\Delta} = 10$													
$(1;11)$	36	10.76%	0.01%	29.68%	6	8.34%	2.56%	16.94%	30	37.50%	1.56%	79.18%	0
$(3;13)$	36	6.51%	0.01%	18.51%	3	4.04%	1.82%	7.43%	32	36.69%	3.14%	78.28%	1
$h = 0.5$	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max	Inst.
$L^{\Delta} = 2$													
$(1;3)$	36	5.07%	0.01%	13.15%	4	2.49%	0.86%	5.57%	32	24.12%	3.92%	56.92%	0
$(3;5)$	36	2.09%	0.00%	6.50%	0	–	–	–	35	21.88%	2.29%	52.91%	1
$L^{\Delta} = 5$													
$(1;6)$	36	16.24%	2.48%	29.60%	18	15.43%	1.58%	26.40%	18	16.77%	1.22%	43.52%	0
$(3;8)$	36	9.10%	0.91%	17.81%	13	9.93%	3.02%	14.86%	21	17.08%	3.31%	42.67%	2
$L^{\Delta} = 10$													
$(1;11)$	36	31.15%	11.25%	56.18%	30	28.84%	2.56%	61.04%	6	10.58%	1.56%	26.51%	0
$(3;13)$	36	20.86%	6.70%	36.31%	26	20.56%	1.82%	38.52%	9	10.81%	0.79%	28.95%	1
$h = 1.0$	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max	Inst.
$L^{\Delta} = 2$													
$(1;3)$	36	8.28%	0.79%	16.28%	12	7.32%	2.54%	11.41%	23	15.42%	1.49%	37.00%	1
$(3;5)$	36	3.82%	0.13%	8.49%	8	3.25%	1.32%	5.86%	27	14.10%	1.59%	34.25%	1
$L^{\Delta} = 5$													
$(1;6)$	36	21.27%	7.87%	38.24%	30	17.91%	1.58%	38.16%	6	7.03%	1.22%	17.42%	0
$(3;8)$	36	12.81%	3.84%	22.35%	24	12.04%	1.44%	22.02%	9	8.74%	3.31%	19.94%	3
$L^{\Delta} = 10$													
$(1;11)$	36	36.61%	6.78%	74.45%	36	35.79%	7.45%	80.10%	0	–	–	–	0
$(3;13)$	36	25.36%	6.03%	47.70%	35	24.46%	6.91%	51.29%	1	0.79%	0.79%	0.79%	0

Table A.1: Single- vs. dual-sourcing cost for $\mu = 10$

	Best single (BS) vs. DIP				Best single (BS) vs. COP				COP vs. DIP			
	No.	$\frac{TRC_{BS}-TRC_{DIP}}{TRC_{DIP}}$			No.	$\frac{TRC_{BS}-TRC_{COP}}{TRC_{COP}}$			No.	$\frac{TRC_{COP}-TRC_{DIP}}{TRC_{DIP}}$		
Poisson	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max
(L^f, L^s)												
$(1, 6)$	12	17.80%	2.48%	38.24%	12	0.65%	-43.52%	38.16%	11	24.71%	0.44%	81.46%
$(3; 8)$	12	9.64%	0.91%	22.07%	11	-6.92%	-42.67%	21.65%	12	21.52%	0.35%	76.00%
h												
0.5	12	10.20%	0.91%	27.27%	12	-12.77%	-43.52%	23.54%	12	32.96%	3.02%	81.46%
1.0	12	17.24%	3.84%	38.24%	11	7.73%	-19.94%	38.16%	11	12.23%	0.35%	30.63%
$\frac{b}{b+h}$												
0.95	12	11.11%	0.91%	31.61%	11	-7.81%	-43.52%	31.04%	12	25.65%	0.44%	81.46%
0.99	12	16.32%	3.63%	38.24%	12	1.47%	-33.52%	38.16%	11	20.20%	0.35%	61.42%
c^f												
102	8	22.94%	11.04%	38.24%	8	20.29%	5.05%	38.16%	7	2.66%	0.35%	5.70%
105	8	12.07%	3.84%	23.82%	7	-3.30%	-19.94%	17.33%	8	16.60%	5.53%	30.63%
110	8	6.14%	0.91%	14.94%	8	-25.94%	-43.52%	-6.23%	8	47.32%	20.90%	81.46%
nbin	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max
(L^f, L^s)												
$(1, 6)$	24	18.58%	3.18%	37.03%	24	6.75%	-30.50%	37.80%	21	14.86%	-0.56%	48.47%
$(3, 8)$	24	10.83%	1.30%	21.79%	22	-0.04%	-30.90%	21.61%	21	14.14%	0.69%	46.60%
CV												
0.49	24	14.27%	1.30%	37.03%	24	1.82%	-30.90%	37.80%	22	15.73%	-0.56%	48.47%
1.05	24	15.14%	2.95%	29.26%	22	5.34%	-23.70%	25.82%	20	13.15%	2.73%	37.75%
h												
0.5	24	12.37%	1.30%	29.26%	23	-3.48%	-30.90%	25.82%	24	18.47%	2.73%	48.47%
1.0	24	17.04%	4.36%	37.03%	23	10.48%	-12.25%	37.80%	18	9.21%	-0.56%	19.22%
$\frac{b}{b+h}$												
0.95	24	12.83%	1.30%	32.31%	24	0.16%	-30.90%	32.47%	21	16.90%	0.69%	48.47%
0.99	24	16.58%	4.57%	37.03%	22	7.15%	-21.62%	37.80%	21	12.10%	-0.56%	35.90%
c^f												
102	16	21.32%	11.52%	37.03%	16	18.60%	4.71%	37.80%	10	3.70%	-0.56%	7.22%
105	16	14.39%	4.36%	26.78%	15	4.25%	-12.25%	21.08%	16	10.35%	4.25%	19.22%
110	16	8.41%	1.30%	19.19%	15	-13.36%	-30.90%	7.23%	16	25.41%	10.25%	48.47%
Gamma	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max
(L^f, L^s)												
$(1, 6)$	36	19.20%	3.75%	37.61%	36	8.36%	-29.05%	37.79%	31	13.30%	0.54%	46.22%
$(3, 8)$	36	11.47%	1.56%	22.35%	34	1.69%	-29.68%	22.02%	33	11.87%	0.37%	44.42%
CV												
0.5	24	14.96%	1.56%	37.61%	24	2.78%	-29.68%	37.79%	21	15.83%	0.85%	46.22%
1.0	24	15.59%	3.06%	29.60%	22	6.13%	-23.01%	26.40%	21	12.14%	0.37%	36.58%
1.5	24	15.44%	4.87%	29.17%	24	6.54%	-17.37%	24.75%	22	9.85%	0.54%	27.89%
h												
0.5	36	13.70%	1.56%	29.60%	35	-0.58%	-29.68%	26.40%	36	16.11%	2.08%	46.22%
1.0	36	16.97%	4.90%	37.61%	35	10.83%	-11.15%	37.79%	28	8.01%	0.37%	18.45%
$\frac{b}{b+h}$												
0.95	36	14.01%	1.56%	33.39%	36	2.35%	-29.68%	33.38%	34	13.88%	0.37%	46.22%
0.99	36	16.65%	4.89%	37.61%	34	8.06%	-20.32%	37.79%	30	11.08%	2.08%	34.07%
c^f												
102	24	19.80%	7.14%	37.61%	24	17.36%	5.53%	37.79%	16	3.14%	0.37%	7.08%
105	24	16.09%	4.90%	27.22%	23	6.87%	-11.15%	22.44%	24	9.24%	3.01%	18.45%
110	24	10.10%	1.56%	21.89%	23	-9.39%	-29.68%	10.74%	24	22.17%	9.27%	46.22%

Table A.2: Single- vs. dual-sourcing cost for $L^{\Delta} = 5$, $h = 0.5$ and 1.0, $\mu = 10$

	Best single (BS) vs. DIP				Best single (BS) vs. COP				COP vs. DIP			
	No.	$\frac{TRC_{BS}-TRC_{DIP}}{TRC_{DIP}}$			No.	$\frac{TRC_{BS}-TRC_{COP}}{TRC_{COP}}$			No.	$\frac{TRC_{COP}-TRC_{DIP}}{TRC_{DIP}}$		
Poisson	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max
(L^f, L^s)												
$(1, 11)$	12	39.71%	11.25%	74.45%	12	31.09%	-26.51%	80.10%	12	10.42%	-3.70%	51.40%
$(3, 13)$	12	25.53%	6.70%	47.70%	12	16.26%	-28.95%	51.29%	11	12.15%	-2.94%	50.19%
h												
0.5	12	26.19%	6.70%	56.18%	12	11.02%	-28.95%	61.04%	12	18.47%	-3.02%	51.40%
1.0	12	39.05%	14.41%	74.45%	12	36.33%	-0.79%	80.10%	11	3.37%	-3.70%	15.32%
$\frac{b}{b+h}$												
0.95	12	27.87%	6.70%	64.19%	12	18.00%	-28.95%	70.51%	12	12.53%	-3.70%	51.40%
0.99	12	37.37%	14.04%	74.45%	12	29.35%	-17.32%	80.10%	11	9.85%	-3.14%	39.94%
c^f												
102	8	49.15%	28.74%	74.45%	8	53.20%	30.18%	80.10%	8	-2.58%	-3.70%	-1.11%
105	8	30.16%	14.41%	50.42%	8	23.60%	-0.79%	52.95%	7	7.03%	-1.65%	15.32%
110	8	18.56%	6.70%	34.66%	8	-5.77%	-28.95%	22.23%	8	28.77%	10.17%	51.40%
nbin	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max
(L^f, L^s)												
$(1, 11)$	24	33.82%	12.07%	60.43%	24	29.65%	-10.77%	66.89%	18	5.55%	-3.87%	25.60%
$(3, 13)$	24	23.31%	7.32%	41.20%	24	18.77%	-15.04%	44.82%	17	6.58%	-2.50%	26.32%
CV												
0.49	24	30.95%	7.32%	60.43%	24	26.24%	-15.04%	66.89%	17	7.05%	-3.87%	26.32%
1.05	24	26.18%	10.25%	45.13%	24	22.18%	-7.70%	45.09%	18	5.10%	-2.35%	19.88%
h												
0.5	24	26.10%	7.32%	53.91%	24	18.14%	-15.04%	54.22%	18	10.75%	-1.41%	26.32%
1.0	24	31.03%	11.34%	60.43%	24	30.28%	7.89%	66.89%	17	1.07%	-3.87%	6.78%
$\frac{b}{b+h}$												
0.95	24	26.93%	7.32%	60.43%	24	22.02%	-15.04%	66.89%	18	6.90%	-3.87%	26.32%
0.99	24	30.20%	11.34%	53.91%	24	26.40%	-4.12%	54.22%	17	5.14%	-3.35%	19.34%
c^f												
102	16	33.88%	11.34%	60.43%	16	35.83%	12.92%	66.89%	10	-2.17%	-3.87%	-1.13%
105	16	31.17%	15.21%	48.80%	16	28.37%	7.89%	49.22%	9	4.16%	0.68%	6.78%
110	16	20.65%	7.32%	37.17%	16	8.44%	-15.04%	32.82%	16	12.25%	3.27%	26.32%
Gamma	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max	Inst.	Avg.	Min	Max
(L^f, L^s)												
$(1, 11)$	36	31.97%	6.78%	59.60%	36	27.92%	-9.10%	65.47%	28	4.85%	-3.55%	24.31%
$(3, 13)$	36	22.17%	6.03%	41.93%	35	18.48%	-13.76%	45.15%	25	5.95%	-2.22%	24.96%
CV												
0.5	24	31.40%	7.77%	59.60%	24	26.76%	-13.76%	65.47%	17	6.70%	-3.55%	24.96%
1.0	24	27.26%	10.84%	46.23%	24	23.45%	-6.51%	46.92%	19	4.65%	-2.46%	19.20%
1.5	24	22.55%	6.03%	40.10%	23	19.44%	6.83%	34.74%	17	4.85%	-1.40%	14.41%
h												
0.5	36	25.88%	7.77%	54.32%	35	19.18%	-13.76%	54.36%	27	9.35%	-1.10%	24.96%
1.0	36	28.26%	6.03%	59.60%	36	27.25%	6.91%	65.47%	26	1.23%	-3.55%	6.46%
$\frac{b}{b+h}$												
0.95	36	26.27%	7.77%	59.60%	35	22.18%	-13.76%	65.47%	28	5.98%	-3.55%	24.96%
0.99	36	27.87%	6.03%	54.32%	36	24.32%	-3.14%	54.36%	25	4.68%	-2.93%	18.24%
c^f												
102	24	27.64%	6.03%	59.60%	24	29.06%	6.91%	65.47%	15	-1.69%	-3.55%	-0.49%
105	24	30.75%	16.00%	49.28%	24	27.97%	8.96%	49.44%	14	3.92%	-0.47%	6.46%
110	24	22.82%	7.77%	40.10%	23	12.32%	-13.76%	34.74%	24	10.63%	3.00%	24.96%

Table A.3: Single- vs. dual-sourcing cost for $L^{\Delta} = 10$, $h = 0.5$ and 1.0, $\mu = 10$

A.2 Combination of the stochastic- and guaranteed-service approach

Four stages

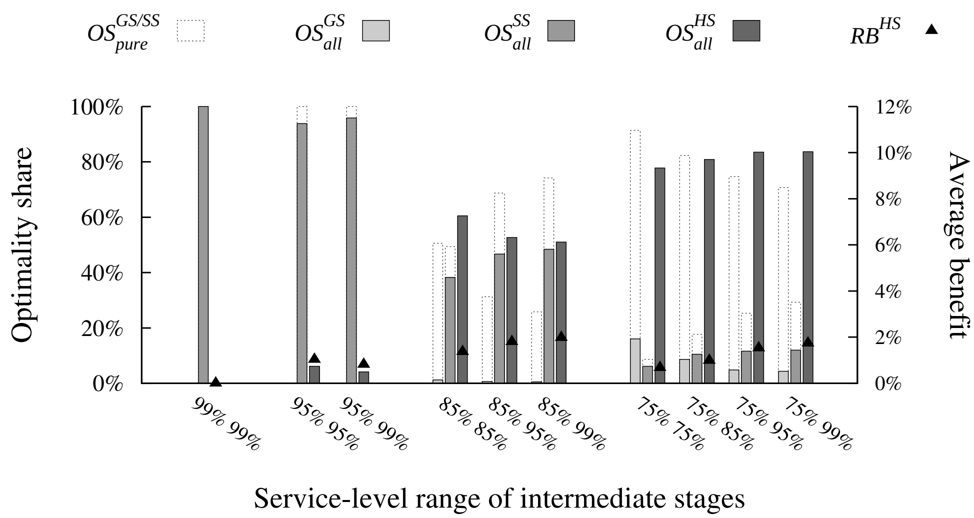


Figure A.1: Optimality share and additional average HS benefit with respect to the internal service-level ranges

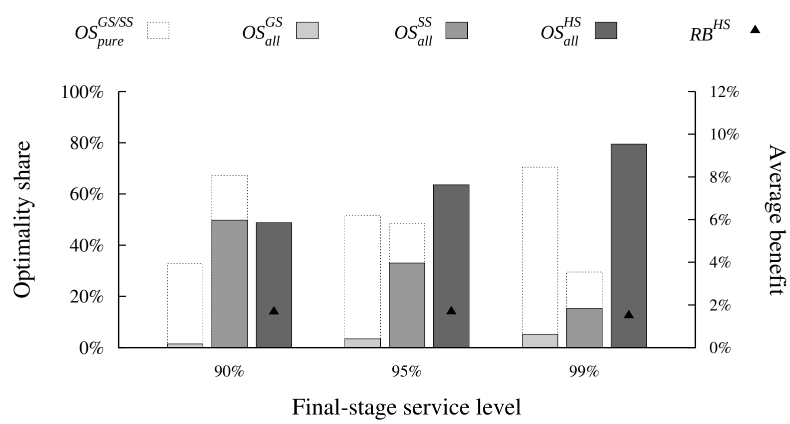


Figure A.2: Optimality share and additional average HS benefit with respect to final-stage service level

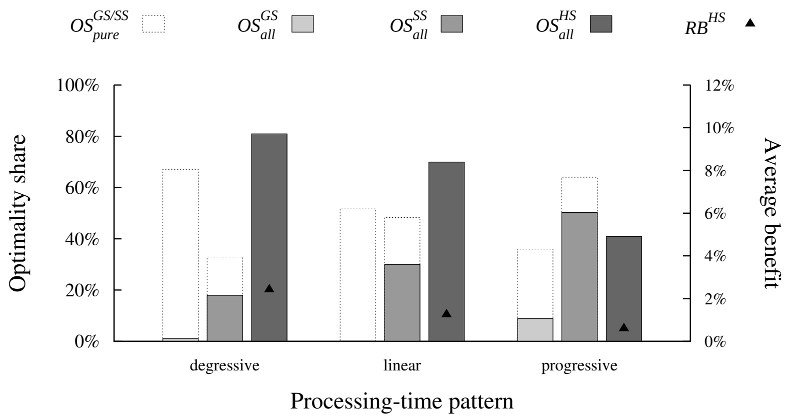


Figure A.3: Optimality share and additional average HS benefit with respect to processing-time pattern

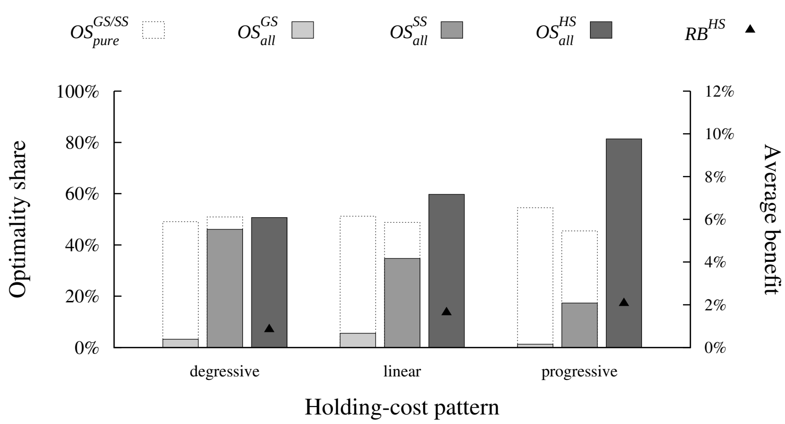


Figure A.4: Optimality share and additional average HS benefit with respect to holding-cost pattern

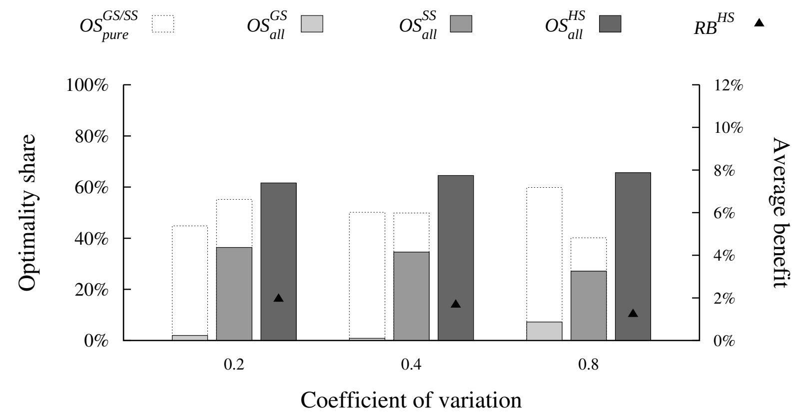


Figure A.5: Optimality share and additional average HS benefit with respect to coefficient of variation of demand

Five stages

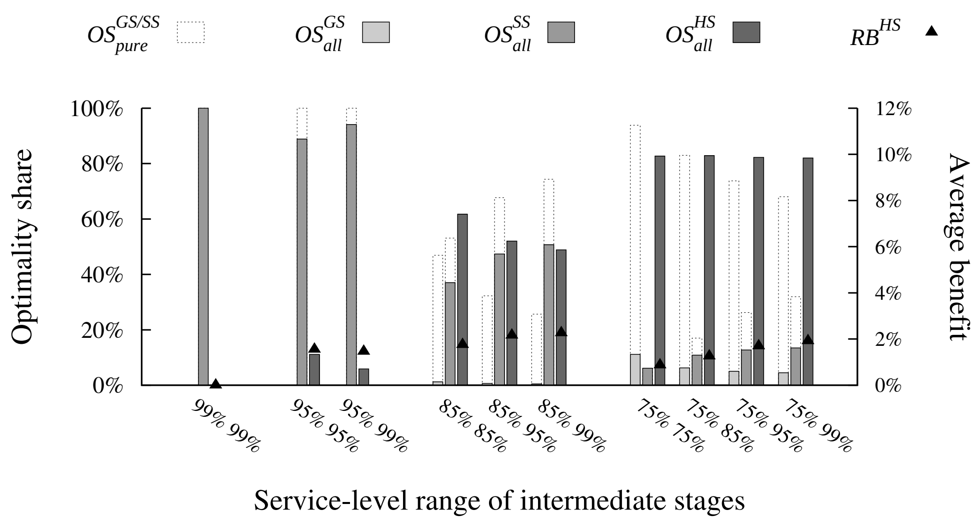


Figure A.6: Optimality share and additional average HS benefit with respect to the internal service-level ranges

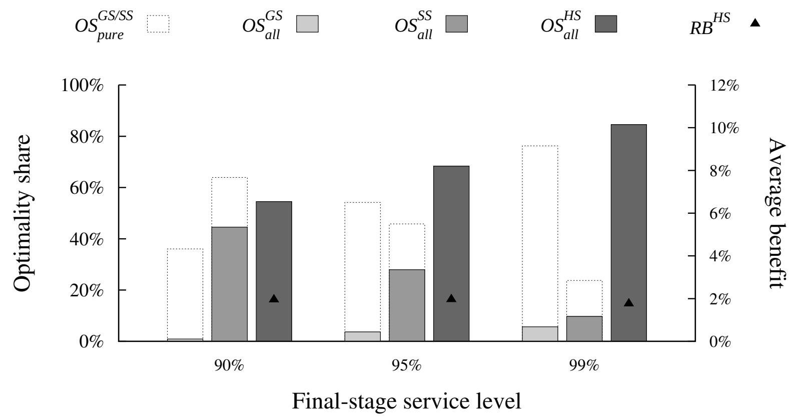


Figure A.7: Optimality share and additional average HS benefit with respect to final-stage service level

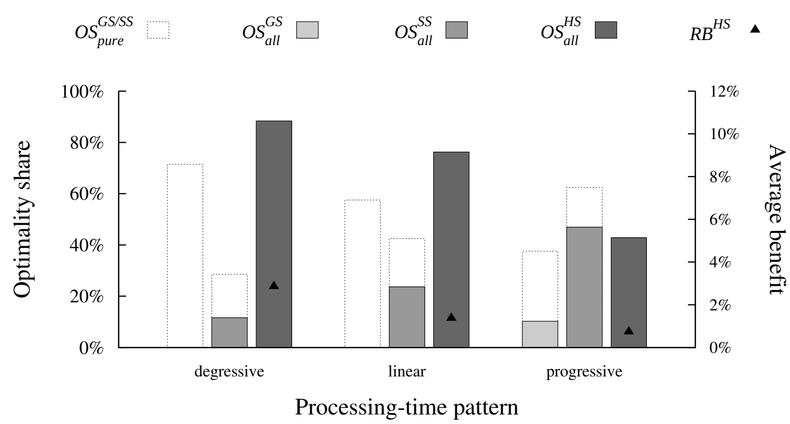


Figure A.8: Optimality share and additional average HS benefit with respect to processing-time pattern

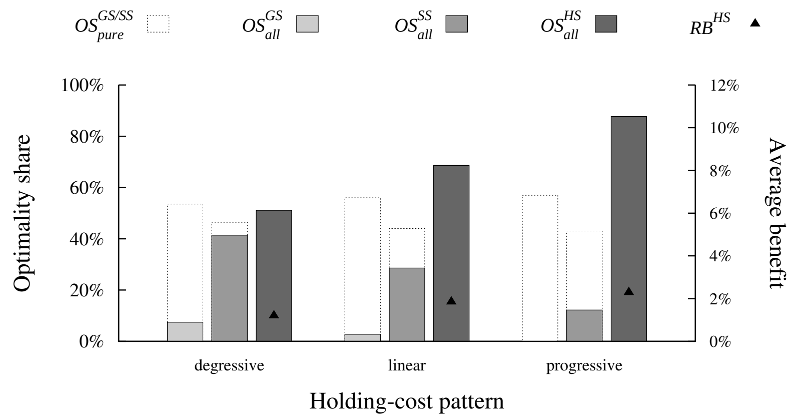


Figure A.9: Optimality share and additional average HS benefit with respect to holding-cost pattern

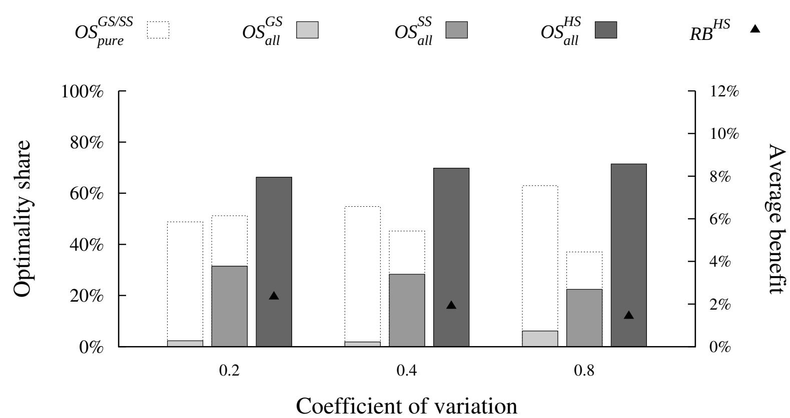


Figure A.10: Optimality share and additional average HS benefit with respect to coefficient of variation of demand

B Proofs

B.1 Lemma 3.3.3.2

Proof:
Given the stationary distribution of $\hat{D}(\Delta)$, (3.33) resembles the net stock calculation in a single-sourcing order-up-to level system with lead-time demand $\hat{D}(\Delta)$ (see (2.44) and (2.45)). Consequently, the optimal B^s is found by (3.35) accordingly. Since the expected fast order quantity is independent of B^s (Lemma 3.3.3.1), this is also the optimal B^s for the SIP for this given Δ. $\square$

B.2 Lemma 3.3.3.3

Proof:
Both parts follow from the definition of $\hat{D}(\Delta)$ and the fact that probabilities are non-negative (and using finite differences). $\square$

B.3 Lemma 3.3.3.4

Proof:
The inventory position recursion is

$$IP_{t+1}^f = IP_t^f + Q_t^f + Q_{t-L^\Delta+1}^s - D_t = B^f + O_t + Q - d_t \quad . \tag{B.1}$$

Assume that in some period, say $t = 0$, $O_0 = 0$. This occurs, e.g., if the inventory process starts with fast orders of size B^f/L^f arriving in periods $1, ..., L^f$ and no slow orders for that timespan. In periods $L^f + 1, ..., L^s$ slow orders of size Q arrive. Hence, $IP_1^f = B^f + Q - d_0 \stackrel{(B.1)}{\Longrightarrow} O_0 = 0$. From (3.69) and (B.1) it follows that

$$O_{t+1} = (O_t + Q - d_t)^+ \quad . \tag{B.2}$$

In period $t = 1$, $O_1 = (Q - d_0)^+$, which is independent of B^f. Assuming that $O_t, \forall t = 2, ..., n-1$ is independent of B^f, it follows from (3.79) that O_n is independent of B^f, and only dependent on Q. □

B.4 Lemma 3.3.3.5

Proof:

From (3.80) and (3.81) it follows that the equations for the stationary overshoot distribution are

$$o_0 = \sum_{j=0}^{\infty} Pr\{D \geq j + Q\} \cdot o_j \tag{B.3}$$

$$o_i = \sum_{i=0}^{\infty} Pr\{D = j + Q - i\} \cdot o_j \qquad i > 0 \quad . \tag{B.4}$$

A recursive representation of the above equations is

$$o_i = \frac{o_{i-1}}{1-p} \qquad i = 2, ..., Q \quad , \tag{B.5}$$

$$o_i = \frac{o_{i-1} - o_{i-Q-1} Pr\{D = 0\}}{1-p} \qquad i = Q+1, ... \quad . \tag{B.6}$$

(B.5) can be rewritten as

$$o_i = o_1 (1-p)^{1-i} \qquad i = 1, ..., Q \quad . \tag{B.7}$$

Insertion yields

$$o_0 = \sum_{i=1}^{\infty} o_i \frac{Pr\{D \geq Q+i\}}{Pr\{D < Q\}} \quad , \tag{B.8}$$

$$o_1 = \sum_{i=0}^{\infty} o_i Pr\{D = Q+i-1\} \quad . \tag{B.9}$$

All state probabilities o_i, $i > 2$ can be expressed in terms of o_0 and o_1 by recursively applying (B.7) and (B.6). For the derivation of these recursive expressions it is important to note that a Q-cycle runs from $xQ+x$ to $(x+1)Q+x$ with $x = 1, 2, \ldots$. By analyzing several of these recursive expressions for different Q-cycles the following general expression in terms of o_0 and o_1 is found:

$$\begin{aligned} o_{xQ+x-1+j} = {} & o_1 \left(\frac{1}{(1-p)^{xQ+x-2+j}} + \sum_{n=1}^{x-1} (-1)^n p^n \frac{\mathcal{X}_x^n(j)}{(1-p)^{xQ+x+j-2-nQ}} + (-1)^x p^x \frac{\mathcal{X}_x^x(j)}{(1-p)^{x+j-2}} \right) \\ & - o_0 \left(\frac{p}{(1-p)^{(x-1)Q+x-1+j}} + \sum_{n=2}^{x-1} (-1)^{n-1} p^n \frac{\mathcal{Y}_x^n(j)}{(1-p)^{xQ+x+j-1-nQ}} \right. \\ & \left. + (-1)^{x-1} p^x \frac{\mathcal{Y}_x^x(j)}{(1-p)^{x+j-1}} \right) \quad \text{for } x \geq 1 \text{ and } 1 \leq j \leq Q+1 \end{aligned} \tag{B.10}$$

where the functions $\mathcal{X}_x^n(j)$ and $\mathcal{Y}_x^n(j)$ are defined as follows for $x \geq 1$

$$\mathcal{X}_x^n(j) = \begin{cases} \sum_{k=1}^{j-1} \mathcal{X}_{x-1}^{n-1}(k+1) & x = n \\ \mathcal{X}_{x-1}^n(Q+1) + \sum_{k=1}^{j} \mathcal{X}_{x-1}^{n-1}(k) & x > n \end{cases} \qquad n \geq 1 \tag{B.11}$$

$$\mathcal{X}_x^0 = 1 \tag{B.12}$$

$$\mathcal{Y}_x^n(j) = \begin{cases} \sum_{k=1}^{j} \mathcal{Y}_{x-1}^{n-1}(k) & x = n, n \geq 2 \\ \mathcal{Y}_{x-1}^n(Q+1) + \sum_{k=1}^{j} \mathcal{Y}_{x-1}^{n-1}(k) & x > n, n \geq 1 \end{cases} \tag{B.13}$$

$$\mathcal{Y}_1^1 = \begin{cases} 0 & \text{if } x \text{ from (B.16) for the first/largest } Q\text{-cycle is equal to 1} \\ 1 & \text{otherwise} \end{cases} \tag{B.14}$$

$$\mathcal{Y}_x^0 = 0 \tag{B.15}$$

In order to apply the general expression (B.10), o_i needs to be translated into $o_{xQ+x-1+j}$, i.e. x and j need to be determined from i. This can be done as follows. Since $1 \leq j \leq Q+1$ (see above),

$$\frac{i-Q}{Q+1} \leq x \leq \frac{i}{Q+1} \quad \Rightarrow \quad x(i) = \left\lfloor \frac{i}{Q+1} \right\rfloor \tag{B.16}$$

where $x(i)$ indicates that x depends on i. Given i and x, j can be computed as

$$j(i,x) = i - x(Q+1) + 1 \quad . \tag{B.17}$$

For ease of presentation, (B.10) is rewritten as

$$o_i = o_1\mathcal{G}(i) - o_0\mathcal{H}(i) \qquad\qquad i \geq Q+1 \tag{B.18}$$

with

$$\begin{aligned}\mathcal{G}(i) = {} & \frac{1}{(1-p)^{x(i)Q+x(i)-2+j(i,x)}} + \sum_{n=1}^{x(i)-1} (-1)^n p^n \frac{\mathcal{X}^n_{x(i)}(j(i,x))}{(1-p)^{x(i)Q+x(i)+j(i,x)-2-nQ}} \\ & + (-1)^{x(i)} p^{x(i)} \frac{\mathcal{X}^{x(i)}_{x(i)}(j(i,x))}{(1-p)^{x(i)+j(i,x)-2}}\end{aligned} \tag{B.19}$$

$$\begin{aligned}\mathcal{H}(i) = {} & \frac{p}{(1-p)^{(x(i)-1)Q+x(i)-1+j(i,x)}} + \sum_{n=2}^{x(i)-1} (-1)^{n-1} p^n \frac{\mathcal{Y}^n_{x(i)}(j(i,x))}{(1-p)^{x(i)Q+x(i)+j(i,x)-1-nQ}} \\ & + (-1)^{x(i)-1} p^{x(i)} \frac{\mathcal{Y}^{x(i)}_{x(i)}(j(i,x))}{(1-p)^{x(i)+j(i,x)-1}}\end{aligned} \tag{B.20}$$

and $x(i)$ and $j(i,x)$ from (B.16) and (B.17), respectively. By using (B.18), (B.9) can be rewritten as

$$\begin{aligned}o_1 &= \sum_{j=0}^{\infty} o_j Pr\{D = Q+j-1\} \\ &= o_0 Pr\{D = Q-1\} + o_1 \sum_{j=1}^{Q} (1-p)^{1-j} Pr\{D = Q+j-1\}\end{aligned}$$

$$+\sum_{j=Q+1}^{\infty}\left[v_{S+1}\mathcal{G}(j)-v_S\mathcal{H}(j)\right]Pr\{D=Q+j-1\}$$

Solving for o_1 gives

$$o_1=o_0\frac{\left[Pr\{D=Q-1\}-\sum_{j=Q+1}^{\infty}\mathcal{H}(j)Pr\{D=Q+j-1\}\right]}{1-\left[\sum_{j=1}^{Q}(1-p)^{1-j}Pr\{D=Q+j-1\}+\sum_{j=Q+1}^{\infty}\mathcal{G}(j)Pr\{D=Q+j-1\}\right]}. \tag{B.21}$$

Furthermore,

$$\sum_{j=0}^{\infty}o_j=1$$

$$o_0+o_1\sum_{j=1}^{Q}(1-p)^{1-j}+\sum_{j=Q+1}^{\infty}\left[o_1\mathcal{G}(j)-o_0\mathcal{H}(j)\right]=1$$

$$\Rightarrow o_0=\frac{1-o_1\left(\sum_{j=1}^{Q}(1-p)^{1-j}+\sum_{j=Q+1}^{\infty}\mathcal{G}(j)\right)}{1-\sum_{j=Q+1}^{\infty}\mathcal{H}(j)}. \tag{B.22}$$

Inserting (B.22) into (B.21) yields

$$o_1=\left(\frac{1}{1-\sum_{j=Q+1}^{\infty}\mathcal{H}(j)}-o_1\frac{\sum_{j=1}^{Q}(1-p)^{1-j}+\sum_{j=Q+1}^{\infty}\mathcal{G}(j)}{1-\sum_{j=Q+1}^{\infty}\mathcal{H}(j)}\right)$$
$$\cdot\frac{\left[Pr\{D=Q-1\}-\sum_{j=Q+1}^{\infty}\mathcal{H}(j)Pr\{D=Q+j-1\}\right]}{1-\left[\sum_{j=1}^{Q}(1-p)^{1-j}Pr\{D=Q+j-1\}+\sum_{j=Q+1}^{\infty}\mathcal{G}(j)Pr\{D=Q+j-1\}\right]} \tag{B.23}$$

$$\Rightarrow o_1=$$
$$\frac{1}{1-\sum_{j=Q+1}^{\infty}\mathcal{H}(j)}$$

$$\cdot \frac{\left[Pr\{D = Q-1\} - \sum_{j=Q+1}^{\infty} \mathcal{H}(j) Pr\{D = Q+j-1\}\right]}{1 - \left[\sum_{j=1}^{Q} (1-p)^{1-j} Pr\{D = Q+j-1\} + \sum_{j=Q+1}^{\infty} \mathcal{G}(j) Pr\{D = Q+j-1\}\right]}$$

$$\cdot \frac{1}{1 + \frac{\sum_{j=1}^{Q}(1-p)^{1-j} + \sum_{j=Q+1}^{\infty} \mathcal{G}(j)}{1-\sum_{j=Q+1}^{\infty} \mathcal{H}(j)} \frac{\left[Pr\{D=Q-1\} - \sum_{j=Q+1}^{\infty} \mathcal{H}(j) Pr\{D=Q+j-1\}\right]}{1-\left[\sum_{j=1}^{Q}(1-p)^{1-j} Pr\{D=Q+j-1\} + \sum_{j=Q+1}^{\infty} \mathcal{G}(j) Pr\{D=Q+j-1\}\right]}}. \tag{B.24}$$

Given o_1, o_0 can be determined via (B.22) and all other state probabilities via (B.7) and (B.18).

□

B.5 Lemma 3.3.3.6

Proof:
If the critical fractile, $\frac{b}{b+h}$ (α-target service level), is such that $B^{s^*}(\delta^s) \geq \mathbb{E}\left[\check{D}(\delta^s)\right] = \left((L^f + 1) + \delta^s L^{\Delta}\right)\mu$ (see (3.145)), this is called a situation with **positive** safety stock. For this case, the following properties can be established:

1. The expected on-hand stock (3.147) is strictly increasing in δ^s, which follows from Lemmata B.5.0.1 and B.5.0.2.

2. The expected backorders (3.148) are strictly increasing in δ^s, which follows from Lemma B.5.0.3.

3. The procurement cost term in TRC_{OSP} is strictly decreasing in δ^s, which is obvious.

Given these properties, it follows immediately that the TRC_{OSP} function is unimodal in δ^s. The lemmata can be derived as follows. (3.147) can be rewritten as

$$\mathbb{E}\left[OH(B^{s^*}(\delta^s))\right] = B^{s^*}(\delta^s) - \mathbb{E}\left[\min\left\{\check{D}(\delta^s), B^{s^*}(\delta^s)\right\}\right] \quad . \tag{B.25}$$

(In the following, the expressions refer to a continuous demand distribution. In case of a discrete demand distribution the integral has to be replaced by a sum.)

It is assumed that the demand random variable is from a non-negative strongly unimodal distribution, i.e. a distribution that is still unimodal after convolution. This holds for all distributions studied in this thesis. In order to prove the positive increase of the expected on-hand stock as δ^s increases, first the behavior in δ^s of the two terms in equation (B.25) is considered separately and afterwards it is shown that the difference between these two terms is always positive as δ^s increases. To this end, define for $\delta_2^s > \delta_1^s$

$$B_\Delta^{s^*} = B^{s^*}(\delta_2^s) - B^{s^*}(\delta_1^s) \tag{B.26}$$

$$\mathbb{E}_\Delta\left[\min\left\{\check{D}, B^{s^*}\right\}\right] = \mathbb{E}\left[\min\left\{\check{D}(\delta_2^s), B^{s^*}(\delta_2^s)\right\}\right] - \mathbb{E}\left[\min\left\{\check{D}(\delta_1^s), B^{s^*}(\delta_1^s)\right\}\right] \tag{B.27}$$

$$\delta_\Delta^s = \delta_2^s - \delta_1^s \tag{B.28}$$

i.e. $B_\Delta^{s^*}$ denotes the change in the optimal order-up-to level (the first term in equation (B.25)), whereas $\mathbb{E}_\Delta[\min\{\check{D}, B^{s^*}\}]$ denotes the change in the second term in equation (B.25) as δ^s increases from δ_1^s to δ_2^s.

Lemma B.5.0.1 *If the critical fractile assumes a value such that a* **positive** *safety stock is required, the change in the optimal order-up-to level $B_\Delta^{s^*}$ due to an increase in δ^s of δ_Δ^s is larger than $\delta_\Delta^s L^\Delta \mu$, but approaches $\delta_\Delta^s L^\Delta \mu$ from above.*

Proof:
From (3.145) and (3.146) it follows that

$$CV(\delta^s) = \frac{\sqrt{\text{VAR}\left[\check{D}(\delta^s)\right]}}{\mathbb{E}\left[\check{D}(\delta^s)\right]} = \frac{\sigma \cdot \sqrt{(L^f+1) + [\delta^s]^2 L^\Delta}}{\left((L^f+1) + \delta^s L^\Delta\right)\mu} \quad . \tag{B.29}$$

Therefore,

$$\lim_{\delta^s \to \infty} CV(\delta^s) = 0 \quad . \tag{B.30}$$

This means that as δ^s increases, the coefficient of variation decreases. The order-up-to level is the sum of the pipeline inventory and the safety stock. The pipeline inventory is a linear function of δ^s, $\delta^s L^\Delta \mu$. The safety stock is sized to cover against the variability of lead-time demand, i.e. it depends on $\sigma \cdot \sqrt{(L^f + 1) + [\delta^s]^2 L^\Delta}$. Due to the square root effect, the additional amount of (positive) safety stock required becomes smaller as δ^s gets larger, i.e. the safety stock and thus optimal order-up-to level is concave in δ^s. Ultimately (i.e. $\delta^s \to \infty$), for a δ^s-increase of δ^s_Δ only an amount of $\delta^s_\Delta L^\Delta \mu$, i.e. no additional safety stock at all needs to be added to the previously optimal order-up-to level, $B^{s^*}(\delta^s_1)$, in order to arrive at the new optimal one, $B^{s^*}(\delta^s_2)$, and still comply with the critical fractile, i.e.

$$\lim_{\delta^s \to \infty} B^{s^*}_\Delta = \delta^s_\Delta L^\Delta \mu \quad . \tag{B.31}$$

□

Lemma B.5.0.2 *If the critical fractile assumes a value such that a* **positive** *safety stock is required, the change of* $\mathbb{E}\left[\min\left\{\check{D}(\delta^s), B^{s^*}(\delta^s)\right\}\right]$ *due to an increase in* δ^s *of* δ^s_Δ *is smaller than* $\delta^s_\Delta L^\Delta \mu$.

Proof:

Define

$$\check{D}_\Delta(\delta^s_\Delta) = \delta^s_\Delta D(L^\Delta) \quad . \tag{B.32}$$

Then, $\mathbb{E}_\Delta\left[\min\left\{\check{D}, B^{s^*}\right\}\right] < \delta^s_\Delta L^\Delta \mu$ can be rewritten as

$$\begin{aligned} &\mathbb{E}\left[\min\left\{\check{D}(\delta^s_1) + \check{D}(\delta^s_\Delta), B^{s^*}(\delta^s_1) + B^{s^*}_\Delta\right\}\right] - \mathbb{E}\left[\min\left\{\check{D}(\delta^s_1), B^{s^*}(\delta^s_1)\right\}\right] \\ &\qquad < \delta^s_\Delta L^\Delta \mu \\ &B^{s^*}_\Delta + \mathbb{E}\left[\min\left\{\check{D}(\delta^s_1) + \check{D}_\Delta(\delta^s_\Delta) - B^{s^*}_\Delta, B^{s^*}(\delta^s_1)\right\}\right] - \mathbb{E}\left[\min\left\{\check{D}(\delta^s_1), B^{s^*}(\delta^s_1)\right\}\right] \\ &\qquad < \delta^s_\Delta L^\Delta \mu. \end{aligned} \tag{B.33}$$

Rearranging terms, yields

$$\begin{aligned} &\mathbb{E}\left[\min\left\{\check{D}(\delta^s_1), B^{s^*}(\delta^s_1)\right\}\right] - \mathbb{E}\left[\min\left\{\check{D}(\delta^s_1) + \check{D}_\Delta(\delta^s_\Delta) - B^{s^*}_\Delta, B^{s^*}(\delta^s_1)\right\}\right] \\ &\qquad > B^{s^*}_\Delta - \delta^s_\Delta L^\Delta \mu. \end{aligned} \tag{B.34}$$

Define $Y := \check{D}(\delta_1^s)$, i.e. the random variable of the first (partial) expectation expression on the left-hand side. Also, let $Z := \check{D}(\delta_1^s) + \check{D}_\Delta(\delta_\Delta^s) - B_\Delta^{s^*}$ denote the random variable expression of the second (partial) expectation expression. Then,

$$\mu_Y = \left((L^f + 1) + \delta_1^s L^\Delta\right) \mu \tag{B.35}$$

$$\sigma_Y = \sigma \cdot \sqrt{(L^f + 1) + [\delta_1^s]^2 L^\Delta} \tag{B.36}$$

and

$$\mu_Z = \left((L^f + 1) + \delta_1^s L^\Delta\right) \mu + \delta_\Delta^s L^\Delta \mu - B_\Delta^{s^*} = \left((L^f + 1) + \delta_1^s L^\Delta\right) \mu - \left(B_\Delta^{s^*} - \delta_\Delta^s L^\Delta \mu\right) \tag{B.37}$$

$$\sigma_Z = \sigma \cdot \sqrt{(L^f + 1) + [\delta_2^s]^2 L^\Delta} \quad . \tag{B.38}$$

If there was no bound in the calculation of the expectations (due to the min-expression), the left-hand side in inequality (B.34) would equal the right-hand side, because

$$\left((L^f + 1) + \delta_1^s L^\Delta\right) \mu - \left[\left((L^f + 1) + \delta_1^s L^\Delta\right) \mu - \left(B_\Delta^{s^*} - \delta_\Delta^s L^\Delta \mu\right)\right] = B_\Delta^{s^*} - \delta_\Delta^s L^\Delta \mu \quad , \tag{B.39}$$

i.e. $\mathbb{E}_\Delta\left[\min\left\{\check{D}, B^{s^*}\right\}\right] = \delta_\Delta^s L^\Delta \mu$. However, due to the bound, the left-hand side in inequality (B.34) is larger than the right-hand side for the following reason. The bound in the min-expression leads to a lowering of the value of these expressions, i.e. $\mathbb{E}\left[\min\left\{\check{D}(\delta_1^s), B^{s^*}(\delta_1^s)\right\}\right] < \mu_Y$ and $\mathbb{E}\left[\min\left\{\check{D}(\delta_1^s) + \check{D}_\Delta(\delta_\Delta^s) - B_\Delta^{s^*}, B^{s^*}(\delta_1^s)\right\}\right] < \mu_Z$. For inequality (B.34) to be true, it must hold that the lowering of the second min-expression is larger than the first one. Formally, this can be stated as follows.

Let X denote a random variable. Further, note that the bound in both min-expressions is identical, namely $B^{s^*}(\delta_1^s)$. Recall that the expectation of any *non-negative* random variable can be calculated as

$$\mathbb{E}[X] = \int_0^\infty [1 - F(x)]\, dx \quad . \tag{B.40}$$

Then,

$$\mathbb{E}\left[\min\left\{X, B^{s^*}(\delta_1^s)\right\}\right] = \int_0^{B^{s^*}(\delta_1^s)} [1 - F(x)]\ dx \tag{B.41}$$

$$= \int_0^{\infty} [1 - F(x)]\ dx - \int_{B^{s^*}(\delta_1^s)}^{\infty} [1 - F(x)]\ dx \tag{B.42}$$

$$= \mathbb{E}[X] - \int_{B^{s^*}(\delta_1^s)}^{\infty} [1 - F(x)]\ dx \quad . \tag{B.43}$$

That means, the lowering is given by

$$\int_{B^{s^*}(\delta_1^s)}^{\infty} [1 - F(x)]\ dx \tag{B.44}$$

and it remains to be shown that

$$\int_{B^{s^*}(\delta_1^s)}^{\infty} [1 - F_{\mu_Y,\sigma_Y}(x)]\ dx < \int_{B^{s^*}(\delta_1^s)}^{\infty} [1 - F_{\mu_Z,\sigma_Z}(x)]\ dx \quad . \tag{B.45}$$

If the distributions of both Y and Z had the same standard deviation, their probability density and cumulative distribution functions would look the same only shifted by the difference between their means. However, since Z has a higher standard deviation than Y, its density function is flatter. Consequently, the slope of the distribution function of Z is larger than that of Y for small values, but decreases for larger ones. From the fact that both distribution functions return the same cumulative probability mass for $B^{s^*}(\delta_1^s)$ it follows that they intersect at this point. Consequently, for lower values than $B^{s^*}(\delta_1^s)$, the function F_{μ_Y,σ_Y} is below F_{μ_Z,σ_Z} and therefore returns lower distribution function values and vice versa (see Figure B.1), i.e.

$$F_{\mu_Y,\sigma_Y}(x) \begin{cases} < F_{\mu_Z,\sigma_Z}(x) & \text{for } x < B^{s^*}(\delta_1^s) \\ = F_{\mu_Z,\sigma_Z}(x) & \text{for } x = B^{s^*}(\delta_1^s) \\ > F_{\mu_Z,\sigma_Z}(x) & \text{for } x > B^{s^*}(\delta_1^s) \end{cases} \quad . \tag{B.46}$$

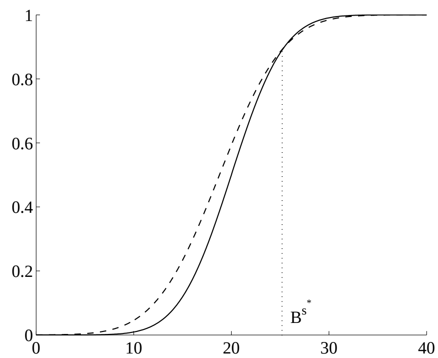


Figure B.1: Distribution function F_{μ_Y,σ_Y} (solid) and F_{μ_Z,σ_Z} (dashed)

Hence, it follows that

$$\int_{B^{s*}(\delta_1^s)}^{\infty} \left[1 - F_{\mu_Y,\sigma_Y}(x)\right] \, dx < \int_{B^{s*}(\delta_1^s)}^{\infty} \left[1 - F_{\mu_Z,\sigma_Z}(x)\right] \, dx \quad . \tag{B.47}$$

Consequently, the deviation of the partial expectation from the expectation is larger in case of F_{μ_Z,σ_Z} than it is in case of F_{μ_Y,σ_Y}. Since the difference between the (unbounded) expectations μ_Y and μ_Z is already equal to $B_\Delta^{s^*} - \delta_\Delta^s L^\Delta \mu$ and thus only gets larger through the lowering, inequality (B.34) is true. □

Lemma B.5.0.3 *If the critical fractile assumes a value such that a* **positive** *safety stock is required, the change in the expected backorders,* $\mathbb{E}_\Delta[BO]$*, due to an increase in* δ^s *of* δ_Δ^s *is positive.*

Proof:

From relation (2.27) it follows that

$$\begin{aligned} \mathbb{E}[BO] &= \mathbb{E}[OH] - \mathbb{E}[NS] \\ &= \mathbb{E}[OH] - SST \\ &= B^{s^*}(\delta^s) - \mathbb{E}\left[\min\left\{\check{D}(\delta^s), B^{s^*}(\delta^s)\right\}\right] - B^{s^*}(\delta^s) + \mathbb{E}\left[\check{D}(\delta^s)\right] \\ &= \mathbb{E}\left[\check{D}(\delta^s)\right] - \mathbb{E}\left[\min\left\{\check{D}(\delta^s), B^{s^*}(\delta^s)\right\}\right] \quad . \end{aligned} \tag{B.48}$$

In terms of change as δ^s increases, the following relation results

$$\mathbb{E}_\Delta\left[BO\right] = \delta^s_\Delta L^\Delta \mu - \mathbb{E}_\Delta\left[\min\left\{\check{D}, B^{s^*}\right\}\right] \quad . \tag{B.49}$$

From Lemma B.5.0.2 it follows that $\mathbb{E}\left[\min\left\{\check{D}(\delta^s), B^{s^*}(\delta^s)\right\}\right] < \delta^s_\Delta L^\Delta \mu$. Consequently, $\mathbb{E}_\Delta\left[BO\right] \geq 0$. □

This completes the entire proof.

□

B.6 Lemma 3.3.3.7

$TRC^{norm}_{OSP}(\delta^s)$ is convex in δ^s, if $\frac{\partial^2 TRC^{norm}_{OSP}}{\partial[\delta^s]^2} \geq 0$.

Proof:

For ease of presentation, define

$$r_{OH} = h \cdot \sigma\left(k\Phi(k) + \phi(k)\right) \tag{B.50}$$

$$r_{BO} = b \cdot \sigma\left(\phi(k) - k(1 - \Phi(x))\right) \quad , \tag{B.51}$$

which are both non-negative for all feasible values of k. Then,

$$\begin{aligned}\frac{\partial TRC^{norm}_{OSP}}{\partial \delta^s} &= r_{OH} \cdot \left((L^f + 1) + [\delta^s]^2 L^\Delta\right)^{-\frac{1}{2}} \delta^s L^\Delta \\ &\quad + r_{BO} \cdot \left((L^f + 1) + [\delta^s]^2 L^\Delta\right)^{-\frac{1}{2}} \delta^s L^\Delta - c \cdot \mu\end{aligned} \tag{B.52}$$

$$\begin{aligned}\frac{\partial^2 TRC^{norm}_{OSP}}{\partial[\delta^s]^2} =& - r_{OH}\left((L^f + 1) + [\delta^s]^2 L^\Delta\right)^{-\frac{3}{2}} [\delta^s]^2 [L^\Delta]^2 \\ &+ r_{OH}\left((L^f + 1) + [\delta^s]^2 L^\Delta\right)^{-\frac{1}{2}} L^\Delta \\ &- r_{BO}\left((L^f + 1) + [\delta^s]^2 L^\Delta\right)^{-\frac{3}{2}} [\delta^s]^2 [L^\Delta]^2 \\ &+ r_{BO}\left((L^f + 1) + [\delta^s]^2 L^\Delta\right)^{-\frac{1}{2}} L^\Delta\end{aligned}$$

$$= r_{OH}\left((L^f+1)+[\delta^s]^2L^\Delta\right)^{-\frac{1}{2}}L^\Delta\cdot\left[1-\frac{[\delta^s]^2L^\Delta}{(L^f+1)+[\delta^s]^2L^\Delta}\right]$$
$$+r_{BO}\left((L^f+1)+[\delta^s]^2L^\Delta\right)^{-\frac{1}{2}}L^\Delta\cdot\left[1-\frac{[\delta^s]^2L^\Delta}{(L^f+1)+[\delta^s]^2L^\Delta}\right] \tag{B.53}$$

The first factor of both summands is obviously non-negative. For the whole expression to be non-negative, the second factor in parenthesis needs to be non-negative, too. This is true, if the denominator is larger than the nominator.

$$[\delta^s]^2L^\Delta \overset{!}{\leq} (L^f+1)+[\delta^s]^2L^\Delta \tag{B.54}$$
$$\Leftrightarrow 0 \leq (L^f+1) \tag{B.55}$$

This condition is always met, because the lead times are assumed to be non-negative.
□

B.7 Derivation of optimal δ^s

In case of $\omega_1=\omega_2=0$, (3.164) can be rewritten as

$$c\cdot\mu\cdot\left((L^f+1)+[\delta^s]^2L^\Delta\right)^{\frac{1}{2}}=\delta^sL^\Delta(r_{OH}+r_{BO}) \qquad |\,()^2 \tag{B.56}$$
$$(c\cdot\mu)^2\left((L^f+1)+[\delta^s]^2L^\Delta\right)=[\delta^s]^2\left(L^\Delta(r_{OH}+r_{BO})\right)^2 \quad . \tag{B.57}$$
$$\Rightarrow\ [\delta^s]^2=\frac{L^f+1}{\left(\frac{L^\Delta(r_{OH}+r_{BO})}{c\cdot\mu}\right)^2-L^\Delta} \tag{B.58}$$
$$\Rightarrow\ \delta^s=\sqrt{\frac{L^f+1}{\left(\frac{L^\Delta(r_{OH}+r_{BO})}{c\cdot\mu}\right)^2-L^\Delta}} \tag{B.59}$$

The feasible region for δ^s has not been taken into account so far. This is done by the conditions specified in (3.167). In case the unconstraint global optimum is negative, (in which case (B.59) would not produce any solution) the best feasible value is $\delta^s=0$. If the unconstraint global optimum is larger than 1, the best feasible value is $\delta^s=1$.

B.8 Lemma 4.2.3.3

Proof:
(4.26) can be rewritten as

$$\begin{aligned}
C_i &= h_i T_i \mu + h_i \mathbb{E}\left[OH_i\right] + c_i^{OF} \mathbb{E}\left[BO_i\right] - h_i \mathbb{E}\left[BO_i\right] \\
&= h_i T_i \mu + h_i \mathbb{E}\left[SST_i\right] + c_i^{OF} \mathbb{E}\left[BO_i\right] \\
&\overset{(4.17)}{=} h_i T_i \mu + h_i \left(B_i(\tau_i) - \tau_i \mu\right) + c_i^{OF} \mathbb{E}\left[BO_i\right] \\
&= h_i \mu \left(T_i - \tau_i\right) + h_i B_i(\tau_i) + c_i^{OF} \mathbb{E}\left[BO_i\right] \\
&= h_i \mu \left(T_i - \tau_i\right) + h_i B_i(\tau_i) + c_i^{OF} \int_{B_i(\tau_i)}^{\infty} \left(x - B_i(\tau_i)\right) f_{\tau_i}(x)\, dx \quad . \qquad \text{(B.60)}
\end{aligned}$$

For ease of presentation, the dependency of B_i on τ_i is not explicitly indicated in the further analysis.

$$\frac{\partial C_i}{\partial B_i} = h_i - c_i^{OF}\left(1 - F_{\tau_i}(B_i)\right) \qquad \text{(B.61)}$$

$$\frac{\partial^2 C_i}{\partial B_i^2} = c_i^{OF} f_{\tau_i}(B_i) \qquad \geq 0 \qquad \text{(B.62)}$$

The optimal B_i can be determined by setting the first derivative equal to 0.

$$\frac{\partial C_i}{\partial B_i} = 0 \;\Leftrightarrow\; B_i^*(\tau_i) = F_{\tau_i}^{-1}\left(1 - \frac{h_i}{c_i^{OF}}\right) \qquad \text{(B.63)}$$

which is feasible since $c_i^{OF} \geq h_i$. □

B.9 Lemma 4.2.3.4

Proof:
From (2.50) and (2.52) it is known that for an α-service level constraint it holds that

$$B_i^*(\tau_i) = F_{\tau_i}^{-1}\left(\alpha_i^{target}\right) \quad . \qquad \text{(B.64)}$$

From (4.28) and (B.64), (4.29) follows immediately. □

B.10 Lemma 4.2.3.6

Proof:

$$\begin{aligned}\frac{\partial C_i}{\partial \tau_i} &= h_i\mu + h_i\frac{k_i\sigma}{2\sqrt{\tau_i}} + c_i^{OF}\left(\phi(k_i) - k_i(1-\Phi(k_i))\right)\frac{\sigma}{2\sqrt{\tau_i}} \\ &= \underbrace{h_i\mu}_{\geq 0} + \underbrace{\frac{\sigma}{2\sqrt{\tau_i}}}_{\geq 0}\underbrace{\big(h_ik_i + c_i^{OF}\left(\phi(k_i) - k_i(1-\Phi(k_i))\right)\big)}_{x}\end{aligned} \tag{B.65}$$

which is positive, if x is positive, i.e.

$$\begin{aligned}\big(h_ik_i + c_i^{OF}\left(\phi(k_i) - k_i(1-\Phi(k_i))\right)\big) &> 0 \\ \frac{h_i}{c_i^{OF}}k_i + \phi(k_i) - k_i(1-\Phi(k_i)) &> 0 \\ k_i\left(\frac{h_i}{c_i^{OF}} - (1-\Phi(k_i))\right) &> -\phi(k_i) \quad .\end{aligned} \tag{B.66}$$

With $k_i = \Phi^{-1}(1 - h_i/c_i^{OF})$ from (4.30) it follows that

$$\begin{aligned}\Phi^{-1}\left(1-\frac{h_i}{c_i^{OF}}\right)\left(\frac{h_i}{c_i^{OF}} - \left(1-\Phi\left(\Phi^{-1}\left(1-\frac{h_i}{c_i^{OF}}\right)\right)\right)\right) &> -\phi\left(\Phi^{-1}\left(1-\frac{h_i}{c_i^{OF}}\right)\right) \\ \underbrace{\Phi^{-1}\left(1-\frac{h_i}{c_i^{OF}}\right)}_{>0}\underbrace{\left(\frac{h_i}{c_i^{OF}} - \left(1-1+\frac{h_i}{c_i^{OF}}\right)\right)}_{>0} &> \underbrace{-\phi\left(\Phi^{-1}\left(1-\frac{h_i}{c_i^{OF}}\right)\right)}_{<0} \quad ,\end{aligned} \tag{B.67}$$

which is true.

$$\frac{\partial^2 C_i}{\partial \tau_i^2} = h_i \left(-\frac{k_i \sigma}{4\tau_i \sqrt{\tau_i}} \right) + c_i^{OF} \left(-\frac{\sigma}{4\tau_i \sqrt{\tau_i}} \right) \left(\phi(k_i) - k_i(1 - \Phi(k_i)) \right)$$

$$= \underbrace{-\frac{\sigma}{4\tau_i \sqrt{\tau_i}}}_{<0} \underbrace{\left(h_i k_i + c_i^{OF} \left(\phi(k_i) - k_i(1 - \Phi(k_i)) \right) \right)}_{>0 \text{ (see above)}} \leq 0 \quad \text{(B.68)}$$

□

B.11 Property 4.4.2.1

Proof:

If it is optimal for the final stage of the GS subnetwork not to hold any stock, i.e. $\tau_{i-1} = 0$, this stock allocation will also be achieved in a different partitioning pattern consisting of a GS subnetwork that runs until stage $i-2$ and an SS subnetwork that comprises of stages $i-1$ to j. In the new partitioning pattern the SS approach can also choose not to place any stock at stage $i-1$, if this is cost-optimal. □

B.12 Property 4.4.2.3

Proof:

$\tau_i = 0$ means that the first GS stage does not hold any stock. If this stock allocation pattern is cost-optimal, it will also be found by optimizing an HS system that consists of an SS subnetwork from l to i and a GS subnetwork from $i+1$ to j. □

B.13 Lemma 4.4.2.5

Proof:

Part (1) is obvious. In order to prove part (2) first note that in the SS approach, the relevant timespan, with regard to which the order-up-to level at a stage is sized, is the processing time. Since it is assumed that the external customer demand needs to be satisfied immediately, the outgoing service time of the HS stage is 0, i.e. the net replenishment time of the comprised GS stage corresponds to its processing time. Further, recall from Section 4.4.2.1 that the order-up-to levels within the HS stage are determined as if all comprised stages belonged to a pure SS subnetwork and the relevant timespan of the final stage is its net replenishment time, which in this case is equal to its processing time. Proof of part (3): From part (1) it follows that the supply chain must consist of at least two stages in order to allow for an HS stage that differs from a GS stage. Part (2) postulates that there must be at least one additional GS stage that either precedes or succeeds an HS stage (consisting of at least two stages) in order to differentiate the solution from a pure SS network. In such a three-stage HS serial supply chain, the *allocation benefit* can be exploited within the HS stage. Furthermore, the *decoupling benefit* of the comprised GS stage (within the HS stage) can be exploited towards a succeeding GS stage or the *decoupling benefit* of a preceding GS stage can be exploited towards a succeeding HS stage. Hence, an HS solution can be superior to the pure approaches. □

B.14 Lemma 4.5.2.1

Proof:

It is sufficient to consider the terms of the cost function that refer to a stockpoint i and its two suppliers, s and f. For ease of presentation and w.l.o.g., it is assumed that the suppliers have only a single supplier themselves and therefore a single incoming service time SI_s and SI_f. Then, the total cost function for case 2 is given as

$$C^{\mathbf{P}^2} = h_s \cdot SST_s^2(SI_s, ST_i^s, \delta_i^s, \alpha_s^{target}) + h_f \cdot SST_f^2(SI_f, ST_i^f, 1-\delta_i^s, \alpha_f^{target}) \\ + h_i \cdot SST_i^2(ST_i^f, ST_i^s, ST_i, \delta_i^s, \alpha_i^{target}) \quad . \tag{B.69}$$

From (4.83) it is obvious that the safety stock quantity depends on the variability of the random variable $\check{D}^2$ defined in (4.81). Since ST_i^f does not impact the safety stock quantity at the slow supplier, this safety stock expression can be excluded from further analysis. For stockpoint f, we get

$$\text{VAR}\left[\check{D}_f^2(SI_f, ST_i^f, 1-\delta_i^s)\right] = (1-\delta_i^s)^2 \left(SI_f + T_f - ST_i^f\right)\sigma_i^2 \quad , \tag{B.70}$$

which becomes smaller as ST_i^f increases. Similarly, for stockpoint i,

$$\begin{aligned}\text{VAR}&\left[\check{D}_i^2(ST_i^f, ST_i^s, ST_i, \delta_i^s)\right] = \\ &\left([\delta_i^s]^2\left(ST_i^s + T_i^s - ST_i\right) + (1-\delta_i^s)^2\left(ST_i - ST_i^f - T_i^f\right)\right)\sigma_i^2 = \\ &\left([\delta_i^s]^2\left(ST_i^s + T_i^s - ST_i\right) + (1-\delta_i^s)^2 ST_i - (1-\delta_i^s)^2\left(ST_i^f + T_i^f\right)\right)\sigma_i^2,\end{aligned} \tag{B.71}$$

which also becomes smaller as ST_i^f increases. Consequently, the entire safety stock quantity and thus total cost decrease as ST_i^f increases.

For normally distributed demand, this effect can be shown be computing the first derivative with respect to ST_i^f. For simplicity reasons, define $r_m = h_m k_m \sigma_m$ for $m = s, f, i$ with $\sigma_s = \delta^s \sigma_i$ and $\sigma_f = (1-\delta^s)\sigma_i$. Then,

$$\begin{aligned}C_{norm}^{\mathbf{P}^2} = {}& r_s\sqrt{SI_s + T_s - ST_i^s} + r_f\sqrt{SI_f + T_f - ST_i^f} \\ &+ r_i\sqrt{[\delta_i^s]^2\left(ST_i^s + T_i^s - ST_i\right) + (1-\delta_i^s)^2\left(ST_i - ST_i^f - T_i^f\right)}\end{aligned} \tag{B.72}$$

and

$$\begin{aligned}\frac{\partial C_{norm}^{\mathbf{P}^2}}{\partial ST_i^f} = {}& -\frac{1}{2} r_f\left(SI_f + T_f - ST_i^f\right)^{-\frac{1}{2}} \\ &- (1-\delta_i^s)^2 r_i\left([\delta_i^s]^2\left(ST_i^s + T_i^s - ST_i\right) + (1-\delta_i^s)^2\left(ST_i - ST_i^f - T_i^f\right)\right)^{-\frac{1}{2}} \qquad \text{(B.73)} \\ \le {}& 0 \quad ,\end{aligned}$$

which completes the proof. □

Bibliography

Abramowitz, M., I.A. Stegun. 1970. *Handbook of Mathematical Functions*. Dover Publications, New York.

Adan, I.J.B.F., M.J.A. Eenige, J.A.C. Resing. 1995. Fitting discrete distributions on the first two moments. *Probability in the Engineering and Informational Sciences* **9** 623–632.

Aggarwal, P.K., K. Moinzadeh. 1994. Order expedition in multi-echelon production/distribution systems. *IIE Transactions* **26**(2) 86–96.

Agrawal, N., S.A. Smith. 1996. Estimating negative binomial demand for retail inventory management with unobservable lost sales. *Naval Research Logistics* **43**(6) 839–862.

Alfredsson, P., J. Verrijdt. 1999. Modeling emergency supply flexibility in a two-echelon inventory system. *Management Science* **45**(10) 1416–1431.

Allon, G., J.A. van Mieghem. 2010. Global dual sourcing: Tailored base surge allocation to near and offshore production. *Management Science* **56**(1) 110–124.

Arts, J., M. van Vuuren, G.P. Kiesmüller. 2009. Efficient optimization of the dual-index policy using Markov Chains. Tech. rep., Beta Working Paper WP291, Eindhoven.

Axsäter, S. 1990. Simple solution procedures for a class of two-echelon inventory problems. *Operations Research* **38**(1) 64–69.

Axsäter, S. 1993. Exact and approximate evaluation of batch-ordering policies for two-level inventory systems. *Operations Research* **41**(4) 777–785.

Axsäter, S. 1998. Evaluation of installation stock based (R, Q)-policies for two-level inventory systems with Poisson demand. *Operations Research* **46**(3) 135–145.

Axsäter, S. 2000. Exact analysis of continuous review (R, Q)-policies in two-echelon inventory systems with compound poisson demand. *Operations Research* **48**(5) 686–696.

Banks, J., J.S. Carson II, B.L. Nelson, D.M. Nicol. 2009. *Discrete-event system simulation*. 5th ed. Prentice Hall.

Barankin, E.W. 1961. A delivery-lag inventory model with an emergency provision. *Naval Research Logistics Quarterly* **8**(3) 285–311.

Beckmann, M.J. 1964. Dynamic programming and inventory control. *Operational Research Society* **15**(4) 389–400.

Beyer, D., J. Ward. 2000. Network server supply chain at HP: A case study. *Hewlett Packard Software Technology Lab Report HPL 2000-84, Palo Alto* .

Bossert, J.M., S.P. Willems. 2007. A periodic review modeling approach for guaranteed service supply chains. *Interfaces* **37**(5) 420–435.

Bradley, J.R. 2004. A Brownian approximation of a production-inventory system with a manufacturer that subcontracts. *Operations Research* **52**(5) 765–784.

Bulinskaya, E. 1964. Some results concerning optimal inventory policies. *Theory of Probability and Its Applications* **9**(3) 389–403.

Burgin, T.A. 1975. The gamma distribution and inventory control. *Operational Research Quarterly* **26**(3) 507–525.

Cachon, G.P. 2001. Exact evaluation of batch-ordering inventory policies in two-echelon supply chains with periodic review. *Operations Research* **49**(1) 79–98.

Chao, X., S.X. Zhou. 2009. Optimal policy for a multiechelon inventory system with batch ordering and fixed replenishment intervals. *Operations Research* **57**(2) 377–390.

Chaudhry, M.L. 1992. Computing stationary queueing-time distributions of $GI/D/1$ and $GI/D/c$ queues. *Naval Research Logistics* **39**(7) 975–996.

Chen, F. 2000. Optimal policies for multi-echelon inventory problems with batch ordering. *Operations Research* **48**(3) 376–389.

Chen, J., D.K.J. Lin, D.J. Thomas. 2003. On the single item fill rate for a finite horizon. *Operations Research Letters* **31**(2) 119–123.

Chiang, C. 2007. Optimal control policy for a standing order inventory system. *European Journal of Operational Research* **182**(2) 695–703.

Chopra, S., P. Meindl. 2007. *Supply Chain Management: Strategy, Planning, & Operation*. 3rd ed. Pearson Education, Inc., Upper Saddle River, New Jersey.

Clark, A.J., H. Scarf. 1960. Optimal policies for a multi-echelon inventory problem. *Management Science* **6**(4) 465–490.

Clark, A.J., H. Scarf. 1962. Approximate solutions to a simple multi-echelon inventory problem. K.J. Arrow, S. Karlin, H. Scarf, eds., *Studies in Applied Probability and Management Science*. Stanford University Press, Stanford, CA, 88–100.

Cooley, J.W., J.W. Tukey. 1965. An algorithm for the machine calculation of complex Fourier series. *Math. Comput.* **19**(90) 297–301.

Croston, J.D. 1972. Forecasting and stock control for intermittent demands. *Operational Research Quarterly* **23**(3) 289–304.

Daniel, J.S.R., C. Rajendran. 2005. Determination of base-stock levels in a serial supply chain: A simulation-based simulated annealing heuristic. *International Journal of Logistics Systems and Management* **1**(2/3) 149–186.

Daniel, J.S.R., C. Rajendran. 2006. Heuristic approaches to determine base-stock levels in a serial supply chain with a single objective and with multiple objectives. *European Journal of Operational Research* **175**(1) 566–592.

Daniel, K.H. 1962. A delivery-lag inventory model with emergency order. H. Scarf, D. Gilford, M. Shelly, eds., *Multi-Stage Inventory Models and Techniques*, chap. 2. Stanford University Press, Stanford, CA.

de Kok, A.G. 1989. A moment-iteration method for approximating the waiting-time characteristics of the $GI/G/1$ queue. *Probability in the Engineering and Informational Sciences* **3**(2) 273–287.

de Kok, A.G., J.C. Fransoo. 2003. Planning supply chain operations: Definition and comparison of planning concepts. A.G. de Kok, S.C. Graves, eds., *Supply Chain Management: Design, Coordination, and Cooperation*, *Handbooks in Operations Research and Management Science*, vol. 11, chap. 12. Elsevier, Amsterdam.

de Kok, A.G., J.W.C.H. Visschers. 1999. Analysis of assembly systems with service level constraints. *International Journal of Production Economics* **59**(1-3) 313–326.

DeCroix, G. 2006. Optimal policy for a multiechelon inventory system with remanufacturing. *Operations Research* **54**(3) 532–543.

DeCroix, G., J.-S. Song, P. Zipkin. 2005. A series system with returns: Stationary analysis. *Operations Research* **53**(2) 350–362.

Diks, E.B. 1997. Controlling divergent multi-echelon systems. Ph.D. thesis, Eindhoven University of Technology.

Diks, E.B., A.G. de Kok. 1998. Optimal control of a divergent multi-echelon inventory system. *European Journal of Operational Research* **111**(1) 75–97.

Diks, E.B., A.G. de Kok. 1999. Computational results for the control of a divergent N-echelon inventory system. *International Journal of Production Economics* **59** 327–336.

Dittmar, D., S. Klosterhalfen, S. Minner. 2009. An integrated approach to safety stock optimization in serial supply chains. Tech. rep., Schumpeter School of Business and Economics, University of Wuppertal.

Dong, L., H.L. Lee. 2003. Optimal policies and approximations for a serial multiechelon inventory system with time-correlated demand. *Operations Research* **51**(6) 969–980.

Dunsmuir, W.T.M., R.D. Snyder. 1989. Control of inventories with intermittent demand. *European Journal of Operational Research* **40**(1) 16–21.

Ellis, S., K. Knickle, P. Manenti. 2009. The modern supply chain: Inventory optimization competitive assessment. URL `http://www.oracle.com/corporate/analyst/reports/industries/aim/idc-manufacturing-insights-io.pdf`.

Eppen, G.D., R.K. Martin. 1988. Determining safety stock in the presence of stochastic lead time and demand. *Management Science* **34**(11) 1380–1390.

Ettl, M., G.E. Feigin, G.Y. Lin, D.D. Yao. 2000. A supply network model with base-stock control and service requirements. *Operations Research* **48**(2) 216–232.

Federgruen, A. 1993. Centralized planning models for multi-echelon inventory systems under uncertainty. S.C. Graves, A.H.G Rinnooy Kan, P.H. Zipkin, eds., *Logistics of Production and Inventory*, *Handbooks in Operations Research and Management Science*, vol. 4, chap. 3. Elsevier, Amsterdam.

Federgruen, A., P.H. Zipkin. 1984. Computational issues in an infinite-horizon, multi-echelon inventory model. *Operations Research* **32**(4) 818–836.

Feng, Q., S.P. Sethi, H. Yan, H. Zhang. 2006a. Are base-stock policies optimal in inventory problems with multiple delivery modes? *Operations Research* **54**(4) 801–807.

Feng, Q., S.P. Sethi, H. Yan, H. Zhang. 2006b. Optimality and nonoptimality of the base-stock policy in inventory systems with multiple delivery modes. *Journal of Industrial and Management Optimization* **2**(1) 19–42.

Fox, E.J., R. Metters, J. Semple. 2006. Optimal inventory policy with two suppliers. *Operations Research* **54**(2) 389–393.

Fukuda, Y. 1964. Optimal policy for the inventory problem with negotiable leadtime. *Management Science* **10**(4) 609–708.

Gallego, G., Y. Jin, A. Muriel, G. Zhang, V. Yildiz. 2007. Optimal ordering policies with convertible lead times. *European Journal of Operational Research* **176**(2) 892–910.

Ganeshan, R. 1999. Managing supply chain inventories: A multiple retailer, one warehouse, multiple supplier model. *International Journal of Production Economics* **59**(1-3) 314–354.

Glasserman, P., S. Tayur. 1995. Sensitivity analysis for base-stock levels in multi-echelon production-inventory systems. *Management Science* **41**(2) 263–281.

Glasserman, P., S. Tayur. 1996. A simple approximation for a multistage capacitated production-inventory system. *Naval Research Logistics* **43**(1) 41–58.

Graves, S.C. 1988. Safety stocks in manufacturing systems. *Journal of Manufacturing and Operations Management* **1**(1) 67–101.

Graves, S.C., S.P. Willems. 2000. Optimizing strategic safety stock placement in supply chains. *Manufacturing & Service Operations Management* **2**(1) 68–83.

Graves, S.C., S.P. Willems. 2003. Supply chain design: Safety stock placement and supply chain configuration. A.G. de Kok, S.C. Graves, eds., *Supply Chain Management: Design, Coordination, and Operation*, chap. 3. Handbooks in Operations Research and Management Science, Elsevier, Amsterdam.

Graves, S.C., S.P. Willems. 2005. Optimizing the supply chain configuration for new products. *Management Science* **51**(8) 1165–1180.

Graves, S.C., S.P. Willems. 2008. Strategic inventory placement in supply chains: Non-stationary demand. *Manufacturing & Service Operations Management* **10**(2) 278–287.

Horst, R., H. Tuy. 1996. *Global optimization*. 3rd ed. Springer, Berlin Heidelberg New York.

Humair, S., S.P. Willems. 2006. Optimizing strategic safety stock placement in supply chains with clusters of commonality. *Operations Research* **54**(4) 725–742.

Humair, S., S.P. Willems. 2010. Optimizing strategic safety stock placement in general acyclic networks. *Operations Research (forthcoming)* .

Inderfurth, K. 1991. Safety stock optimization in multi-stage inventory systems. *International Journal of Production Economics* **24**(1-2) 103–113.

Inderfurth, K., S. Minner. 1998. Safety stocks in multi-stage inventory systems under different service measures. *European Journal of Operations Research* **106**(1) 57–73.

Janakiraman, G., J.A. Muckstadt. 2009. A decomposition approach for a class of capacitated serial systems. *Operations Research* **57**(6) 1384–1393.

Janssen, F.B.S.L.P., A.G. de Kok. 1999. A two-supplier inventory model. *International Journal of Production Economics* **59**(1-3) 395–403.

Janssen, F.B.S.L.P., R.M.J. Heuts, A.G. de Kok. 1998. On the (R, s, Q) inventory model when demand is modelled as a compound bernoulli process. *European Journal of Operational Research* **104**(3) 423–436.

Johansen, S.G., A. Thorstenson. 2008. Pure and restricted base-stock policies for the lost-sales inventory problem with periodic review and constant lead times. Tech. Rep. 42760, CORAL - Centre for Operations Research Applications in Logistics, Department of Business Studies, Aarhus School of Business, Aarhus University.

Karlin, S., H. Scarf. 1958. Inventory models of the Arrow-Harris-Marschak type with time lag. K.J. Arrow, S. Karlin, H. Scarf, eds., *Studies in the Mathematical Theory of Inventory and Production*, chap. 9. Stanford University Press, Stanford, CA, 155–178.

Kiesmüller, G.P. 2003. A new approach for controlling a hybrid stochastic manufacturing/remanufacturing system with inventories and different leadtimes. *European Journal of Operational Research* **147**(1) 62–71.

Kimball, G.E. 1988. General principles of inventory control. *Journal of Manufacturing and Operations Management* **1** 119–130.

Klemm, H. 1973. On the operating characteristic 'Service Level'. A. Prékopa, ed., *Inventory Control and Water Storage*. North-Holland, Amsterdam - London, 169–178.

Klosterhalfen, S., G.P. Kiesmüller, S. Minner. 2010a. A comparison of the constant-order and dual-index policy for dual sourcing. *International Journal of Production Economics* **accepted**.

Klosterhalfen, S., S. Minner. 2010. Safety stock optimization in distribution systems – A comparison of two competing approaches. *International Journal of Logistics: Research and Applications* **13**(2) 99–120.

Klosterhalfen, S., S. Minner, S.P. Willems. 2010b. Safety stock optimization in assembly systems with dual sourcing. Tech. rep., University of Mannheim, Mannheim.

Kwon, I.-H., S.-S. Kim, J.-G. Baek. 2006. A simulation based heuristic for serial inventory systems under fill-rate constraints. *The International Journal of Advanced Manufacturing Technology* **31**(3-4) 297–304.

Langenhoff, L.J.G., W.H.M. Zijm. 1990. An analytical theory of multi-echelon production/distribution systems. *Statistica Neerlandica* **44**(3) 149–174.

Law, A.M., W.D. Kelton. 2000. *Simulation modeling and analysis*. 3rd ed. McGraw Hill Higher Education.

Lawson, D.G., E.L. Porteus. 2000. Multistage inventory management with expediting. *Operations Research* **48**(6) 878–893.

Lee, H.L., C. Billington. 1993. Material management in decentralized supply chains. *Operations Research* **41**(5) 835–847.

Lee, H.L., M. Wolfe. 2003. Supply chain security without tears. *Supply Chain Management Review* **January-February** 12–20.

Lesnaia, E. 2004. Optimizing safety stock placement in general network supply chains. Ph.D. thesis, Sloan School of Management. URL `http://web.mit.edu/sgraves/www/LesnaiaThesis.pdf`. Accessed on 11/06/2006.

Levi, R., M. Pl, R.O. Roundy, D.B. Shmoys. 2007. Approximation algorithms for stochastic inventory control models. *Mathematics of Operations Research* **32**(2) 284–302.

Levi, R., R.O. Roundy, D.B. Shmoys, V.A. Truong. 2008. Approximation algorithms for capacitated stochastic inventory control models. *Operations Research* **56**(5) 1184–1199.

Levi, R., R.O. Roundy, V.A. Truong. 2006. Provably near-optimal balancing policies for multi-echelon stochastic inventory control models. *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*. ACM Press.

Magnanti, T.L., Z.-J.M. Shen, J. Shu, D. Simchi-Levi, C.-P. Teo. 2006. Inventory placement in acyclic supply chain networks. *Operations Research Letters* **34**(2) 228–238.

Minner, S. 1997. Dynamic programming algorithms for multi-stage safety stock optimization. *OR Spektrum* **19**(4) 261–271.

Minner, S. 2000. *Strategic Safety Stocks in Supply Chains*. Springer-Verlag, Berlin Heidelberg New York.

Minner, S. 2003. Multiple-supplier inventory models in supply chain management: A review. *International Journal of Production Economics* **81-82**(1) 265–279.

Minner, S., E.B. Diks, A.G. de Kok. 2003. A two-echelon inventory system with supply lead time flexibility. *IIE Transactions* **35**(2) 117–129.

Moinzadeh, K., P.K. Aggarwal. 1997. An information based multiechelon inventory system with emergency orders. *Operations Research* **45**(5) 694–701.

Moinzadeh, K., C.P. Schmidt. 1991. An $(S-1, S)$ inventory system with emergency orders. *Operations Research* **39**(2) 308–321.

Muckstadt, J.A. 2005. *Analysis and Algorithms for Service Parts Supply Chains*. 2nd ed. Springer.

Muckstadt, J.A., R.O. Roundy. 1993. Analysis of multistage products and locations. S.C. Graves, A.H.G Rinnooy Kan, P.H. Zipkin, eds., *Logistics of Production and Inventory*, *Handbooks in Operations Research and Management Science*, vol. 4. Elsevier, Amsterdam, 59–132.

Muckstadt, J.A., L.J. Thomas. 1980. Are multi-echelon inventory methods worth implementing in systems with low-demand-rate items? *Management Science* **26**(5) 483–494.

Muharremoglu, A., J.N. Tsitsiklis. 2003. Dynamic leadtime management in supply chains. Tech. rep.

Muharremoglu, A., J.N. Tsitsiklis. 2008. A single-unit decomposition approach to multiechelon inventory systems. *Operations Research* **56**(5) 1089–1103.

Neuts, F. 1964. An inventory model with optimal time lag. *SIAM Journal on Applied Mathematics* **12**(1) 179–185.

Parker, R.P., R. Kapuscinski. 2004. Optimal policies for a capacitated two-echelon inventory system. *Operations Research* **52**(5) 739–755.

Paterson, C., G.P. Kiesmüller, R. Teunter, K. Glazebrook. 2009. Inventory models with lateral transshipments: A review. Tech. rep., Beta Working Paper WP287, Technical University Eindhoven.

Puterman, M.L. 1994. *Markov Decision Processes*. Wiley, New York.

Rao, U., A. Scheller-Wolf, S. Tayur. 2000. Development of a rapid-response supply chain at Caterpillar. *Operations Research* **48**(2) 189–204.

Rosenshine, M., D. Obee. 1976. Analysis of a standing order inventory system with emergency orders. *Operations Research* **24**(6) 1143–1155.

Rosling, K. 1989. Optimal inventory policies for assembly systems under random demands. *Operations Research* **37**(4) 565–579.

Scheller-Wolf, A., S. Veeraraghavan, G.-J. van Houtum. 2007. Effective dual sourcing with a single index policy. Tech. rep., Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA.

Schneider, H. 1981. Effect of service-levels on order-points or order-levels in inventory models. *International Journal of Production Research* **19**(6) 615–631.

Schoenmeyr, T., S.C. Graves. 2009a. Strategic safety stocks in supply chains with capacity constraints. Tech. rep., Massachusetts Institute of Technology, Cambridge, MA.

Schoenmeyr, T., S.C. Graves. 2009b. Strategic safety stocks in supply chains with evolving forecasts. *Manufacturing & Service Operations Management* **11**(4) 657–673.

Sethi, S.P., H. Yan, H. Zhang. 2003. Inventory models with fixed costs, forecast updates, and two delivery modes. *Operations Research* **51**(2) 321–328.

Shang, K.H. 2008. Note: A simple heuristic for serial inventory systems with fixed order costs. *Operations Research* **56**(4) 1039–1043.

Shang, K.H., J.-S. Song. 2006. A closed-form approximation for serial inventory systems and its application to system design. *Manufacturing & Service Operations Management* **8**(4) 394–406.

Shang, K.H., J.-S. Song. 2007. Serial supply chains with economies of scale: Bounds and approximations. *Operations Research* **55**(5) 843–853.

Shang, K.H., S.X. Zhou. 2009a. Optimal and heuristic echelon (r, nQ, T) policies in serial inventory systems with fixed costs. *Operations Research (forthcoming)* .

Shang, K.H., S.X. Zhou. 2009b. A simple heuristic for echelon (r, nQ, T) policies in serial supply chains. *Operations Research Letters (accepted)* .

Sheopuri, A., G. Janakiraman, S. Seshadri. 2010. New policies for the stochastic inventory control problem with two supply sources. *Operations Research (Articles in Advance)* .

Silver, E.A., D.P. Bischak. 2010. The exact fill rate in a periodic review base stock system under normally distributed demand. Tech. rep., Haskayne School of Business, University of Calgary, Alberta, Canada.

Silver, E.A., D.F. Pyke, R. Peterson. 1998. *Inventory Management and Production Planning and Scheduling*. 3rd, Wiley, New York.

Simchi-Levi, D., P. Kaminsky, E. Simchi-Levi. 2008. *Designing and managing the supply chain: Concepts, strategies, and case studies*. 3rd ed. McGraw-Hill/Irwin.

Simchi-Levi, D., Y. Zhao. 2005. Safety stock positioning in supply chains with stochastic lead times. *Manufacturing & Service Operations Management* **7**(4) 295–318.

Simpson, K.F. Jr. 1958. In-process inventories. *Operations Research* **6**(6) 863–873.

Song, J.S., P.H. Zipkin. 2009. Inventories with multiple supply sources and network of queues with overflow bypasses. *Management Science* **55**(3) 362–372.

Steward, W.J. 2009. *Probability, Markov Chains, Queues, and Simulation: The Mathematical Basis of Performance Modeling*. Princeton University Press, Princeton, New Jersey.

Teunter, R.H., A.A. Syntetos, M.Z. Babai. 2010. Determining order-up-to levels under periodic review for compound binomial (intermittent) demand. *European Journal of Operational Research* **203**(3) 619–624.

Thomas, D.J. 2005. Measuring item fill-rate performance in a finite horizon. *Manufacturing & Service Operations Management* **7**(1) 74–80.

Thomas, D.J., J.E. Tyworth. 2006. Pooling lead-time risk by order splitting: A critical review. *Transportation Research Part E* **42**(4) 245–257.

Tijms, H.C. 1994. *Stochastic Models: An algorithmic approach*. John Wiley & Sons, Chichester.

Tijms, H.C., H. Groenevelt. 1984. Simple approximations for the reorder point in periodic and continuous review (s, S) inventory systems with service level constraints. *European Journal of Operational Research* **17**(2) 175–190.

Trebilcock, B. 2009a. 2009: Top 20 supply chain management software suppliers. URL `http://www.mmh.com/article/356221-2009_Top_20_supply_chain_management_software_suppliers.php?q=top+20+supply+chain+software`.

Trebilcock, B. 2009b. Supply chain management: Inventory optimization delivers the goods. URL `http://www.mmh.com/blog/Company_Briefings/22809-Supply_chain_management_Inventory_optimization_delivers_the_goods.php`.

Tyworth, J.E., Y. Guo, R. Ganeshan. 1996. Inventory control under gamma demand and random lead time. *Journal of Business Logistics* **17**(1) 291–304.

van der Heijden, M.C., E.B. Diks, A.G. de Kok. 1997. Stock allocation in general multi-echelon distribution systems with (R, S) order-up-to-policies. *Int. J. Production Economics* **49**(2) 157–174.

van Houtum, G.-J., A. Scheller-Wolf, J. Yi. 2007. Optimal control of serial inventory systems with fixed replenishment intervals. *Operations Research* **55**(4) 674–687.

van Houtum, G.J., K. Inderfurth, W.H.M. Zijm. 1996. Materials coordination in stochastic multi-echelon systems. *European Journal of Operational Research* **95**(1) 1–23.

van Houtum, G.J., W.H.M. Zijm. 1991. Computational procedures for stochastic multi-echelon production systems. *International Journal of Production Economics* **23**(1-3) 223–237.

van Houtum, G.J., W.H.M. Zijm. 1997. Incomplete convolutions in production and inventory models. *OR Spectrum* **19**(2) 97–107.

van Houtum, G.J., W.H.M. Zijm. 2000. On the relation between cost and service models for general inventory systems. *Statistica Neerlandica* **54**(2) 127–147.

Veeraraghavan, S., A. Scheller-Wolf. 2008. Now or later: Dual index policies for capacitated dual sourcing systems. *Operations Research* **56**(4) 850–864.

Veinott, A.F., Jr. 1966. The status of mathematical inventory theory. *Management Science* **12**(11) 745–777.

Viswanathan, N. 2007. The supply chain innovator's technology footprint 2007 – A benchmark report on companies' technology investment plans for gaining immediate and strategic payback. URL `http://www.aberdeen.com/summary/report/benchmark/3981-RA-Supply-Chain.asp`.

Viswanathan, N. 2009. Inventory management: 3 keys to freeing working capital. URL `http://www.aberdeen.com/summary/report/benchmark/5965-RA-inventory-management-capital.asp`.

Whittemore, A.S., S.C. Saunders. 1977. Optimal inventory under stochastic demand with two supply options. *SIAM Journal on Applied Mathematics* **32**(2) 293–305.

Willems, S.P. 2008. Data set – Real-world multiechelon supply chains used for inventory optimization. *Manufacturing & Service Operations Management* **10**(1) 19–23.

Yazlali, Ö., F. Erhun. 2009. Dual-supply inventory problem with capacity limits on order sizes and unrestricted ordering costs. *IIE Transactions* **41**(8) 716–729.

Yi, J., A. Scheller-Wolf. 2003. Dual sourcing from a regular supplier and a spot market. Tech. rep., Carnegie Mellon University, Pittsburgh, PA.

Zhang, V.L. 1996. Ordering policies for an inventory system with three supply modes. *Naval Research Logistics* **43**(5) 691–708.

Zhang, V.L., W.H. Hausman. 1994. An inventory model with dual supply modes. Tech. rep., Stanford University, CA.

Zhao, Y. 2008a. Evaluation and optimization of installation base-stock policies in supply chains with compound poisson demand. *Operations Research* **56**(2) 437–452.

Zhao, Y. 2008b. Performance analysis of acyclic supply chains. Tech. rep., Rutgers University, Newark, NJ.

Zipkin, P.H. 2000. *Foundations of Inventory Management*. Jeffrey J. Shelstad, The McGraw-Hill Companies, Inc.

Ehrenwörtliche Erklärung

Ich versichere, dass ich diese Arbeit ohne Hilfe Dritter und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel angefertigt und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Mannheim, den 16. April 2010

Steffen Klosterhalfen

Curriculum Vitae

Steffen Klosterhalfen
University of Mannheim
Business School
Email: klosterhalfen@bwl.uni-mannheim.de

Education and Academic Experience

01/2009 – 04/2009	Boston University, USA Visiting researcher, School of Management, Department of Operations & Technology Management (Assoc. Prof. Sean P. Willems)
02/2006 – 12/2008	University of Mannheim, Germany Research assistant, Department of Logistics (Prof. Dr. Stefan Minner)
01/2008 – 03/2008	Eindhoven University of Technology, The Netherlands Visiting researcher, Subdepartment of Operations, Planning, Accounting and Control (Prof. Dr. Gudrun P. Kiesmüller)
10/2000 – 09/2005	University of Mannheim, Germany Studies in Business Administration (Diploma)
09/2002 – 06/2003	University of Warwick, Great Britain Studies in Business Administration as an exchange student

Work Experience

02/2006 – present	Graduate School of Economic & Social Sciences (GESS), University of Mannheim, Germany Center Manager of the Center for Doctoral Studies in Business (CDSB) of the graduate school