Efficient optimization of the Dual-Index policy using Markov Chain approximations

Arts, J.J.

Award date:
2009
Efficient optimization of the Dual-Index policy using Markov Chain approximations

by

Joachim J. Arts

BSc Industrial Engineering & Management Science — Eindhoven University of Technology 2007
Student identity number 0574754

in partial fulfilment of the requirements for the degree of

Master of Science
in Operations Management and Logistics

Supervisors:
Dr. G.P. Kiesmüller, TU/e, OPAC
Prof.dr.ir. G.J.J.A.N. van Houtum, TU/e, OPAC
Dr.ir. M. van Vuuren, CQM bv, Chain Management
Subject headings: Markov Chain, Inventory management, Dual sourcing, Stochastic lead times, Approximations
Efficient optimization of the Dual-Index policy using Markov Chain approximations

Joachim J Arts
0574754

July 27, 2009
Preface

This Master Thesis is the result of my graduation project in the Master program Operations Management & Logistics at the school of Industrial Engineering, Eindhoven University of Technology (TU/e). The project took place at the Centre for Quantitative Methods BV (CQM) from February to July 2009.

Conducting my graduation project at CQM placed me in the convenient position of having two daily supervisors, Gudrun Kiesmüller (TU/e) and Marcel van Vuuren (CQM). I am indebted to both for valuable ideas, guidance and the investment of considerable amounts of time. I thank Gudrun Kiesmüller for patiently listening to many ideas, most of which did not make it into this thesis, and providing valuable and prompt feedback. I thank Marcel van Vuuren in particular for encouraging me to model/approximate by means of a discrete time Markov chain, even though I was reluctant to pursue this approach at first. I also thank Marcel for suggesting ways to develop intuition with simulation results. I acknowledge my second supervisors Geert-Jan van Houtum (TU/e) and Jan van Doremalen (CQM) for their input relating to technical details and application to practice.

Finally, I would like to thank my family for their love and support. My mother taught me to value knowledge and education. My wife Heidi supported by encouraging me and giving up sleep to care for our son, Liam.

Joachim Arts
Eindhoven, July 2009
Abstract

We consider the inventory control of a single product in one location with two supply sources facing stochastic demand. A premium is paid for each product ordered from the faster ‘emergency’ supply source. Unsatisfied demand is backordered and ordering decisions are made periodically. The optimal control policy for this type of system is known to be complex. For this reason we study a type of base-stock policy known as the Dual-Index policy as control mechanism for this inventory system. Under this policy ordering decisions are based on a regular and an emergency inventory position and their order-up-to-levels. In each period first an order is placed with the emergency supplier so that the emergency inventory position (on-hand stock + outstanding orders arriving within the emergency lead time - backorders) meets its order-up-to-level. Next a regular order is placed to raise the regular inventory position (on-hand stock + all outstanding orders - backorders) to its order-up-to-level. Previous work on this type of policy assumes deterministic lead times and uses simulation to find the optimal order up to levels. We extend the policy to accommodate stochastic regular lead times and provide an approximate evaluation method so that optimization can be done without simulation. A numerical study shows that this approach yields excellent results for deterministic lead times and good results for stochastic lead-times. We also give managerial insights into the use of the model with deterministic lead-time, when in fact lead times are stochastic.

Keywords

Markov Chain, Inventory management, Dual sourcing, Stochastic lead times, Approximations
Summary

Research into inventory systems is mostly done under the assumption that only one supplier or supply mode exists to procure, manufacture or ship goods. While many useful results have been obtained under this assumption, these models nevertheless omit an important aspect of many real inventory systems, namely that inventories can be replenished in more than one way. For example, it is common that one item can be procured from different suppliers or manufactured in different plants. Alternatively an item may be shipped over sea or by air (expediting). Even within the production environment of a single plant the production lead time can be decreased by producing in overtime. In all these examples there are multiple ways to replenish inventory with different lead times and costs. We refer to these systems as dual-sourcing systems.

In this thesis we study a general model for the inventory control in dual-sourcing systems. Specifically we consider the inventory control of a single product in one location with two supply sources facing stochastic demand. The lead times of both sources are assumed to be integer multiples of the review period. The regular lead time can be either stochastic or deterministic while the emergency lead time is deterministic. The faster supply source will be referred to as the emergency supplier while the slower supply source will be referred to as the regular supplier. Units procured from the emergency supplier incur additional cost. Ordering from the regular channel may represent manufacturing somewhere in Asia, while ordering through the emergency channel may represent ordering from a more expensive local supplier. Other applications include, but are not limited to, shipping goods by sea (‘regular’) or air (‘emergency’) freight and manufacturing with (‘emergency’) or without (‘regular’) overtime. The problem we consider is the minimization of inventory holding and ordering costs subject to a service level constraint.

As control mechanism for this inventory system we study the Dual-Index policy (DIP). Under this policy ordering decisions are based on a regular and an emergency inventory position and their order-up-to-levels. In each period first an order is placed with the emergency supplier so that the emergency inventory position (on-hand stock + outstanding orders arrival
ing within the emergency lead time - backorders) meets its order-up-to-level. Next a regular order is placed to raise the regular inventory position (on-hand stock + all outstanding orders - backorders) to its order-up-to-level. As opposed to other order-up-to-policies in single sourcing problems, the emergency inventory position can exceed its order-up-to-level. This excess is called the overshoot. Veeraraghavan and Scheller-Wolf (2008) show that the distribution of the overshoot depends only on the difference between the regular and emergency order-up-to-level and that the optimal DIP can be obtained by solving a simple newsvendor equation once the overshoot distribution is known.

As yet however, the overshoot distribution can only be obtained through simulation. We provide an alternate proof of this separability result. An insight from this proof leads to an alternate way to model the overshoot using an exact one dimensional Markov Chain. For this one-dimensional Markov Chain we propose approximations for the transition probabilities so that it can be used to approximate the overshoot distribution. Using this approximation we can optimize the DIP without the use of extensive simulation. A numerical study shows that this approach yields excellent results for deterministic lead times.

Next we generalize the Dual-Index policy to accommodate stochastic regular lead times and show that the separability results holds also in this case, using the same method of proof as we used for deterministic lead times. We propose another one dimensional Markov Chain for this situation and provide approximations for the transition probabilities. A numerical study shows that this approach yields close to optimal solutions.

A numerical investigation of the application of the model with deterministic lead times when lead times are in fact stochastic leads to the following results:

1. It is essential to incorporate lead time variance in the model when lead time variability is large relative to demand variability, especially when meeting service level requirements is imperative;

2. Using the maximum lead time or a high percentile of the lead time distribution as deterministic lead time parameter gravely increases both costs and service and is very far from optimal;

3. When lead time variability is small compared to demand variability the model with deterministic lead times yields reasonable results provided the mean lead time is used as deterministic lead time parameter and service level requirements are not strict.
List of Tables

2.1 Summary of notations ................................................. 10
2.2 Test-bed of instances with deterministic regular lead time ......... 18
3.1 Test-bed of instances with stochastic regular lead time ............. 33
3.2 Distribution types for $L$ ............................................. 33
3.3 Accuracy of cost and service estimates for stochastic lead times .... 35
3.4 Quality of solutions compared to optimal DIP ........................ 37
4.1 Quality of solutions compared to optimal DIP for $l_{Det} = \mathbb{E}[L]$ ............................................... 41
4.2 Quality of solutions compared to optimal DIP for $l_{Det} = L_{max}$ .... 43
## List of Figures

2.1 Graphical representation of model with deterministic regular lead times . . . 8
2.2 Overshoot distributions for a few illustrative cases as determined by simulation and the corresponding Markov Chain approximations . . . . . . . . . . . . . . 18
2.3 Accuracy of Markov Chain approximation . . . . . . . . . . . . . . . . . . . 19
2.4 Quality of obtained approximate solutions . . . . . . . . . . . . . . . . . . . 20

3.1 Graphical representation of model with stochastic regular lead times . . . . 23
3.2 Types of distributions used for $L$ . . . . . . . . . . . . . . . . . . . . . . . . 34
3.3 Accuracy of Markov Chain approximation . . . . . . . . . . . . . . . . . . . 36

4.1 Comparison of models . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 40

5.1 Example of an assembly system with dual-sourcing . . . . . . . . . . . . . . 46

B.1 Examples of quasi-uniform probability mass functions . . . . . . . . . . . . . 56
Chapter 1

Introduction

Research into inventory systems is mostly done under the assumption that only one supplier or supply mode exists to procure, manufacture or ship goods. While many useful results have been obtained under this assumption (e.g. News-vendor type results for many systems, see van Houtum (2006) for an overview), these models nevertheless omit an important aspect of many real inventory systems, namely that inventories can be replenished in more than one way. For example, it is common that one item can be procured from different suppliers or manufactured in different plants. Alternatively an item may be shipped over sea or by air (expediting). Even within the production environment of a single plant the production lead time can be decreased by producing in overtime. In all these examples there are multiple ways to replenish inventory with different lead times and costs.

The situations described in the previous subparagraph can be approached in roughly two ways. The first approach is to carefully select one of the supplier/supply modes and than source all inventory from that supplier/supply mode. We refer to the problem of making this decision as vendor selection. The second approach is to use both suppliers/supply modes simultaneously. This paper is concerned with the second approach which we refer to as dual sourcing.

Suppliers are becoming more willing to offer different supply modes to their customers. Paccar parts in Eindhoven, for example, which handles spare-part logistics for DAF trucks N.V., makes a distinction between regular and emergency delivery modes for shipping parts to different locations throughout Europe. Another situation where multiple supply modes occur naturally is in remanufacturing systems. In this setting serviceable products can be produced from raw materials or from remanufacturing returned items. These two modes of inventory replenishment are naturally associated with different costs and lead-times. A
similar situation also occurs in the inventory control of spare parts. Spare parts are kept on stock so that a capital good can readily be made available upon failure of a part. The failed part is then sent into normal or emergency repair with associated different lead times and costs.

In this thesis we study a general model for the inventory control in dual-sourcing systems. We consider the inventory control of a single product in one location that is reviewed periodically and has two supply sources with different lead times. The lead times are assumed to be integer multiples of the review period. The faster supply source will be referred to as the emergency supplier while the slower supply source will be referred to as the regular supplier. Units procured from the emergency supplier incur additional cost. Ordering from the regular channel may represent manufacturing somewhere in Asia, while ordering through the emergency channel may represent ordering from a more expensive local supplier. Other applications include, but are not limited to, shipping goods by sea (‘regular’) or air (‘emergency’) freight and manufacturing with (‘emergency’) or without (‘regular’) overtime. The problem we shall consider is the minimization of holding and ordering costs subject to a service level constraint.

Models for the situations described above are difficult to analyze. Under specific restrictive assumptions the analysis can become tractable such as the assumption of a unit lead time difference for which the optimal policy is known but the application area is very narrow. When lead time differences are more than one period the optimal policy is known to be complex, difficult to implement and computationally hard to obtain (Whittmore & Saunders (1977), Feng et al. (2006a) and Feng et al. (2006b)). In this thesis we consider exactly this context. For this reason we do not study the globally optimal policy but consider a class of base-stock type policies and optimize within this class. Specifically we consider the Dual-Index policy (DIP) that has the attractive property of reducing to the optimal policy when the lead time difference is only one period. This policy, originally proposed for manufacturing systems, is easily implementable and performs very close to the optimal policy (Veeraraghaven & Scheller-Wolf (2008)). Until now the DIP has resisted analytical or even approximate analytical optimization so that resort had to be taken to simulation based procedures.

The DIP policy tracks two inventory positions: a regular inventory position (on-hand stock + all outstanding order - backlog) and an emergency inventory position (on-hand stock + outstanding orders that will arrive within the emergency lead time - backlog). In each period ordering decisions are made to raise both inventory positions to their order-up-
to-levels. Under this policy the emergency inventory position can, and indeed usually does, exceed its corresponding order-up-to-level. This excess is called the overshoot and plays a central role in the analysis of the DIP. Despite its relatively simple form, optimization of the DIP still requires substantial computational effort because it requires determining several overshoot distributions. In principle the overshoot distribution can be obtained exactly by solving a multidimensional discrete time Markov chain (DTMC). However this approach suffers from the curse of dimensionality and consequently the usual approach is to determine the overshoot distribution by simulation. Veeraraghaven & Scheller-Wolf (2008) prove a separability result that drastically decreases the amount of simulation needed, but the computational time remains substantial.

In this thesis we revisit the model compared of Veeraraghavan & Scheller-Wolf (2008), and generalize it by incorporating stochastic regular lead times. We provide an alternate proof of the aforementioned separability result for both deterministic and stochastic lead times. An insight from this proof is used to construct a one-dimensional DTMC that describes the overshoot process. By approximating the transition probabilities for this DTMC we obtain a computationally efficient optimization procedure. Because this method is computationally more efficient than simulation it can be used to incorporate dual-sourcing in more complex settings such as multi-echelon inventory systems. With the availability of models for both stochastic and deterministic lead times we also investigate the consequences of applying the model with deterministic lead times to situations where lead times are in fact stochastic.

This thesis is organized as follows. In Section 1.1 we review the literature on dual-sourcing research and position our results with respect to earlier results. We then present our model with deterministic lead times in Chapter 2. In Chapter 3 we generalize our model to accommodate stochastic regular lead times. Chapter 4 provides an extensive numerical study on the application of the model with deterministic lead times in settings where the regular lead time is in fact stochastic and provides managerial insights. Finally Chapter 5 gives conclusions and directions for further research.

1.1. Literature review

Minner (2003) provides a review of the literature pertaining to many different issues surrounding multiple supply sources. Broadly speaking the research in multiple sourcing is divided into the strategic approach, which studies issues such as exchange rate volatility, risk management and vendor selection, and the operational approach that mainly studies
the inventory control of such systems. Among the different perspectives we focus on operational/tactical control of multiple sourcing systems. One body of research focusses on the number of supply sources as a decision variable and usually assumes that different sources are identical. In these situations replenishment orders are split among the different supply sources and optimal order splitting is the object of study. Another body of research considers a situation with two suppliers that have different lead times. Replenishing inventory from the faster supplier incurs additional cost. This paper contributes to this body of research. As Minner (2003) provides an excellent review of research up to around 2001 we briefly discuss key results from before that time. Then we discuss relevant research since that time.

Early research focusses on the structure of the optimal policy for periodic inventory systems with dual-sourcing. Barankin (1961) considers the single period problem with instantaneous emergency delivery and a regular lead time of one period. Fukuda (1964) formulates the problem as one of negotiable lead-time for the infinite horizon case and gives an analytical derivation of the optimal policy by discounted dynamic programming. He considers a system that operates in discrete time, that has two suppliers whose lead-times are deterministic and differ by exactly one period. Sethi et al. (2003) extend Fukuda's (1964) model with fixed ordering costs and demand forecast updates and show that the optimal policy is of the \((s, S)\)-type. Yazlali & Erhun (2009) extend Fukuda's (1964) model with minimum and maximum capacity requirements for both suppliers, and derive the optimal policy. The assumption that the lead times of both suppliers differs by only one period is crucial to obtaining optimal policies with a simple structure. In 1977 Whittmore & Saunders and more recently Feng et al. (2006a) and Feng et al. (2006b) showed that in the optimal policy ordering decisions depend on the entire vector of outstanding orders for general lead time differences. Thus the optimal policy is complex and not of the base-stock type when the lead time difference is more than one period.

Despite the fact that the optimal policy for general lead time differences has been known to be complex since 1977, the focus on good policies with a simpler structure is rather recent. Scheller-Wolf et al. (2003) consider the same setting as Whittmore & Saunders (1977) and propose the single index policy under which ordering for both the emergency and regular supplier are based on a single state parameter: the inventory position. This policy is simple and can easily be optimized when demand distributions are mixtures of erlangian distributions. When the lead-time difference is one period the single-index policy also reduces to the optimal policy. Kiesmüller (2003) proposes the use of a policy that tracks two inventory positions associated with different lead-times in the context of a remanufacturing system. The
1.1. LITERATURE REVIEW

The key idea here is that the decision on the amount to order at the emergency supplier should not be based on information about orders that will arrive after this order. Veeraraghavan & Scheller-Wolf (2008) study this policy and call it the Dual-Index policy. They provide the aforementioned separability result for deterministic lead times. This separability result separates the optimization of the DIP, which is a two-dimensional optimization problem to two one-dimensional optimization problems.

A completely different policy for this problem setting are standing order or constant order policies. In these policies the regular supplier delivers a fixed quantity every period while the emergency supplier may be controlled using various types of policies. Recent contributions in this area are Chiang (2007) who derives the optimal policy structure given that the regular order quantity is fixed and Allon & van Mieghem (2008) who approximate the related Tailored Base Surge policy using Brownian motions.

A closely related problem is the expedition of orders after they have entered the pipeline. Lawson & Porteus (2001) study this problem in a serial multi-echelon periodic review context. They show that a type of base-stock policy, called a “top down base-stock policy” is optimal when orders can be expedited and delayed at will in the entire supply chain. Gallego et al. (2007) study a single stock-point in continuous time with the possibility of expediting existing orders and derive the optimal policy under the assumption of Poisson demand.

All literature in dual-sourcing assumes deterministic lead times except for Song & Zipkin (2009) and Gaukler et al. (2008). Song and Zipkin study a model of a stock-point facing Poisson demand operating in continuous time. They assume a $(S - 1, S)$-type ordering policy and show how to model this system as a network of queues with one or more overflow bypasses. Gaukler et al. (2008) also consider a single stock-point operating in continuous time and propose a policy based on the classical $(Q, R)$-policy. They show how to find optimal parameter settings under a set of specific assumptions.

The setting we consider is similar to the settings in Fukuda (1964), Whittmore & Saunders (1977) and Veeraraghavan & Scheller-Wolf (2008). Our two most important contributions are (i) the development of an efficient approximation for the overshoot distribution so that optimization of the DIP becomes computationally more feasible and (ii) the incorporation of stochastic lead times in the periodic review setting.
Chapter 2

Dual sourcing with deterministic lead times

Before considering dual-sourcing in the presence of stochastic lead times in section 3 we first study the special case of deterministic lead times in the present chapter. As remarked in the introduction, the optimal policy for dual sourcing systems is highly state dependent and therefore we consider the Dual-Index policy (DIP) for the control of such systems. The approach adopted for studying this policy will be used again when we study systems with stochastic lead times and this approach differs from previous approaches to analyze the dual index policy. This section is organized as follows. Our model and the operation of the DIP is introduced in section 2.1 and followed by our analysis in section 2.2. In section 2.3 we provide numerical results.

2.1. Model

Our model is similar to the model studied by Veeraraghavan & Scheller-Wolf (2008). We consider the inventory control of a single product in one location with two supply sources facing stochastic demand. A premium $c$ is paid for each product ordered from the faster ‘emergency’ supply source. Unsatisfied demand is backordered and ordering decisions are made periodically. Without loss of generality we assume the length of a review period is one. Demand per period is a sequence of non-negative i.i.d. discrete random variables \( \{ D_n \} \) with \( n \) a period index. We assume that \( \Pr(D > 0) > 0 \) and \( \Pr(D < \infty) = 1 \). The net inventory (stock on-hand - backlog) at the beginning of period \( n \) will be denoted \( I_n \). Any on-hand stock \( I_n^+ \) at the beginning of a period \( n \) incurs a holding cost of \( h \) per SKU. (We
use the standard notations \( x^+ = \max(0, x) \) and \( x^- = \max(0, -x) \). We denote the backlog at the beginning of a period \( B_n = I_n^- \). Orders placed at the regular (emergency) channel arrive after a deterministic lead-time \( l^r (l^e) \) and we assume \( l := l^r - l^e, l \geq 1 \). Lead-times are assumed to be an integer multiple of the review period. The regular (emergency) order placed in period \( n \) is denoted \( Q^r_n \) \( (Q^e_n) \). Later (in Chapter 3) we will relax the assumption that \( l^r \) is deterministic. A schematic representation of the situation described above is given in figure 2.1.

As control mechanism for this inventory system we study the Dual-Index policy (DIP) defined by two parameters \((S_e, S_r)\) which operates as follows. At the beginning of each period \( n \) we review the emergency inventory position

\[
IP^e_n = I_n + \sum_{i=n-l^r}^{n-1} Q^r_i + \sum_{i=n-l^e}^{n-1} Q^e_i
\]  

and place an emergency order \( Q^e_n \) to raise the emergency inventory position to its order-up-to-level \( S_e \),

\[
Q^e_n = (S_e - IP^e_n)^+.
\]  

After placing the emergency order we inspect the regular inventory position

\[
IP^r_n = I_n + \sum_{i=n-l^r}^{n-1} Q^r_i + \sum_{i=n-l^e}^{n-1} Q^e_i = IP^e_n + Q^e_n + \sum_{i=n+1-l}^{n-1} Q^r_i
\]

and place a regular order \( Q^r_n \) to raise the regular inventory position to its order-up-to-level \( S_r \),

\[
Q^r_n = S_r - IP^r_n.
\]

After ordering, shipments are received and demand for the period is satisfied or backordered. Thus within a period \( n \) the sequence of events can be summarized as follows:

1. review the on-hand inventory and incur holding costs \( hI^+_n \)
2. review the emergency inventory position and place an emergency order $Q^e_n$; emergency ordering costs are incurred as $cQ^e_n$

3. review the regular inventory position and place a regular order $Q^r_n$

4. receive shipments $Q^e_{n-1}$ and $Q^r_{n-1}$

5. demand $D_n$ occurs and is satisfied except for possible back-orders $B_n$.

Note that the emergency inventory position under this policy can, and indeed often does, exceed the emergency order-up-to-level. The amount by which the emergency inventory position exceeds the emergency order-up-to-level is called the overshoot. After ordering the emergency inventory position is given by $S^e_n + O_n$ where $O_n \in \{0, 1, ..., S^r - S^e\}$ is the overshoot and satisfies

$$O_n = IP^e_n + Q^e_n - S^e_n = (IP^e_n - S^e_n)^+.$$  \hspace{1cm}(2.5)

Determining the stationary distribution of the overshoot $O$ will play a key role in evaluating the performance of a given policy $(S^r, S^e)$.

Our objective will be to minimize average cost subject to a modified fill-rate constraint. The modified fill-rate is defined as

$$\gamma = 1 - E[B]/E[D].$$  \hspace{1cm}(2.6)

The modified fill-rate is closely related to the regular fill-rate often denoted $\beta$. The regular fill-rate is the fraction of demand that can be satisfied from stock on-hand. The $\gamma$ service levels also bears on the possibility that back-orders take more than a single period to be filled when a backlog does occur. When service-levels are high the difference between the regular and modified fill rate is minimal.

The average costs related to our problem are the costs of emergency ordering and holding costs given by

$$C(S^e, S^r) = hE[I^+] + cE[Q^e].$$  \hspace{1cm}(2.7)

We are now in a position to formulate the optimization problem $P$:

$$(P) \quad \text{min} \quad C(S^e, S^r)$$

$$\text{s.t.} \quad \gamma(S^e, S^r) \geq \gamma_0$$

$$S^e, S^r \in \mathbb{Z}.$$ \hspace{1cm}(2.8)

Here $\gamma_0$ denotes the target service level. The integrality constraint on $S^e$ and $S^r$ is the consequence of the discrete nature of demand. This problem is a non-linear integer programming problem (NLIP).
An overview of all introduced notations and some notations that will be introduced in later sections is given in Table 2.1.

<table>
<thead>
<tr>
<th>notation</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_n$</td>
<td>Amount of ordered products that will not arrive within the emergency lead-time in period $n$ after ordering ($:= \sum_{i=n+1}^{n+l_r} Q_i^r$)</td>
</tr>
<tr>
<td>$B_n$</td>
<td>Backlog in period $n$</td>
</tr>
<tr>
<td>$c$</td>
<td>Premium to buy one product at the emergency supplier</td>
</tr>
<tr>
<td>$C(S_e, S_r)$</td>
<td>Average holding and incremental ordering costs for policy $(S_e, S_r)$</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Modified fill rate</td>
</tr>
<tr>
<td>$\gamma_0$</td>
<td>Target modified fill-rate</td>
</tr>
<tr>
<td>$D_n$</td>
<td>Demand in period $n$, random variable</td>
</tr>
<tr>
<td>$\Delta$</td>
<td>Difference between regular and emergency order-up-to-level ($:= S_r - S_e$)</td>
</tr>
<tr>
<td>$h$</td>
<td>Inventory holding cost per period per SKU</td>
</tr>
<tr>
<td>$I_n$</td>
<td>The net inventory (on-hand stock - backlog) at the beginning of period $n$</td>
</tr>
<tr>
<td>$IP_n^r$</td>
<td>Regular inventory position at the beginning of period $n$ after ordering at the emergency supplier</td>
</tr>
<tr>
<td>$IP_n^e$</td>
<td>Emergency inventory position at the beginning of period $n$ before ordering</td>
</tr>
<tr>
<td>$l^e$</td>
<td>Replenishment lead-time for emergency orders</td>
</tr>
<tr>
<td>$l^r$</td>
<td>Replenishment lead-time (deterministic) for regular orders</td>
</tr>
<tr>
<td>$\delta$</td>
<td>Difference between regular and emergency replenishment lead-time ($:= l^r - l^e$)</td>
</tr>
<tr>
<td>$n$</td>
<td>Period index</td>
</tr>
<tr>
<td>$O_n$</td>
<td>Overshoot in period $n$ ($:= (IP_n^e - S_e)^+$)</td>
</tr>
<tr>
<td>$S_r$</td>
<td>Regular order-up-to-level</td>
</tr>
<tr>
<td>$S_e$</td>
<td>Emergency order-up-to-level</td>
</tr>
<tr>
<td>$Q_n^r$</td>
<td>Regular order quantity placed in period $n$</td>
</tr>
<tr>
<td>$Q_n^e$</td>
<td>Emergency order quantity placed in period $n$</td>
</tr>
</tbody>
</table>

### 2.2. Analysis

This section is organized as follows. In section 2.2 we present the separability result and show how it can be exploited to find the optimal DIP if the overshoot distribution can be determined. In section 2.2 we present an exact one-dimensional Discrete Time Markov Chain (DTMC) that describes the overshoot. Following in section 2.2 we provide approximations for the transition probabilities such that this DTMC can be utilized to approximate the overshoot distribution.

Throughout the analyses in this thesis for any random variable $X_n$ we define the stationary expectation and distribution as

$$E[X] = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_n \quad \text{and} \quad \Pr(X \leq x) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} I\{X_n \leq x\} \quad (2.9)$$

where $I\{x\}$ is the indicator function of the event $x$. Whenever we drop the index of a random variable we are referring to the stationary random variables with mean and distribution
2.2. ANALYSIS

defined above. Additionally we denote the \( k \)-fold convolution of a random variable \( X \) as \( X^{(k)} \) and the squared coefficient of variance of a random variable \( X \) as \( c_X^2 := \frac{\text{Var}[X]}{\mathbb{E}[X]} \).

Optimization

In our analysis we shall see that the difference between \( S_r \) and \( S_e \) plays an important role. Therefore, we define \( \Delta := S_r - S_e \). This definition allows for the specification of a DIP as either \((S_e, S_r)\) or \((S_e, S_e + \Delta)\). For the analysis it will be more convenient to consider \( S_e \) and \( \Delta \) as decision variables. In this section we show how to find the optimal \( S_e \) for fixed \( \Delta \). This allows for a simple search procedure over \( \Delta \) to find the optimal DIP.

First we investigate an interesting property of the DIP. Consider the pipeline stock that will not arrive within the emergency lead time and denote this quantity \( A_n \) in period \( n \) after ordering:

\[
A_n = \sum_{i=n+1-l}^{n} Q_r^i. \quad (2.10)
\]

Lemma 2.1. (Key functional relation) Suppose that \( IP_r^k \leq S_r \) for some \( k \in \mathbb{N}_0 \). Then for all \( n \geq k \) the Dual-Index policy ensures that the following identity holds

\[
\Delta = O_n + A_n. \quad (2.11)
\]

Proof. Reconsider the regular inventory position as given in equation (2.3),

\[
IP_r^n = IP_e^n + Q_e^n + \sum_{i=n+1-l}^{n-1} Q_r^i. \quad (2.12)
\]

Now we substitute equation (2.5) and add \( Q_r^n \) to both sides of this equation,

\[
IP_r^n + Q_r^n = S_e + O_n + \sum_{i=n+1-l}^{n} Q_r^i. \quad (2.13)
\]

By supposition \( IP_r^n \leq S_r \), so \( Q_r^n = S_r - IP_r^n \) and thus we obtain.

\[
S_r = S_e + O_n + A_n. \quad (2.14)
\]

Rearrangement and substitution of the identity \( \Delta = S_r - S_e \) yields the result. \( \square \)

The supposition that \( n \geq k \) and \( IP_r^n \leq S_r \) is no real restriction because the number of periods that \( IP_r^n \) can be greater than \( S_r \) is finite with probability 1 under the earlier assumption that the demand distribution is non-negative and \( \Pr(D > 0) > 0 \). Since we are looking at the behavior of this system in stationary state, this finite period in which \( IP_r^n \geq S_r \)
has no effect. Lemma 2.1 essentially states that $A_n$ and $O_n$ are complements so that any knowledge regarding $A_n$ implies knowledge regarding $O_n$. The identity $\Delta = O_n + A_n$ also completely describes the operation of the DIP as we used the entire operation of the DIP to prove this identity. Before establishing our separability result we need one more lemma which is originally due to Veeraraghaven & Scheller Wolf (2008).

**Lemma 2.2.** (Recursions for $O_n$, $Q^e_n$ and $Q^r_n$) The overshoot $O_n$, emergency order quantity $Q^e_n$ and regular order quantity $Q^r_n$ satisfy the following recursions:

\[
O_{n+1} = (O_n - D_n + Q^r_{n+1-l})^+,
\]
\[
Q^e_{n+1} = (D_n - O_n - Q^r_{n+1-l})^+,
\]
\[
Q^r_{n+1} = D_n - Q^e_{n+1}.
\]

**Proof.** The proof of this lemma appears in Veeraraghavan & Scheller-Wolf (2008) as lemma 4.1 and corollary 4.1 and in appendix A of this thesis. □

The recursions (2.15)-(2.17) are quite intuitive. Equation (2.15) describes that the overshoot diminishes each period with the demand and increases with the regular order that enters the information horizon of the emergency inventory position. The emergency order quantity can also be thought of as the ‘undershoot’, i.e., $Q^e_n = (S_e - IP^e_n)^+$ from which relation (2.16) follows. Relation (2.17) follows from the property that in each period the total order amount equals the demand in the previous period. With these results we now establish the separability result part of which also appears as proposition 4.1 in Veeraraghavan & Scheller-Wolf (2008). We remark again that our proof is different.

**Lemma 2.3.** (Separability result) The distributions of $O$, $Q^r$ and $Q^e$ depend on $S_r$ and $S_e$ only through their difference $\Delta = S_r - S_e$.

**Proof.** Consider the recursions in lemma 2.2. To make these equations independent of the starting conditions we substitute the identity for $O_n$ in lemma 2.1. This substitution also makes the operation of the DIP explicit:

\[
O_{n+1} = (\Delta - D_n - \sum_{i=n-l+2}^{n} Q^r_i)^+,
\]
\[
Q^e_{n+1} = (D_n + \sum_{i=n-l+2}^{n} Q^r_i - \Delta)^+,
\]
\[
Q^r_{n+1} = D_n - Q^e_{n+1}.
\]

For the summation $\sum_{i=n-l+2}^{n} Q^r_i$ we read zero whenever $l < 2$. Recursions (2.18)-(2.20) completely determine the stochastic processes $\{O_n\}$, $\{Q^r_n\}$ and $\{Q^e_n\}$ once the sequence $\{D_n\}$
has been specified. Since the stochastic processes \{O_n\}, \{Q^n_r\} and \{Q^n_e\} can be described completely using \(S_r\) and \(S_e\) only through their difference, it follows that their stationary distributions are functions of \(S_r\) and \(S_e\) only through their difference.

\[\square\]

**Remark** The assumption that \(\{D_n\}\) is an i.i.d. sequence was not used to prove lemma 2.3. Thus this result holds for all stochastic processes \(\{D_n\}\), including but not limited to autocorrelated sequences.

Let us define \(O^\Delta\) as the stationary random variable \(O\) for a given \(\Delta\). Lemma 2.3 can be exploited to obtain the optimal DIP for fixed \(\Delta\).

**Theorem 2.4.** (On the optimal choice for \(S_e\)) For fixed \(\Delta\) the optimal \(S_e\) is the smallest integer that satisfies the following inequality:

\[
\sum_{k=0}^{\Delta} \mathbb{E} \left[ (D^{(L_e+1)} - S_e - k)^+ \right] \Pr(O^\Delta = k) \leq (1 - \gamma_0) \mathbb{E}(D).
\]

(2.21)

**Proof.** As a consequence of lemma 2.3 the cost term related to emergency ordering \(c \mathbb{E}[Q^e]\) becomes a fixed constant when \(\Delta\) is fixed. Thus, for fixed \(\Delta\) the relevant cost function is given by \(\tilde{C}(S_e) = h \mathbb{E}[I^+]\) and the problem reduces to a one-dimensional optimization problem we shall call \(Q\):

\[
(Q) \quad \min \quad \tilde{C}(S_e) \\
\text{s.t.} \quad \gamma(S_e, S_e + \Delta) \geq \gamma_0 \\
S_e \in \mathbb{Z}.
\]

(2.22)

Now by the identity \(\gamma = 1 - (\mathbb{E}[B]/\mathbb{E}[D])\) the service level constraint can be modified into a constraint on \(\mathbb{E}[B]\). The expected backlog can be found by conditioning on the emergency inventory position after ordering, using that demand is an i.i.d. sequence and recalling that by lemma 2.3 the distribution of \(O\) is already fixed:

\[
\mathbb{E}[B] = \sum_{k=0}^{\Delta} \mathbb{E} \left[ (D^{(L_e+1)} - S_e - k)^+ \right] \Pr(O^\Delta = k) \leq (1 - \gamma_0) \mathbb{E}[D].
\]

(2.23)

The objective function

\[
h \mathbb{E}[I^+] = \sum_{k=0}^{\Delta} \mathbb{E} \left[ (S_e + k - D^{(L_e+1)})^+ \right] \Pr(O^\Delta = k)
\]

(2.24)

is non-decreasing in \(S_e\) as can easily be shown by recalling that probabilities are non-negative and using finite differences. This implies that the smallest integer \(S_e\) that satisfies inequality (2.21) is the optimal solution to \(Q\). \(\square\)
Remark It is also easy to show that $E[B]$ is a non-increasing function of $S_e$. Thus the optimal $S_e$ given $\Delta$ can easily be found using a simple method such as a bisection search.

The above result provides a simple way to find the optimal DIP if the distribution of $O$ and $E[Q^e]$ can be determined for fixed $\Delta$. If this can be done one may simply perform a search procedure over $\Delta$ to find the globally optimal DIP. To evaluate the cost term $cE[Q^e]$ for the objective function of problem $P$ we note that the first moment of $O$ completely determines the first moment of $Q^e$ through the relations

$$E[Q^r] = \frac{E[A]}{l} = \frac{\Delta - E[O]}{l} \tag{2.25}$$

and

$$E[D] = E[Q^r] + E[Q^e]. \tag{2.26}$$

Thus from the distribution of $O$ it is easy to determine the cost term $cE[Q^r]$. In the next two subsections we describe a one-dimensional Discrete Time Markov Chain (DTMC) that describes the overshoot. Moreover, we provide approximations for its transition probabilities such that the overshoot can be approximated efficiently.

A one-dimensional Markov Chain for the Overshoot

Lemma 2.1 gives insight into the behavior of $O_n$. Instead of studying $O_n$ we may study $A_n$ that has a straightforward physical interpretation as the pipeline stock that will not arrive within the short lead time $l_e$. $A_n$ obeys the following recurrence relation:

$$A_{n+1} = \Delta - O_{n+1} \quad \text{by lemma 2.1}$$

$$= \Delta - (\Delta - D_n - \sum_{i=n+2-l}^n Q^r_i) \quad \text{by substituting (2.18)}$$

$$= \Delta - (\Delta - D_n - A_n + Q^r_{n+1-l}) \quad \text{by substituting (2.10)}$$

$$= \min(\Delta, A_n - Q^r_{n+1-l} + D_n). \tag{2.27}$$

In principle $A_n$ can be modeled by a DTMC. To construct this Markov Chain for $A_n$ however, we would need to store the last $l$ regular order quantities in the state information. This leads to an $l$-dimensional Markov Chain. From equation (2.27) we retrieve that $Q^r_n \in \{0, 1, ..., \Delta\}$ and so this DTMC would have $\sum_{k=0}^{\Delta} \sum_{x_1=0}^k \sum_{x_2=0}^{k-x_1} \sum_{x_1=0}^{k-x_2} \sum_{x_1=0}^{k-x_2} \sum_{x_1=0}^{k-x_2} (x_2, x_2, ..., x_2)$ states. It is computationally infeasible to find the equilibrium distribution of this DTMC for most practical instances. To remedy this we study a compacted DTMC with only one dimension and $\Delta + 1$ states. Observe that the recurrence relation (2.27) completely defines a Markov Chain for $A_n$ if the probability mass functions of $D$ and $\{Q^r_{n+1-l} \mid A_n\}$ are known. This DTMC is
defined by the transition probabilities \( p_{ij} = \Pr(A_{n+1} = j|A_n = i) \) that can be obtained by distinguishing the cases \( j < \Delta \) and \( j = \Delta \). First consider the case \( j < \Delta \), we have

\[
p_{ij} = \Pr(A_{n+1} = j|A_n = i) \\
= \Pr(A_n - Q_{n+1-l} + D_n = j|A_n = i) \\
= \sum_{k=0}^{j} \Pr(Q_{n+1-l} = A_n + D_n - j|A_n = i, D_n = k) \Pr(D_n = k) \\
= \sum_{k=0}^{j} \Pr(Q_{n+1-l} = i + k - j|A_n = i) \Pr(D = k).
\] (2.28)

The case \( j = \Delta \) is very similar:

\[
p_{i\Delta} = \Pr(A_{n+1} = \Delta|A_n = i) \\
= \Pr(A_n - Q_{n+1-l} + D_n \geq \Delta|A_n = i) \\
= \sum_{k=0}^{i} \Pr(D_n \geq \Delta + Q_{n+1-l} - A_n|A_n = i, Q_{n+1-l} = k) \Pr(Q_{n+1-l} = k|A_n = i) \\
= \sum_{k=0}^{i} \Pr(Q_{n+1-l} = k|A_n = i) \Pr(D \geq \Delta + k - i).
\] (2.29)

Now we organize these transition probabilities in the transition matrix \( P \):

\[
P = \begin{pmatrix}
  p_{00} & \cdots & p_{0\Delta} \\
  \vdots & \ddots & \vdots \\
  p_{\Delta0} & \cdots & p_{\Delta\Delta}
\end{pmatrix}.
\] (2.30)

If we define \( \theta_{x,y} = \Pr(Q_{n+1-l} = x|A_n = y) \), \( \zeta_x = \Pr(D = x) \) and \( \eta_x = \Pr(D \geq x) \) the transition matrix of this Markov chain is given by the product of a left lower triangular stochastic matrix \( L \) and a right upper triangular stochastic matrix \( U \):

\[
P = LU, \quad L = \begin{pmatrix}
  \theta_{0,0} & 0 \\
  \theta_{1,1} & \theta_{0,1} \\
  \vdots & \ddots \\
  \theta_{\Delta,\Delta} & \theta_{\Delta-1,\Delta} & \cdots & \theta_{0,\Delta}
\end{pmatrix}, \quad U = \begin{pmatrix}
  \zeta_0 & \cdots & \zeta_{\Delta-1} & \eta_{\Delta} \\
  \vdots & \ddots & \vdots \\
  \zeta_0 & \eta_1 \\
  0 & \eta_0
\end{pmatrix}.
\] (2.31)

The matrix \( L \) describes the downward transitions due to a regular order leaving the range where it will not arrive within \( l^e \), while the matrix \( U \) describes the upward transitions due to new demand being ordered and entering the pipeline.

If we let \( \pi(x) \) denote \( \Pr(A = x) \), \( \pi = [\pi(0), \ldots, \pi(\Delta)] \) and \( e = [1, 1, \ldots, 1]^T \), then the stationary distribution \( \pi \) can be found by solving the set of linear equations

\[
\pi P = \pi, \quad \pi e = 1.
\] (2.32)

The distribution of \( D \) is assumed to be known, but the distribution of \( \{Q_{n+1-l}^r|A_n\} \) is in fact unknown. In the next section we construct an approximation for this distribution so that the introduced one-dimensional DTMC can be used to approximate the overshoot distribution.
Approximations for the transition probabilities

To determine the transition probabilities in the DTMC of the previous section we need the probability mass functions of \( D \) and \( \{Q_{r,n+1-l}^r|A_n\} \). The latter can be approximated using the following (limiting) result.

**Proposition 2.5.** The following statements hold:

(i) As \( \Delta \to \infty \), \( \Pr(Q_{r,n+1}^r = x) \to \Pr(D_n = x) \).

(ii) As \( \Delta \to \infty \), \( \Pr\left(Q_{r,n+1-l}^r = x|A_n = y\right) \to \Pr\left(D_{n+1-l} = x|\sum_{i=n+1-l}^n D_i = y\right) \).

(iii) For \( \Delta = 1 \), \( \Pr\left(Q_{r,n+1}^r = x|A_n = y\right) = \Pr\left(D_{n+1} = x|\sum_{i=n+1}^n D_i = y\right) \).

**Proof.** We rewrite equation (2.27) to

\[
A_{n+1} = \min(\Delta, A_n - Q_{n+1-l}^r + D_n)
\]

\[= \min(\Delta, A_{n+1} - Q_{n+1}^r + D_n). \tag{2.33}
\]

Now if we let \( \Delta \to \infty \) and recall the condition \( \Pr(D < \infty) = 1 \) we immediately retrieve part (i) of the proposition. For part (ii) to hold we need to show that when \( \Delta \to \infty \), \( Q_n^r \) becomes an i.i.d. sequence. This follows from an induction argument on part (i) of this proposition and the assumption that \( D_n \) is an i.i.d. sequence. For part (iii) we distinguish between the condition \( A_n = 0 \) and \( A_n = 1 \). Part (iii) holds trivially under the conditioning \( A_n = 0 \). For the conditioning \( A_n = 1 \) we need only show that \( \Pr(Q_{n+1-l}^r = 1|A_n = 1) = \Pr(D_{n+1-l} = 1|\sum_{i=n+1-l}^n D_i = 1) \) because \( \Pr(Q_{n+1-l}^r = 0|A_n = 1) \) is the complement of \( \Pr(Q_{n+1-l}^r = 1|A_n = 1) \). Recall the definition of \( A_n \) as the sum of \( l \) regular orders. When \( A_n = 1 \), there is exactly one order of one SKU, and it can be any of the regular orders included in \( A_n \) with probability \( 1/l \), i.e., \( \Pr(Q_{n+1-l}^r = 0|A_n = 1) = 1/l \). Now we have

\[
\Pr\left(D_{n+1-l} = 1|\sum_{i=n+1-l}^n D_i = 1\right) = \frac{\Pr(D = 1) \Pr(D^{l-1} = 0)}{\Pr(D^l = 1)} = \frac{\Pr(D = 1) \Pr(D = 0)^{l-1}}{l \Pr(D = 1) \Pr(D = 0)^{l-1}} = \frac{1}{l}. \tag{2.34}
\]

But \( 1/l = \Pr(Q_{n+1-l}^r = 0|A_n = 1) \) as required.

Intuitively parts (i) and (ii) of proposition 2.5 are obvious because \( \Delta = \infty \) corresponds to single sourcing with the regular supplier, in which case \( Q_{n+1}^r = D_n \). Parts (ii) and (iii) of proposition 2.5 suggest that \( \Pr\left(D_{n+1-l} = x|\sum_{i=n+1}^n D_i = y\right) \) can be used to approximate
Pr \left( Q_{n+1-l}^r = x|A_n = y \right) as this approximation is exact for extremely small \( \Delta (\Delta = 1) \) and extremely large \( \Delta (\Delta \rightarrow \infty) \). Thus an approximation for Pr \left( Q_{n+1-l}^r = x|A_n = y \right) is given by

\[
\theta_{x,y} = \Pr \left( Q_{n+1-l}^r = x|A_n = y \right) \\
\approx \frac{\Pr \left( D_{n+1-l} = x|\sum_{i=n+1-l}^{n} D_i = y \right)}{\Pr \left( D = x \right) \Pr \left( D^{(l-1)} = y - x \right)}
\]

Using this approximation for \( \theta_{x,y} \) we can compute an approximation for \( \Pr(A = x) \) by solving the set of linear equations (2.32). Then by using relation (2.1) we obtain an approximation for the distribution of \( O \) as \( \Pr(O = x) = \Pr(A = \Delta - x) \).

Numerical experiments indicate that this approximation works well in a wide range and not only closely approximates the first two moments of \( O \) but also the often unusual shape of its distribution. To illustrate the unusual shape of the overshoot distribution four typical examples are given in figure 2.2, where the overshoot distribution as determined by simulation is shown in conjunction with the approximation based on the above analysis.

**Remark** Note that the above approximation is also exact when \( l = 1 \). Since the DIP is the optimal policy for \( l = 1 \) this approach yields the globally optimal policy whenever \( l = 1 \).

### 2.3. Numerical study

In this section we report on a numerical study to test the accuracy of the Markov Chain approximation that we propose. To this end a test bed of 960 instances of problem \( P \) was created that is a full factorial design of the parameter settings summarized in table 2.2. The demand distributions we used are the discrete two moment fits suggested by Adan et al. (1996).

First we performed the optimization using the approximate Markov Chain approach. This approach found the parameters \( S^*_{r,MC} \) and \( S^*_{e,MC} \) to be optimal. We then evaluated the cost function and service level for \( S^*_{r,MC} \) and \( S^*_{e,MC} \) by both simulation and our Markov Chain approach. The simulation was ran such that the half-width of a 99%-confidence interval was less then 1% of the obtained estimate. Figure 2.3 displays histograms of the quantities

\[
\delta_{\gamma} := \gamma_{MC}(S^*_{r,MC}, S^*_{e,MC}) - \gamma_{sim}(S^*_{r,MC}, S^*_{e,MC})
\]

and

\[
\Delta_C := \frac{C_{MC}(S^*_{r,MC}, S^*_{e,MC}) - C_{sim}(S^*_{r,MC}, S^*_{e,MC})}{C_{sim}(S^*_{r,MC}, S^*_{e,MC})} \cdot 100\%
\]
Figure 2.2: Overshoot distributions for a few illustrative cases as determined by simulation and the corresponding Markov Chain approximations.

Table 2.2: Test-bed of instances with deterministic regular lead time

<table>
<thead>
<tr>
<th>Parameter</th>
<th>settings</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E[D]$</td>
<td>5, 50</td>
</tr>
<tr>
<td>$c_D^2$</td>
<td>$\frac{1}{4}$, $\frac{1}{2}$, 1, $\frac{3}{2}$, 2</td>
</tr>
<tr>
<td>$i^c$</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td>$l$</td>
<td>2, 4, 6, 8</td>
</tr>
<tr>
<td>$h$</td>
<td>1</td>
</tr>
<tr>
<td>$c$</td>
<td>2, 4, 8, 16</td>
</tr>
<tr>
<td>$\gamma_0$</td>
<td>0.95, 0.98</td>
</tr>
</tbody>
</table>
2.3. NUMERICAL STUDY

where the subscripts indicate whether the estimate was found by simulation or our Markov Chain approach.

Figure 2.3: Accuracy of Markov Chain approximation

The histogram of $\delta_\gamma$ seems to indicate a slight bias of the Markov Chain to overestimate the service level. This is no real concern however, because the deviations are in the order of magnitude of $10^{-3}$. Moreover, the integer nature of the decision variables ensures that the required service level is usually exceeded. In our test bed the Markov Chain solution was never significantly below the target service level $\gamma_0$. Further the relative deviation of the cost estimate from the Markov Chain is almost always within 1% which is also the tolerance of a simulation estimate. Thus this method is much faster than simulation without significant loss of accuracy.

Next we wish to test whether the DIP we find by using our Markov Chain approach is indeed the optimal DIP or at least very close to optimal. We performed the optimization over the same test-bed using the simulation based method proposed by Veeraraghavan & Scheller-Wolf (2008) and found the parameters $S^*_{r,sim}$ and $S^*_{e,sim}$ to be optimal. Then we compared the optimal costs found through simulation optimization with the solutions found with the Markov Chain approach, i.e., we considered the measure

$$\Delta C^* := \frac{C_{sim}(S^*_{r,MC}, S^*_{e,MC}) - C_{sim}(S^*_{r,sim}, S^*_{e,sim})}{C_{sim}(S^*_{r,sim}, S^*_{e,sim})} \cdot 100\%.$$  (2.38)

The histogram of $\Delta C^*$ shown in figure 2.4 indicates that our solution is indeed optimal in more than half of the cases and that it deviates from optimality by more than 1% only in rare cases. The histogram also shows that in some cases the Markov Chain solution performs better than the optimal simulation solution. This is due to simulation estimate tolerance.
As a closing note we remark that the DTMC based optimization procedure was faster than the simulation based procedure by a factor between 20-30. Thus we find that the Markov Chain approach is much faster than simulation and performs almost indistinguishable from simulation optimization.
Chapter 3

Dual sourcing with stochastic lead times

In many practical situations lead times are not deterministic. In the setting with two suppliers, the lead time of the emergency supplier is usually quite small compared to the regular supplier. Furthermore the variance of the emergency lead time is usually small compared to a review period and consequently the emergency lead time can be assumed to be deterministic. The lead time of the regular supplier however is likely to vary over multiple review periods so that in many instances this lead time should be considered to be stochastic. For example, if the regular supplier represents manufacturing in the far east, then lead times are influenced by many factors such as capacity utilization at the plant, border control and weather conditions on the shipping route. In this chapter revisit the Dual Index policy, but we will relax the assumption that $l^r$ is deterministic. In section 3.1 we define our model followed by its analysis in section 3.2. We conclude this chapter with some numerical results in section 3.3.

3.1. Model

The model we consider is identical to the model described in section 2.1 with one exception: now we assume that $L^r_n$ is a stochastic integer and that the lower bound of its support is $l^e + 1$. Define the random variable $L_n$ as

$$L_n := L^r_n - l^e.$$  \hfill (3.1)

The support of $L$ is constituted by the positive integers $\mathbb{N}$. We can think of $L^r_n$ as consisting of a deterministic part $l^e$ and a stochastic part $L_n$. We will assume that $\{L_n\}$ is an i.i.d.
sequence and \( \Pr(L = i) = q_i \). This implies that order crossover is possible and places us in a setting similar to Robinson et al. (2001). Further we let \( l_n \) and \( l'_n \) denote realizations of the random variables \( L_n \) and \( L'_n \). The situation as described above is shown schematically in figure 3.1.

Inventory positions are defined using set notation. Let \( X_n \) be the set of all period indices such that at the beginning of period \( n \) before ordering, the regular orders from these periods have not yet arrived in stock,

\[
X_n = \{k | k \geq n - l'_k, \ k < n\}.
\]  

(3.2)

Additionally let \( Y_n \) be the set of all period indices such that at the beginning of period \( n \) before ordering the regular orders from these periods have not yet arrived in stock but will do so within the emergency lead time,

\[
Y_n = \{k | k \geq n - l'_k, \ k \leq n - l_k\}.
\]  

(3.3)

Obviously it holds that \( Y_n \subseteq X_n \). A graphical representation of these sets is given in figure 3.1, along with two sets that are defined later. Using these sets we can again define the emergency and regular inventory positions as

\[
IP^e_n = I_n + \sum_{i \in Y_n} Q^r_i + \sum_{i=n-l^e}^{n-1} Q^e_i
\]  

(3.4)

and

\[
IP^r_n = I_n + \sum_{i \in X_n} Q^r_i + \sum_{i=n-L_e}^{n} Q^e_i = IP^e_n + Q^e_n + \sum_{i \in X_n \setminus Y_n} Q^r_i.
\]  

(3.5)

Notice that these definitions reduce to the earlier definitions in case of deterministic regular lead times. We also remark that we do not necessarily need to know the realizations of \( L'_n \) up to time \( n \) for the inventory positions to be well defined. The only necessary information needed is to know in real time when the order from period \( k \) will arrive within the emergency lead time \( l^e \), i.e., we need to know when \( k = n - l_k \). In essence the random variable \( L'_n \) consists of a random component \( L_n \) and a deterministic component \( l^e \). We assume that that the random component becomes known before or at the time the remaining lead time of a regular order is \( l^e \).

There are multiple ways for this information to become available in practice. First we may know what the regular lead time will be as soon as we place an order. Second we may know when a regular order will arrive within the emergency lead time because this time is naturally associated with known events such as a shipment harboring at the port. Less
probable is the third option of this information being available through the use of modern information technology such as RFID. In the context of manufacturing in overtime or other use of flexible capacity this information may be available by simple inspection of the job floor.

Ordering decisions are still given by

\[ Q^e_n = (S_e - IP^e_n)^+, \quad Q^r_n = S_r - IP^r_n \]  (3.6)

and overshoot still satisfies the original definition:

\[ O_n = IP^e_n + Q^e_n - S_e = (IP^e_n - S_e)^+. \]  (3.7)

With these new definitions we can analyze the DIP when regular lead times are stochastic.

3.2. Analysis

Our analysis will proceed along the same lines as the analysis for deterministic regular lead times, i.e., we show how to find the optimal DIP for fixed \( \Delta \) (Section 3.2) and provide a one
dimensional DTMC that describes the overshoot (Section 3.2). We provide approximations for the transition probabilities of this DTMC in Section 3.2.

Optimization

Let us turn again to the amount of pipeline stock that will not arrive within the emergency lead time, $A_n$. Let $U_n$ be the set of all period indices such that in period $n$ after ordering the regular orders from these periods will not arrive within the emergency lead time, this set is also shown in figure 3.1,

$$U_n = \{k|k \geq n - l_k + 1, \ x \leq n\}. \quad (3.8)$$

Now the definition of $A_n$ can be written as

$$A_n = \sum_{i \in U_n} Q_i. \quad (3.9)$$

Lemma 3.1. (Key functional relation) Suppose that $IP_r^r \leq S_r$ for some $k \in \mathbb{N}_0$. Then for all $n \geq k$ the Dual-Index policy ensures that the following identity holds

$$\Delta = O_n + A_n. \quad (3.10)$$

Proof. Reconsider the regular inventory position as given in equation (3.5),

$$IP_r^n = IP_e^n + Q_e^n + \sum_{i \in X_n \setminus Y_n} Q_i. \quad (3.11)$$

Now we substitute the definition of the overshoot (from equation (3.7)) and add $Q_r$ to both sides of this equation,

$$IP_r^n + Q_r^n = S_e + O_n + \sum_{i \in X_n \setminus Y_n \cup \{n\}} Q_i. \quad (3.12)$$

By supposition $IP_r^n \leq S_r$ so $Q_r^n = S_r - IP_r^n$ and the left-hand side of (3.12) becomes $S_r$. When we take a closer look at the set over which the sum in (3.12) runs it is straightforward to verify that $U_n = X_n \setminus Y_n \cup \{n\}$ so that we can substitute the definition of $A_n$ to obtain

$$S_r = S_e + O_n + A_n. \quad (3.13)$$

Rearrangement and substitution of the identity $\Delta = S_r - S_e$ yields the result. $\square$

The supposition that $n \geq k$ and $IP_r^r \leq S_r$ is no real restriction because the number of periods that $IP_r^n$ can be greater than $S_r$ is finite with probability 1 under the earlier
3.2. ANALYSIS

assumption that the demand distribution is non-negative and \( \Pr(D > 0) > 0 \). Since we are looking at the behavior of this system in stationary state this finite period in which \( IP^r_k \geq S_r \) has no effect. Lemma 3.1 is a direct generalization of lemma 2.1 and essentially states that \( A_n \) and \( O_n \) are direct compliments also in the presence of stochastic regular lead times. Note also that lemma 3.1 holds for all stochastic processes \( \{L_n\}_{n \in \mathbb{N}_0} \), not just i.i.d. sequences.

Now we introduce \( V_n \) the set of period indices such that at the beginning of period \( n \) after ordering the regular orders from these periods will enter the information horizon of the emergency inventory position in period \( n + 1 \),

\[
V_n = \{k| k = n - l_k + 1\}. \tag{3.14}
\]

This set is also shown in figure 3.1. We emphasize that the sets \( X_n \) and \( Y_n \) are defined before ordering while \( U_n \) and \( V_n \) are defined after ordering. It is immediate that \( V_n \subseteq U_n \). As before we now turn attention to recursions for \( O_n \), \( Q^e_n \) and \( Q^r_n \) and then establish our separability result.

**Lemma 3.2.** (Recursions for \( O_n \), \( Q^e_n \) and \( Q^r_n \)) The overshoot \( O_n \), emergency and regular order quantities satisfy the following recursions:

\[
O_{n+1} = (O_n - D_n + \sum_{i \in V_n} Q^r_i)^+ \tag{3.15},
\]

\[
Q^e_{n+1} = (D_n - O_n - \sum_{i \in V_n} Q^r_i)^+ \tag{3.16},
\]

\[
Q^r_{n+1} = D_n - Q^e_{n+1}. \tag{3.17}
\]

**Proof.** The proof for this lemma is given in appendix A. \( \square \)

Recursions (3.15)-(3.17) have intuitive physical meaning. Equation (3.15) describes that the overshoot diminishes each period with the demand and increases with the regular order(s), if any, that enter the information horizon of the emergency inventory position. The emergency order quantity can also be thought of as the ‘undershoot’, i.e. \( Q^e_n = (S_e - IP^e_n)^+ \) from which relation (3.16) follows. Relation (3.17) follows from the property that in each period the total order amount equals the demand in the previous period.

With these results we can prove the same separability result that was shown to hold for deterministic regular lead times.

**Lemma 3.3.** (Separability result) The distributions of \( O \) and \( Q^e \) and \( Q^r \) depend on \( S_r \) and \( S_e \) only through their difference \( \Delta = S_r - S_e \).
Proof. Consider the recursions in lemma 3.2. To make these equations independent of the starting conditions we substitute the identity for $O_n$ in lemma 3.1. This substitution also makes the operation of the DIP explicit:

\[
O_{n+1} = \left( \Delta - D_n - \sum_{i \in U_n \setminus V_n} Q_r^i \right)^+, \tag{3.18}
\]

\[
Q_{n+1}^r = \left( D_n + \sum_{i \in U_n \setminus V_n} Q_r^i - \Delta \right)^+, \tag{3.19}
\]

\[
Q_{n+1}^e = D_n - Q_{n+1}^e. \tag{3.20}
\]

For the summation $\sum_{i \in U_n \setminus V_n} Q_r^i$ we read 0 whenever $U_n \setminus V_n = \emptyset$. These recursions completely determine the stochastic processes \{\(O_n\)\}, \{\(Q_n^r\)\} and \{\(Q_n^e\)\} once the stochastic sequences \{\(D_n\)\}, and \{\(L_n\)\} have been specified. Since the stochastic processes \{\(O_n\)\} and \{\(Q_n^r\)\} and \{\(Q_n^e\)\} can be described completely using \(S_r\) and \(S_e\) only through their difference, it follows that their stationary distributions are functions of \(S_r\) and \(S_e\) only through their difference.

Remark In establishing lemma 3.3 we did not require that either \{\(D_n\)\} or \{\(L_n\)\} are i.i.d. sequences. In principle the stationary overshoot distribution is well defined when \(\Delta\) is fixed for all processes \{\(D_n\)\} and \{\(L_n\)\} such that $D_n \in \mathbb{N}_0$ and $L_n \in \mathbb{N}_0$ for all $n \in \mathbb{N}_0$. We do use that \{\(D_n\)\} and \{\(L_n\)\} are i.i.d. in sections 3.2 and 3.2 to construct an efficient approximation for $\Pr(O = x)$. However the distribution of $O$, $Q^e$ or $Q^r$ can be determined by simulation for more general processes \{\(D_n\)\} and/or \{\(L_n\)\}.

Let us define $O^\Delta$ as the stationary random variable $O$ for a given $\Delta$. Lemma 3.3 leads to the following theorem on the optimal choice for $S_e$ for fixed $\Delta$.

**Theorem 3.4.** (On the optimal choice for $S_e$) For fixed $\Delta$ the optimal $S_e$ is the smallest integer that satisfies the following inequality

\[
\sum_{k=0}^{\Delta} \mathbb{E} \left[ (D^{(L_e+1)} - S_e - k)^+ \right] \Pr(O^\Delta = k) \leq (1 - \gamma_0)\mathbb{E}(D). \tag{3.21}
\]

Proof. As a consequence of lemma 3.3 the cost term related to emergency ordering, $c\mathbb{E}[Q^e]$, becomes a fixed constant when $\Delta$ is fixed. Thus, for fixed $\Delta$ the relevant cost function is given by $\tilde{C}(S_e) = h\mathbb{E}[L^+]$ and the problem reduces to a one-dimensional optimization problem we shall call $\mathcal{Q}$.

\[
(\mathcal{Q}) \quad \min_{\gamma(S_e, S_e + \Delta) \geq \gamma_0} \tilde{C}(S_e) \quad \text{s.t.} \quad S_e \in \mathbb{Z}. \tag{3.22}
\]
Now by the identity \( \gamma = 1 - (E[B]/E[D]) \) the service level constraint can be modified into a constraint on \( E[B] \). The expected backlog can be found by conditioning on the emergency inventory position after ordering, using that demand is an i.i.d. sequence and recalling that by lemma 3.3 the distribution of \( O \) is already fixed:

\[
E[B] = \sum_{k=0}^{\Delta} E \left[ (D^{(L_e+1)} - S_e - k)^+ \right] \Pr(O^\Delta = k) \leq (1 - \gamma_0)E[D]. \tag{3.23}
\]

The objective function

\[
hE[I^+] = \sum_{k=0}^{\Delta} E \left[ (S_e + k - D^{(L_e+1)})^+ \right] \Pr(O^\Delta = k) \tag{3.24}
\]

is non-decreasing in \( S_e \) as can easily be shown by recalling that probabilities are non-negative and using finite differences. This implies that the smallest integer \( S_e \) that satisfies inequality (3.23) is the optimal solution to \( Q \), which completes the proof.

**Remark** It is also easy to show that \( E[B] \) is a non-increasing function of \( S_e \). Thus the optimal \( S_e \) given \( \Delta \) can be found using a simple method such as a bisection search.

The above result provides a simple way to find the optimal DIP if the distribution of \( O \) and \( E[Q^e] \) can be determined for fixed \( \Delta \). If this can be done one may simply perform a search procedure over \( \Delta \) to find the globally optimal DIP. To evaluate the cost term \( cE[Q^e] \) for the objective function of problem \( \mathcal{P} \) we note that the first moment of \( O \) completely determines the first moment of \( Q^e \) through the relations

\[
E[Q^r] = \frac{E[A]}{E[L]} = \frac{\Delta - E[O]}{E[L]} \tag{3.25}
\]

and

\[
E[D] = E[Q^r] + E[Q^e]. \tag{3.26}
\]

Thus from the distribution of \( O \) it is easy to determine the cost term \( cE[Q^e] \). In the next two sections we describe a one-dimensional Discrete Time Markov Chain (DTMC) that describes the overshoot and provide approximations for its transition probabilities so that the overshoot can be approximated efficiently.

**A one-dimensional Markov Chain for the overshoot**

Lemma 3.1 gives insight into the behavior of \( O_n \). Instead of studying \( O_n \) we may study \( A_n \) that has a straightforward physical interpretation as the pipeline stock that will not arrive
within the short lead time \( l_e \). \( A_n \) obeys the following recurrence relation

\[
A_{n+1} = \Delta - O_{n+1}
\]

\[
= \Delta - \left( \Delta - D_n - \sum_{i \in V_n \setminus V} Q_i^r \right)^+
\]

\[
= \Delta - \left( \Delta - D_n - A_n + \sum_{i \in V_n} Q_i^r \right)^+
\]

\[
= \min (\Delta, A_n - \sum_{i \in V_n} Q_i^r + D_n)
\]

(3.27)

In principle \( A_n \) can be modeled by a DTMC. To construct this Markov Chain for \( A_n \) however, we would need to store all the regular order quantities for which \( L_k \leq n - k \) in the state information, as well as their respective lead times. This leads to an Markov Chain that has as many dimensions as the maximum possible \( L \), \( L_{\text{max}} \) say. Note that \( Q_i^r \in \{0,1,\ldots,\Delta\} \) and so this DTMC would have \( \sum_{k=0}^{\Delta} \sum_{x_1=0}^k \sum_{x_2=0}^{k-1} \cdots \sum_{x_{L_{\text{max}}}=0}^{k-L_{\text{max}}+1} x_i \) states. It is computationally infeasible to find the equilibrium distribution of this DTMC for most practical instances. To remedy this we study a more compact DTMC with only one dimension and \( \Delta + 1 \) states. Observe that the recurrence relation (3.27) completely defines a Markov Chain for \( A_n \) if the probability mass functions of \( D \) and \( \{\sum_{i \in V_n} Q_i^r | A_n\} \) are known. This Markov Chain is defined by the transition probabilities \( p_{ij} = \Pr (A_{n+1} = j | A_n = i) \) that can be obtained by distinguishing the cases \( j < \Delta \) and \( j = \Delta \). First consider the case \( j < \Delta \), we have

\[
p_{ij} = \Pr (A_{n+1} = j | A_n = i)
\]

\[
= \Pr (A_n - \sum_{i \in V_n} Q_i^r + D_n = j | A_n = i)
\]

\[
= \sum_{k=0}^{j} \Pr (\sum_{i \in V_n} Q_i^r = A_n + D_n - j | A_n = i, D_n = k) \times \Pr (D_n = k)
\]

(3.28)

The case \( j = \Delta \) is very similar:

\[
p_{i\Delta} = \Pr (A_{n+1} = \Delta | A_n = i)
\]

\[
= \Pr (A_n - \sum_{i \in V_n} Q_i^r + D_n \geq \Delta | A_n = i)
\]

\[
= \sum_{k=0}^{i} \Pr (D_n \geq \Delta + \sum_{i \in V_n} Q_i^r - A_n | A_n = i, \sum_{i \in V_n} Q_i^r = k) \times \Pr (\sum_{i \in V_n} Q_i^r = k | A_n = i)
\]

(3.29)

Now we organize these transition probabilities in the transition matrix \( P \):

\[
P = \begin{pmatrix}
p_{00} & \cdots & p_{0\Delta} \\
\vdots & \ddots & \vdots \\
p_{\Delta 0} & \cdots & p_{\Delta \Delta}
\end{pmatrix}
\]

(3.30)
If we define $\vartheta_{x,y} = \Pr(\sum_{i \in V_n} Q_i^r = x | A_n = y)$, $\zeta_x = \Pr(D = x)$ and $\eta_x = \Pr(D \geq x)$ the transition matrix of this Markov chain is given by the product of a left lower triangular stochastic matrix $L$, that describes the transitions downward due to $\sum_{i \in V_n} Q_i^r$ leaving the set $U_n$, and a right upper triangular stochastic matrix $U$, that describes the transitions upward due to new regular orders entering the pipeline:

$$P = LU$$

$$L = \begin{pmatrix}
\vartheta_{0,0} & \vartheta_{0,1} & 0
\vartheta_{1,1} & \vartheta_{1,2} & \ddots
\vdots & \ddots & \ddots
\vartheta_{\Delta,\Delta} & \vartheta_{\Delta-1,\Delta} & \ldots & \vartheta_{0,\Delta}
\end{pmatrix}$$

$$U = \begin{pmatrix}
\zeta_0 & \cdots & \zeta_{\Delta-1} & \eta_{\Delta}
\cdots & \ddots & \ddots & \cdots
\zeta_0 & \eta_1
0 & \eta_0
\end{pmatrix}.$$  \hfill (3.31)

If we let $\pi(x)$ denote $\Pr(A = x)$, $\pi = [\pi(0), \ldots, \pi(\Delta)]$ and $e = [1, 1, \ldots, 1]^T$, then the stationary distribution $\pi$ can be found by solving the set of linear equations

$$\pi P = \pi \quad \pi e = 1 \hfill (3.32)$$

The distribution of $D$ is assumed to be known, but the distribution of $\{\sum_{i \in V_n} Q_i^r | A_n\}$ is in fact unknown. In the next section we construct an approximation for this distribution so that the introduced one-dimensional Markov Chain can be used to approximate the overshoot distribution.

**Approximations for the transition probabilities**

To determine the transition probabilities in the DTMC of the previous section we need the probability mass functions of $D$ and $\{\sum_{i \in V_n} Q_i^r | A_n\}$. The latter can be approximated using the following limiting result.

**Proposition 3.5.** The following statements hold

(i) As $\Delta \to \infty$, $\Pr(Q_{n+1}^r = x) \to \Pr(D_n = x)$

(ii) As $\Delta \to \infty$, $\Pr(\sum_{i \in V_n} Q_i^r = x | A_n = y) \to \Pr\left(\sum_{i=n-|V_n|+1}^{n} D_i = x | \sum_{i=n-|U_n|+1}^{n} D_i = y\right)$

**Proof.** We rewrite equation (3.27) to

$$A_{n+1} = \min\left(\Delta, A_n - \sum_{i \in V_n} Q_i^r + D_n\right)$$

$$= \min\left(\Delta, \sum_{i \in U_n} Q_i^r - \sum_{i \in V_n} Q_i^r + D_n\right)$$

$$= \min\left(\Delta, \sum_{i \in U_n \cup \{n+1\}} Q_i^r - Q_{n+1}^r + D_n\right)$$

$$= \min\left(\Delta, A_{n+1} - Q_{n+1}^r + D_n\right).$$  \hfill (3.33)
CHAPTER 3. DUAL SOURCING WITH STOCHASTIC LEAD TIMES

Now if we let $\Delta \to \infty$ and recall the condition $\Pr(D < \infty) = 1$ we immediately retrieve part (i) of the proposition. For part (ii) to hold we need to show that when $\Delta \to \infty$, $Q^r_n$ becomes an i.i.d. sequence, so that the distribution of $A_n$ depends only on the stationary distribution of $|U_n|$. This follows from an induction argument on part (i) of this proposition and the assumption that $D_n$ is an i.i.d. sequence.

Remark When considering deterministic lead times we already showed in proposition 2.5 that the approximation we propose is exact also for $\Delta = 1$. For stochastic $L_n$ this is no longer the case. The numerical results in Section 3.3 reflect this fact.

Part (ii) of proposition 2.5 suggests that $\Pr \left( \sum_{i=n-|V_n|+1}^{n} D_i = x \mid \sum_{i=n-|V_n|+1}^{n} D_i = y \right)$ can be used to approximate $\Pr \left( \sum_{i \in V_n} Q^r_i = x \mid A_n = y \right)$. For practical application of this approximation however we are also interested in how fast this convergence occurs, i.e., how large should $\Delta$ be for this approximation to work properly. While the convergence under part (i) of proposition 2.5 is rather slow, the convergence under part (ii) is much faster because the conditioning truncates the tail probabilities of $D$ above $\Delta$ where $D$ and $Q^r$ are known to differ the most for finite $\Delta$. Furthermore we know from previous research (Scheller-Wolf et al. (2003)) that most products are supplied by the regular supplier in typical cases for the optimal DIP, which corresponds to a large $\Delta$. Thus an approximation for $\Pr \left( \sum_{i \in V_n} Q^r_i = x \mid A_n = y \right)$ is given by:

$$\vartheta_{x,y} = \Pr \left( \sum_{i \in V_n} Q^r_i = x \mid A_n = y \right) \approx \Pr \left( \sum_{i=n-|V_n|+1}^{n} D_i = x \mid \sum_{i=n-|V_n|+1}^{n} D_i = y \right).$$

The computation of $\Pr \left( \sum_{i=n-|V_n|+1}^{n} D_i = x \mid \sum_{i=n-|V_n|+1}^{n} D_i = y \right)$ however is not straightforward because it requires knowledge of the random variables $|U_n|$ and $|V_n|$ which in turn depend on the process $\{L_n\}$. Indeed for the computation of this probability we digress to study the joint stationary distribution of $|U_n|$ and $|V_n|$ when $L_n$ is assumed to be a sequence of i.i.d random variables with finite support. In appendix B.1 a fitting procedure is described to obtain random variables with finite support. In principle one may study the joint distribution of $|U_n|$ and $|V_n|$ for different lead time processes $\{L_n\}$.

Let $K_n$ denote the number of orders in the pipeline that will not arrive within the emergency lead time in period $n$ after ordering,

$$K_n = |U_n|. \quad (3.34)$$
Further let $\Lambda_n$ denote the number of orders that are about to enter the information horizon of the emergency inventory position,

$$\Lambda_n = |V_n|.$$ (3.35)

We wish to determine the joint stationary distribution of these two quantities $\Pr(K = \kappa \cap \Lambda = \lambda)$. We do this recursively. Recall that the distribution of $L$ is given by $q_\nu = \Pr(L = \nu)$, $i \in \{1,2,\ldots,L_{\text{max}}\}$. Further we define

$$\varphi_{\kappa,\lambda,\nu} = \Pr(K = \kappa \cap \Lambda = \lambda | \text{orders were placed the last } \nu \text{ periods only (not before)}) \quad (3.36)$$

Obviously this definition means that the distribution needed is given by

$$\Pr(K = \kappa \cap \Lambda = \lambda) = \varphi_{\kappa,\lambda,L_{\text{max}}} := \varphi_{\kappa,\lambda},$$ (3.37)

since orders that were were placed more than $L_{\text{max}}$ period ago cannot belong to the sets $U_n$ or $V_n$. The probabilities $\varphi_{\kappa,\lambda,\nu}$ can be computed recursively as follows:

$$\varphi_{\kappa,\lambda,\nu} = \varphi_{\kappa-1,\lambda-1,\nu-1} q_\nu + \varphi_{\kappa-1,\lambda,\nu-1} \sum_{m=\nu+1}^{L_{\text{max}}} q_m + \varphi_{\kappa,\lambda,\nu-1} \sum_{m=1}^{\nu-1} q_m.$$ (3.38)

The initial probabilities are straightforwardly seen to be

$$\varphi_{1,0,1} = \sum_{m=2}^{L_{\text{max}}} q_m, \quad \varphi_{1,1,1} = q_1, \quad \varphi_{\kappa,\lambda,1} = 0 \text{ otherwise.}$$ (3.39)

This concludes our derivation of the joint stationary distribution of $|U_n|$ and $|V_n|$.

**Remark** The process $K_n$ can also be thought of as the number of customers in a discrete time $D/G/L_{\text{max}}/L_{\text{max}}$-queue. Each period $n$ a customer arrives (order is placed) and that customer is immediately enters service for a random time $L_n$ (order stays in the set $U$ for $L_n$ periods). Thus this $D/G/c/c$-queue has the special property that the service distribution has a finite support on $\{1,\ldots,L_{\text{max}}\}$ while the interarrival time is 1. In general the evaluation of the steady state distribution of $D/G/c/c$-queues cannot be done in polynomial time if it can be done at all. For this specific case the evaluation can be done in $\mathcal{O}(L_{\text{max}}^3)$ time. To see this note that the number of times we compute recursion (3.38) including the initialization before we obtain $\varphi_{\kappa,\lambda}$ is given by

$$\sum_{i=2}^{L_{\text{max}}+1} \sum_{x=2}^i x^2 = \sum_{i=2}^{L_{\text{max}}+1} \frac{2i^3-i^2}{6} = \frac{1}{2} \sum_{i=2}^{L_{\text{max}}+1} i^2 + \frac{1}{2} \sum_{i=2}^{L_{\text{max}}+1} i - L_{\text{max}}$$ (3.40)

Thus the complexity is $\mathcal{O}(L_{\text{max}}^3)$. 

Now that the joint distribution of $K$ and $\Lambda$ is known, we can compute the approximation for $\Pr(\sum_{i \in V_n} Q_i^r = x|A_n = y)$ by conditioning on the values of $|U_n|$ and $|V_n|$:

$$\vartheta_{x,y} = \Pr(\sum_{i \in V_n} Q_i^r = x|A_n = y) \approx \Pr\left(\sum_{i=n-\Lambda+1}^{n} D_i = x|\sum_{i=n-K+1}^{n} D_i = y\right)$$

$$= \sum_{\kappa=1}^{L_{\text{max}}} \sum_{\lambda=1}^{\kappa} \Pr(\Lambda = \lambda|K = \kappa) \Pr\left(\sum_{i=n-K+1}^{n} D_i = y\right) \times \Pr\left(\sum_{i=n-\lambda+1}^{n} D_i | \sum_{i=n-K+1}^{n} D_i = y\right)$$

$$= \sum_{\kappa=1}^{L_{\text{max}}} \sum_{\lambda=1}^{\kappa} \Pr(\Lambda = \lambda|K = \kappa) \Pr\left(\sum_{i=n-K+1}^{n} D_i = y\right) \times \frac{\Pr\left(D^{(\lambda)} = x\right) \Pr\left(D^{(\kappa-\lambda)} = y-x\right)}{\Pr\left(D^{(\kappa)} = y\right)} \times \sum_{\zeta=1}^{L_{\text{max}}} \Pr\left(D^{(\zeta)} = y\right) \Pr\left(K = \zeta\right).$$  (3.41)

In expression (3.41) the probability $\Pr(K = \kappa|\sum_{i=n-K+1}^{n} D_i = y)$ is obtained by applying Bayes’ theorem:

$$\Pr\left(\sum_{i=n-K+1}^{n} D_i = y\right) = \frac{\Pr\left(D^{(K)} = y|K = \kappa\right) \Pr(K = \kappa)}{\Pr\left(D^{(K)} = y\right)} = \frac{\Pr\left(D^{(\kappa)} = y\right) \Pr(K = \kappa)}{\sum_{\zeta=1}^{L_{\text{max}}} \Pr\left(D^{(\zeta)} = y\right) \Pr(K = \zeta)},$$  (3.42)

while the probabilities $\Pr(\Lambda = \lambda|K = \kappa)$ are easily obtained from $\varphi_{\kappa,\lambda}$, the joint density of $\Lambda$ and $K$.

Using this approximation for $\vartheta_{x,y}$ we can compute an approximation for $\Pr(A = x)$ by solving the set of linear equations (3.32). Then by using relation (3.10) we obtain an approximation for the distribution of $O$ as $\Pr(O = x) = \Pr(A = \Delta - x)$.

### 3.3. Numerical study

In this section we report on a numerical study to test the accuracy of the Markov Chain approximation that we propose. To this end a test bed of 1440 instances to problem $P$ was created that is a full factorial design of the parameter settings summarized in table 3.1. The demand distributions we used are the discrete two moment fits suggested by Adan et al. (1996), while the different types of distributions for $L$ are defined in table 3.2 and shown in figure 3.2.

First we performed the optimization using the approximate Markov Chain approach. This approach found the parameters $S^r_{r,MC}$ and $S^e_{e,MC}$ to be optimal. We then evaluated the
3.3. NUMERICAL STUDY

Table 3.1: Test-bed of instances with stochastic regular lead time

<table>
<thead>
<tr>
<th>Parameter settings</th>
<th>E[D]</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>c_D^2</td>
<td>\frac{1}{4}, \frac{1}{2}, 1, \frac{3}{2}, 2</td>
<td></td>
</tr>
<tr>
<td>t_e</td>
<td>1, 2</td>
<td></td>
</tr>
<tr>
<td>E[L]</td>
<td>4, 8, 12</td>
<td></td>
</tr>
<tr>
<td>h</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>10, 20, 30, 40</td>
<td></td>
</tr>
<tr>
<td>\gamma_0</td>
<td>0.95, 0.98</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.2: Distribution types for L

<table>
<thead>
<tr>
<th>Distribution Type \ x</th>
<th>Pr(L = x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>U1 (uniform1)</td>
<td>0</td>
</tr>
<tr>
<td>U2 (uniform2)</td>
<td>\frac{1}{5}</td>
</tr>
<tr>
<td>S1 (symmetric1)</td>
<td>0</td>
</tr>
<tr>
<td>S2 (symmetric2)</td>
<td>\frac{1}{10}</td>
</tr>
<tr>
<td>LS (left skewed)</td>
<td>0</td>
</tr>
<tr>
<td>RS (right skewed)</td>
<td>\frac{1}{10}</td>
</tr>
</tbody>
</table>

cost function and service level for S_{r,MC}^* and S_{e,MC}^* by both simulation and our Markov Chain approach. The simulation was ran such that the half-width of a 99%-confidence interval was less than 1% of the obtained estimate. We considered the measures

\[ \Delta_C := \frac{|C_{MC}(S_{r,MC}^*, S_{e,MC}^*) - C_{sim}(S_{r,MC}^*, S_{e,MC}^*)|}{C_{sim}(S_{r,MC}^*, S_{e,MC}^*)} \cdot 100\% \quad (3.43) \]

and

\[ \delta_\gamma := |\gamma_{MC}(S_{r,MC}^*, S_{e,MC}^*) - \gamma_{sim}(S_{r,MC}^*, S_{e,MC}^*)|. \quad (3.44) \]

Table 3.3 shows the averages of these quantities along with their relative frequencies (in %) for the indicated intervals. We remark that in most cases the Markov Chain estimate was an underestimation of the costs.

The Markov Chain approximation is quite accurate (4.16%) on average but substantial deviations do occur. The table shows that estimates become more accurate as demand
variability increases. An explanation for this is that as demand variability increases, the system becomes more chaotic and the effects that we do not model explicitly become less important.

Our approximation also becomes more accurate when the emergency lead time increases. This is in line with expectation because holding cost can also be written as $h\mathbb{E}[S_e + O - D^{(f+1)}]$, from which we see that the demand distribution (which we know exactly) influences holding cost more when $l^e$ is large. Accuracy also increases when the expedition premium goes up. This is because expediting becomes less attractive when $c$ goes up, so that $\Delta$ becomes larger and our approximation works better. That our approximation becomes less accurate as $\mathbb{E}[L]$ increases can be explained again by inspecting the holding cost $h\mathbb{E}[S_e + O - D^{(f+1)}]$. The contribution of $O$, which we know only approximately, compared to $D^{(f+1)}$ becomes smaller when $\mathbb{E}[L]$ decreases. This is because $\Delta$ (and therefore also $O$) increases with $\mathbb{E}[K] = \mathbb{E}[L]$ (by Little’s law). The target service level and different distributions for $L$ do not influence accuracy much.

Next we are interested in how close to optimal the approximate Markov Chain solution we found is. To this end we performed the same optimization but determined the overshoot distribution through simulation. This simulation optimization procedure yielded $S_{r,sim}^*$ and
### 3.3. NUMERICAL STUDY

Table 3.3: Accuracy of cost and service estimates for stochastic lead times

<table>
<thead>
<tr>
<th>Distribution type</th>
<th>Error in Cost, ( \Delta C ) [%]</th>
<th>Deviation in service, ( \delta_\gamma ) ( \times 10^{-3} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( c_{2D} )</td>
<td>avg. 0-2 2-4 4-6 6-8 &gt; 8</td>
<td>avg. 0-2 2-4 4-6 6-8 &gt; 8</td>
</tr>
<tr>
<td>( \frac{1}{4} )</td>
<td>8.52 5.2 7.3 13.2 17.0 57.3</td>
<td>4.96 16.3 34.0 19.1 12.2 18.4</td>
</tr>
<tr>
<td>( \frac{1}{2} )</td>
<td>5.37 2.1 24.0 35.4 30.9 7.6</td>
<td>3.55 25.0 42.0 16.3 12.8 3.8</td>
</tr>
<tr>
<td>1</td>
<td>3.22 10.1 68.8 21.2 0.0 0.0</td>
<td>2.50 46.2 36.5 13.9 3.1 0.3</td>
</tr>
<tr>
<td>( \frac{3}{2} )</td>
<td>2.11 42.7 56.9 0.3 0.0 0.0</td>
<td>1.71 67.7 26.7 5.6 0.0 0.0</td>
</tr>
<tr>
<td>2</td>
<td>1.56 80.9 19.1 0.0 0.0 0.0</td>
<td>1.72 67.7 26.4 5.2 0.3 0.3</td>
</tr>
<tr>
<td>( r^c )</td>
<td>avg. 0-2 2-4 4-6 6-8 &gt; 8</td>
<td>avg. 0-2 2-4 4-6 6-8 &gt; 8</td>
</tr>
<tr>
<td>1</td>
<td>4.47 22.9 36.7 15.4 9.7 15.3</td>
<td>2.93 44.0 32.6 12.5 6.4 4.4</td>
</tr>
<tr>
<td>2</td>
<td>3.85 33.5 33.8 12.6 9.4 10.7</td>
<td>2.84 45.1 33.6 11.5 5.0 4.7</td>
</tr>
<tr>
<td>( c )</td>
<td>avg. 0-2 2-4 4-6 6-8 &gt; 8</td>
<td>avg. 0-2 2-4 4-6 6-8 &gt; 8</td>
</tr>
<tr>
<td>10</td>
<td>5.13 19.7 33.1 15.6 12.5 19.2</td>
<td>3.41 40.0 31.7 11.9 8.3 8.1</td>
</tr>
<tr>
<td>20</td>
<td>4.30 25.3 36.4 14.7 9.4 14.2</td>
<td>3.06 41.7 33.9 13.9 4.7 5.8</td>
</tr>
<tr>
<td>30</td>
<td>3.80 31.4 35.8 13.3 8.1 11.4</td>
<td>2.61 48.1 33.1 11.1 4.7 3.1</td>
</tr>
<tr>
<td>40</td>
<td>3.40 36.4 35.6 12.5 8.3 7.2</td>
<td>2.47 48.6 33.9 11.1 5.0 1.4</td>
</tr>
<tr>
<td>( \gamma_0 )</td>
<td>avg. 0-2 2-4 4-6 6-8 &gt; 8</td>
<td>avg. 0-2 2-4 4-6 6-8 &gt; 8</td>
</tr>
<tr>
<td>0.95</td>
<td>4.28 24.7 37.8 14.0 10.4 13.1</td>
<td>3.91 29.7 30.1 19.6 11.4 9.2</td>
</tr>
<tr>
<td>0.98</td>
<td>4.04 31.7 32.6 14.0 8.8 12.9</td>
<td>1.86 59.4 36.1 4.4 0.0 0.0</td>
</tr>
<tr>
<td>( E[L] )</td>
<td>avg. 0-2 2-4 4-6 6-8 &gt; 8</td>
<td>avg. 0-2 2-4 4-6 6-8 &gt; 8</td>
</tr>
<tr>
<td>4</td>
<td>2.61 47.3 34.8 12.5 4.2 1.3</td>
<td>1.91 61.5 30.8 6.5 0.8 0.4</td>
</tr>
<tr>
<td>8</td>
<td>4.51 21.9 35.2 16.7 11.7 14.6</td>
<td>2.97 39.4 37.7 12.5 6.9 3.5</td>
</tr>
<tr>
<td>12</td>
<td>5.35 15.4 35.6 12.9 12.9 23.1</td>
<td>3.79 32.9 30.8 17.1 9.4 9.8</td>
</tr>
<tr>
<td>Distribution type</td>
<td>avg. 0-2 2-4 4-6 6-8 &gt; 8</td>
<td>avg. 0-2 2-4 4-6 6-8 &gt; 8</td>
</tr>
<tr>
<td>U1</td>
<td>3.83 31.7 34.6 12.9 9.6 11.3</td>
<td>2.76 47.5 32.1 8.8 7.1 4.6</td>
</tr>
<tr>
<td>U2</td>
<td>4.28 27.5 34.2 14.6 10.0 13.8</td>
<td>2.88 44.2 33.3 12.1 7.1 3.3</td>
</tr>
<tr>
<td>S1</td>
<td>3.98 28.8 35.8 14.2 9.2 12.1</td>
<td>2.82 46.7 30.8 14.2 3.3 5.0</td>
</tr>
<tr>
<td>S2</td>
<td>4.56 24.2 36.7 13.8 9.2 16.3</td>
<td>3.03 41.3 35.4 10.4 8.3 4.6</td>
</tr>
<tr>
<td>LS</td>
<td>4.41 27.5 33.8 14.2 9.6 15.0</td>
<td>2.96 42.5 36.3 12.1 4.6 4.6</td>
</tr>
<tr>
<td>RS</td>
<td>3.88 29.6 36.3 14.6 10.0 9.6</td>
<td>2.88 45.4 30.8 14.6 3.8 5.4</td>
</tr>
<tr>
<td>Total</td>
<td>avg. 0-2 2-4 4-6 6-8 &gt; 8</td>
<td>avg. 0-2 2-4 4-6 6-8 &gt; 8</td>
</tr>
<tr>
<td></td>
<td>4.16 28.2 35.2 14.0 9.6 13.0</td>
<td>2.89 44.6 33.1 12.0 5.7 4.6</td>
</tr>
</tbody>
</table>

\( S^*_e,\text{sim} \) as optimal order-up-to-levels. We considered the cost deviation from the optimal DIP with \( S^*_r,\text{sim} \) and \( S^*_e,\text{sim} \), i.e.,

\[
\Delta C^* = \frac{C_{\text{sim}}(S^*_r,MC, S^*_e,MC) - C_{\text{sim}}(S^*_r,\text{sim}, S^*_e,\text{sim})}{C_{\text{sim}}(S^*_r,\text{sim}, S^*_e,\text{sim})} \cdot 100\%.
\]

(3.45)

Figure 3.3 shows a scatter plot of \( \Delta C^* \) versus \( \gamma_{\text{sim}}(S^*_r,MC, S^*_e,MC) - \gamma_0 \), along with their marginal histograms. Points close to the origin indicate that the solution is indeed very close to optimal. In fact solutions in the second quadrant indicate that The Markov Chain approach yielded solutions with better cost performance while still meeting the required
service level. As was the case in section 2.3, this is due to simulation tolerance. Most solutions however lie in the first quadrant indicating that our Markov Chain approach is usually on the safe side with respect to the required service level, at the expense of a little additional cost. In general we see that the quality of the solutions obtained by the Markov Chain approach is quite good. When accurate cost estimates are very important, it is possible to simulate the solution found by the Markov Chain approach such that a more accurate cost estimate is obtained.

Figure 3.3: Accuracy of Markov Chain approximation

To see how different problem parameters influence the quality of Markov Chain solution we inspect the distribution of $\Delta C^*$ for different problem parameters. Table 3.3 shows how $\Delta C^*$ is distributed for different problem parameters. The most prominent effect is that the quality of solutions deteriorates when demand variability (as measured by $c_D^2$) decreases. We already saw that the Markov Chain approximation of the cost function deteriorates as demand variability decreases. This would explain why solutions to these problems are of a lesser quality. We see that the maximum and minimum deviation in our entire test-bed
both occur when \( c_D^2 = \frac{1}{4} \). Since dual-sourcing is a way to buffer against demand variability, it is convenient that the Markov Chain approach appears to be better able to approach the optimum as demand variability increases.

We also see, as we did for \( \Delta C \) and presumably for the same reasons, that the quality of obtained solutions increases as:

- emergency lead time \( l^e \) increases

### Table 3.4: Quality of solutions compared to optimal DIP

<table>
<thead>
<tr>
<th>( c_D^2 )</th>
<th>avg.</th>
<th>min</th>
<th>max</th>
<th>&lt;</th>
<th>-1 -</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>&gt;5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \frac{1}{4} )</td>
<td>2.58</td>
<td>-5.36</td>
<td>6.51</td>
<td>7.6</td>
<td>4.5</td>
<td>11.8</td>
<td>11.1</td>
<td>15.6</td>
<td>19.1</td>
<td>18.4</td>
<td>11.8</td>
<td></td>
</tr>
<tr>
<td>( \frac{1}{2} )</td>
<td>1.31</td>
<td>-2.78</td>
<td>4.30</td>
<td>2.4</td>
<td>11.5</td>
<td>22.2</td>
<td>35.1</td>
<td>23.3</td>
<td>5.2</td>
<td>0.3</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.46</td>
<td>-3.25</td>
<td>2.97</td>
<td>4.5</td>
<td>22.2</td>
<td>50.0</td>
<td>20.5</td>
<td>2.8</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>( \frac{3}{2} )</td>
<td>0.40</td>
<td>-1.10</td>
<td>2.07</td>
<td>1.0</td>
<td>25.0</td>
<td>56.9</td>
<td>16.7</td>
<td>0.3</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.22</td>
<td>-1.41</td>
<td>1.52</td>
<td>1.4</td>
<td>33.0</td>
<td>58.7</td>
<td>6.9</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
</tbody>
</table>

### Distribution type

<table>
<thead>
<tr>
<th>avg.</th>
<th>min</th>
<th>max</th>
<th>&lt;</th>
<th>-1 -</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>&gt;5</th>
</tr>
</thead>
<tbody>
<tr>
<td>U1</td>
<td>0.88</td>
<td>-3.25</td>
<td>5.27</td>
<td>4.2</td>
<td>21.3</td>
<td>39.2</td>
<td>16.3</td>
<td>10.0</td>
<td>6.3</td>
<td>1.7</td>
<td>1.3</td>
</tr>
<tr>
<td>U2</td>
<td>0.98</td>
<td>-4.10</td>
<td>5.86</td>
<td>3.8</td>
<td>18.8</td>
<td>40.4</td>
<td>17.9</td>
<td>6.3</td>
<td>7.1</td>
<td>3.8</td>
<td>2.1</td>
</tr>
<tr>
<td>S1</td>
<td>1.01</td>
<td>-2.78</td>
<td>6.40</td>
<td>4.6</td>
<td>17.1</td>
<td>40.0</td>
<td>20.0</td>
<td>6.7</td>
<td>4.6</td>
<td>4.6</td>
<td>2.5</td>
</tr>
<tr>
<td>S2</td>
<td>1.13</td>
<td>-2.36</td>
<td>6.26</td>
<td>2.5</td>
<td>19.6</td>
<td>37.5</td>
<td>16.3</td>
<td>10.8</td>
<td>6.7</td>
<td>2.9</td>
<td>3.8</td>
</tr>
<tr>
<td>LS</td>
<td>1.06</td>
<td>-1.23</td>
<td>6.51</td>
<td>2.1</td>
<td>20.4</td>
<td>38.8</td>
<td>19.6</td>
<td>7.9</td>
<td>2.1</td>
<td>6.7</td>
<td>2.5</td>
</tr>
<tr>
<td>RS</td>
<td>0.90</td>
<td>-5.36</td>
<td>5.42</td>
<td>3.3</td>
<td>18.3</td>
<td>43.8</td>
<td>18.3</td>
<td>8.8</td>
<td>2.5</td>
<td>2.9</td>
<td>2.1</td>
</tr>
</tbody>
</table>

### Total

<table>
<thead>
<tr>
<th>avg.</th>
<th>min</th>
<th>max</th>
<th>&lt;</th>
<th>-1 -</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>&gt;5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.99</td>
<td>-5.36</td>
<td>6.51</td>
<td>3.4</td>
<td>19.2</td>
<td>39.9</td>
<td>18.1</td>
<td>8.4</td>
<td>4.9</td>
<td>3.8</td>
<td>2.4</td>
<td></td>
</tr>
</tbody>
</table>
• the emergency premium $c$ increases

• the stochastic part of $E[L']$, $E[L]$ decreases

The service level and the distribution type of $L$ do not influence the quality of solutions markedly.

Finally we remark that the Markov Chain approach was between 30-50 times faster than the simulation based approach. The Markov Chain approach is just as fast as it was for the setting with deterministic lead times because the bottleneck computation remains the solution of linear systems of equations. Simulation needs more time than before because two sources of randomness have to be simulated such that random number generation increases as well as the simulation time needed before the overshoot distribution converges.
Chapter 4

Comparison of models

In chapters 2 and 3 we presented models where the regular lead times were deterministic and stochastic respectively. Implementing the model with deterministic regular lead times is less involved than the model with stochastic regular lead times. Thus one is justified in asking what the benefits are of implementing the more complicated model. In this Section we address this question with a numerical study, and provide managerial insights.

Within the class of base stock policies for periodic inventory control it is commonplace to assume deterministic lead times. One reason for this is that base stock type policies are optimal in many settings with deterministic lead times. When lead times are stochastic, base stock policies are not usually optimal. Due to their appealing structure however base stock policies are also implemented in these situations which are very common in practice. Robinson et al. (2001) study how lead time stochasticity can be incorporated for single sourcing base stock policies, and show that properly incorporating lead time variability significantly improves the quality of solutions. In this section we provide a numerical study that investigates how inclusion of lead time variability in the model affects the quality of solutions. We also investigate how we should set $l_{Det}$, the deterministic lead time difference, when we know the distribution of $L$, but opt to use the model with deterministic lead times.

4.1. Experiment and results

Consider again the test-bed defined in section 3.3 in tables 3.1 and 3.2. For each case in this test-bed we determine $S_e$ and $S_r$ under the assumption that $L$ is deterministic and given by $l_{Det} = E[L]$ or $l_{Det} = L_{max}$ and denote the optimal parameter setting as $S_{r,MC,l_{Det}}=E[L]$, $S_{e,MC,l_{Det}}=E[L]$ and $S_{r,MC,l_{Det}}=L_{max}$, $S_{e,MC,l_{Det}}=L_{max}$ respectively. Then we simulate the perfor-
mance of these solutions for the original case (that includes stochastic regular lead times). We compared the performance of this policy against the optimal DIP for this situation, found via simulation optimization. To compare costs we considered the relative deviation from the optimal costs

\[
\Delta C_\ast = \frac{C_{\text{sim}}(S_{e,MC,a}, S_{e,MC,a}) - C_{\text{sim}}(S_{r,\text{sim}}, S_{e,\text{sim}})}{C_{\text{sim}}(S_{r,\text{sim}}, S_{e,\text{sim}})} \cdot 100\%. \tag{4.1}
\]

Here \(a\) is either \(l_{\text{Det}} = \mathbb{E}[L]\) denoting that the deterministic model was used with the mean lead time as deterministic lead time parameter, or \(l_{\text{Det}} = L_{\text{max}}\) indicating that the maximum lead time was used as deterministic lead time parameter. Further when \(a = L\) the model with stochastic lead times was used. Note that the case \(a = L\) was reported on in Section 3.3.

Figure 4.1: Comparison of models

Figure 4.1 shows plots of \(\Delta C_\ast\) versus \(\gamma_{\text{sim}}(S_{e,MC,a}, S_{e,MC,a}) - \gamma_0\). The scales for \(L\) and \(l_{\text{Det}} = \mathbb{E}[L]\) are identical to facilitate comparison while the scale for \(l_{\text{Det}} = L_{\text{max}}\) is different because its performance differs drastically from the other two settings. First of all we note that the setting with \(l_{\text{Det}} = L_{\text{max}}\) performs rather poorly. Both costs and service levels are
4.1. EXPERIMENT AND RESULTS

extremely high compared to the optimal DIP. In the scatter plot associated with \( l_{\text{Det}} = L_{\text{max}} \) we clearly see two groups of data points. These groups correspond to problem instances with \( \gamma_0 = 0.95 \) and \( \gamma_0 = 0.98 \) because the maximum possible deviation for these groups are 0.05 and 0.02, as \( \gamma \in (-\infty, 1) \).

The performance for \( l_{\text{Det}} = E[L] \) is much better than for \( l_{\text{Det}} = L_{\text{max}} \) but there are many instances where the realized service level is much to low compared to the target service level. The figure clearly shows that there is a tendency to have lower costs than the optimal DIP at the expense of not reaching the required service level.

The performance is best for our Markov Chain approach that does incorporate stochastic lead times. In this case required service levels are usually met and never missed by more than 0.0082 while cost deviations are all within 6.51% (see table 3.3).

<table>
<thead>
<tr>
<th>( c^2_{ij} )</th>
<th>( \gamma_0 )</th>
<th>( E[L] )</th>
<th>Distribution type</th>
<th>U1</th>
<th>U2</th>
<th>S1</th>
<th>S2</th>
<th>LS</th>
<th>RS</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>avg</td>
<td>min</td>
<td>max</td>
<td>(&lt; -10 )</td>
<td>(-10 - -8 )</td>
<td>(-8 - -6 )</td>
<td>(-6 - -4 )</td>
<td>(-4 - -2 )</td>
<td>(-2 - 0 )</td>
<td>(0 - 2 )</td>
<td>(&gt; 2 )</td>
</tr>
<tr>
<td>( \frac{1}{4} )</td>
<td>-0.16</td>
<td>-22.89</td>
<td>6.59</td>
<td>15.6</td>
<td>7.6</td>
<td>6.9</td>
<td>11.5</td>
<td>17.0</td>
<td>18.4</td>
<td>13.2</td>
</tr>
<tr>
<td>( \frac{1}{2} )</td>
<td>-2.04</td>
<td>-11.34</td>
<td>4.43</td>
<td>1.4</td>
<td>1.7</td>
<td>6.3</td>
<td>13.5</td>
<td>19.4</td>
<td>35.4</td>
<td>13.9</td>
</tr>
<tr>
<td>1</td>
<td>-1.25</td>
<td>-7.18</td>
<td>4.06</td>
<td>0.0</td>
<td>0.0</td>
<td>0.3</td>
<td>4.5</td>
<td>21.9</td>
<td>55.2</td>
<td>14.6</td>
</tr>
<tr>
<td>( \frac{1}{2} )</td>
<td>-0.53</td>
<td>-2.96</td>
<td>2.91</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>5.6</td>
<td>72.6</td>
<td>18.8</td>
</tr>
<tr>
<td>2</td>
<td>-0.42</td>
<td>-2.52</td>
<td>2.46</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.7</td>
<td>76.4</td>
<td>21.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( l^c )</th>
<th>( c )</th>
<th>( \gamma_0 )</th>
<th>( E[L] )</th>
<th>U1</th>
<th>U2</th>
<th>S1</th>
<th>S2</th>
<th>LS</th>
<th>RS</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>avg</td>
<td>min</td>
<td>max</td>
<td>(&lt; -10 )</td>
<td>(-10 - -8 )</td>
<td>(-8 - -6 )</td>
<td>(-6 - -4 )</td>
<td>(-4 - -2 )</td>
<td>(-2 - 0 )</td>
<td>(0 - 2 )</td>
<td>(&gt; 2 )</td>
</tr>
<tr>
<td>1</td>
<td>-1.73</td>
<td>-22.89</td>
<td>6.59</td>
<td>3.5</td>
<td>1.4</td>
<td>2.8</td>
<td>6.9</td>
<td>13.1</td>
<td>50.8</td>
<td>16.3</td>
</tr>
<tr>
<td>2</td>
<td>-1.63</td>
<td>-20.97</td>
<td>5.80</td>
<td>3.3</td>
<td>2.4</td>
<td>2.6</td>
<td>4.9</td>
<td>12.8</td>
<td>52.4</td>
<td>16.4</td>
</tr>
<tr>
<td>10</td>
<td>-0.43</td>
<td>-10.77</td>
<td>6.59</td>
<td>0.3</td>
<td>0.8</td>
<td>0.3</td>
<td>3.6</td>
<td>7.8</td>
<td>51.4</td>
<td>29.4</td>
</tr>
<tr>
<td>20</td>
<td>-1.33</td>
<td>-16.86</td>
<td>5.80</td>
<td>1.7</td>
<td>1.7</td>
<td>1.9</td>
<td>6.4</td>
<td>12.8</td>
<td>52.8</td>
<td>16.4</td>
</tr>
<tr>
<td>30</td>
<td>-2.16</td>
<td>-22.67</td>
<td>4.97</td>
<td>4.7</td>
<td>2.2</td>
<td>4.2</td>
<td>6.4</td>
<td>12.5</td>
<td>56.1</td>
<td>8.9</td>
</tr>
<tr>
<td>40</td>
<td>-2.79</td>
<td>-22.89</td>
<td>5.25</td>
<td>6.9</td>
<td>2.8</td>
<td>4.4</td>
<td>7.2</td>
<td>18.6</td>
<td>46.1</td>
<td>10.6</td>
</tr>
<tr>
<td>0.95</td>
<td>-2.00</td>
<td>-22.89</td>
<td>6.59</td>
<td>5.0</td>
<td>1.9</td>
<td>3.5</td>
<td>6.5</td>
<td>12.5</td>
<td>49.7</td>
<td>15.7</td>
</tr>
<tr>
<td>0.98</td>
<td>-1.35</td>
<td>-22.77</td>
<td>5.76</td>
<td>1.8</td>
<td>1.8</td>
<td>1.9</td>
<td>5.3</td>
<td>13.3</td>
<td>53.5</td>
<td>16.9</td>
</tr>
</tbody>
</table>

Table 4.1: Quality of solutions compared to optimal DIP for \( l_{\text{Det}} = E[L] \)
As before we now turn attention to the way various problem parameters influence the cost performance of obtained solutions. We remind the reader that table 3.3 contains the relative cost deviations of the Markov Chain approach that incorporates stochastic lead times, $\Delta_{MC,L}^\ast$.

Table 4.1 shows the distributions of $\Delta_{C}^\ast$ for all problem parameters when $l_{Det} = E[L]$. We see that performance improves relatively as $c_2^D$ increases. An explanation for this effect is that demand variability plays a more dominant role compared to lead time variance as $c_2^D$ increases, and the deterministic model does incorporate this variability.

The same argument applies for $E[L]$. As $E[L]$ decreases, $c_2^L$ increases, because the variance of lead time distributions remains constant. Thus as $E[L]$ decreases the amount of lead time variability increases relative to the amount of demand variability. In this situation also the Markov Chain approach that incorporates lead time variability is better able to approach the optimal DIP.

Another feature of this table stands out. The only positive cost deviation on average occurs when the lead time has a U2 distribution. Maximum cost deviation and percentage of deviation above 2% is also high for this distribution type. Distribution type U2 has the highest variance among the lead time distributions in the test bed. This is a typical finding. It implies that when lead time variability is high, the Markov Chain approach with $l_{Det} = E[L]$ finds solutions with greater costs than the optimal DIP, often without reaching the required service level. In this situation the Markov Chain approach that does incorporate stochastic lead times is especially valuable.

In the general the effects of other problem parameters is not very pronounced. Small effect that we do observe are (i) performance improves as $l^e$ increases and (ii) performance deteriorates as $c$ increases. Part (i) can be explained as before by inspecting the holding costs given by $hE \left( (S_e + O - D(l^e+1))^+ \right)$. As $l^e$ increases the contribution of $O$, whose distribution is known only approximately, becomes less compared to the contribution of $D(l^e+1)$ which we know exactly. Part (ii) can be explained by observing that the Markov Chain approach with $l_{Det} = E[L]$ usually finds a smaller $\Delta_{l_{Det}=E[L]} = S_{e,MC,l_{Det}=E[L]} - S_{e,MC,} = E[L]$ to be optimal than the true optimal $\Delta^* = S_{e,sim}^* - S_{e,sim}^*$. As a result $E[A]$ decreases relative to optimum and so $E[Q^*] = E[D] - \frac{E[A]}{E[L]}$ increases. Since emergency ordering costs are given by $cE[Q^*]$, emergency ordering costs when $l_{Det} = E[L]$ decrease relative to the optimum as $c$ increases.

Using a model with deterministic lead times where $l_{Det} = L_{max}$ performs very poorly as figure 4.1 shows. Table 4.1 shows that the ability to approach the optimal DIP improves
as $c^2_D$ increases and $c^2_L$ decreases. The reasons for this are analogous to the reasons for the same effect when $l_{Det} = \mathbb{E}[L]$. Overall however, the performance is so poor that using the deterministic model with $l_{Det} = L_{\text{max}}$ is very inadvisable.

### 4.2. Conclusion

The numerical results in this section have three implications:

1. It is essential to incorporate lead time variance in the model when lead time variability is large relative to demand variability, especially when meeting service level requirements is imperative;

2. Using the maximum lead time or a high percentile of the lead time distribution as
deterministic lead time parameter gravely increases both costs and service and is very far from optimal;

3. When lead time variability is small compared to demand variability the model with deterministic lead times yields reasonable results provided the mean lead time is used as deterministic lead time parameter and service level requirements are not strict.

Implication (1) is important from a practical point of view. Many results in inventory theory are based on the assumption of deterministic lead times. In practice this assumption often does not hold and this significantly affects the performance of the inventory system, in a negative manner. Since the managers of many inventory systems are bound to Service Level Agreements (SLA’s) with customers they oftentimes use the maximum or a high percentile of the lead time distribution as a deterministic lead time parameter in inventory models. Implication (2) states that this practice leads to very poor performance in terms of cost and usually renders higher service than the SLA requires. Properly dealing with this situation requires modeling lead time stochasticity explicitly. The circumstances under which one may opt for a model with deterministic lead times, and the manner of use are prescribed by implication (3).
Chapter 5

Conclusion

In this thesis we presented two models. The first model deals with the Dual-Index policy for a single stage dual sourcing inventory system facing stochastic demand with deterministic lead times and uses an efficient approximate Markov Chain. Our main contributions here are to (i) provide an approximate evaluation method of the dual index policy using Markov Chains that does not require simulation, thus making optimization more efficient and (ii) providing an alternate and insightful proof of the separability result that reduces the optimization of the DIP to two one-dimensional optimization problems.

The second model we presented was a generalization of the first by allowing regular lead times to be stochastic. In this situation we defined a Dual-Index policy with mild informational requirements on the realizations of regular lead times. We proved that the same separability result holds as for the model with deterministic lead times. An approximate evaluation method using Markov Chains was also constructed for this more complex system.

Finally we compared the models with and without stochastic lead times for a setting with stochastic lead times. We found that models with deterministic lead times should not normally be used in settings where lead time is stochastic. We also identified situations where a model with deterministic lead times suffices even when lead time stochasticity is present and provided guidelines on how to apply the model with deterministic lead times.

5.1. Directions for future research

The research in this thesis can be extended in several important ways. The most obvious and possibly useful extension is to define and analyze the dual-index policy for multi-echelon inventory systems. Consider for example a serial supply chain. Clark & Scarf (1960) showed
that base-stock policies are optimal for this system and that the optimal base-stock levels can be obtained by successively solving newsvendor equations. This decomposition result relies on the fact that all stock points in a serial supply chain face the same demand process. When the most downstream stock point is the only stock point with two sources, this property is retained. In that case finding the optimal echelon-DIP should be a straightforward task using the results in this thesis.

When stock points other than the most downstream stock point have two sources the property that each stock point essentially faces the same demand process is not preserved, because some of the demand is ordered via the second source. Inventory control for this type of system is an interesting new research direction.

Apart from serial supply chains one may consider assembly systems (convergent chain structure) or distribution systems (divergent chain structure) with dual sourcing.

Assembly systems are especially interesting because synchronization of orders is possible even in the presence of dual sourcing. We refer to Rosling (1989) and Schmidt & Nahmias (1985) for an exposition on synchronization in assembly systems. To see that synchronization is also possible in an assembly system with dual sourcing consider the example in figure 5.1 If only regular lead times were present in this system, ordering component 2 would be

Figure 5.1: Example of an assembly system with dual-sourcing

synchronized with previous ordering decisions for component 1. With the availability of two sources however this can also be done the other way around. When an order is placed at the regular source of component 2, the emergency source of component 1 can be used to synchronize with this order. The definition and analysis of a synchronized Dual-Index type policy for this type of system is a promising research direction.

In the single stock point setting several important extensions are possible. One may consider other lead time processes that do not allow for crossover of orders, or non-stationarity of demand distributions. Finally one may investigate the use of similar approximation tech-
niques in the context of remanufacturing systems.
Bibliography


Appendix A

Proofs

Proof of lemma 2.2

The emergency inventory position satisfies the following by definition:

\[ IP_{n+1}^e = IP_n^e + Q_n^e - D_n + Q_{n+1-l}^r \]  \hspace{1cm} \text{by rewriting (2.1)} \tag{A.1}

\[ = S_e + O_n - D_n + Q_{n+1-l}^r \]  \hspace{1cm} \text{by substitution of equation (2.5)}.

Now for the overshoot we have by rewriting its definition (equation (2.5))

\[ O_{n+1} = (IP_{n+1}^e - S_e)^+ \]
\[ = (S_e + O_n - D_n + Q_{n+1-l}^r - S_e)^+ \tag{A.2} \]
\[ = (O_n - D_n + Q_{n+1-l}^r)^+ . \]

Similarly for the emergency order quantity we have by rewriting (2.2):

\[ Q_{n+1}^e = (S_e - IP_{n+1}^e)^+ \]
\[ = (S_e - S_e - O_n + D_n - Q_{n+1-l}^r)^+ \tag{A.3} \]
\[ = (D_n - O_n - Q_{n+1-l}^r)^+ . \]

The identity \( Q_{n+1}^r = D_n - Q_{n+1-l}^e \) follows immediately from the fact that the DIP ensures that in each period the total amount ordered equals demand from the previous period. \( \square \)

Proof of lemma 3.2

The emergency inventory position satisfies

\[ IP_{n+1}^e = IP_n^e + Q_n^e - D_n + \sum_{i \in V_n} Q_i^r \]
\[ = S_e + O_n - D_n + \sum_{i \in V_n} Q_i^r . \]  \hspace{1cm} \tag{A.4}
Rewriting the definition of the overshoot we obtain

\[ Q^e_{n+1} = (S_e - IP^e_{n+1})^+ \]
\[ = (S_e - S_e - O_n + D_n - \sum_{i \in V_n} Q^r_i)^+ \]
\[ = (D_n - O_n - \sum_{i \in V_n} Q^r_i)^+. \]  

(A.5)

Similarly for the emergency order quantity we have by rewriting (3.6):

\[ Q^e_{n+1} = (S_e - IP^e_{n+1})^+ \]
\[ = (S_e - S_e - O_n + D_n - \sum_{i \in V_n} Q^r_i)^+ \]
\[ = (D_n - O_n - \sum_{i \in V_n} Q^r_i)^+. \]  

(A.6)

The identity \( Q^r_{n+1} = D_n - Q^e_{n+1} - l \) follows immediately from the fact that the DIP ensures that in each period the total amount ordered equals demand from the previous period. \( \square \)
Appendix B

Fitting a quasi-uniform distribution

In this appendix we describe how one may fit a quasi-uniform distribution on the first two moments of a discrete random variable. For the model in Chapter 3 the regular lead time is modeled as having a discrete distribution with finite support on $\mathbb{N}$. In some instances this discrete distribution may be estimated from past regular lead time realizations, i.e., one may use the empirical regular lead time distribution. It occurs regularly however that he only available information concerning the regular lead times are the first two moments. In such a situation one would like to fit a distribution with finite support on these moments. Furthermore one may want to investigate the effect of lead time variability on the performance of a dual-sourcing inventory system. In this situation it is convenient if lead time distributions with different means and variance and with finite support can be constructed routinely.

The choice of fitting a quasi uniform distribution has three justifications. First, it is the author’s experience that lead time data from practice often closely resembles a uniform distribution. Second the performance of inventory systems is often relatively insensitive to the exact shape of the distributions of random variables involved, (e.g. Naddor (1978) and Fortuin (1980)). Third the quasi-uniform distribution has finite support which is essential in lending tractability to the model in chapter 3.

Now we introduce the quasi-uniform random variable. The random variable $U \in \mathbb{Z}$ is said to have a quasi-uniform distribution between $a \in \mathbb{Z}$ and $b \in \mathbb{Z}$ ($a \leq b$) if

$$
\Pr(U = u) = \begin{cases} 
  p_a, & \text{if } u = a; \\
  p_b, & \text{if } u = b; \\
  \frac{1-p_a-p_b}{b-a-1}, & \text{if } a < u < b; \\
  0, & \text{otherwise.}
\end{cases}
$$

(B.1)

Notice that the distribution of $U$ is uniform except possibly at its boundaries $a$ and $b$, hence
Let \( \lfloor x \rfloor \) (\( \lceil x \rceil \)) denote \( x \) rounded down (up) to the nearest integer. Suppose the random variable \( X \) has mean \( \mu \) and squared coefficient of variance \( c^2_X \). Then the first two moments of \( U \) match the first two moments of \( X \) if \( a, b, p_a \) and \( p_b \) are chosen such that

\[
\begin{align*}
a &= \mu - \frac{1}{2} \left\lfloor \mu \right\rfloor \left( \sqrt{1 + 12 c^2_X \mu^2} \right) \\
b &= \mu + \frac{1}{2} \left\lceil \mu \right\rceil \left( \sqrt{1 + 12 c^2_X \mu^2} \right) \tag{B.2} \\
p_b &= \frac{E - BD/A}{B + C} \tag{B.4} \\
p_a &= \frac{D/A + p_b}{B + C} \tag{B.5}
\end{align*}
\]

where

\[
\begin{align*}
A &= \frac{a - b}{2} \tag{B.6} \\
B &= a^2 - \left(\frac{-1}{6}a - \frac{1}{6} + \frac{1}{6}b\right)(2a^2 + a + 2ab - b + 2b^2) \tag{B.7} \\
C &= b^2 - \left(\frac{-1}{6}a - \frac{1}{6} + \frac{1}{6}b\right)(2a^2 + a + 2ab - b + 2b^2) \tag{B.8} \\
D &= \mu - \frac{a + b}{2} \tag{B.9} \\
E &= (c^2_X + 1)\mu^2 - \left(\frac{-1}{6}a - \frac{1}{6} + \frac{1}{6}b\right)(2a^2 + a + 2ab - b + 2b^2) \tag{B.10} \\
\end{align*}
\]

There does not exist a discrete distribution for every combination of \( \mu \) and \( c^2_X \). By using the fact that for any mean \( \mu \in \mathbb{R} \) the discrete distribution on \( \mathbb{Z} \) that has minimal variance is concentrated on the integers \( \lfloor \mu \rfloor \) and \( \lceil \mu \rceil \) it is easy to verify that a quasi-uniform (or indeed any discrete distribution on the integers) can only be fitted on the moments \( \mu \) and \( c^2_X \) if

\[
c^2_X \mu^2 \geq (\mu - \lfloor \mu \rfloor)(1 + \mu - \lceil \mu \rceil)^2 + (\mu - \lfloor \mu \rfloor)^2(1 + \mu - \lceil \mu \rceil). \tag{B.11}
\]
The fitting procedure in this appendix works for all $\mu$ and $c_X^2$ that satisfy the conditions described in the previous paragraph. However for the application in this thesis we also require that the distribution has a support on $\mathbb{N}$ only. This condition is satisfied if $a \geq 1$. For most practical applications this condition is verified.