

Efficient optimization of the dual-index policy using Markov chains

Citation for published version (APA):

Arts, J. J., Vuuren, van, M., & Kiesmüller, G. P. (2009). *Efficient optimization of the dual-index policy using Markov chains*. (BETA publicatie : working papers; Vol. 291). Technische Universiteit Eindhoven.

Document status and date:

Published: 01/01/2009

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Efficient Optimization of the Dual-Index Policy Using Markov Chains

Joachim Arts^{1,*}, Marcel van Vuuren², Gudrun Kiesmüller¹

¹Eindhoven University of Technology, School of Industrial Engineering
P.O. Box 513, 5600 MB Eindhoven, the Netherlands

²Centre for Quantitative Methods,
P.O. Box 414, 5600 AK Eindhoven

October 8, 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 system is known to be complex. For this reason we study a type of base-stock policy known as the dual-index policy (DIP) as control mechanism for this inventory system. Under this policy ordering decisions are based on a regular and an emergency inventory position and their corresponding order-up-to-levels. Previous work on this policy assumes deterministic lead times and uses simulation to find the optimal order-up-to levels. We provide an alternate proof for the result that separates the optimization of the DIP in two one-dimensional problems. An insight from this proof allows us to generalize the model to accommodate stochastic regular lead times and provide an approximate evaluation method based on limiting results so that optimization can be done without simulation. An extensive numerical study shows that this approach yields excellent results for deterministic lead times and good results for stochastic lead times.

Keywords: inventory, dual-sourcing, dual-index policy, Markov Chain, approximation, lead times, D/G/c/c queue

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.

*Corresponding author, E-mail: j.j.arts@tue.nl

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 then 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 latter 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 paper 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 called 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 paper we investigate exactly this context. For this reason we consider a class of base-stock type policies and optimize within this class. Specifically we study 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 re-manufacturing systems, is easily implementable and performs very close to the optimal policy (Veeraraghavan & 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. Veeraraghavan & Scheller-Wolf (2008) prove a separability result that drastically decreases the amount of simulation needed, but the computational time remains substantial.

In this paper 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 based on limiting results we obtain a computationally efficient optimization procedure.

This paper is organized as follows. In Section 2 we review the literature on dual-sourcing and position our results with respect to earlier work. We then present the model with deterministic lead times in Section 3 and introduce the dual-index policy formally. In section 4 we analyze this policy and give limiting results to easily find approximately optimal settings. Next we generalize our model to accommodate stochastic regular lead times in Sections 5 and 6. Section 7 provides

an extensive numerical study on the accuracy of our approximation. We give conclusions and directions for further research in Section 8.

2. 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 situations 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 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 in the context of two supplier models. 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. This type of policy was first studied by Rosenshine & Obee (1976). 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

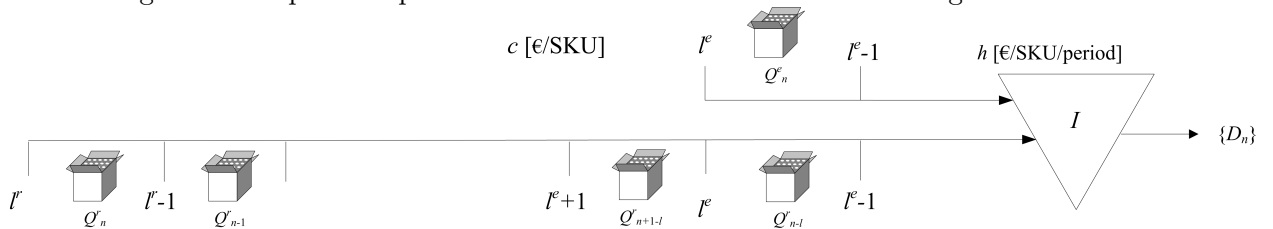
on the classical (Q, R) -policy. They show how to find optimal parameter settings under a set of specific assumptions.

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.

3. Model with deterministic lead times

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 to be 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_n^r (Q_n^e). Later (in Section 5) we will relax the assumption that l^r is deterministic. A schematic representation of the situation described above is given in Figure 1.

Figure 1: Graphical representation of model with deterministic regular lead times



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

$$IP_n^e = I_n + \sum_{i=n-l^r}^{n-1} Q_i^r + \sum_{i=n-l^e}^{n-1} Q_i^e \quad (1)$$

and if necessary place an emergency order Q_n^e to raise the emergency inventory position to its order-up-to-level S_e ,

$$Q_n^e = (S_e - IP_n^e)^+. \quad (2)$$

After placing the emergency order we inspect the regular inventory position

$$IP_n^r = I_n + \sum_{i=n-l^r}^{n-1} Q_i^r + \sum_{i=n-l^e}^n Q_i^e = IP_n^e + Q_n^e + \sum_{i=n+1-l}^{n-1} Q_i^r \quad (3)$$

and place a regular order Q_n^r to raise the regular inventory position to its order-up-to-level S_r ,

$$Q_n^r = S_r - IP_n^r. \quad (4)$$

After ordering, shipments are received and demand for the period is satisfied or backordered if there is no stock available. 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_n^e ; emergency ordering costs are incurred as cQ_n^e ; (3) review the regular inventory position and place a regular order Q_n^r ; (4) receive shipments $Q_{n-l^e}^e$ and $Q_{n-l^r}^r$; (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 S_e . 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 $S_e + O_n$ where $O_n \in \{0, 1, \dots, S_r - S_e\}$ is the overshoot and satisfies

$$O_n = IP_n^e + Q_n^e - S_e = (IP_n^e - S_e)^+. \quad (5)$$

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 is the minimization of the long run average cost subject to a modified fill-rate constraint. The modified fill-rate is defined as

$$\gamma = 1 - \mathbb{E}[B]/\mathbb{E}[D]. \quad (6)$$

The modified fill-rate is closely related to the regular fill-rate often denoted β . The γ service level 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 modified fill-rate is a very tight lower bound of the regular fill-rate.

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

$$C(S_e, S_r) = h\mathbb{E}[I^+] + c\mathbb{E}[Q^e]. \quad (7)$$

We are now in a position to formulate the optimization problem \mathcal{P} :

$$\begin{aligned} (\mathcal{P}) \quad & \min C(S_e, S_r) \\ & \text{s.t.} \quad \gamma(S_e, S_r) \geq \gamma_0 \\ & \quad S_e, S_r \in \mathbb{Z}. \end{aligned} \quad (8)$$

Here γ_0 denotes the target service level. This problem is a non-linear integer programming problem (NLIP). The integrality constraint on S_e and S_r is the consequence of the discrete nature of demand. Note that continuous demand distributions can also be used, but discretization has to be applied. An overview of all introduced notations and some notations that will be introduced in later sections is given in Table 1.

4. Analysis of model with deterministic lead times

This section is organized as follows. In section 4.1 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 4.2 we present an exact one-dimensional Discrete Time Markov Chain (DTMC) that describes the overshoot. Following in section 4.3 we provide approximations for the transition probabilities such that this DTMC can be utilized to approximate the overshoot distribution.

Throughout the analyses in this paper for any random variable X_n we define the stationary expectation and distribution as $\mathbb{E}[X] = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_n$ and $\Pr(X \leq x) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{X_n \leq x\}$ where $\mathbb{I}\{x\}$ 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 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 $c_X^2 := \frac{\text{Var}[X]}{\mathbb{E}^2[X]}$.

4.1 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

Table 1: Summary of notations

notation	description
A_n	Amount of ordered products that will not arrive within the emergency lead-time in period n <i>after</i> ordering ($:= \sum_{i=n+1-l}^n Q_i^r$)
B_n	Backlog in period n
c	Premium to buy one product at the emergency supplier
$C(S_e, S_r)$	Average holding and incremental ordering costs for policy (S_e, S_r)
γ	Modified fill rate
γ_0	Target modified fill-rate
D_n	Demand in period n , random variable
Δ	Difference between regular and emergency order-up-to-level ($:= S_r - S_e$)
h	Inventory holding cost per period per SKU
I_n	The net inventory (on-hand stock - backlog) at the beginning of period n
IP_n^r	Regular inventory position at the beginning of period n after ordering at the emergency supplier
IP_n^e	Emergency inventory position at the beginning of period n before ordering
l^e	Replenishment lead-time for emergency orders
l^r	Replenishment lead-time (deterministic) for regular orders
l	Difference between regular and emergency replenishment lead-time ($:= l^r - l^e$)
n	Period index
O_n	Overshoot in period n ($:= (IP_n^e - S_e)^+$)
S_r	Regular order-up-to-level
S_e	Emergency order-up-to-level
Q_n^r	Regular order quantity placed in period n
Q_n^e	Emergency order quantity placed in period n

$(S_e, S_e + \Delta)$. For the analysis it will be more convenient to consider S_e and Δ as decision variables. In this section we show how to find the optimal S_e for fixed Δ . This allows for a simple search procedure over Δ 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:

$$A_n = \sum_{i=n+1-l}^n Q_i^r. \quad (9)$$

Lemma 4.1. (Key functional relation) *Consider the dual-index policy for a system with deterministic lead times and A_n as defined in equation (9). Suppose that $IP_k^r \leq S_r$ for some $k \in \mathbb{N}_0$. Then for all $n \geq k$ the dual-index policy ensures that the following identity holds*

$$\Delta = O_n + A_n. \quad (10)$$

Proof. This is a special case of lemma 6.1; we defer the proof to there. \square

Lemma 4.1 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 is evident from the proof. Before establishing our separability result we need one more lemma which is originally due to Veeraraghavan & Scheller Wolf (2008).

Lemma 4.2. (Recursions for O_n , Q_n^e and Q_n^r) *Consider the model with deterministic lead times operated by the dual-index policy. The overshoot O_n , emergency order quantity Q_n^e and regular order quantity Q_n^r satisfy the following recursions:*

$$O_{n+1} = (O_n - D_n + Q_{n+1-l}^r)^+, \quad (11)$$

$$Q_{n+1}^e = (D_n - O_n - Q_{n+1-l}^r)^+, \quad (12)$$

$$Q_{n+1}^r = D_n - Q_{n+1}^e. \quad (13)$$

Proof. The proof of this lemma appears in Veeraraghavan & Scheller-Wolf (2008) as lemma 4.1 and corollary 4.1. This lemma is also a special case of lemma 6.2; we defer the proof to there. \square

The recursions (11)-(13) are quite intuitive. Equation (11) 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., $Q_n^e = (S_e - IP_n^e)^+$ from which relation (12) follows. Relation (13) 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 4.3. (Separability result) *Consider the model with deterministic lead times operated by the dual-index policy. 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. This lemma is a special case of lemma 6.3; we defer the proof to there \square

Let us define O^Δ as the stationary random variable O for a given Δ . Lemma 4.3 can be exploited to obtain the optimal DIP for fixed Δ .

Theorem 4.4. (On the optimal choice for S_e) *Consider the dual-index policy for the control of our model with deterministic lead times. For fixed Δ the optimal S_e is the smallest integer that satisfies*

the following inequality:

$$\sum_{k=0}^{\Delta} \mathbb{E} \left[\left(D^{(L_e+1)} - S_e - k \right)^+ \right] \Pr(O^\Delta = k) \leq (1 - \gamma_0) \mathbb{E}(D). \quad (14)$$

Proof. As a consequence of lemma 4.3 the cost term related to emergency ordering, $c\mathbb{E}[Q^e]$, becomes a fixed constant when Δ is fixed. Thus, for fixed Δ 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 \mathcal{Q} .

$$\begin{aligned} (\mathcal{Q}) \quad & \min \quad \tilde{C}(S_e) \\ & \text{s.t.} \quad \gamma(S_e, S_e + \Delta) \geq \gamma_0 \\ & \quad \quad S_e \in \mathbb{Z}. \end{aligned} \quad (15)$$

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 4.3 the distribution of O is already fixed:

$$\mathbb{E}[B] = \sum_{k=0}^{\Delta} \mathbb{E} \left[\left(D^{(L_e+1)} - S_e - k \right)^+ \right] \Pr(O^\Delta = k) \leq (1 - \gamma_0) \mathbb{E}[D]. \quad (16)$$

The objective function

$$h\mathbb{E}[I^+] = \sum_{k=0}^{\Delta} \mathbb{E} \left[\left(S_e + k - D^{(L_e+1)} \right)^+ \right] \Pr(O^\Delta = k) \quad (17)$$

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 (16) is the optimal solution to \mathcal{Q} , which completes the proof. \square

Remark It is also easy to show that $\mathbb{E}[B]$ is a non-increasing function of S_e . Thus the optimal S_e given Δ 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 $\mathbb{E}[Q^e]$ can be determined for fixed Δ . If this can be done one may simply perform a search procedure over Δ to find the globally optimal DIP. To evaluate the cost term $c\mathbb{E}[Q^e]$ 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 $\mathbb{E}[Q^r] = \frac{\mathbb{E}[A]}{t} = \frac{\Delta - \mathbb{E}[O]}{t}$ and $\mathbb{E}[D] = \mathbb{E}[Q^r] + \mathbb{E}[Q^e]$. Thus from the distribution of O it is easy to determine the cost term $c\mathbb{E}[Q^e]$. 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.

4.2 A one-dimensional Markov Chain for the Overshoot

Lemma 4.1 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:

$$\begin{aligned}
A_{n+1} &= \Delta - O_{n+1} \\
&= \Delta - (\Delta - D_n - \sum_{i=n+2-l}^n Q_i^r)^+ \\
&= \Delta - (\Delta - D_n - A_n + Q_{n+1-l}^r)^+ \\
&= \min(\Delta, A_n - Q_{n+1-l}^r + D_n).
\end{aligned} \tag{18}$$

In principle A_n can be modeled by a DTMC. To construct this DTMC 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 (18) we retrieve that $Q_n^r \in \{0, 1, \dots, \Delta\}$ and so this DTMC would have $\sum_{k=0}^{\Delta} \sum_{x_1=0}^k \sum_{x_2=0}^{k-x_1} \dots \sum_{x_l=0}^{k-\sum_{i=1}^{l-1} x_i} \binom{k}{x_1, x_2, \dots, x_l}$ states. It is computationally infeasible to find the equilibrium distribution of this DTMC for most practical instances. To remedy this we study a compacter DTMC with only one dimension and $\Delta + 1$ states. Observe that the recurrence relation (18) completely defines a Markov Chain for A_n if the probability mass functions of D and $\{Q_{n+1-l}^r | A_n\}$ 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

$$\begin{aligned}
p_{ij} &= \Pr(A_{n+1} = j | A_n = i) \\
&= \Pr(A_n - Q_{n+1-l}^r + D_n = j | A_n = i) \\
&= \sum_{k=0}^j \Pr(Q_{n+1-l}^r = A_n + D_n - j | A_n = i, D_n = k) \Pr(D_n = k) \\
&= \sum_{k=0}^j \Pr(Q_{n+1-l}^r = i + k - j | A_n = i) \Pr(D = k).
\end{aligned} \tag{19}$$

The case $j = \Delta$ is very similar:

$$\begin{aligned}
p_{i\Delta} &= \Pr(A_{n+1} = \Delta | A_n = i) \\
&= \Pr(A_n - Q_{n+1-l}^r + D_n \geq \Delta | A_n = i) \\
&= \sum_{k=0}^i \Pr(D_n \geq \Delta + Q_{n+1-l}^r - A_n | A_n = i, Q_{n+1-l}^r = k) \Pr(Q_{n+1-l}^r = k | A_n = i) \\
&= \sum_{k=0}^i \Pr(Q_{n+1-l}^r = k | A_n = i) \Pr(D \geq \Delta + k - i).
\end{aligned} \tag{20}$$

Now we organize these transition probabilities in the transition matrix P :

$$P = \begin{pmatrix} p_{00} & \dots & p_{0\Delta} \\ \vdots & \ddots & \vdots \\ p_{\Delta 0} & \dots & p_{\Delta\Delta} \end{pmatrix}. \tag{21}$$

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

$$\pi P = \pi, \quad \pi e = 1. \quad (22)$$

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 subsection we construct an approximation for this distribution based on limiting results so that the introduced one-dimensional DTMC can be used to approximate the overshoot distribution.

4.3 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 $\{Q_{n+1-l}^r | A_n\}$. The latter can be approximated using the following (limiting) result.

Proposition 4.5. *The following statements hold:*

- (i) As $\Delta \rightarrow \infty$, $\Pr(Q_{n+1}^r = x) \rightarrow \Pr(D_n = x)$.
- (ii) As $\Delta \rightarrow \infty$, $\Pr(Q_{n+1-l}^r = x | A_n = y) \rightarrow \Pr(D_{n+1-l} = x | \sum_{i=n+1-l}^n D_i = y)$.
- (iii) For $\Delta = 1$, $\Pr(Q_{n+1-l}^r = x | A_n = y) = \Pr(D_{n+1-l} = x | \sum_{i=n+1-l}^n D_i = y)$.

Proof. Part (i) and (ii) are special cases of proposition 6.5; we defer the proof to there. 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

$$\begin{aligned} \Pr(D_{n+1-l} = 1 | \sum_{i=n+1-l}^n D_i = 1) &= \frac{\Pr(D = 1) \Pr(D^{(l-1)} = 0)}{\Pr(D^{(l)} = 1)} \\ &= \frac{\Pr(D = 1) \Pr(D = 0)^{l-1}}{l \Pr(D = 1) \Pr(D = 0)^{l-1}} \\ &= 1/l. \end{aligned} \quad (23)$$

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

Intuitively parts (i) and (ii) of proposition 4.5 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 4.5 suggest that $\Pr(D_{n+1-l} = x | \sum_{i=n+1-l}^n D_i = y)$ can be used to approximate $\Pr(Q_{n+1-l}^r = x | A_n = y)$ as this approximation is exact for extremely small Δ ($\Delta = 1$) and extremely large Δ ($\Delta \rightarrow \infty$). Thus an approximation for $\Pr(Q_{n+1-l}^r = x | A_n = y)$ is given by

$$\begin{aligned} \Pr(Q_{n+1-l}^r = x | A_n = y) &\approx \Pr(D_{n+1-l} = x | \sum_{i=n+1-l}^n D_i = y) \\ &= \frac{\Pr(D = x) \Pr(D^{(l-1)} = y - x)}{\Pr(D^{(l)} = y)}. \end{aligned} \quad (24)$$

Using this approximation for $\Pr(Q_{n+1-l}^r = x | A_n = y)$ we can compute an approximation for $\Pr(A = x)$ by solving the set of linear equations (22). Then by using relation (4.1) we obtain an approximation for the distribution of O as $\Pr(O = x) = \Pr(A = \Delta - x)$.

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 $l = 1$.

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, where the overshoot distribution as determined by simulation is shown in conjunction with the approximation based on the above analysis. Results of a detailed numerical study are presented in Section 7.

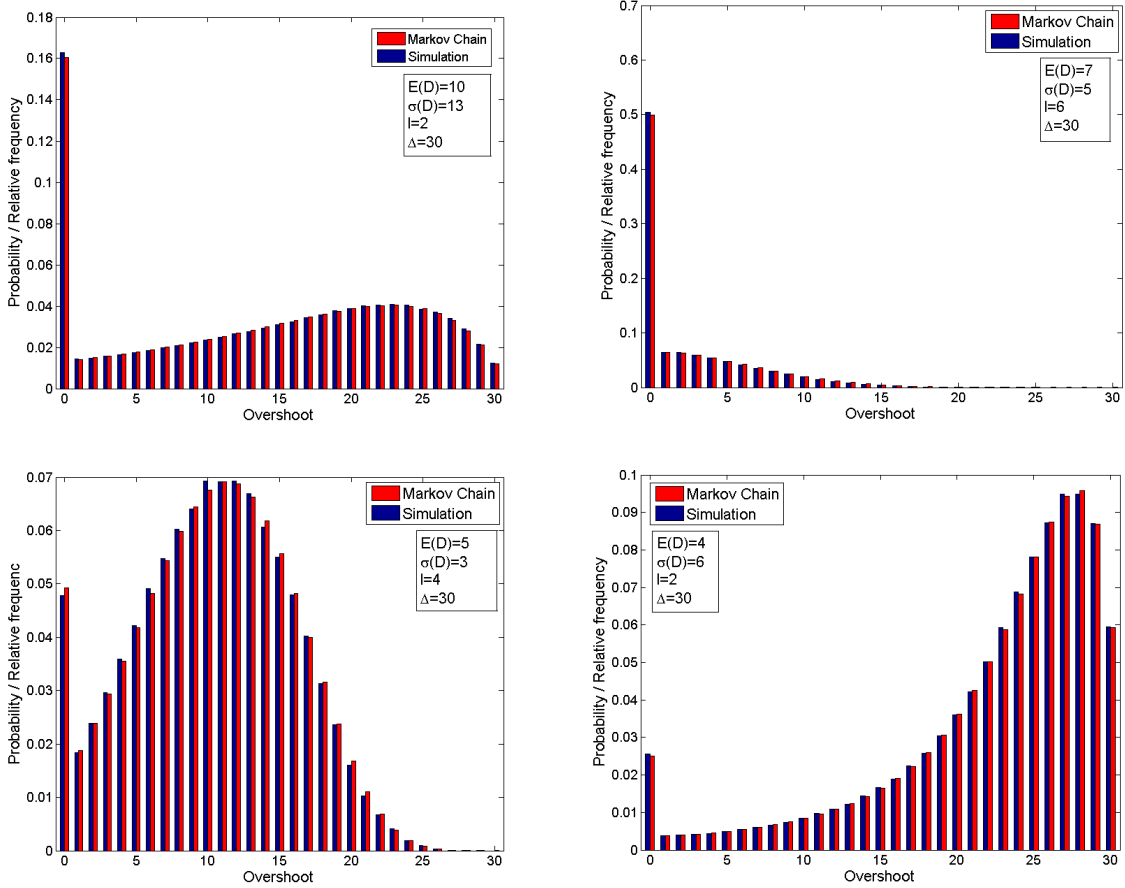
5. Model with stochastic lead times

The model we consider is identical to the model described in section 3 with one exception: now we assume that L_n^r is a stochastic integer and that the lower bound of its support is at least $l^e + 1$. Define the random variable L_n as

$$L_n := L_n^r - l^e. \quad (25)$$

The support of L is constituted by the positive integers \mathbb{N} . We can think of L_n^r 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 $\Pr(L = \nu) = q_\nu$. This implies that order crossover is possible and places us in a setting similar to that of Robinson et al. (2001). Further we let l_n and l_n^r denote realizations of the random variables L_n and L_n^r .

Figure 2: Overshoot distributions for a few illustrative cases as determined by simulation and the corresponding Markov Chain approximations



Inventory positions are now 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,

$$X_n = \{k | k \geq n - l_k^r, k < n\}.$$

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:

$$Y_n = \{k | k \geq n - l_k^r, k \leq n - l_k^e\}.$$

Using these sets we can again define the emergency and regular inventory positions as

$$IP_n^e = I_n + \sum_{i \in Y_n} Q_i^r + \sum_{i=n-l^e}^{n-1} Q_i^e \quad (26)$$

and

$$IP_n^r = I_n + \sum_{i \in X_n} Q_i^r + \sum_{i=n-L_e}^n Q_i^e = IP_n^e + Q_n^e + \sum_{i \in X_n \setminus Y_n} Q_i^r. \quad (27)$$

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^r 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^r 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. 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 equations (2) and (4) and the overshoot still satisfies the original definition in equation (5). We now proceed to analyze the DIP when regular lead times are stochastic.

6. Analysis of model with stochastic lead times

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 Δ (Section 6.1) and provide a one dimensional DTMC that describes the overshoot (Section 6.2). We provide approximations for the transition probabilities of this DTMC in Section 6.3.

6.1 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:

$$U_n = \{k | k \geq n - l_k + 1, k \leq n\}. \quad (28)$$

Now the definition of A_n can be written as

$$A_n = \sum_{i \in U_n} Q_i^r. \quad (29)$$

Lemma 6.1. (Key functional relation) *Consider the model with stochastic lead times operated by the dual-index policy and A_n as defined in equation (29). Suppose that $IP_k^r \leq S_r$ for some $k \in \mathbb{N}_0$. Then for all $n \geq k$ the dual-index policy ensures that the following identity holds*

$$\Delta = O_n + A_n. \quad (30)$$

Proof. Reconsider the regular inventory position as given in equation (27),

$$IP_n^r = IP_n^e + Q_n^e + \sum_{i \in X_n \setminus Y_n} Q_i^r. \quad (31)$$

Now we substitute the definition of the overshoot (from equation (5)) and add Q_n^r to both sides of this equation,

$$IP_n^r + Q_n^r = S_e + O_n + \sum_{i \in X_n \setminus Y_n \cup \{n\}} Q_i^r. \quad (32)$$

By supposition $IP_n^r \leq S_r$ so $Q_n^r = S_r - IP_n^r$ and the left-hand side of (32) becomes S_r . When we take a closer look at the set over which the sum in (32) 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

$$S_r = S_e + O_n + A_n. \quad (33)$$

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

Lemma 6.1 is a direct generalization of lemma 4.1 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 6.1 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 $n + 1$,

$$V_n = \{k | k = n - l_k + 1\}. \quad (34)$$

We emphasize that the sets X_n and Y_n are defined *before* ordering while U_n and V_n are defined *after* ordering. As before we now turn attention to recursions for O_n , Q_n^e and Q_n^r and then establish our separability result.

Lemma 6.2. (Recursions for O_n , Q_n^e and Q_n^r) *Consider the model with stochastic lead times as defined in Section 5. The overshoot O_n , emergency and regular order quantities satisfy the following recursions:*

$$O_{n+1} = (O_n - D_n + \sum_{i \in V_n} Q_i^r)^+, \quad (35)$$

$$Q_{n+1}^e = (D_n - O_n - \sum_{i \in V_n} Q_i^r)^+, \quad (36)$$

$$Q_{n+1}^r = D_n - Q_{n+1}^e. \quad (37)$$

Proof. The emergency inventory position satisfies

$$\begin{aligned} IP_{n+1}^e &= IP_n^e + Q_n^e - D_n + \sum_{i \in V_n} Q_i^r \\ &= S_e + O_n - D_n + \sum_{i \in V_n} Q_i^r. \end{aligned} \quad (38)$$

Rewriting the definition of the overshoot (equation (5)) we obtain

$$\begin{aligned} O_{n+1} &= (IP_{n+1}^e - S_e)^+ \\ &= (S_e + O_n - D_n + \sum_{i \in V_n} Q_i^r - S_e)^+ \\ &= (O_n - D_n + \sum_{i \in V_n} Q_i^r)^+. \end{aligned} \quad (39)$$

Similarly for the emergency order quantity we have by rewriting (2):

$$\begin{aligned} Q_{n+1}^e &= (S_e - IP_{n+1}^e)^+ \\ &= (D_n - O_n - \sum_{i \in V_n} Q_i^r)^+. \end{aligned} \quad (40)$$

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

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

Lemma 6.3. (Separability result) *Consider the model with stochastic lead times as defined in Section 5. 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. Recall the recursions in lemma 6.2. To make these equations independent of the starting conditions we substitute the identity for O_n in lemma 6.1. This substitution also makes the operation of the DIP explicit:

$$O_{n+1} = \left(\Delta - D_n - \sum_{i \in U_n \setminus V_n} Q_i^r \right)^+, \quad (41)$$

$$Q_{n+1}^e = \left(D_n + \sum_{i \in U_n \setminus V_n} Q_i^r - \Delta \right)^+, \quad (42)$$

$$Q_{n+1}^r = D_n - Q_{n+1}^e. \quad (43)$$

For the summation $\sum_{i \in U_n \setminus V_n} Q_i^r$ 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 $\{Q_n^e\}$ and $\{Q_n^r\}$ 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 In establishing lemma 6.3 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 Δ 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 6.2 and 6.3 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^Δ as the stationary random variable O for a given Δ . Lemma 6.3 leads to the following theorem on the optimal choice for S_e for fixed Δ

Theorem 6.4. (On the optimal choice for S_e) *Consider the model with stochastic lead times as defined in Section 5. For fixed Δ the optimal S_e is the smallest integer that satisfies the following inequality*

$$\sum_{k=0}^{\Delta} \mathbb{E} \left[\left(D^{(L_e+1)} - S_e - k \right)^+ \right] \Pr(O^\Delta = k) \leq (1 - \gamma_0) \mathbb{E}(D). \quad (44)$$

Proof. The proof is analogous to the proof of Theorem 4.4 and therefore omitted. \square

The optimal DIP for the system with stochastic lead times can also be found by a search procedure over Δ . To find the cost term $c\mathbb{E}[Q^e]$ for the objective function of problem \mathcal{P} in this more general situation, we make use of the identities $\mathbb{E}[Q^r] = \frac{\Delta - \mathbb{E}[O]}{\mathbb{E}[L]}$ and $\mathbb{E}[D] = \mathbb{E}[Q^r] + \mathbb{E}[Q^e]$. In the next two sections we describe a one-dimensional DTMC and transition probability approximations for our generalized model.

6.2 A one-dimensional Markov Chain for the overshoot

As was the case for the model with deterministic lead times, lemma 6.1 allows us to study A_n to find the distribution of O . A_n still has the appealing physical interpretation as the pipeline stock

that will not arrive within the short lead time l_e and obeys the following recurrence relation

$$\begin{aligned}
A_{n+1} &= \Delta - O_{n+1} \\
&= \Delta - \left(\Delta - D_n - \sum_{i \in U_n \setminus V_n} Q_i^r \right)^+ \\
&= \Delta - \left(\Delta - D_n - A_n + \sum_{i \in V_n} Q_i^r \right)^+ \\
&= \min(\Delta, A_n - \sum_{i \in V_n} Q_i^r + D_n)
\end{aligned} \tag{45}$$

It will be evident from the model with discrete lead times that an exact DTMC for this problem suffers even more from the curse of dimensionality. For this reason we again turn our attention to a one-dimensional DTMC that can be constructed in a manner analogous to that in Section 4.2.

This DTMC is given by the transition probabilities $p_{ij} = \Pr(A_{n+1} = j | A_n = i)$:

$$p_{ij} = \begin{cases} \sum_{k=0}^j \Pr(\sum_{i \in V_n} Q_i^r = i + k - j | A_n = i) \Pr(D = k), & \text{if } j < \Delta; \\ \sum_{k=0}^i \Pr(\sum_{i \in V_n} Q_i^r = k | A_n = i) \Pr(D \geq \Delta + k - i), & \text{if } j = \Delta. \end{cases} \tag{46}$$

To make this one-dimensional DTMC of use, it remains to find the distribution of $\{\sum_{i \in V_n} Q_i^r | A_n\}$ or an approximation thereof. This will be done in the next subsection.

6.3 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 6.5. *The following statements hold*

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

Proof. We rewrite equation (45) to

$$\begin{aligned}
A_{n+1} &= \min(\Delta, A_n - \sum_{i \in V_n} Q_i^r + D_n) \\
&= \min(\Delta, \sum_{i \in U_n} Q_i^r - \sum_{i \in V_n} Q_i^r + D_n) \\
&= \min(\Delta, \sum_{i \in U_n \setminus V_n \cup \{n+1\}} Q_i^r - Q_{n+1}^r + D_n) \\
&= \min(\Delta, A_{n+1} - Q_{n+1}^r + D_n).
\end{aligned} \tag{47}$$

Now if we let $\Delta \rightarrow \infty$ 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 $\Delta \rightarrow \infty$, Q_n^r 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. \square

Remark When considering deterministic lead times we already showed in proposition 4.5 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 7 reflect this fact.

Part (ii) of proposition 4.5 suggests that $\Pr\left(\sum_{i=n-|V_n|+1}^n D_i = x \mid \sum_{i=n-|U_n|+1}^n D_i = y\right)$ can be used to approximate $\Pr\left(\sum_{i \in V_n} Q_i^r = x \mid A_n = y\right)$. The computation of this approximation is however 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 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,

$$K_n = |U_n|. \quad (48)$$

Further let Λ_n denote the number of orders that are about to enter the information horizon of the emergency inventory position,

$$\Lambda_n = |V_n|. \quad (49)$$

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 $q_\nu = \Pr(L = \nu)$, $\nu \in \{1, 2, \dots, L_{max}\}$. Further we define

$$\varphi_{\kappa, \lambda, \nu} = \Pr(K = \kappa \cap \Lambda = \lambda \mid \text{orders were placed the last } \nu \text{ periods only (not before)}) \quad (50)$$

Obviously, this definition means that the distribution needed is given by

$$\Pr(K = \kappa \cap \Lambda = \lambda) = \varphi_{\kappa, \lambda, L_{max}} := \varphi_{\kappa, \lambda}, \quad (51)$$

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

$$\varphi_{\kappa, \lambda, \nu} = \varphi_{\kappa-1, \lambda-1, \nu-1} q_\nu + \varphi_{\kappa-1, \lambda, \nu-1} \cdot \sum_{m=\nu+1}^{L_{max}} q_m + \varphi_{\kappa, \lambda, \nu-1} \cdot \sum_{m=1}^{\nu-1} q_m. \quad (52)$$

The initial probabilities are straightforwardly seen to be

$$\varphi_{1,0,1} = \sum_{m=2}^{L_{max}} q_m, \quad \varphi_{1,1,1} = q_1, \quad \varphi_{\kappa,\lambda,1} = 0 \quad \text{otherwise.} \quad (53)$$

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 $D/G/L_{max}/L_{max}$ -queue. Each period n a customer arrives (order is placed) and that customer 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 $\{1, \dots, L_{max}\}$ 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 $\mathcal{O}(L_{max}^3)$ time. To see this, note that the number of times we compute recursion (52) including the initialization before we obtain $\varphi_{\kappa,\lambda}$ is given by $\sum_{i=2}^{L_{max}+1} \sum_{x=2}^i x = \frac{1}{6}L_{max}^3 + L_{max}^2 + \frac{5}{6}L_{max}$. Thus the complexity is $\mathcal{O}(L_{max}^3)$.

Now that the joint distribution of K and Λ 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 $|V_n|$:

$$\begin{aligned} \Pr(\sum_{i \in V_n} Q_i^r = x | A_n = y) &\approx \Pr(\sum_{i=n-\Lambda+1}^n D_i = x | \sum_{i=n-K+1}^n D_i = y) \\ &= \sum_{\kappa=1}^{L_{max}} \sum_{\lambda=1}^{\kappa} \Pr(\Lambda = \lambda | K = \kappa) \Pr(K = \kappa | \sum_{i=n-K+1}^n D_i = y) \\ &\quad \times \Pr(\sum_{i=n-\lambda+1}^n D_i | \sum_{i=n-\kappa+1}^n D_i = y) \\ &= \sum_{\kappa=1}^{L_{max}} \sum_{\lambda=1}^{\kappa} \Pr(\Lambda = \lambda | K = \kappa) \Pr(K = \kappa | \sum_{i=n-K+1}^n D_i = y) \\ &\quad \times \frac{\Pr(D^{(\lambda)} = x) \Pr(D^{(\kappa-\lambda)} = y-x)}{\Pr(D^{(\kappa)} = y)}. \end{aligned} \quad (54)$$

In expression (54) the probability $\Pr(K = \kappa | \sum_{i=n-K+1}^n D_i = y)$ is obtained by applying Bayes' theorem:

$$\begin{aligned} \Pr(K = \kappa | \sum_{i=n-K+1}^n D_i = y) &= \frac{\Pr(D^{(K)} = y | K = \kappa) \Pr(K = \kappa)}{\Pr(D^{(K)} = y)} \\ &= \frac{\Pr(D^{(\kappa)} = y) \Pr(K = \kappa)}{\sum_{z=1}^{L_{max}} \Pr(D^{(z)} = y) \Pr(K = z)}, \end{aligned} \quad (55)$$

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

Using this approximation for $\Pr(\sum_{i \in V_n} Q_i^r = x | A_n = y)$ we can compute an approximation for $\Pr(A = x)$ by finding the equilibrium distribution of the DTMC for A_n . Then by using relation (30) we obtain an approximation for the distribution of O as $\Pr(O = x) = \Pr(A = \Delta - x)$.

7. Numerical results

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 1680 instances of problem \mathcal{P} was created that is a full factorial design of the parameter settings summarized in Table 2. 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.

Table 2: Test-bed of problem instances \mathcal{P}

Parameter	settings						
$\mathbb{E}[D]$	25						
c_D^2	$\frac{1}{4}$,	$\frac{1}{2}$,	1,	$\frac{3}{2}$,	2		
l^e	1, 2						
$\mathbb{E}[L]$	4, 8, 12						
h	1						
c	10,	20,	30,	40			
γ_0	0.95, 0.98						
Distribution type of L	U1,	U2,	S1,	S2,	LS,	RS,	DET

Table 3: Distribution types for L

Distribution Type \ x	$\Pr(L = x)$				
	$\mathbb{E}[L] - 2$	$\mathbb{E}[L] - 1$	$\mathbb{E}[L]$	$\mathbb{E}[L] + 1$	$\mathbb{E}[L] + 2$
U1 (uniform1)	0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	0
U2 (uniform2)	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$
S1 (symmetric1)	0	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$	0
S2 (symmetric2)	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{4}{10}$	$\frac{2}{10}$	$\frac{1}{10}$
LS (left skewed)	0	$\frac{4}{10}$	$\frac{3}{10}$	$\frac{2}{10}$	$\frac{1}{10}$
RS (right skewed)	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{3}{10}$	$\frac{4}{10}$	0
DET (deterministic)	0	0	1	0	0

For each instance we performed the optimization using the simulation approach suggested by Veeraraghavan & Scheller-Wolf (2008), yielding the DIP (S_e^{sim}, S_r^{sim}) to be optimal. We also performed the optimization with our approach involving approximate Markov Chains, which found the DIP (S_e^{MC}, S_r^{MC}) to be optimal. We then evaluated the total cost and modified fill-rate for both solutions using simulation. The simulation was run such that the halfwidth of 99%-confidence interval for cost was less than 1% of the point estimate. As measures of optimality we considered

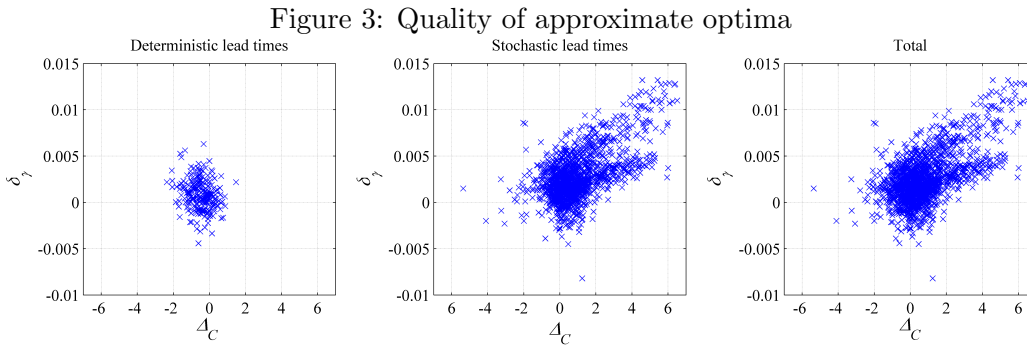
the relative deviation from the optimal DIP

$$\Delta_C = \frac{C_{sim}(S_r^{MC}, S_e^{MC}) - C_{sim}(S_r^{sim}, S_e^{sim})}{C_{sim}(S_r^{sim}, S_e^{sim})} \cdot 100\%$$

and the absolute deviation from the target modified fill-rate

$$\delta_\gamma = \gamma_{sim}(S_r^{MC}, S_e^{MC}) - \gamma_0.$$

Figure 3 shows scatter-plots of Δ_C versus δ_γ the problem instances with deterministic lead times, stochastic lead times and the entire test-bed. In these Figures points close to the origin are close to optimal. In fact points in the second quadrant indicate that solutions found using our approximate method outperform simulation optimization. Thus this Figure shows that the approximation is at least as good as simulation for deterministic lead times. For stochastic lead times there is a small tendency to find solutions that are more expensive than optimal at an increase in service relative to the target level. An explanation of the superior performance of the approximation for deterministic lead times over stochastic lead times is to be found in the fact that the approximation for deterministic lead times is based on limiting results for $\Delta \rightarrow \infty$ and $\Delta = 1$. These Figures show that the performance of our approximate method is excellent.



To understand how different problem parameters influence the performance of our approximation we tabulated the average, minimum, maximum and relative frequencies of Δ_C for different problem parameters in Table 4. We see that the approximation improves as demand variability increases. This is a convenient property because dual-sourcing is a way o buffer demand variability.

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 $h\mathbb{E}[(S_e + O - D^{(l^e+1)})^+]$, 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 Δ 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 $h\mathbb{E}[(S_e + O - D^{(l^e+1)})^+]$. The contribution of O , which we know only approximately, compared to $D^{(l^e+1)}$ becomes smaller when $\mathbb{E}[L]$ decreases. This is because Δ (and therefore also O) increases with $\mathbb{E}[K] = \mathbb{E}[L]$ (by Little's law). The target service level and different non-stochastic distributions for L do not influence accuracy much. We do see as before that the approximation performs much better when lead times are deterministic.

Table 4: Quality of approximate solutions for different problem parameters

c_D^2	Relative cost deviation from optimal DIP $[\Delta_C]$										
	avg.	min	max	< -1	-1 - 0	0 - 1	1 - 2	2 - 3	3 - 4	4 - 5	> 5
$\frac{1}{4}$	2.13	-5.36	6.51	10.42	11.61	12.50	9.82	13.39	16.37	15.77	10.12
$\frac{1}{2}$	1.05	-2.78	4.30	5.36	16.96	22.92	30.06	19.94	4.46	0.30	0.00
1	0.32	-3.25	2.97	7.14	27.98	44.94	17.56	2.38	0.00	0.00	0.00
$\frac{3}{2}$	0.40	-1.10	2.07	1.04	25.00	56.94	16.67	0.35	0.00	0.00	0.00
2	0.06	-2.37	1.52	4.43	41.67	48.70	5.21	0.00	0.00	0.00	0.00
l^e	avg.	min	max	< -1	-1 - 0	0 - 1	1 - 2	2 - 3	3 - 4	4 - 5	> 5
1	0.80	-3.25	6.51	5.8	24.2	36.8	16.7	7.5	3.6	3.6	1.9
2	0.76	-5.36	6.36	5.7	26.1	37.1	14.4	6.9	4.8	2.9	2.1
c	avg.	min	max	< -1	-1 - 0	0 - 1	1 - 2	2 - 3	3 - 4	4 - 5	> 5
10	1.35	-5.36	6.40	1.2	14.8	36.7	21.9	11.2	4.8	5.5	4.0
20	0.84	-2.06	6.51	5.0	23.6	38.3	16.7	6.9	4.8	2.4	2.4
30	0.59	-3.25	5.95	6.7	29.0	37.6	13.1	5.2	4.0	3.3	1.0
40	0.35	-4.10	5.95	10.2	33.1	35.2	10.5	5.5	3.1	1.7	0.7
γ_0	avg.	min	max	< -1	-1 - 0	0 - 1	1 - 2	2 - 3	3 - 4	4 - 5	> 5
0.95	0.79	-4.10	6.51	7.4	24.4	34.9	17.0	6.4	3.2	3.6	3.1
0.98	0.77	-5.36	6.01	4.2	25.8	39.0	14.0	8.0	5.1	2.9	1.0
$\mathbb{E}[L]$	avg.	min	max	< -1	-1 - 0	0 - 1	1 - 2	2 - 3	3 - 4	4 - 5	> 5
4	0.03	-5.36	4.72	11.4	38.0	38.6	8.8	2.5	0.5	0.2	0.0
8	0.95	-2.37	6.09	3.9	21.1	36.6	18.9	9.5	4.6	4.3	1.1
12	1.37	-1.56	6.51	2.0	16.3	35.7	18.9	9.6	7.3	5.2	5.0
Distribution type	avg.	min	max	< -1	-1 - 0	0 - 1	1 - 2	2 - 3	3 - 4	4 - 5	> 5
U1	0.88	-3.25	5.27	4.2	21.3	39.2	16.3	10.0	6.3	1.7	1.3
U2	0.98	-4.10	5.86	3.8	18.8	40.4	17.9	6.3	7.1	3.8	2.1
S1	1.01	-2.78	6.40	4.6	17.1	40.0	20.0	6.7	4.6	4.6	2.5
S2	1.13	-2.36	6.26	2.5	19.6	37.5	16.3	10.8	6.7	2.9	3.8
LS	1.06	-1.23	6.51	2.1	20.4	38.8	19.6	7.9	2.1	6.7	2.5
RS	0.90	-5.36	5.42	3.3	18.3	43.8	18.3	8.8	2.5	2.9	2.1
DET	-0.48	-2.37	1.48	20.0	60.4	19.2	0.4	0.0	0.0	0.0	0.0
Total	avg.	min	max	< -1	-1 - 0	0 - 1	1 - 2	2 - 3	3 - 4	4 - 5	> 5
	0.78	-5.36	6.51	5.8	25.1	37.0	15.5	7.2	4.2	3.2	2.0

Computational times for our approximation are also much shorter than for the simulation based procedure. For this test-bed the optimization method based on the approximation was around 50 times faster than the simulation based method for problem instances with deterministic lead times

and around 70 times faster for problem instances with stochastic lead times.

8. Conclusion and directions for future research

In this paper 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 controlled by the dual-index policy. Our main contributions here are to (i) provide an alternate and insightful proof of the separability result that reduces the optimization of the DIP to two one-dimensional optimization problems and (ii) provide an approximate evaluation method of the dual-index policy using Markov Chains based on limiting results that does not require simulation, thus making optimization more efficient.

The second model we presented was a generalization of the first by allowing regular lead times to be stochastic. In this situation we (i) defined a dual-index policy with mild informational requirements on the realizations of regular lead times; (ii) proved that the same separability result holds as for the model with deterministic lead time and (iii) developed an approximate evaluation method using Markov Chains based on limiting results again making optimization much more efficient.

In an extensive numerical study we showed that the approximations we suggest perform very well in finding a close to optimal dual-index policy and are faster by a factor 50-70 than the simulation based procedure.

The research in this paper 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 paper.

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.

References

- [1] Adan, I.J.B.F., van Eenige, M.J.A., Resing, J.A.C. (1996) Fitting discrete distributions on the first two moments, *Probability in the Engineering and Informational Sciences*, **9**, 623-632
- [2] Allon, G., van Mieghem, J.A. (2008) Global dual sourcing: tailored base surge allocation to near and offshore production, *Working Paper, Kellogg school of management, Northwestern university, Evanston Illinois*
- [3] Barankin, E.W. (1961) A delivery-lag inventory model with an emergency provision (the single-period case), *Naval Research Logistics Quarterly*, **8**, 285-311
- [4] Chiang, C. (2007) Optimal control policy for a standing order inventory system, *European Journal of Operational Research*, **182**, 695-703
- [5] Clark, A.J., Scarf, H. (1960) Optimal policies for a multiechelon inventory problem, *Management Science*, **6**(4), 475-490
- [6] Feng, Q., Sethi, S.P., Yan, H., Zhang, H. (2006a) Optimality and nonoptimality of the base-stock policy in inventory problems with multiple delivery modes, *Journal of Industrial and Management Optimization*, **2**(1), 19-42
- [7] Feng, Q., Sethi, S.P., Yan, H., Zhang, H. (2006b) Are base-stock policies optimal in inventory problems with multiple delivery modes?, *Operations Research*, **54**(4), 801-807
- [8] Fukuda, Y. (1964) Optimal policies for the inventory problem with negotiable leadtime, *Management Science*, **10**(4), 690-708
- [9] Gallego, G., Jin, Y., Muriel, A., Zhang, G., Yildiz, V. (2007) Optimal ordering policies with convertible lead times, *European Journal of Operational Research*, **176**, 892-910
- [10] Gaukler, G.M., Özer, Ö., Hausman, W.H. (2008) Order Progress Information: Improved Dynamic Emergency Ordering Policies, *Production and Operations Management*, **17**(6), 599-613
- [11] Lawson D.G., Porteus, E.L. (2000) Multi-stage inventory management with expediting, *Operations Research*, **48**(6), 878-893
- [12] Kiesmüller, G.P. (2003) A new approach for controlling a hybrid stochastic manufacturing/re-manufacturing system with inventories and different lead-times, *European Journal of Operational Research*, **147**, 62-71

- [13] Minner, S. (2003) Multiple supplier inventory models in supply chain management: A review, *International Journal of Production Economics*, **81-82**, 265-279
- [14] Robinson, L.W., Bradley, J.R., Thomas, L.T. (2001) Consequences of order crossover under order-up-to inventory policies, *Manufacturing & Service Operations Management*, **3(3)**, 175-188
- [15] Rosenshine, M., Obee, D., (1976) Analysis of a standing order system with emergency orders, *Operations Research*, **24(6)**, 1143-1155
- [16] Sethi, S.P., Yan, H., Zhang, H. (2003) Inventory models with fixed costs, forecast updates and two delivery modes, *Operations Research*, **51(2)**, 321-328
- [17] Scheller-Wolf, A., Veeraraghavan, S.K., van Houtum, G.J.J.A.N. (2003) Effective dual sourcing with a single index policy, *Working paper Tepper School of Business, Carnegie Mellon University, Pittsburgh*
- [18] Song, J.S., Zipkin, P. (2009) Inventories with multiple supply sources and networks of queues with overflow bypasses, *Management Science* **55(3)**, 362-372
- [19] van Houtum, G.J. (2006) Multiechelon production/inventory systems: optimal policies, heuristics and algorithms, *Tutorials in Operations Research INFORMS 2006*, 163-199
- [20] van Houtum, G.J., Zijm, W.H.M. (2000) On the relation between cost and service models for general inventory systems, *Statistica Neerlandica*, **54(2)** 127-147
- [21] Veeraraghavan, S., Scheller-Wolf, A. (2008) Now or later: a simple policy for effective dual sourcing in capacitated systems, *Operations Research*, **56(4)**, 850-864
- [22] Whittmore, A.S., Saunders, S.C. (1977) Optimal inventory under stochastic demand with two supply options, *SIAM journal of applied mathematics*, **32(2)**, 293-305
- [23] Yazlali, Ö., Erhun, F. (2009) Dual-supply inventory problem with capacity limits on order sizes and unrestricted ordering costs, *IIE Transactions*, **41**, 716-729