PRODUCTION AND STOCK MANAGEMENT UNDER MANUFACTURER AND CUSTOMER DRIVEN SUBSTITUTION

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

NURŞEN TÖRE

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
INDUSTRIAL ENGINEERING

AUGUST 2015

Approval of the thesis:

PRODUCTION AND STOCK MANAGEMENT UNDER MANUFACTURER AND CUSTOMER DRIVEN SUBSTITUTION

submitted by **NURŞEN TÖRE** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Industrial Engineering Department, Middle East Technical University** by,

Prof. Dr. Gülbin Dural Ünver ______________
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Murat Köksalan ______________
Head of Department, **Industrial Engineering**

Assoc. Prof. Dr. Seçil Savaşaneril Tüfekci ______________
Supervisor, **Industrial Engineering, METU**

Examining Committee Members:

Assoc. Prof. Dr. Z. Pelin Bayındır
Industrial Engineering Department, METU ______________

Assoc. Prof. Dr. Seçil Savaşaneril Tüfekci
Industrial Engineering Department, METU ______________

Prof. Dr. Nesim Erkip
Industrial Engineering Department, Bilkent University ______________

Assoc. Prof. Dr. Serhan Duran
Industrial Engineering Department, METU ______________

Assist. Prof. Dr. Banu Yüksel Özkaya
Industrial Engineering Department, Hacettepe University ______________

Date: ______________

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: NURŞEN TÖRE

Signature :

ABSTRACT

PRODUCTION AND STOCK MANAGEMENT UNDER MANUFACTURER AND CUSTOMER DRIVEN SUBSTITUTION

Töre, Nurşen
Ph.D., Department of Industrial Engineering
Supervisor : Assoc. Prof. Dr. Seçil Savaşaneril Tüfekci

August 2015, 316 pages

In this thesis, the stock management problem of a manufacturer is studied in the presence of product substitution flexibility. The study consists of two parts. In the first part, the manufacturer's joint production scheduling and product substitution decisions are analyzed. It is assumed that there are multiple substitutable items and for each item, and inventory is kept separately. If upon demand arrival, stock is not available, then demand is backordered. Depending on the inventory level, it is possible to meet the demand via a substitute item. However, it takes a time and cost to substitute the item with another one. The production and substitution decisions are given dynamically over time depending on the net inventory level of the products. It is shown that for the two-product setting optimal production policy is characterized by a hedging point and a switching curve, and the substitution policies are characterized by threshold substitution levels. For the two-product setting the benefit of substitution, and the value of stock level information are quantified. For the multiple item setting, several heuristics are proposed. Through numerical analysis, the conditions under which each heuristic performs well are identified.

In the second part of the thesis, the manufacturer's response to substitution behavior of the customer is studied. The manufacturer can manipulate the customer's substitution behaviour by price and availability of the products. The manufacturer decides on the optimal stock levels, the initial price of the products, and the market markdown price after the demand is realized. It is assumed there are two substitutable products and for each item, initial stock levels are determined separately. Since the products the manufacturer produces are similar, uniform pricing is used. Then according to price determined by the manufacturer, random demands for products occur. After the initial selling period, price is updated, that results in customer demand shift to the other product. The customer's initial demand and the demand spill-over due to price and stock-out based substitution are characterized through the Representative Consumer Theorem framework. We adopt a two-phase solution methodology where in the first we determine the optimal price after the demand realization, and then determine the initial price and stock levels. While doing so we gain an understanding on how the manufacturer should determine the optimal prices. To determine the initial price and the stock levels, we offer a solution approach for finding the KKT points of the corresponding model. Through numerical analysis, we obtain insights on how stock levels are determined, how the price decisions and the profit respond to the substitution behavior of the customers. Finally, through numerical analysis we quantify the value of exploiting the stockout based substitution via markdown pricing, and the value of the optimal solution.

Keywords: Markov Decision Processes, Continuous Time Markov Chains, Priority Queues, Multi-product Make-to-Stock Queues, Index policies, Value of substitution, Value of stock level information, Customer-driven substitution, price-based substitution, Stockout-based substitution, Markdown pricing

ÖZ

ÜRETİCİ VE MÜŞTERİ KAYNAKLI İKAME ALTINDA ÜRETİM VE STOK YÖNETİMİ

Töre, Nurşen
Doktora, Endüstri Mühendisliği Bölümü
Tez Yöneticisi : Doç. Dr. Seçil Savaşaneril Tüfekci

Ağustos 2015 , 316 sayfa

Bu tezde, üreticinin stok yönetim problemi ürün ikame esnekliği altında incelenmiştir. Bu çalışma iki bölümden oluşmaktadır. İlk bölümde üreticinin birarada verdiği üretim çizelgeleme ve ürün ikame kararları incelenmiştir. Bir ürünün birden fazla üründen herhangi biri tarafından ikame edilebileceği varsayılmıştır. Her ürün için ayrı stok tutulduğu, talep geldiğinde elde stok bulunmazsa ardısmarlandığı varsayılmıştır. Stok seviyelerine bağlı olarak, ürün yerine başka bir ürün ikame edilebilir. Ancak bir ürün yerine başka bir ürünü ikame etmek, zaman ve maliyete neden olur. Üretim ve ikame kararları zaman üzerinde dinamik şekilde stok düzeylerine bağlı olarak verilir. İki ürünlü sistem için en iyi üretim politikası risk kontrol noktalı, ve ürün değişim çizgisi ile karakterize edilebilir, ikame politikaları ise ürünün stoğuna bağlı olan eşik seviyeli bir karar olarak karakterize edilir. İki ürünlü sistem için ikamenin yararı ve stok seviyesi bilgisinin değeri ölçülmüştür. Çoklu ürün için farklı sezgiseller önerilmiştir. Sayısal çalışmalar kullanarak, hangi sezgiselin hangi durum altında iyi çalıştığı araştırılmıştır. İki ürün varsayımı altında, ürünler için yarı-mamul stok tutma kararının getirileri sayısal olarak incelenmiş, ikamenin ya da ortak stok tutma kararının hangi durumlarda daha karlı olduğu, sayısal olarak çalışılmışıtr.

Tezin ikinci kısmında, üreticinin kararlarını alırken müşterinin ikame davranışlarını da gözönünde bulundurduğu varsayılmıştır. Üretici müşterinin ikame davranışına fiyatlandırma ve stok bulundurma yoluyla etki edebilir. Üretici ilk stok düzeyleri, ilk fiyatlandırma ve sonradan yapılan indirim fiyatını belirler. İki adet ikame edilebilen ürün olduğu ve bunların stok seviyelerinin ayrı ayrı belirlendiği farzedilmiştir. Üretici yakın ürünler üretir ve sezon içinde ürünler aynı fiyattan satılır. Üretici fiyatı belirledikten ürünler için rassal talepler belli olur. İlk satış periyodundan sonra, fiyat indirimine gidilir, bu durum, ürünün stokta kalmaması ile birlikte müşteri talebinin diğer ürüne kaymasına neden olur. Müşterinin talep yapısı Temsili Tüketici Teoremi çerçevesinde modellenmiştir. Optimal fiyatlar ve stok seviyeleri analitik ve doğrusal olmayan programlama yaklaşımı ile belirlenmiş, kararların nasıl verildiğine dair bir anlayış geliştirilmiştir. Sayısal analizler ile elde edilen anlayış pekiştirilmiştir. Son olarak, stokta bulunmama kaynaklı ikame ve fiyat indiriminin getirilerinin yanısıra, optimal politikanın getirileri incelenmiştir.

Anahtar Kelimeler: Markov Karar Süreçleri, Süreli zamanlı Markov Zincirleri, Öncelik politikalı kuyruklar, Çok ürünlü stoğa üretmeli kuyruklar, Dizin politikaları, İkamenin değeri, Stok düzeyi bilgisinin değeri, Müşteri güdümlü ikame, fiyat kaynaklı ikame, Stokta bulunmama kaynaklı ikame, Fiyat düşürmeli fiyatlandırma

To my mom and dad

ACKNOWLEDGMENTS

The completion of this thesis gives me so much pleasure. I would like to expand my deepest gratitude to all those who have directly and indirectly guided me in writing this thesis.

First of all, I would like to express my sincere gratefulness to my thesis supervisor Assoc. Prof. Dr. Seçil Savaşaneril Tüfekci. Her inspiring advises and meticulous guidance made this study possible. I am grateful to her for this study and for her mentorship for many aspects in my life.

I am grateful to Prof. Dr. Yasemin Serin for her contribution. I am thankful to the member of thesis supervising committee Prof. Dr. Nesim Erkip, Assoc. Prof. Dr. Z. Pelin Bayındır and former comittee member Prof. Dr. Sinan Kayalıgil for their guidance throughout the stages of this study and their advises. I am also thankful to the other members of the thesis examining committee Assoc. Prof. Dr. Serhan Duran and Assist. Prof. Dr. Banu Yüksel Özkaya for their invaluable comments.

In Department of Industrial Engineering, METU, many people, especially my classmates and assistants, have made valuable suggestions on this study which gave me an inspiration to improve my work.

I would like to thank Fulya, Serkan, Merve, and my spanish class-mates for their friendship and encouragements.

Without the support of my mother, this study would not be possible. For this reason she deserves the dedication of my thesis. She is always there for me, I am grateful for that. I am thankful to my sister Gülşen because of her encouragements and guidance. I also thank Bahadır for being mental support.

This research is supported by the Tübitak research grants 108M004 and 112M431.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

CHAPTER 1

INTRODUCTION

Flexibility is a valuable asset for manufacturers operating in business environments characterized with uncertain demand. Manufacturers may benefit from various forms of flexibility. Sethi and Sethi (1990) makes a comprehensive review of types of manufacturing flexibility, ranging from operation flexibility, process flexibility to material handling flexibility. Operation flexibility is the ability to produce a part in a flexible sequence of operations, or via alternative operations. Process flexibility is the ability to produce a large set of parts without major setups. These definitions imply that manufacturing flexibility is characterized by the abilities of the components, and availability of choices provided by the machines during the course of manufacturing. Another form of flexibility that may benefit the manufacturers, and which is addressed in this study, is the flexibility in fulfilling the demand. This type of flexibility helps improve the "availability" and can be exemplified by product substitution, delayed differentiation, modularity in product design, or delivery flexibility. The manufacturing and demand fulfillment flexibility are intertwined, and the latter cannot be achieved without the former.

In this thesis, we aim to quantify the value of flexibility attained through product substitution in a manufacturing setting, where substitution is the practice of providing an alternative product in order to meet the demand for another product. In general, substitution can take place at the product (end-product) level or at the component level, can be static (anticipative) or dynamic (stock-out-based).

Dynamic substitution is triggered by the occurrence of demand when the requested product is not in stock. Zhao et al. (2008) studies operating policy of an individual spare parts dealer in a centralized network which very well can be implemented in automotive industry. Here, dealers share inventory information, and based on real-time data, if they are out of stock, they can ask spare parts from other dealers.

Anticipative substitution is a planned-in-advance substitution and often times the aim is to achieve scale economies in cycle stock. For example, if the manufacturing (or replenishment) is cyclic due to set-ups, then the firm may plan to meet the demand with a substitute product until the cycle turn of the required product comes. Planned substitution balances the savings in holding cost and set-up cost with that of the substitution costs. In stock-out based substitution, demand is substituted when the product requested is currently not available. Electricity Meters subdivision of a major oil field services company manufactures residential and commercial electricity meters in two different grades: a basic version and expensive RF version, the company has the ability to supply RF-enabled meters in place of the cheaper traditional meters in order to alleviate the setup costs. (Dawande et al., 2010). In manufacturing-driven substitution, under make-to-order environments, the manufacturer provides the customer with a product that is superior to the one originally requested. Here, the cost of substitution to the manufacturer is the additional material/effort required to produce the relatively superior product. In downgrading there is no goodwill cost associated with substitution. In the two-way substitution, the principal cost of substitution to the manufacturer is that associated with lost opportunity; inability to capture a higher margin and/or loss of goodwill.

End item level substitution is substituting a finished product request by a different finished product, where substitution of RF-enabled meters and traditional meters is one possible example (Dawande et al., 2010). Component substitution is being able to use different type of components in bill of materials of a product. Balakrishnan and Geunes (2000) discuss on example where aluminum bloom components in varying dimensions are substituted during the conversion of the blooms into the end-products.

There are in between examples, for example in Li et al. (2010) wireless access point product will be replaced with its next generation successor. The company did not buy many old products on the End-of-Product-Life notice, hoping that the successor would cover the demand. However until the notice, there is more demand for the product, so that its successor products need to be re-assembled as the product itself.

Practice of product substitution can be observed in various sectors. For example, in computer manufacturing the products are usually classified with respect to their processing capabilities. When demand occurs for a certain product, which is not available at the time, the manufacturer may offer a superior product at the price of the lower-grade product (Shumsky and Zhang, 2009). This will lead to savings in inventory holding and lost sales cost, but will result in loss of profit margin. Similar examples can be found in manufacturing of various technology products. In Gallego et al. (2006), downgrading is studied in the context of semiconductor industries. In steel industry, demand for beams of lesser strength can be substituted by beams of higher strength (Bassok et al., 1999). Another example is in aluminum tube manufacturing industry, where aluminum blooms of various sizes can be substituted in the process of obtaining the final tubes (Balakrishnan and Brown, 1996).

In logistics, the flexibility in operations can be enabled by location-based substitution, which is considered as an example of demand fulfillment flexibility. Location-based substitution is defined as meeting the demand at one location with a product at another location, via lateral transshipment. Here the substitution costs involve the actual relocation costs of the products instead of goodwill costs. The transshipment may take place in the occasion of a stock-out, or may just serve the purpose of balancing the safety stock at the locations. Archibald et al. (1997), Axsäter (2003) are examples of such substitution schemes.

Substitution is a practice observed also in retail. In retail, the substitution takes place when the consumer replaces the brand/product requested with the offered (and closest) brand/product. This implies a distinction between manufacturer-driven substitution and customer-driven substitution. In manufacturer-driven substitution, the manufacturer has control on how the demand will be met, and it is assumed that the manufacturer may prefer not to meet the demand (e.g. backlog if possible), or may offer a substitute product depending on its inventory level, production status, or suitability of substitution, and incur an associated cost.

In customer-driven substitution, the consumer has a preference among the available products and substitution takes place according to these preferences. The preference of a customer is a characteristics of the customer, which cannot be influenced by the retailer. In contrast to the manufacturer-driven substitution, the substitution is not directly controlled by the retailer (seller). Given the substitution behavior of the customer base, the retailer may decide on the product assortment and how much inventory to keep for each product. There are two types of customer-driven substitution, one is static (assortment-based), and the other is dynamic (stock-out based). In assortment-based substitution, the consumer selects his first-choice within the assortment (given that the choice makes his utility non-negative), and if the choice is not available due to stock out, does not opt for a second choice.

In stock-out based substitution, the consumer makes a selection in the assortment considering the inventory availability. Stock-out based substitution can be modeled as either via an exogenous demand approach or a utility based approach. In exogenous models customer chooses his favorite product with a probability, in the case of unavailability, a substitute product will be taken following to a probability function.

In addition to the presence effect there can be other effects stimulating the customer substitution. Substitution is both related to capacity levels, and demand. So the factors affecting demand also contributes to substitution, such as price, other producers price, availability of product substitutes and complements, time window that product is unavailable, quality of the product, and customer benefits.

Substitution can also be discussed in it relation with “canibalization”, a concept in marketing. Ccannibalization refers to a reduction in sales volume, sales revenue, or market share of one product as a result of the introduction of a new product by the same producer. If there is single supplier incorporating the customer substitution should be considered, in order to hedge against cannibalism. In Germany the yoghurt category displays a high degree of differentiation. As a result, cannibalization is a critical issue for firms that offer multiple products within a particular product category (Schröder, 2012). This is because, the unmet demand can spill to competitive products. Gruen et al. (2002) examine consumer response to stockouts across eight categories at retailers worldwide, and report that 26% of customers substitute with a different brand, 19% substitute with the same brand, 15% delay purchase, 31% switch to another store, and 9% never buy that product.

We study two research problems to contribute to the stream on manufacturer-driven substitution and customer-driven substitution. In the first problem, we study a manufacturer that produces n products in a single facility, where the products are substitutable. The manufacturer operates under a produce-to-stock scheme, keeping real-time status data. It keeps stock of n products, and backorders the demand, there is inventory holding and backorder costs associated with them. The cost of substitution depends on the requested product and the substitute product. The cost structure allows for one-way or two-way substitution, or for any general substitution scheme. Demand is stochastic and the manufacturer balances the risk of overstocking with the risk of backordering by production scheduling (switchover) decisions, where production times are random variables and production decisions are preemptive and there are no setup costs associated with switch-overs. In addition to production scheduling decisions, the manufacturer also can control the inventory levels by substitution decisions. We assume that there is a time and cost associated with each substitution decision. Time it takes for substitution may, for instance, correspond to the time it takes to make necessary modifications on the substitute product. Thus, the manufacturer faces the problem of setting the stock levels for the products, dynamically determining which product to produce next, and which product to substitute with which other product, considering the limited capacity, and the associated costs. Here manufacturer benefits from two types of flexibilities: process flexibility (Sethi and Sethi, 1990) and substitution flexibility. Process flexibility of a manufacturing system relates to the set of product types that the system can produce without major setups. The substitutions are both anticipative and dynamic in nature. They are anticipative because it balances the savings in holding cost with that of the substitution costs. They are also dynamic because real-state information is used for substitution decisions.

In the first part of the thesis the described problem is studied under two-product and n product settings. For two product setting, we characterize the structure of the optimal policy. It is shown that the optimal policies are characterized by an idling (hedging) point, a switching curve and a substitution curve. Then, through numerical analysis we quantify the benefit of substitution, idling and production scheduling.

We then turn to the analysis of n-product case. Identifying the optimal dynamic production, inventory and substitution decision in the presence of n products is challenging due to the need to keep the state information on all n products. For that reason several heuristic policies that possess the dynamic behavior are proposed. The heuristic policies are contrasted with each other, and with a benchmark lower bound and upper bound, to assess their performances. We identify in terms of the product characteristics, which heuristic policy performs out the others under which conditions. We look at the impact of substitution on the performance of the proposed policies.

In the first part of the thesis, the customer substitution characteristics are not incorporated into the decision-making process of manufacturer. In the second part, the manufacturer (or seller) takes the customers purchasing behavior into consideration. It is assumed that there are two products for which the manufacturer faces stochastic demand. Customer substitution behavior is affected by the price and stock availability. There are two periods. At the beginning of period 1, the manufacturer decides on the quantity to produce for each of the two products. Also at this time decides on prices of the products for the first period. Demand is realized by the end of the period. Depending on the quantities produced, there might be over-stock for some of the products and under-stock for others. In the second period, based on the over-stock and under-stock decision manufacturer decide on the price in the subsequent term, based on the realization of the demand functions. Second period problem is deterministic, because the manufacturer at this time has the knowledge of the demand and tries to just allocate the remaining stock by pricing and considering the substitution effects. Under this problem, we aim to find optimal stock levels and the prices that influence the demand. We aim to identify the benefit of exploiting substitution behavior of the customer. We determine the optimal profit under the two period two product setting. We also consider the variations of this problem to address the relevant research questions.

Structure of the thesis

This work is organized as follows: In Chapter 2, the studies on policies in the presence of manufacturer-driven and customer-driven substitution are reviewed. We also review the studies on dynamic scheduling of continuous-time Markovian queues, which is the base setting adopted in our models in Chapters 3,4 and 5.

Chapter 3 opens the discussion of manufacturer-driven substitution policies. In this chapter, a characterization of the problem for two-product setting is provided. The results on value of substitution and the value of information on net inventory status are presented. Chapter 4 and Chapter 5 follow by extending the analysis of the problem introduced in Chapter 3 by considering multiple products, and by proposing different heuristics. Chapter 6 gives a detailed numerical study for the heuristics proposed.

In Chapter 7, the seller's response to customer driven substitution is discussed. Specifically, stock keeping and pricing policies of the manufacturer as an hedge against the customer's substitution behavior are studied. Insights on the benefit of considering the substitution behavior in determining the stocking and pricing decisions are presented.

In Chapter 8, through numerical analysis, insights on the benefit of considering the substitution behavior in determining the stocking and pricing decisions are presented.

Chapter 9 presents the conclusions and discussion on possible extensions.

CHAPTER 2

LITERATURE REVIEW

In literature, substitution policies are studied in two frameworks; manufacturer-driven substitution policies and manufacturer (retailer or seller) operating policies which takes into account customers substitution behavior.

Customer-driven substitution is further categorized based on the behavior of the customer. The customer may either substitute when his favorite product is out of stock, or based on the assortment, without considering the inventory level of the favorite product. The former is called as dynamic (stock-out based) substitution, whereas the latter as static (assortment-based) substitution. In the following, a brief literature review is made on the relevant studies and the current study is positioned in the literature.

In this part, in addition to these, literature on substitution in logistics sector will also be discussed briefly. Lastly, our study's differences with current work on production scheduling and inventory management is put forward.

2.1 Studies on Manufacturer-driven Substitution

Shumsky and Zhang (2009), Li et al. (2010), Bassok et al. (1999), Dawande et al. (2010), Gallego et al. (2006), Rao et al. (2004) are some of the studies that consider operating policies in the presence of manufacturer-driven substitution. In literature manufacture driven substitution is usually used in the presence of a superior product that can fill the demands of the products that are inferior to it. Manufacturer-driven substitution policies might be studied under stochastic or deterministic demand. Substitution policies can be examined under two classes: planned substitution and dynamic substitution.

In planned substitution, to minimize inventory handling, production and setup costs, the substitution option is considered, the aim is to balance these costs by substitution costs. In Dawande et al. (2010), Balakrishnan and Geunes (2000) such problems are considered.

Dawande et al. (2010) analyze the trade-off between the substitution costs and the production changeover costs in the presence of a superior product which can fill both of the demand types. Demand is periodic and deterministic. The setting for the model is discrete time two product lot-sizing model with changeover, inventory holding and substitution costs. Their key finding is that, the extent of the substitution is dependent on the ratio of the substitution cost to changeover cost.

Balakrishnan and Geunes (2000) studies a problem in which different products can be produced by substitutable components in the context of aluminum manufacturing. The authors evaluate substitution policies using data that is taken from an aluminum tube producer, and conclude that when there are demand fluctuations throughout periods, the substitution policies will reduce the system costs. Demand is periodic and deterministic, the authors model the problem as an integer program, and they devise a solution method based on dynamic programming.

In our study, we consider stochastic unit demand modeled as a Poisson process. We consider a flexible manufacturing setting where setup costs are negligible. Thus, in contrast to the studies by Dawande et al. (2010) and Balakrishnan and Geunes (2000), setup cost is not a factor affecting the substitution decisions. The decisions are rather stimulated by dynamic and planned substitutions, whereas under deterministic demand “dynamic substitution” is not under consideration. Furthermore, in this work, holding and backordering costs are the major cost components.

Bassok et al. (1999) consider a setting with full downgrading in which excess demand for lower grade product can be satisfied using higher grade product at manufacturers will. They model a two stage profit maximization problem for multi-product substitution. In the first stage the authors decide on what quantities of products to order and in the second stage the allocation of products to demands take place. For allocation, they show that a greedy allocation policy is optimal under quite general assumptions.

Rao et al. (2004) consider a single period multi-product inventory model with stochastic demand and setup costs under the possibility of downgrading. The decisions are which products to produce, how much to produce and how to allocate/substitute the products to satisfy the demands. They model a two-stage integer stochastic program with recourse and propose solution methods to this problem. Unlike Bassok et al. (1999) the unit substitution cost need not be identical.

Shumsky and Zhang (2009) works on determining the one-way substitution policies under stochastic demand, and examine a problem that resembles the problem of Bassok et al. (1999) for multi-period. In Shumsky and Zhang (2009), there are multiple product types that are prioritized. Upgrading is considered. If customer demand cannot be met with the product he wants, he can be given an upgrade product. In the study only a single-level upgrade is considered. Capacity purchase/allocation, the rationing (substitution) between products decisions are determined by the help of Markov decision processes. The authors consider a single opportunity to invest in capacity, before any demand is realized. There are multiple periods, in which the customers arrive. Demand not satisfied each period is lost. The optimal policy is if there is inventory of the product type, supply the customers who asks for it. If a product inventory is depleted, upgrade the customer's from higher priority product up to some threshold for inventory. This policy type also holds for repeated capacity investment. They show certain monotonicity results for threshold values, and find protection limit bounds. In this model, capacity is invested before any demand is realized. The authors only consider manufacturer based substitution.

In our study, there exists a single resource that is used to manufacture the substitutable products and thus scheduling of the resource is explicitly considered over an infinite horizon. Bassok et al. (1999) and Rao et al. (2004) on the other hand consider single-period problem. Due to the assumptions of problem, backorders are not allowed whereas in our setting we allow for backorder. Our aim is to contrast different production schemes and substitution policies according to whether the policies use dynamic status information or not, and derive heuristic policies that perform well for the multi-product setting.

In Li et al. (2010) a new product is introduced to market in the presence of an old product. During the transition from the old product to the new product, the old product can have shortages. If there is shortage, the producer can offer the new product. The new product is an upgrade on the old product, and the manufacturer offers the customer a better product. The authors consider the initial inventory levels for both products for the deterministic problem, then study the stochastic problem, for optimal ordering quantities, They also investigate the problem of delaying the introduction of a new product, given the initial scheduled introduction and the inventory level of the old product. The authors show that the optimal substitution policy is a time-varying threshold policy. The optimal delay in new product introduction is determined, given the initial inventory level of the old product.

Gallego et al. (2006) study the inventory management production system that results in multiple grade parts in semiconductor industry in which a higher grade product can fill the demand for lower grade product. They model this problem as an inventory cost minimization problem with service constraints. They define the critical part, for which service constraint is tight. They propose heuristics based on myopic allocation utilizing the critical part notion.

Xu et al. (2011) study a substitution model that combines supplier-driven and consumer-driven models. The supplier may choose to offer a substitution at a discount price, and customer has choice to accept the offer. The customer's decision is reflected via acceptance probabilities. The authors study an inventory system in which a supplier meets the demand using two mutually substitutable products over finite periods, with a single replenishment opportunity at the beginning. Demand is assumed to be compound Poisson and Markov decision process is used to model the problem. They show that the optimal substitution follows a threshold rule.

Our study differs from the above-mentioned studies (Li et al., 2010; Gallego et al., 2006; Xu et al., 2011) in that we explicitly consider the scheduling decision, planned and dynamic substitution over a finite planning horizon. The research questions we aim to answer are related to the interplay between production, stocking and substitution decisions.

2.2 Studies on dynamic scheduling of multi product base-stock type

In this thesis, the problems studied under manufacturer-driven substitution (Chapters 3,4,and 5) adopt a multi-product make-to-stock queue setting. In this setting optimal production scheduling, stocking and substitution policies are determined. Actually, in this setting determining the optimal production scheduling and stocking policies is a well-known problem class in literature. We extend the existing studies by adding the substitution decisions.

In the literature there exists a stream of work that study dynamic production scheduling policies for the multi-product production/inventory systems. In these systems, generally demand is assumed to be Poisson and production times exponential. Planning horizon is infinite horizon. Specifically, these problems assume the following setting: there is a capacitated server which can produce for n different product types. For each product type, the production is towards inventory. For each product type, demand arrives independently following Poisson distribution with possibly different rates. If there is no inventory with the associated product, there can be backorders or lost sales. Under real-time information on net-inventory levels of the products and in a continuous time setting, which product must be produced when (production switching) and the inventory levels (hedging points) for each product are decided upon. In this problem type, production setup times are negligible.

Veatch and Wein (1996) studies this problem under exponential service times, the aim is to minimize the inventory holding and backorder costs. Each product has different and independent production rate and customer arrival rate. There exists a single server that does the production. For n products with non-identical production rates, for this problem optimal inventory and production scheduling policy is not characterized yet in the literature. Veatch and Wein (1996) develop heuristic policies. In this study, numerical analysis part is limited and the performance of the heuristic policies are not measured well.

Ha (1997b) under two-product setting and under the assumption that the service rates for the products is equal, shows the optimal policy structure is "hedging point, switching curve" type. By using numerical analysis, he put forward the importance of the production scheduling decisions. The author shows the insensitivity of the total cost of the system to the production scheduling policies. Findings show that meaningful production policies, although they are not optimal, can result in low costs.

Under two-product setting, De Vericourt et al. (2002) partially characterize the optimal structure of the problem under the assumption of different production rates for different products. When the product with the higher $b_i\mu_i$ is backordered, the production priority is given to this product. De Vericourt et al. (2002) partially characterize the switching curve in closed-form.

Zheng and Zipkin (1990) analyze this two product make-to-stock problem to determine the value of the centralized inventory information. In their decentralized information model, production scheduling is done under FCFS scheme, i.e., inventory levels are not considered for production scheduling. Their centralized information model is the longest queue model, in which the server serves the product which fall behind in inventory level. They give explicit solution for the identical product longest queue model. Their work analyze different production schedules, and their effects.

Perez and Zipkin (1997) consider the limiting control problem that stems from the multi-product make-to-stock problem and they are able to characterize the optimal solution to the limiting control problem. They further elaborate on the model by defining heuristics that utilize the results from limiting control problem and myopic allocations of production priority.

Zipkin (1995) consider FCFS and longest queue production scheduling. The author claim that the standard deviation of the queue lengths under these policies are good proxies for system cost hence they give approximations for defining these standard deviations.

In our study, we elaborate on multi-product environment and we assume all the products have the same production rate. For the multi-product inventory production problem, defining the optimal policies is hard so we choose to suggest heuristics for the problem. While we form heuristics, we utilize the information for the heuristic approaches in production scheduling literature which are mentioned above.

Our context is quite similar to those considered in the studies on dynamic scheduling of multiple products with stochastic demand. We differentiate from the aforementioned studies in that we consider the substitution decisions together with scheduling and stocking decisions. We propose heuristics that borrow approaches from the studies above, and propose a new approach to determine a well-performing scheduling, stocking and substitution policy under certain environments.

2.3 Studies on one-way substitution in hybrid remanufacturing/ manufacturing systems

Management of substitutable items naturally arise in the context of remanufacturing. In many cases the remanufactured items can be regarded as a downgrade of the brand-new item. Meeting the demand for the remanufactured item with the new item constitutes the case for one-way substitution. It is expected that price of the remanufactured item is lower than that of the new item, thus customers would not object to such substitution while manufacturer would incur an opportunity cost. Bayındır et al. (2005) and Bayındır et al. (2007) study the benefit of substitution when the manufactured items share a single resource versus are produced separately. Inderfurth (2004) studies a newsvendor problem with two products (remanufactured item and newly manufactured item) where demand of the remanufactured item is met with the new item in case of a shortage of the remanufactured product. Manufacturing lead times as well as return process of the items are taken into consideration when deriving the optimal inventory levels. Ahiska and Kurtul (2014) analyze similar problem to that of Bayındır et al. (2005) under a periodic-review inventory policy. Benefit of substitution is quantified. Pineyro and Viera (2010) model one-way substitution of a remanufactured item with a new item in economic lot sizing context. Demands for both items are deterministic and dynamic, and the model is NP-hard. Li et al. (2006) also study the production planning problem in hybrid remanufacturing/manufacturing environment. Economic lot-sizing model is constructed and the aim is to minimize the total of setup cost, manufacturing and remanufacturing cost, substitution and holding cost. A dynamic-programming based approximation algorithm is developed to obtain near-optimal solutions and managerial insights are obtained on the effect of substitution and remanufacturing.

2.4 Studies on Customer-driven Substitution

Studies on customer-driven substitution will be reviewed under assortment-based substitution and stockout-based substitution. Under both approaches, there are studies that assume centralized decision-making or a competitive setting.

2.4.1 Studies on assortment-based substitution

Kok et al. (2008) exemplifies the demand models that are used to model customer-driven substitution as Multinomial Logit (MNL), locational choice and exogenous demand models. Multinomial Logit is a utility based model in which there is also the “no purchase” option. Utilities have random components. Individuals choose the product with the highest utility among the set of available choices. But the choice is also a random variable. MNL is one example of utility based models. Another utility based model is the locational model, in which product’s utility is defined as a function of its price and location (i.e., differentiation). Exogenous demand models state the demand for each product and what an individual does when the product he or she demands is not available in the assortment. To be exact, customer chooses his favorite product with a probability, if that product is not in the assortment, he can choose to substitute it with another product with some probability.

Modeling the substitution behavior of the customer is important especially when determining the optimal product variety and assortment policies. In our model in Chapter 7, we focus on such policies. We address the problem of a seller (a manufacturer or a retailer) that sells two substitutable products. The manufacturer decides on initial stock level and selling prices over a planning horizon of two-periods. Since the products are substitutable, if one of the items are out of stock at the end of the first period, customers may substitute the product with the other. We explicitly reflect the stock-out based substitution behavior of the customers while excluding the assortment-based substitution in the demand model.

To work with customer-driven substitution models, there is a need to understand the customer behavior and estimate the substitution probabilities. Certain studies discuss how to use the available data to make such estimates.

In Anupindi et al. (1998), the authors model the substitution as follows. If a product is unavailable, then this increases the other products' likelihood of selection. Sales of products that stock out affect the sales data. Knowledge of true demand rates and substitution rates is important. A model is developed in a vending context. The authors analyze the model for perpetual data and periodic data. For perpetual data, the authors develop maximum likelihood estimators. For periodic data the authors develop an algorithm to find MLEs.

Another work that combines the theory with the empirical data is the study by Kök and Fisher (2007). The authors use Multinomial Logit (MNL) to model the consumer choice. It is assumed that the consumers might accept substitutes when their favorite product is unavailable. They first present a procedure for estimating the parameters of substitution and demand using real-life data. Next they develop an iterative scheme for solving the assortment problem. Finally they establish structural properties based on a heuristic.

Utility based models are quite general models in which customer choose highest utility product. Very special and detailed formulations can be created as in Gumus et al. (2011). MNL is a special case of utility based model, and Dong et al. (2009), Ryzin and Mahajan (1999), Maddah and Bish (2007) are some of the studies that utilize MNL. A utility based model that utilizes locational choice is used in Gaur and Honhon (2006).

Gumus et al. (2011) model a complex demand behavior based on the utilities from the products. For every product there is a customer pool. A customer of a product is labeled either loyal or switcher. Loyal customers do not want other product. A switcher prefers a product but may switch to other product if the preferred product is too expensive based on the valuation. If switcher's valuation exceeds product price for a preferred product, she might select other product which the value of the other product that do not exceed the other products valuation. There is also temporal part in customer demand, in which she might linger in the market for one more period to wait for more suitable product. Their empirical results show, and their theoretical results support

(i) Stores that devote more capacity to stock a product category offer deeper discounts compared to stores that devote comparatively less aisle space to a category;

(ii) Stores offer deeper (shallower) discounts on expensive products, compared to inexpensive products within a category, if the extent of inter-product substitution in their assortment is low (high);

(iii) Stores sequentially (simultaneously) promote products within the same category, if the extent of inter-product substitution in their assortment is low (high).

Ryzin and Mahajan (1999) consider joint assortment and inventory decisions for a product line under MNL consumer choice model. The authors consider a static choice model, in which the customer is to decide which product to choose without knowing about the assortment. This implies substitution does not take place. They also assume all the products of the product line have the same unit cost and price. They show the optimal assortment consists of k products having the largest consumer valuations.

Maddah and Bish (2007) consider MNL based utility function in modeling the nd static choice behavior of the customer. Their study is extension of Ryzin and Mahajan (1999) in that the authors assume, a customer would not substitute if her favorite choice is not in the assortment, but consider the affect of the prices on the utilities of the customers. They derive the structure of the optimal policy and propose that dominance relationship simplifies the search for optimal assortment.

Gaur and Honhon (2006) employ a locational choice model. They consider single period assortment and inventory management problem. They contrast their findings with the models that use MNL.

Akçay and Tan (2008) consider a competitive market for substitutable products. The substitution is characterized by an exogenous demand model. Firms' assortment decisions leads to the substitution behavior. The authors consider cooperation of independent firms that offer combined set of products to customers. A supplier gives the product to other supplier in cooperation with discounted price, so that both parties will profit. Such a cooperation is beneficial between symmetric firms, but for asymmetric firms, in order such a mechanism to be beneficial to both parties requires some conditions. One of the important findings is commonality in product assortments has an adverse effect.

In our study, in Chapter 7 we consider joint stocking and pricing decisions assuming that product line items are already determined. We assume a there exists a single decision-maker. Similar to the studies above we consider perishable fashion products with relatively long lead times that necessitates a single stocking opportunity. However, in contrast to the studies above, assortment decision is not under consideration. Customers may substitute the products, and the substitution behavior is affected by price and product availability. We consider a dynamic pricing policy that exploits the substitutability of the products.

2.4.2 Studies on stock-out-based substitution

Past work that relates the customer's substitution behavior to the stock-out occasions may either take an exogenous demand approach or a utility based approach. The papers that use exogenous demand models under centralized decision making and stock-out based substitution are Nagarajan and Rajagopalan (2008) and Parlar (1988). And game theoretic models that use exogenous demand model is Li and Ha (2008), Lippman and McCardle (1997), Netessine and Rudi (2003), Zhao and Atkins (2008) and Parlar (1988). Studies that assume a utility-based demand model and a stock-out based substitution behavior are Mahajan and van Ryzin (2001), Van Ryzin and Mahajan (1999), Dong et al. (2009) and Suh and Aydin (2011). Among these studies, Li and Ha (2008), Lippman and McCardle (1997), Zhao and Atkins (2008), Van Ryzin and Mahajan (1999) take a game-theoretic approach while others assume either only centralized decision-making, or both.

In Nagarajan and Rajagopalan (2008), there are two substitutable products, with total demand D (which can be random), the individual demand proportions are p and (1-p) (p is random) (hence the two products are negatively correlated). The retailer decides on stocking levels at the beginning of a period, if there is unmet demand for a product, its γ proportion will purchase other product if it is available. They analyze both single and multi-period problems. Their important finding is, for γ values that are not too large, there is a special optimal policy pattern, which is called partially decoupled policies, which means the optimal base stock of the product is independent of the other products inventory level, if the latter is above its base-stock. By a numerical study, they show that, heuristics based on partially decoupled policies worked good. In our study, we consider a completely different demand model. Inspired from Atkins and Zhao (2009), we propose a unifying demand model that ensures the consistency between the price-based and stock-out-based substitution behavior. We jointly consider stocking and pricing decisions, and optimally determine those decision variables. Parlar (1988) studies two newsboys that decide on stock levels for two substitutable products. A single period problem is considered. After the demand is realized a portion of the demand that is under-stocked at one location can be substituted by the overstock at another location. Both centralized and competitive settings are analyzed. Substitution probabilities are exogenous, and substitution is contingent on stock out occasion. The existence and uniqueness of Nash equilibrium is shown. The author shows that centralized decision-making lowers the ordering quantities compared to independent or competitive setting. In Li and Ha (2008) studies and extension of the competitive problem of Parlar (1988). The manufacturer decides on capacity levels at the beginning of the term. If there is more demand than capacity, the firm can increase the capacity until reactive capacity, with some cost. Substitution is decided, after firm's reactive capacity is depleted. They study the problem in a competitive environment. It is shown there is Nash equilibrium for this model. Their findings are that additional reactive capacity or lower cost of reactive capacity has positive competitive effect on a firm, but a negative effect on the competitor.

In Lippman and McCardle (1997), for a duopoly, the firms must choose an inventory (or production) level for a perishable product. After initial demand occurrence (which is a random variable), a splitting rule specifies how initial industry demand is allocated among competing firms and how excess demand is allocated among the firms with remaining inventory. Deterministic splitting, simple random splitting, incremental random splitting and independent random demands are considered. They examine whether Nash equilibrium is present and unique under different splitting rules, and their conditions. Their important result is if all excess demand is allocated, then competition never leads to a decrease in industry inventory. Netessine and Rudi (2003) is the extension of Parlar (1988) to n-product setting. The authors consider optimal stocking of substitutable products. They model the substitution as a fraction of unmet demand will select other firm, if this firm cannot supply, demand is lost. They show that for centralized problem when the demand is multivariate normal, the total profit is decreasing in demand correlation. For centralized problem, the cost function may not be concave in one of the variables. But they provide first order optimality conditions. For multiproduct problem, they show the existence of Nash equilibrium in competitive model. Our study (in Chapter 8) can be considered as an extension of Parlar (1988) and Li and Ha (2008). We consider a two-product setting, where the stocking decisions are given at the beginning of the period. At the end of the period, if demand exceeds stock levels then substitution takes place. We additionally consider price-based substitution and dynamic pricing decisions, and determine the optimal level of those decisions. In contrast to Parlar (1988) and Li and Ha (2008) we study the problem under centralized decision-making. Our model is not as general as Netessine and Rudi (2003) since we study two-products and assume the demand uncertainty takes a very simplistic form. On the other hand we introduce dynamic pricing in the presence of substitutable products and ask research questions towards value of considering substitutability in the decisions. We consider a totally different demand model than Lippman and McCardle (1997), include pricing decisions and study a centralized setting.

In Zhao and Atkins (2008) the demand for a product is a function of the price vector. There are n products. Unmet demand can spill to other product demand at a fraction. The firms compete in price and inventory. They show the quasiconcavity of the cost function under some regularity conditions. They show Nash equilibrium exists under reasonable assumptions. They derive the conditions for Nash equilibrium to be unique. In Zhao (2008) the author studies the same setting with the extension of characterizing the coordination among the newsvendors. In Zhao and Atkins (2009), the authors compare two competing strategies among two retailers: in the first, demand mismatch is handled via a transshipment agreement and in the second no such agreement takes place and customers substitute arbitrarily. Nash equilibrium is characterized and conditions under which transshipment agreement is preferred are identified.

Our study is similar in spirit to the problems considered in Zhao and Atkins (2008), Atkins and Zhao (2009), and Zhao and Atkins (2009), in that we study optimal stocking and pricing decisions. Our setting is a centralized one rather than a competitive one. We adopt the same approach to model the demand spill in the occasion of mismatch between supply and demand (Zhao and Atkins, 2009). This model allows for the consistency of price-based and stock-out based substitution. However, we consider a two-period setting rather than a single period setting to explicitly model stock-out based substitution and to exclude assortment based substitution in our demand model. Furthermore, we study the impact of dynamic pricing. We obtain the values of optimal stock levels and prices and aim to obtain insights on how those decisions are made, whereas those are not the questions addressed in the above mentioned studies. Our research questions are completely different in that we focus on the effect of substitution on the decisions and the cost of ignoring substitution.

Mahajan and van Ryzin (2001) consider stock-out-based customer-driven substitution using a utility model. The study combines assortment-based and stock-out based substitution, in the sense that the customers modify their choice probabilities depending on the inventory level over the time. Customer demand is modeled dynamically based on utility maximization. Consumer choice is based on availability, rather than price and quality. There is a utility vector for a customer. Customer buys the highest utility factor product, until the product is finished or her demand is depleted, and then turns on second highest utility product. The firm must choose the initial inventory levels and the assortment. Using sample path analysis, the authors provide the structural properties of the expected profit function.

In Van Ryzin and Mahajan (1999), the same problem is studied in a competitive setting. Each firm chooses its inventory level however the inventory levels of the firms affect each other's due to substitution. The authors use sample path analysis to show Nash equilibrium. Their important finding is there is a bias towards overstocking caused by competition. They propose a method to compute the equilibrium.

Dong et al. (2009) study dynamic pricing and inventory control of substitute multi-products. They are not considering capacity assigning problem, instead, they take the initial inventory levels as exogeneous. Control is done by pricing. This study can be regarded as an extension of Mahajan and van Ryzin (2001). The authors use pricing as a means to change the substitution probabilities dynamically over time, depending on the remaining selling periods and the inventory level of the products. In Mahajan and van Ryzin (2001) price is not a decision variable and there is no time horizon.

Suh and Aydin (2011) consider the problem of pricing of two substitutable products over a finite selling period. They have exogeneous initial inventory levels. They use Multinomial Logit model to model the customer choice. In contrast to Akcay et al. (2010) and Dong et al. (2009), they restrict to a two-product problem, which allows them to establish a number of analytical results about how the optimal prices behave with respect to time and inventory levels. They show the optimal prices themselves may not be monotonic in inventory levels, but the optimal price gaps and purchase probabilities show this behavior.

Our study is related to the studies above in that those studies also consider dynamic pricing of substitutable products in the presence of inventory. Substitution behaviour is affected by the prices and the availability of the product in the assortment. To model the choice behavior of the customer the authors use MNL model. Inventory is stocked only at the beginning of the horizon and cannot be replenished throughout the selling horizon. In the studies, it is assumed that the prices can be updated (prices could be marked-down or marked-up) after each customer arrival, which may not be possible in practice. The usage of MNL model provides a basis for consistency of price-based and availability-based substitution: the same choice dynamics govern those substitution behavior. In our model, we do not allow for very frequent price updates, prices are updated only twice throughout the selling horizon. Through our demand model, we differentiate between assortment-based and stock-out based substitution behavior. We also make dynamic pricing and stock level decisions jointly, rather than assuming an exogenous inventory level.

2.5 On the demand models in the presence of substitutable products

Apart from utility based models, a line of research assumes linear demand functions to model "substitute products". The model considered in Chapter 8 is a linear demand model. Linear demand models of multiple products allow for modeling the substitutability or complementarity of the products. We review the studies in operations management literature that assume linear demand function to address product substitutability. As noted in Lus and Muriel (2009), consumer choice models such as the Multinomial Logit (MNL) model, are typically used in the revenue management and product assortment literature, more simplistic models may be appropriate for strategic decisions to gain analytical tractability.

One such model is linear demand model in which price is linear function of the quantities offered by firm, or equivalently quantities sold are linear function of the price of the products. In the following a group of papers are reviewed that has the following properties: The capacity allocation decision is under consideration, the capacity can be flexible or dedicated, the products are "strategic substitutes": price (or some other property such as lead-time) of a product results in a change in both own demand and the other product's demand, decentralized competitive or centralized decision-making might be undertaken. Game theoretic models that use linear demand model are Goyal and Netessine (2007), Lus and Muriel (2009), while models with centralized decision-making are Goyal and Netessine (2011), Bish and Suwandechochai (2010), Chod and Rudi (2005).

Goyal and Netessine (2007) use a linear type demand model, in order to model customer-driven substitution. They consider three type of decisions for competing firms: choice of technology (technology game), capacity investment (capacity game), and production quantities (production game). The technology and capacity games occur while the demand realization, and the production game is postponed until after the demand is realized. They conclude that flexibility is not always the best response to competition.

Lus and Muriel (2009) argue that price as a function of production quantities models the substitutability effect better than quantity as a function of product prices. They study the optimal investment mix in flexible and dedicated capacities in both monopoly and oligopoly settings. They find the optimal investment in manufacturing flexibility is likely to decrease as the products become closer substitutes.

In Goyal and Netessine (2011) the price of the product depends on the quantity of that product in the market as well as the quantity of the other product. Product flexibility is being able to manufacture multiple products on the same capacity. Volume flexibility is the ability to produce above/below the installed capacity. They model different flexibility combinations. Their findings are the value of volume flexibility is a function of demand correlation between products. The value of product flexibility decreases with demand correlation. The volume flexibility better combats aggregate demand uncertainty, while product flexibility is better at mitigating individual demand uncertainty for each product.

Bish and Suwandechochai (2010) use a linear type demand model, in order to model customer-driven substitution. Their main topic is how the degree of substitution and level of operational postponement effect the optimal capacity and the resulting expected profit. They study two forms of operational postponement, quantity postponement in which product capacities can be assigned after demand realization and price postponement in which price is decided after demand is realized.

Chod and Rudi (2005) study two types of flexibility, resource flexibility and responsive pricing. With Bish and Suwandechochai (2010) terminology, quantity and price postponement. They further analyze the effect of demand variability and correlation. Their finding is value of flexibility is most significant when the demands are highly variable and negatively correlated.

In the thesis (in Chapter 7) we assume demand is characterized by a linear function of the prices of the products. The demand function is derived from the utility maximization problem of a "representative consumer". Following the "representative consumer theory" (Anderson et al., 1992), we derive the demand function in the presence of price and stockout-based substitution. In contrast to the studies above, we do not study resource allocation. However, the way we model the product substitution is similar.

2.6 Studies on substitution in logistics

Substitution decision defining policies in production sector also enlightens the substitution decisions in logistics. In this respect, our work also is related to the works on substitution in logistics sector. In logistics substitution policies are examined in two headings: “lateral transshipment” and “inventory pooling”. Substitution between products and transshipment between two places resembles in approach. Works on logistics starts with METRIC model suggested by Sherbrooke (1968) where minimization of back-order levels in a two-echelon recoverable parts system is aimed under a budget constraint. They aim at determining the places of the bases and the stock levels associated with them. Although in this study a centralized system is under consideration with no lateral transshipment, they deal with the redistribution problem.

Lee (1987) extends Sherbrooke (1968) work by including lateral transshipment among members of pooling groups. They develop a continuous-review multi-echelon model for repairable products when emergency lateral transshipment between identical bases are allowed.

While the inventory and substitution decisions are taken, real time inventory state information can be considered. The studies that uses real time state information are Archibald et al. (1997), Axsäter (2003), Yang and Qin (2007) and Zhao et al. (2008). Common denominator in these studies is while stocking and substitution decisions is held, dynamic state information is used.

In Archibald et al. (1997) multiple depot inventory systems with stock transfer is analyzed when demand is high relative to capacity. They try to find both an optimal policy for reordering at review epochs and an optimal policy for satisfying unmet demand. They are able to show the "control limit" characteristics of the transfer policies.

Axsäter (2003) is a single echelon multi-warehouse problem, which considers lateral transshipment. Given a set of alternative decisions, their decision rule minimizes the expected costs under the assumption that no further transshipment will take place. Although they use the real state information, this myopic rule is then used repeatedly as a heuristic, instead of elaborating fully on an MDP.

Yang and Qin (2007) uses MDP formulation for dealing with lateral transshipment. They study the optimal control of a firm with two capacitated manufacturing plants situated in two distinct geographical regions. They derive the structural properties of the policies, under no back-order assumption.

Zhao et al. (2008) aim to characterize the optimal operating policies in a centralized spare parts dealer network where lateral transshipments are allowed. The authors show that optimal rationing levels and stocking levels are dynamically dependent on the inventory status at the dealers. However these work with dynamical state information setting, either stock replenishment decisions or production scheduling decisions are taken. They are not considered together. In the studies other than Axsäter (2003), two product setting is utilized. In our study, we plan to contribute to literature by defining dynamic policies for multi-product setting.

2.7 Studies on different flexibility types

In Sethi and Sethi (1990), different flexibility types are discussed in manufacturing in addition to definitions, their purposes, how to obtain them and the means for valuation are considered also. The relevant flexibility types are product flexibility in which new parts can be substituted by other parts with ease, process flexibility which is changing product to produce without major setups, volume flexibility: ability to change the amount produced and finally market flexibility: which reflects itself with ability change the manufacturing environment according to market.

Bish et al. (2005) is about managing Flexible Capacity in a make to order environment. Their paper's main concern is how swings in production induce larger order variability at upstream suppliers with different capacity allocation decisions. Their models are single echelon, variability propagating upstream in supply chain is total production. In our two-echelon model, we also model the upstream of the supply chain, so that we can further elaborate on pooling the resources in upstream benefits.

De Groote (1994) considers a general framework for modeling and analysis of flexibility. The paper defines an incomplete order of flexibility over different manufacturing settings. They show that different flexibility settings can be beneficial in different environments, operation strategies and strategic interfaces. They show what properties a manufacturing firm needs to have to be accepted as more flexible. In our two echelon problem, we discuss different flexibilities that a system can have, and their benefits over different environments.

Gupta and Benjaafar (2004)discusses, delaying product differentiation can be beneficial to firm, however, associated costs should be considered. The advantage of delaying differentiation can be dependent on capacity utilization, product variety, and order delay requirement. In our two-echelon model, by pooling the upstream, delaying differentiation occurs. In numerical analysis part, we consider delayed and not delayed differentiation modes. Their finding is their model exhibits much greater benefit from process flexibility than inventory flexibility.

Like our model, Iravani et al. (2014) study process flexibility and inventory flexibility. They model a dynamically controlled two-product, make to stock system with stochastic processing times, and stochastic demand. The main difference between our models is their model is pure make to stock, while our model is hybrid of make-to-stock and make-to-order environments. In make-to-stock environment, inventory flexibility, substitution act as an hedge against the risk of lost sales but in the make-to-order environment, besides acting as an hedge against back order, it is a component that decrease, back-order costs.

Iravani et al. (2011) study different flexibility models, and they devise a general approach Capability Flexibility methodology to define the relative flexibility of different models.

Tanrisever et al. (2012) distinguish between process flexibility (the ability to produce multiple products on multiple lines) and operational flexibility (dynamically change the capacity allocation decisions among product families). They quantify the benefits of operational flexibility.

2.8 Positioning our study in the literature

We aim to contribute to the body of work in manufacturer-driven substitution and customer-driven substitution.

2.8.1 Manufacturer- Driven Substitution

We aim to contribute to the literature in a couple of directions. In the first problem (Chapters 3, 4, 5, and 6) considered in the thesis, we study an infinite horizon problem under stochastic demand. Besides stock replenishment and substitution decisions, the production scheduling decisions are also addressed. After the manufacturer decides on substitution, a time elapses until substitution is realized and there is cost associated with the substitution realization. We assume there is a process in order to substitute a product and there is dissatisfaction of customers due to substitution. Substitution decisions can be used to change the stock level of any product before or after the demand is realized. We study inventory holding decisions and substitution together. Production scheduling and substitution decisions are determined under inventory and backorder costs.

The contribution of the study is treatment of production and substitution decisions in a two-product and multi-product environment under dynamic setting. Determining the dynamic production and substitution decisions requires the use of stock information. Results of the study for the two-product setting permit to quantitatively determine the value of stock level information. Characterizing the optimal production, scheduling and substitution policy in the two-product continuous-time Markovian setting under identical service times is one contribution of our study. Structural results show that the production and scheduling decisions under optimal policy are similar to those characterized in Ha (1997b). Substitution decisions also follow the scheduling decisions in that they are characterized by two switching curves, one for each product. Ha (1997b) further studies some heuristics to determine the switching curve in the presence of optimal hedging point for the two-product system. We extend it to multi-product system and include the problem of determining well-performing hedging points. Comparison of different production and substitution schemes in the presence of stochastic demand is an unaddressed problem in the literature. Recently Iravani et al. (2014) has addressed a similar problem where production, scheduling and substitution decisions are under consideration for a two-product setting. However, the authors consider a pure inventory system with lost sales, whereas we allow for backorders. Furthermore, in Iravani et al. (2014) substitution takes place in the case of stock-out occasion only. We consider both planned and dynamic substitution. Authors compare the benefit of *production flexibility* and *substitution flexibility* and derive heuristics for switching curves. We quantify the benefit of stock level information when making scheduling and substitution decisions.

Under multi-product setting, because of the size of the state space, determining the optimal policies is not possible. In our study, for multi-product problem, the heuristics that define the policies quickly are suggested, and we measure the performances of these heuristics. Our problem fall into the class of determining optimal hedging and switching curves in a infinite horizon continuous-time Markovian system. Previous studies on this class of problem include Ha (1997b); Perez and Zipkin (1997); Zipkin (1995); Veatch and Wein (1992, 1996); De Vericourt et al. (2002). As an extension to this class of problems, we consider substitution policies. For the multi-period setting, we develop heuristics and identify the conditions under which each heuristic performs well. In the above mentioned studies on the other hand authors come up with a unified heuristic that would perform well on a range of settings. We on the other hand exploit the product hierarchy structure when developing one of our heuristics. As a result, each set of heuristics will perform well under different settings. Furthermore, we show that existence of substitution decisions may have an affect on the performance of the heuristics.

2.8.2 Customer Driven substitution

In the second problem considered in the thesis (Chapter 7,8), the aim is to incorporate customer-driven substitution into the operating policies of the manufacturer. Specifically, we aim to study the manufacturer's response to customer-driven substitution. We assume that for each product the manufacturer (retailer) makes inventory and pricing decisions. For each product the demand is determined by the price of the product and the price of the other product, which is due to the substitutable nature of the products. The inventory decisions are made before the demand is realized. The manufacturer may further clear out the stock-outs and over-stocks by changing the initial price of the products, and thus affecting the substitution between the products.

In our model, there are two products that have a substitutable characteristic defined by the consumer demand. The manufacturer responds to this structure both pro-actively and reactively. Proactive decisions include stock level and price decisions. Reactive decisions include mark-down pricing to affect the residual demand for the products.Through the linear demand model which is extended to two periods, we explicitly reflect stockout-based substitution: the customers who do not find their desired item in stock, wait for another period to purchase the substitute product at a possibly lower price. While modeling the stock-out based substitution, we do not use exogenous substitution probabilities but relate the cross-price effect between the products to predict the spillover rate of the demand. This approach is borrowed from Atkins and Zhao (2009). However, in their approach the actual demand and the demand spillover must take place within the same period whereas we introduce the price update decision in between. We assume demand is stochastic, and thus manufacturer should make the stocking decisions considering both the substitution behavior of the customer and stochastic nature of the demand. The model under consideration allows us to study the interaction of pricing and stocking decisions in the presence of substitution due to price differential and product availability. Considering several forms of substitution jointly, and explicitly considering the stock level decisions of a manufacturer is an unaddressed problem in the literature.

The following questions are addressed under this setting. What is the value of ability to clear out or reduce stock-outs via pricing? How much the profit deteriorates if the manufacturer gives the stocking or pricing decisions ignoring the customer's substitution behavior? How are the optimal quantity and pricing decisions made and what is the affect of substitution on the decisions?

To address these questions several variations of the main two-product two-period model are also studied.We contribute to the literature on customer-driven substitution, since we incorporate two types of substitution behaviour, price differential based and stockout based (dynamic) substitution, in the demand model. Furthermore, the manufacturer's response in terms of stocking and pricing decisions are analyzed jointly. In the literature either pricing and inventory decisions are given under static choice, ie., without considering substitution decisions, or static, assortment-based substitution is considered without inventory and/or pricing decisions. There are studies that consider dynamic pricing, as in our study, but in those studies assortment-based and stock-out-based substitution are not differentiated. Price is different for each customer and inventory levels are exogenously set. We do not consider assortment/variety decisions, but study price-based and stock-out based substitution. We aim to model stock-out based substitution while excluding assortment-based substitution. Dynamic pricing and inventory decisions are made jointly and there are two selling periods and two prices, rather than individualized prices for each customer. Incorporating such aspects of the problem, we aim to attain deeper insights on management of substitutable products.

CHAPTER 3

MANUFACTURER-DRIVEN PRODUCTION AND SUBSTITUTION POLICIES FOR THE SINGLE-ECHELON SYSTEM

In this chapter and in the following chapter, a manufacturer's joint production, scheduling and substitution problem is studied. The manufacturer produces $N = \{1, 2, \cdots, n\}$ products in a make-to stock queue setting. The products share a single manufacturing resource. The stock for each product is kept separately. When demand arrives for a product, either the demand is met, or if stock is not available the demand is either backordered, or substituted by another product.

In this chapter, first in § 3.1 the n-product problem is introduced and modelled. Then, for the two-product setting the optimal policy is characterized in § 3.2. The optimal production and substitution policies make complete usage of the state information when taking decisions. To quantify the benefit of state information, a numerical analysis is conducted. Specifically, the conditions that make the state information valuable for production and/or substitution decisions are discussed.

In the following chapter, § 4, the analysis is extended to the general n-product setting. Identifying the optimal policy for the n-product setting is challenging due to the curse of dimensionality. For this reason, several heuristic policies are proposed and the performances of the heuristics are compared. Through benchmark policies that provide upper and lower bounds, the performances of these heuristic policies are assessed.

3.1 Description of the problem

Table 3.1: Table of notation for Chapter 3

$X_i(t)$	State variable that shows the net inventory level for product i at time t.
$c(x)$	The cost incurred at state x.
h_i	Holding cost for product i.
b_i	Backorder cost for product i.
μ_i	Production rate for product i.
λ_i	Arrival rate for product i.
μ_{ij}	Substitution rate of product j with product i backorder.
c_{ij}	Cost of substitution for product j with product i backorder.
e_j	Unit vector with j'th component equal to 1.
O_1	Production operator.
O_2	Substitution operator.
$S_i(x_j)$	Idling threshold for product i as a function of product j state.
$K_i(x_j)$	Switchover threshold for product i as a function of product j state.
$S_i(x_j)$	Substitution threshold for product i as a function of product j state.

Demand for product $i \in N$, arrives independently with rate λ_i following a Poisson process. The products are stored separately, but are produced by a single production facility one unit at a time. The production rate for product i is μ_i. Production times are exponentially distributed random variables. Switching from one item to the other is assumed to take no time or money. Upon demand arrival for a certain product, if the product is available in stock, the demand is immediately met. Otherwise, demand is either backlogged, or is met by a substitute product. Substitution of product j with product i requires processing on product i and thus is assumed to occur with an exponential time with rate μ_{ij}. It is assumed that the manufacturer may substitute product j with product i before a demand is realized as an hedge against the risk of product i stock-out. At a given time a product may be processed to substitute only one other product. In other words, for each product substitution occurs one at a time. Note that under the stability condition $\sum_k \rho_k < 1$ where $\rho_k = \lambda_k \min\{1/\mu_k, \min_i\{1/\mu_i + 1/\mu_{ki}\}\}$, there exists a production and substitution policy under which the system is stable. For each i product in stock incurs a cost of h_i and for each backordered demand of product i incurs a unit cost of b_i is incurred per unit time. The cost of substituting product j with product i is c_{ij}. It is assumed that one unit of product i substitutes one unit of product j.

The system state is denoted with $\mathbf{X}(t) = \{X_i(t)\}_{i=0}^{N}$, where $X_i(t)$ is the net inventory level for product i at time t. The manufacturer decides which product to schedule, and whether to substitute product j with product i. We assume that the production and substitution decisions can be cancelled at any time (e.g., upon a state change), given that the process is not completed. This is a viable assumption for assembly environment where a product component is taken out and inserted. Thus, production, scheduling, and substitution decisions can be taken at any point in time. The production scheduling decision is whether to produce or not, and which product to produce. Substitution decision for the substitute product i is, which product should be substituted by i. Substitution decision is taken only if $X_i(t) > 0$. However, net inventory level of product j to be substituted, $X_j(t)$, can possibly be positive.

An admissible policy $U(t) = U(X(t)) = \{U_0(t), U_1(t), .., U_k(t)\}$ is defined as follows. $U_0(t)$ is the production scheduling policy which becomes 0 if production stops or takes value i if product i is produced. The substitution decision for product k is denoted with $U_k(t)$, which takes value 0 if no substitution takes place or takes value $i \in \{i : i \neq k\}$, if the product i is to be substituted. Because the system has memoryless property, the optimal policy is a state dependent Markovian policy. Thus we omit the index t and denote the state with net inventory level in the system $x = \{x_1, \cdots, x_n\}$.

The objective is to minimize the sum of holding, backorder and substitution costs, and the problem is an optimal control problem that can be modeled as a continuous-time Markov decision process (MDP). In state x, the cost is incurred at rate $c(x) = \sum_i (h_i x_i^+ + b_i x_i^-)$. Average cost criterion is used to determine the optimal policy. Let the uniform rate be $\Lambda = \sum_i \lambda_i + \bar{\mu} + \sum_i \bar{\mu}_i$, where $\bar{\mu} = \max_i \mu_i$ and $\bar{\mu}_i = \max_j \mu_{ij}$. The functional equation is expressed as follows:

$$v(x) + \frac{g}{\Lambda} = Tv(x), \qquad (3.1.1)$$

where

$$\begin{aligned} Tv(x) &= \frac{1}{\Lambda}\Big(c(x) + \sum \lambda_i v(x - e_i) + \bar{\mu} v(x) + \min\{0, \min_i \{\mu_i \triangle_i v(x)\}\} \\ &\quad + \sum_i (\bar{\mu}_i v(x) + I(x_i > 0) \min\{0, \min_{\{j : j \neq i\}} \{\mu_{ij} (\triangle_{ij} v(x) + c_{ij})\}\})\Big). \end{aligned} \qquad (3.1.2)$$

In Eq. 3.1.2 $\triangle_j v(x) = v(x + e_j) - v(x)$, and e_j is unit vector with j^{th} component equal to one and $\triangle_{ij} v(x) = v(x - e_i + e_j) - v(x)$. In (Eq. 3.1.1), g is the cost incurred per unit time, $v(x)$ is the relative value of state x under the optimal policy. In Eq. 3.1.2 the indicator function $I(x_i > 0)$ denotes whether net inventory level of product i is positive. Note in Eq. 3.1.2 it is assumed that the substitution process might be initiated for several products simultaneously, i.e., there does not exist a resource capacity on substitution. However, each product substitution takes place one at a time. In other words, when product i is decided to substitute another product, optimal substitution will be selected at a given time among all possible substitutions.

It is assumed that substitution takes time and cost, but does not consume any production capacity. This implies the substitution operation does not get to scheduled in the regular production process. The underlying assumption is substitutions are handled separately via an external resource and the cost c_{ij} reflects the amount paid for this service. In the literature, in the studies on manufacturer-driven substitution various assumptions exist on how substitutions are executed. Balakrishnan and Geunes (2000) consider joint production, scheduling and substitution decisions. An uncapacitated setting is under consideration, i.e., neither substitutions nor production consume any resource. Substitution takes place in the form of converting one component into a substitute finished product. Production and substitution takes equal amount of time. In Dawande et al. (2010), substitution takes place at a cost, resources are infinite. In Hsu et al. (2005) again similar problems as in Balakrishnan and Geunes (2000) and Dawande et al. (2010) is under consideration. A resource restriction does not exist. However in contrast to the above studies, Hsu et al. (2005) consider physical conversion of one product to the other. Only one-way substitution is allowed, and substitution takes cost and time, but does not consume any resource. Furthermore the time it takes for physical conversion is independent of the number of physical conversions undertaken. A number of previous work study single period inventory management problems (Bassok et al., 1999; Rao et al., 2004) where substitution is the defined as meeting the demand for one product from stock of another product. Substitution cost is included in the models, however, since production process is not considered, neither production/substitution time, nor resource consumption are not under consideration. Studies that assume a multi-period setting such as Shumsky and Zhang (2009) or Gallego et al. (2006), also do not explicitly consider production decisions, and substitution takes place instantaneously. Iravani et al. (2014) studies a similar problem to the one studied in this thesis. Two substitutable products face Poisson demand, separate inventory is kept for the items. There exists a single capacitated resource that produces the items and production policy is characterized by hedging point and switching curves. Upon arrival of demand if the item is out-of-stock then demand may immediately be met by the stock of the available item. Dynamic and one-way substitution is under consideration, which means substitution is not defined as physical conversion of an item but as meeting the demand with a higher quality item. In contrast to the studies above (except Hsu et al. (2005)), we explicitly take into consideration the time and cost it takes to convert one product to the other.

For the rest of the analysis, we assume identical service rate μ in order to characterize the solution. In literature, there are studies that use this assumption, in order to ease analysis. In Ha (1997b), the author characterize the optimal dynamic scheduling policy for make-to-stock queue for identical service times. In De Vericourt et al. (2002); Carr and Duenyas (2000) each customer class is served by the same product that is produced with a single service rate and its inventory is rationed. In these papers, also exponential r.v. service times and Poisson arrivals of the customers are used.

In the following, we first characterize the structure of the optimal production and substitution policy. We quantify the benefits of production decisions and substitution for the two-product setting. When the production or substitution rate differs among the products, the characterization of the structure is not possible. Thus, in the remainder of the study we assume those rates are identical for all products. For the n-product general setting, the characterization of the structure of the optimal policy was not possible, and as the number of products increases, even determining the optimal policy via functional equations becomes computationally burdensome. Thus, in the sequel we characterize the structure of the policy for the two-product setting, under certain assumptions.

3.2 Characterization of the optimal policy for the two-product setting

In the two-product setting, preliminary analysis show that optimal policy characterization is possible under the assumption of identical production rate and substitution rate parameter. Thus, it is assumed that production rate is μ and the substitution rate is μ_s. In this setting when a substitution decision is taken, either product 1 substitutes 2, or vice versa. Thus the functional equation (Eq. 3.2.1) below reflects this. For $n = 2$, let $\Lambda = \mu + \lambda_1 + \lambda_2 + \mu_s$. The functional equation under average reward criteria is

$$\begin{aligned} v(x_1, x_2) + \frac{g}{\Lambda} &= \frac{1}{\Lambda}(c(x) + \lambda_1 v(x_1 - 1, x_2) \\ &+ \lambda_2 v(x_1, x_2 - 1) \\ &+ \mu \mathbf{O}_1 v(x_1, x_2) + \mu_s \mathbf{O}_2 v(x_1, x_2) \end{aligned} \quad (3.2.1)$$

where $c(x) = \sum_{i \in \{1,2\}} h_i x_i^+ + b_i x_i^-$, $\mathbf{O}_1$ is the replenishment operator, and $\mathbf{O}_2$ is the substitution operator. These operators are defined on real-valued functions as follows,

$$\mathbf{O}_1 v(x_1, x_2) = \min\{v(x_1 + 1, x_2), v(x_1, x_2 + 1), v(x_1, x_2)\}, \tag{3.2.2}$$

$$\mathbf{O}_2 v(x_1, x_2) = \min\{v(x_1, x_2), v(x_1 - 1, x_2 + 1) + c_{12}, v(x_1 + 1, x_2 - 1) + c_{21}\}.$$

We characterize the structure of the optimal policy in Theorem 3.2.1.

Theorem 3.2.1. *The optimal policy can be characterized by idling, production switchover and substitution threshold functions. There exist $S_1(x_2)$, $K_1(x_2)$, $T_1(x_2)$ for product 1, and their respective counterparts for product 2; where $S_1(x_2)$ is the net-inventory threshold below which producing product 1 is more profitable than not producing for a given x_2. Also, $K_1(x_2)$ is the production switchover threshold below which producing product 1 is profitable over product 2 for a given x_2, and above which is reverse is true; $T_1(x_2)$ is the substitution threshold above which substituting product 2 with product 1 is more profitable than not substituting, and below which substituting product 2 with product 1 is not beneficial.*

Furthermore the following hold:

1. *$S_1(x_2)$ is decreasing with x_2.*
2. *The decrease in $S_1(x_2)$ occurs in one-step jumps in that, one unit increase in x_2 results in at most one unit decrease in $S_1(x_2)$.*
3. *$K_1(x_2)$ and $T_1(x_2)$ are increasing with x_2.*

The same properties hold for their respective counterparts.

Proof. See Appendix A.

The structure of the policy can be seen in Figure 3.1, the structure implies that the production policy is to produce if $x_1 < S_1(x_2)$ or $x_2 < S_2(x_1)$. If production takes place, and if $x_1 < K_1(x_2)$ product 1 is produced. Substitution policy is, if $x_1 > T_1(x_2)$ substitute product 2 with product 1, to balance the inventory levels and the risks. Note that $T_1(x_2) \geq 0$ implies that substitution can only take place if the substitute product has positive inventory. However, the substituted product might have a positive inventory level. The higher the rate of substitution, the lower will be the level at which substitution requirement occurs for the substituted product. The intersection of the switching curve $(K_1(x_2))$ with idling curves ($S_1(x_2)$ and $S_2(x_1)$) defines a hedging point. Hedging point (S_1, S_2) is the base stock level; for every state that satisfies $(x_1, x_2) < (S_1, S_2)$, production takes place. Under the average reward criterion, any state that doesn't satisfy $(x_1, x_2) < (S_1, S_2)$ is a transient state.

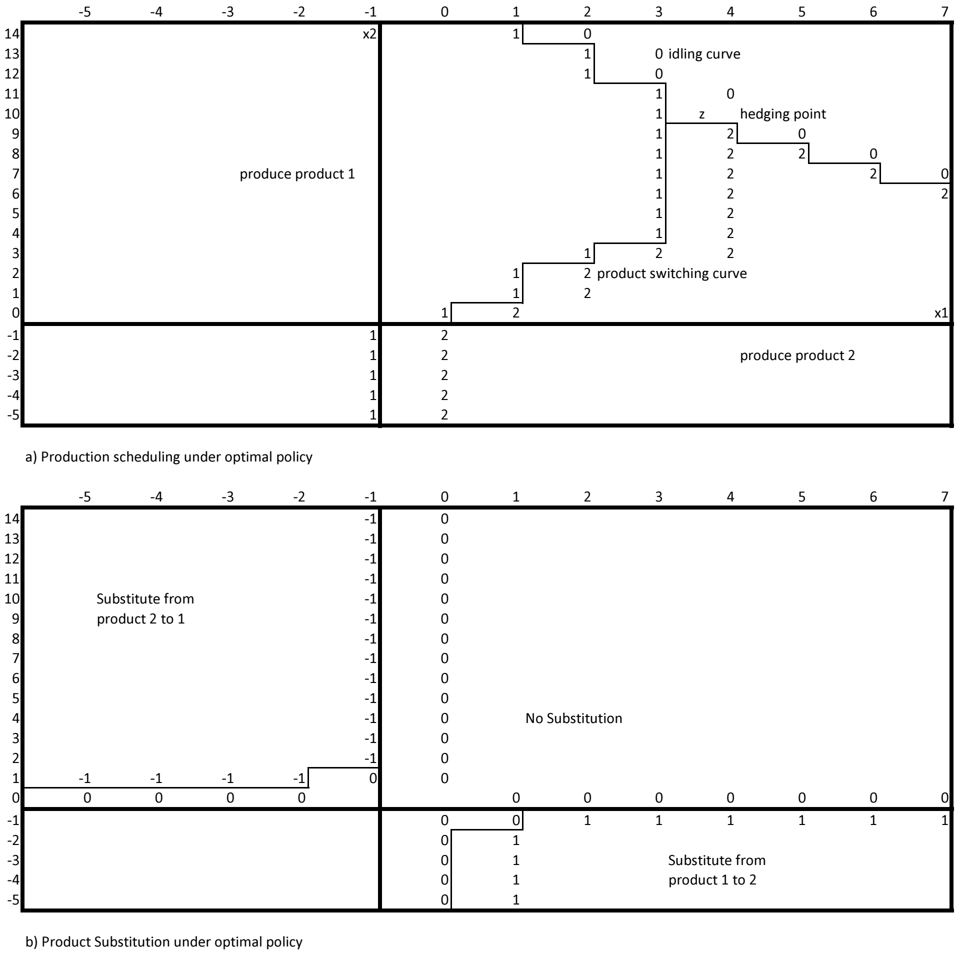


Figure 3.1: Scheduling, production and substitution decisions under the optimal policy

Effects of parameters on decisions

We expect the following effects will take place if the parameters and structure of the system change.

- As we increase the number of products for symmetric system, when $\sum_i \lambda_i$ stays the same, we expect the cost of the system increase, but the gain due to substitutions increase.

- Increase in λ_i increases base stock levels of both products. For substitution curve, product i substitutes for product j demand for only higher inventories of product i and product j substitutes for product i demand for lower inventories of product j also.

- Increase in μ decreases basestock levels of both products. For substitution curve, for both ways, shift the substitution curve for higher inventory values, as there is less need for substitutions.

- Increase in h_i decreases the basestock for product i, while increase the basestock for product j. For substitution curve, product i substitutes for product j demand for only higher inventories of product i and product j substitutes for product i demand for lower inventories of product j also.

- Increase in c_{ij} decrease the basestock product i, while increase the basestock for product j.

- Increase in b_i increase the basestock for product i, increase the basestock for product j. For substitution curve, product i substitutes for product j demand for only higher inventories of product i and product j substitutes for product i demand for lower inventories of product j also. For substitution curve, product i substitutes for product j demand for only higher inventories of product i.

- If products can be ordered with respect to their unit holding and backordering costs, then switching curve would be skewed towards higher unit holding cost product. If products are symmetrical, switching curve would line on the 45$\circ$ line.

- For a multiserver system, similar analysis can be made, we expect similar results with single server system: If the products have hierarchy, the policy tends to act as a priority system, and if the products are symmetric, the policy tends to act as LQ based policies.

3.3 Numerical Analysis

We analyze the impact of production scheduling and substitution decisions on certain performance measures. Optimal policy characterized in § 3.2 fully uses the state information when taking production and substitution decisions; in other words, the production and substitution are done dynamically. To quantify the benefit of state information, the optimal policy is contrasted with those that partially use the state information, namely static production or substitution policies. We define three static policies depending on whether production or substitution uses state information. Models considered can be found in Table 3.2

Table 3.2: Settings Considered

Setting	Prod Dec.	Subs. Dec.
(i)	dynamic	static
(ii)	static	dynamic
(iii)	static	static
(iv)	dynamic	no
(v)	static	no
(opt)	dynamic	dynamic

Under static production, production is assumed to be done under a “randomized" scheme, until the net inventory level of a product reaches its basestock level. Each product is managed independently without taking into account the net inventory level of the other product. Thus, the state information is only partially used. Under randomized production the production capacity is allocated in proportion to the arrival rate of the demand for the products. The production rates are independent of the state (unless a basestock level is hit). Note that if substitution does not exist, randomized production closely approximates a system where each product is managed as an independent $M/M/1$ queue. The only slight difference is that, when a product is at its base-stock level, we assume the capacity is fully dedicated to the other product.

Under static substitution policy, we assume that substitution takes place when the inventory level of the substituting product is higher than a constant level, and substituted product has a negative inventory level. So substitution decision is given only based on a static level, regardless of the inventory level of the product to-be-substituted. Under static stocking and static substitution, decision variables are basestock levels (S_1, S_2) and the substitution levels, (K_1, K_2). The values are determined by exhaustive search over state space. If either production or substitution decisions are given dynamically, then the dynamic policy is obtained optimally for given static levels.

We present the findings of the numerical analysis in the following. We assume the products are identical in terms of cost values and production rates, as follows: $\mu = 1$, $\mu_s = 10$, substitution cost is $c = 15$, holding costs $h_1 = h_2 = 3$, and backordering costs $b_1 = b_2 \in \{35, 80, 125\}$. For the identical product setting arrival rates are $\lambda_1 = \lambda_2 \in \{0.05, 0.1, 0.15, \cdots, 0.4\}$, for the non-identical setting arrival rates are $\lambda_1, \lambda_2 \in \{0.05, 0.1, \cdots, 0.4\}$ where $\lambda_1 + \lambda_2 = 0.8$.

Observation 3.3.1. *When products are identical, the benefit of substitution under static production is higher than the benefit of substitution under dynamic production, especially under high traffic.*

We analyze the benefit of substitution under dynamic and static optimal production policies. For this purpose, the following performance measures are defined as in Table 3.3.

Table 3.3: Performance measures considered

Performance Measure	Expression	Explanation
% Savings due to substitution under dynamic production	$\frac{Cost_{iv} - Cost_{opt}}{Cost_{iv}}$	Dynamic substitution compared with no substitution under dynamic production
%Savings due to substitution under static production	$\frac{Cost_{v} - Cost_{ii}}{Cost_{v}}$	Dynamic substitution compared with no substitution under static production

Figure 3.2 shows that under dynamic production policy as traffic intensity in-

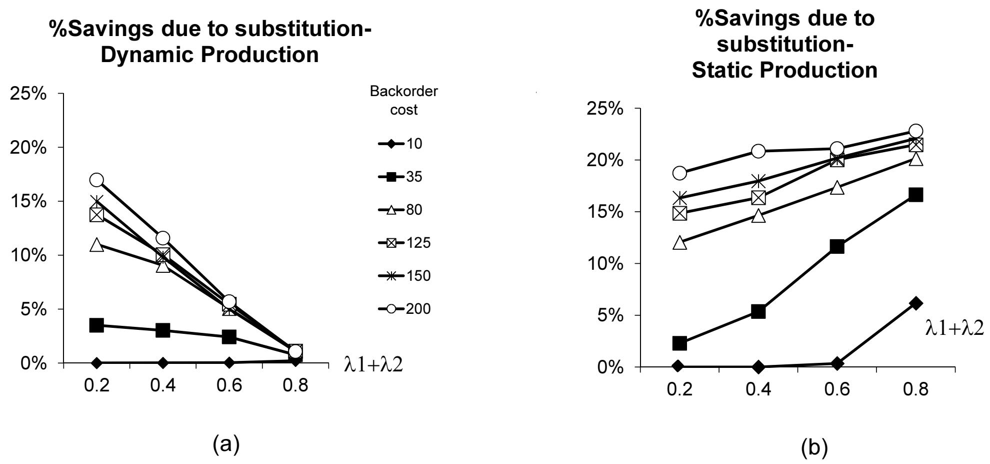


Figure 3.2: Savings due to (dynamic) substitution under production schemes.($\mu_s = 10, C_s = 15$, $\lambda_1 = \lambda_2$, $h = \{1, 3, 5\}$ (results averaged over h).

creases, the benefit of substitution decreases. On the other hand, under static production, substitution benefit increases with the traffic intensity. This implies under high traffic intensity effective usage of the resource gains importance and dynamic scheduling manages effective usage successfully. Substitution further improves the matching of the resource with the need over time. The results show that dynamic scheduling already performs the matching and the need for substitution is minimal. On the other hand, as traffic intensity increases, the performance of static production deteriorates, and substitution makes up for the ineffective allocation of the resources over the demand of the products.

We observe that benefit of substitution increases with an increase in backorder cost. However, for high traffic intensity this may not necessarily be the case under dynamic production. In other words, a change in backorder level does not impact the benefit of substitution when traffic intensity is high and production scheme is dynamic. This implies, dynamic production can keep the level of backorder under control whereas static production is not successful at controlling the backorder level, and needs the support of substitution.

Observation 3.3.2. *As arrival rates become asymmetric, the benefit of substitution under static production decreases whereas under dynamic production the benefit increases.*

First of all, it is observed that under both dynamic and static production, as products become more asymmetric (in terms of arrival rates) the optimized cost decreases. This can be attributed to the increase in the capacity pooling affect. Furthermore, the benefit of substitution under any setting increases with the unit backorder cost, since high backorder cost leads to higher flow due to substitution between the products.

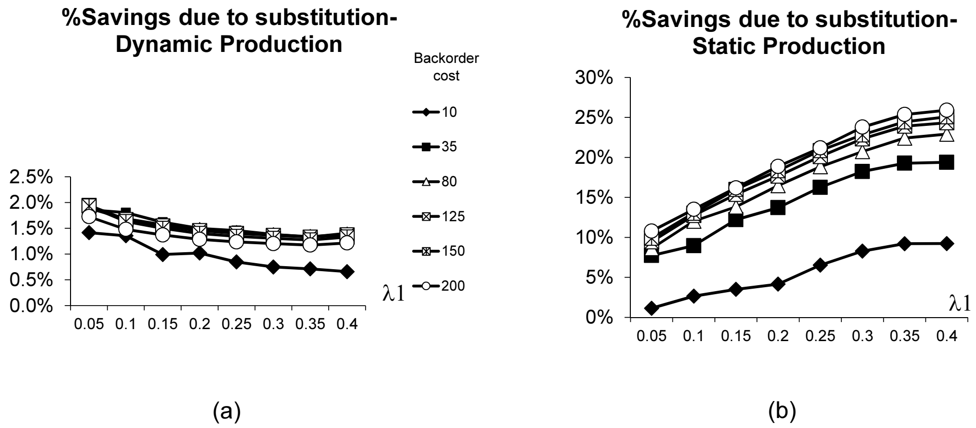


Figure 3.3: Savings due to (dynamic) substitution under production schemes when arrival rates are asymmetric.($\mu_s = 10, C_s = 15$, $\lambda_1 + \lambda_2 = 0.8$, $h = 3$.)

Under dynamic production, as products become more asymmetric, the gain due to substitution increases. When one product's arrival rate is low, the policy tends to lower down the basestock level for that product, and meets the demand from inventory of the other product through substitution. As arrival rates become further apart flow due to substitution increases, and so is the benefit of substitution. As shown in Theorem 3.2.1, substitution and production scheduling decisions are given under similar dynamics. Thus, the settings that favor dynamic scheduling (i.e., asymmetric arrival rates) also favor the substitution. As a result, we observe an increase in benefit of substitution whenever the cost under dynamic scheduling decreases (see Figure 3.3(a)). In scalar terms, we observe that as products become similar, the cost reduction due to substitution decreases.

In the case of static production (randomized policy), we observe a reverse situation. The benefit of substitution decreases as the products become more dissimilar. Under unequal arrival rates, both the basestock levels and the capacity allocations are asymmetric under the randomized policy. Thus, the pooling of resources is already achieved due to the asymmetric nature of the arrivals. Actually, under the most asymmetric arrival rates the cost values under dynamic and static production are close to each other. Thus the percent benefit of substitution is also at the same level under the most asymmetric setting. As arrival rates are closer to each other, performance of randomized policy deteriorates and dynamic substitution contributes significantly to pooling of resources, and to reducing cost (see Figure 3.3(b)). Under dynamic production, when products are similar the benefit of substitution is limited since dynamic scheduling achieves resource pooling to a considerable extend.

Observation 3.3.3. *The benefit of dynamic production scheduling is highest under no substitution. When substitution takes place, the benefit of dynamic scheduling is relatively high under moderate traffic intensities.*

To measure the benefit of dynamic production, three measures are analyzed. (Table 3.4)

Table 3.4: Performance measures considered

Performance Measure	Expression	Explanation
% Savings due to dynamic production under no substitution	$\frac{Cost_v - Cost_{iv}}{Cost_v}$	Dynamic production compared with static production under no substitution
% Savings due to dynamic production under static substitution	$\frac{Cost_{iii} - Cost_{iv}}{Cost_{iii}}$	Dynamic production compared with static production under static substitution
% Savings due to dynamic production under dynamic substitution	$\frac{Cost_{ii} - Cost_{opt}}{Cost_{ii}}$	Dynamic production compared with static production under dynamic substitution

As one might expect, the benefit of dynamic production (over static production) is highest under no substitution, and increases with the traffic intensity. Actually, as discussed in Observation 3.3.1, dynamic production is more powerful under high traffic intensity. It is further observed that regardless of substitution cost and rate, the benefit of dynamic production is the same under dynamic or static substitution, ie. the benefit of dynamic production does not differ among substitution types. (Figure 3.4).

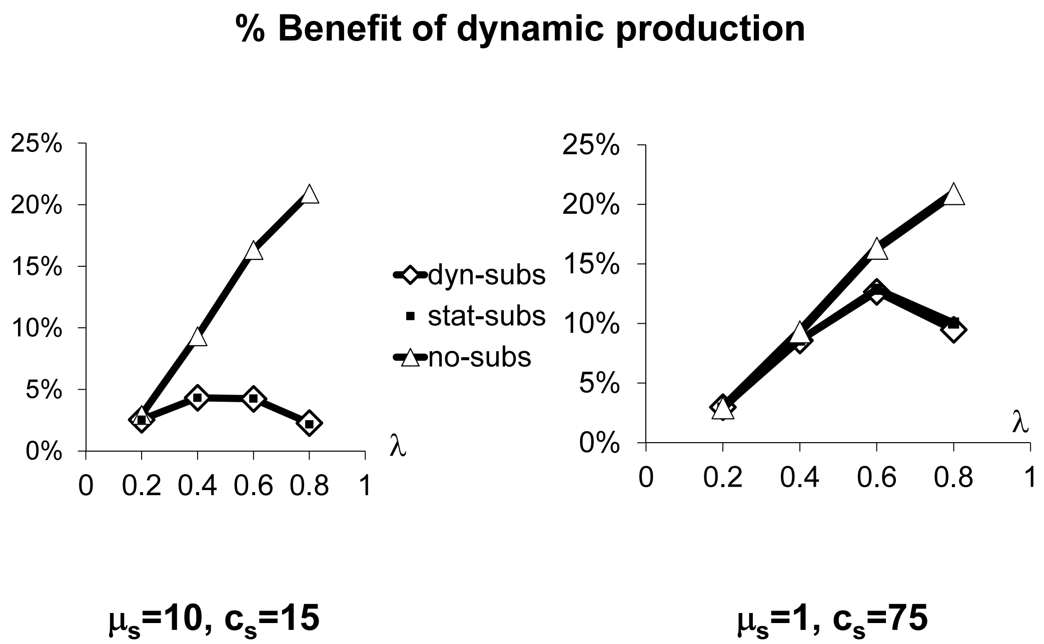


Figure 3.4: Savings due to (dynamic) production under substitution schemes.

In Observation 3.3.1, it is stated that benefit of substitution over dynamic production is limited under high traffic intensity. We observe the complement of this observation for dynamic production. Once dynamic substitution (or any form of substitution for that matter) exists, this limits the contribution of dynamic production. Under low arrival rates, substitution do not take place, and benefit of dynamic production is limited. Under high arrival rates, through substitution a substantial resource pooling can be achieved, and the additional cost reduction due to dynamic production is limited. For the moderate traffic intensity, dynamic production provides relatively the highest savings (Figure 3.4(a)).

Summary of Findings

The benefit of dynamic substitutions over no substitution is highest for:

- For dynamic production when arrival rate for both products is equal, high backorder cost, low total arrival rate. Maximum benefit over all instances is %18.

- For static production when arrival rate for both products is equal, high backorder, high total arrival rate. Maximum benefit over all instances is %25.

- For dynamic production, when total arrival rate stays the same, high backorder cost, arrival rates are more asymmetric($|\lambda_1 - \lambda_2|$ is highest). Maximum benefit over all instances is %2.

- For static production, when total arrival rate stays the same, high backorder cost, arrival rates are more symmetric($|\lambda_1 - \lambda_2|$ is lowest). Maximum benefit over all instances is %28

The benefit of dynamic production over static production is highest for:

- For no substitution case when arrival rate for both products is equal, high total arrival rate. Maximum benefit over all instances is %24.

- For static and dynamic substitution when arrival rate for both products is equal, medium total arrival rate. For bothstatic and dynamic substitution, maximum benefit over all instances is %6 when substitutions are easy($\mu_s = 10, c_s = 15$) and maximum benefit over all instances is %14 when substitutions are hard($\mu_s = 1, c_s = 75$).

3.4 Extensions

We extend the problem to the two echelon system, where a Work-in-Progress(WIP) stock is kept at the upper echelon. This WIP is converted into the final product through another stage of the production process. We study the production, conversion and substitution policies in a continuous time Markovian setting. See Appendix B

CHAPTER 4

POLICIES FOR THE SINGLE-ECHELON N-PRODUCT SYSTEM

In previous section, the structural results and computational analysis for two-product single-echelon production-substitution problem is presented. For the multiple product system, computing the optimal policy is not possible in reasonable time, due to the scale of the problem. The problem under consideration is as follows: A multi-product make-to-stock queue where stock of each product is held separately. Products share a common production resource. Demand for each product occur independently following a Poisson demand with rate λ_i. Production facility is modeled by a single server in which the production times are exponentially distributed with rate μ. If demand for a product cannot be met from stock, it is backordered. The inventory holding cost is h_i and backorder cost is b_i for product i. Depending on the net stock level, a product may be substituted with another product. It is assumed that, if product j is to be substituted with product i, then substitution takes a time which is an exponential random variable with parameter μ_{ij}. A cost c_{ij} is incurred whenever the substitution process is completed. It is assumed that substitutions are handled separately, and there is a dedicated infinite resource for substitution operations. The objective is to decide on the stock level for each product, the production schedule and the substitution policy that will minimize the average cost per unit time.

In this section, we devise heuristic policies for the multi-product setting. After introducing and defining the heuristics, through numerical analysis we identify the conditions under which each heuristic performs well. A majority of the heuristics are derived based on longest-queue-based (LQ-based) idling policy. LQ-based idling policy is a production idling policy proposed by Zipkin (1995) for symmetric products in the context of two-item make-to-stock queue. The policy is adopted by many authors when deriving heuristic policies such as Veatch and Wein (1996).

The following heuristics are proposed:

1. Priority-queue based dynamic heuristic (Prio-DH)
2. Service time look-ahead with longest-queue based idling policy (LQ-STLA)
3. Service and substitution time look-ahead with longest-queue based idling policy (LQ-SSTLA)
4. Busy cycle time look-ahead with longest-queue based idling policy (LQ-BTLA)
5. Substitutions based on inventory level (SI-LEVEL), substitutions based on cost of substitution (SI-COST)

In the following, the heuristics are explained in detail. At times numerical results are provided to improve the understanding. Finally, two benchmark policies that provide a lower bound and an upper bound are defined. To better assess the performance of the heuristics, the cost values obtained under each are compared with the lower bound and the upper bound. The benchmark policies considered are:

1. First-come first-served (FCFS) production policy with no substitutions (upper bound)
2. Pooled inventory with rationing under costless and immediate substitutions (lower bound)

These bounds are not necessarily tight, however they help us to understand the structure of the optimal policy. If for instance a heuristic performs close to either the lower or the upper bound, then we can infer the operating characteristics of the policy.

Chapters 4 and 5 are organized as follows. In this chapter the Prio-DH policy is described. The remaining heuristics are described in Chapter 5.

4.1 Description of the Prio-DH policy

This heuristic is derived from a *preemptive priority make-to-stock queue (MTS queue)*. A preemptive priority-queue is defined as a queue with a single server serving multiple classes of customers. The classes are ordered with respect to their priorities. Suppose there are n customer classes and 1 denotes the class with the highest priority, i.e., $1 \prec 2 \ldots \prec n$. If the server is currently serving a customer from class k, then this means there does not exist any customer from classes $1 \ldots, k-1$ in the queue. If a customer from class $1, \ldots, k-1$ arrives, the server stops service and starts serving the newly arrived customer. Server does not start serving customers from class k until he finishes all customers with higher priorities. In the remainder of the text, the term *preemptive* is not explicitly used.

A priority MTS queue is a priority queue with a basestock level defined for each item. Stock of each item is held separately, but items are produced by a single server. The shortfall queue for each item (shortfall is the difference between the basestock level and the net inventory level) is managed as in a priority-queue. This implies if there are two or more products with net inventory levels below their basestock levels, the production is scheduled according to a priority order. Note that a “priority-queue” can be thought of as a “priority make-to-stock queue” with all basestock levels equal to 0. In the remainder of the text, it is assumed without loss of generality that product priorities are defined by product indices. (Optimal) Basestock level of item k is denoted with S_k. The production policy under priority MTS queue can be classified under the set of production policies that partially uses the state information of the system. When deciding on which item to produce next, the only information necessary is on whether the product is below its basestock level and whether the products that have higher priorities are below their basestock level.

Let's now look at the policy implied by the priority make-to-stock queue. The production policy is within the class of a "hedging point" production policies: there exists a hedging point (the set of basestock levels) at which the production stops. The production schedule follows a priority scheme. Substitution policy is not to substitute any product. The proposed heuristic policy, Prio-DH, takes the policy implied by priority MTS queue as an initial policy, and applies one-step of policy improvement algorithm. The resulting policy keeps the basestock levels as they are, while improves on the production scheduling and the substitution policies. Given λ_k, h_k and b_k values for each item, priority MTS-queue will result in an expected cost per unit time. Improving upon the priority MTS-queue will lower down this initial expected cost. We conjecture that the lower the initial cost, the lower the cost after improvement. Thus, the aim is first to obtain an optimal policy within the class of priority MTS-queue policies. Then, the objective is to find the optimal basestock levels, and the optimal priority order given the parameters λ_k, h_k and b_k for all items. After the priority order and basestock levels are determined, Prio-DH will be obtained via the one-step policy improvement algorithm.

In the following, Prio-DH is described in details through the steps below:

1. **Determining the production idling policy for a given priority-order (Section 4.2)**

   For the priority MTS-queue, given a priority order we determine the optimal basestock levels. These basestock levels are also the hedging points or the idling points for Prio-DH. In this part, first the steady-state distribution of shortfall is analyzed. The steady-state short-fall distribution has a special characteristic: For priority policy, for any given basestock level, the shortfall from basestock level has the same steady-state distribution, independent of the basestock level. The optimal basestock levels are found by marginal analysis.

2. **Obtaining the production scheduling and substitution policy by**

policy improvement (Section 4.3)

We apply one-step of policy improvement to the policy implied by the priority MTS-queue. In order to do this, we need the bias differences of the states. Normally, this is done by policy evaluation for the given policy. However, due to the curse of dimensionality, evaluating a given policy and obtaining the bias differences is not possible. We use an alternative approach. Bias difference is the asymptotic difference in expected total costs due to starting in different states, which can be defined for any two states. Utilizing this definition, we are able to calculate the bias differences by sample path analysis. After bias differences are calculated, one step of policy iteration is applied, and Prio-DH is obtained.

3. **Numerical results on the performance of the policy improvement step (Section 4.4)**

   An interim numerical analysis is made to compare the performances of Prio-DH and the optimal policy. We analyze how different substitution parameters affect the performance of the heuristic. Finally, we present the performance of the heuristic under various priority orders to see the effect of priority order on the performance of Prio-DH.

4. **Determining the priority order (Section 4.5)**

   Different priority orders result in different Prio-DH policies. As priority order changes so will the performance of the heuristic, so we need to determine a "good" starting priority order. One might try to find the best priority order, i.e. the order that gives the lowest cost for the priority order. However, finding the best priority order is a combinatorial problem. We devise heuristics for determining a "good" starting priority order.

5. **Numerical results on the performance of the priority order (Section 4.6)**

   We compare the cost under the priority order heuristics with the cost under the optimal priority order (that is giving least cost for the priority policy). We also numerically analyze the effect of priority order on the performance of Prio-DH.

4.2 Determining the optimal basestock levels for the priority MTS queue

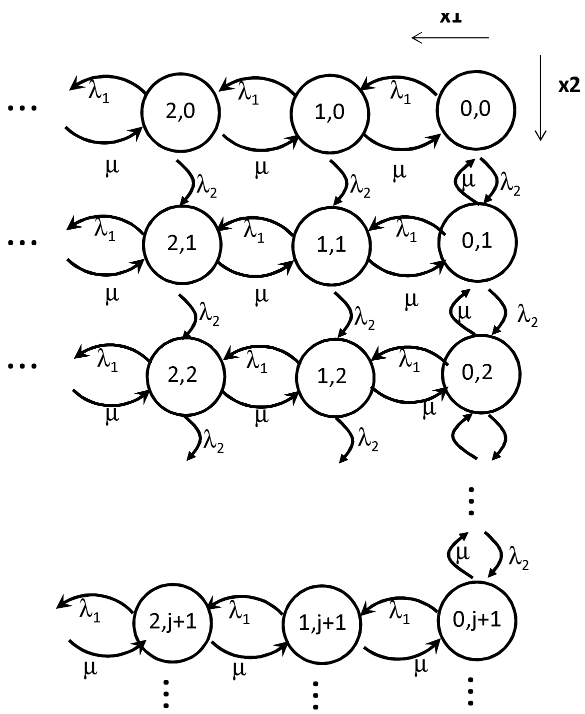


Figure 4.1: Transition diagram for a two-product priority MTS queue (x_i denotes the shortfall level for product i)

In this part the aim is to determine the optimal idling policy for the priority MTS queue, given λ_k, h_k and b_k values for product k, $k = 1, 2, \cdots, n$. Service times are exponential with parameter μ. In this part, we assume the priority order is already given. Without loss of generality suppose product indices define the order: $1 \prec 2 \ldots \prec n$. The steady-state distribution of shortfall from the basestock level is independent of the base stock level [1]. So when deriving the steady-state probabilities, it is natural to work with shortfall queue instead of inventory levels. For this purpose, system state is described by an n-dimensional vector $\mathbf{X(t)} = (X_1(t), X_2(t), ..., X_n(t))$. Here, $X_k(t)$ is the number of outstanding jobs, or the shortfall level, for product k at time t. Due to Markovian property the time dependence is ignored, and state is defined as $x = (x_1, x_2, ..., x_n)$. For each k, x_k takes values in Z^+. Here, note that the state of the product with

[1] Other policies where steady-state probabilities are independent of the basestock levels are, for instance, FCFS production policy or the longest queue production policy.

highest priority is denoted with x_1.

Let steady-state joint probability of shortfall, $(x_1, x_2, \ldots, x_n)$for product k will be $x_k,\ k = 1, 2, \ldots, n$ be denoted with $p_{x1,x2,\ldots,xn}$. Given S_k denotes the optimal basestock level, let p_i^k denote the marginal probability of having a shortfall level of i for item k in steady-state. Note p_i^k is defined as follows:

$$p_i^k = \sum_{x_1=0}^{\infty} \cdots \sum_{x_{k-1}=0}^{\infty} \sum_{x_{k+1}=0}^{\infty} \cdots \sum_{x_n=0}^{\infty} p_{x_1,\ldots,x_k=i,\ldots,x_n}.$$

Given the marginal probabilities, the cost incurred for each product can be calculated independent of other products. This implies, the basestock level for each product can be calculated independent of other products. In the following, in Lemma 4.2.1 a recursive expression is derived for the marginal probabilities of each product. In Proposition 4.2.1 optimal basestock level for each product is obtained. In Proposition 4.2.2, it is shown that for any two identical products, the product with higher priority operates under a lower optimal basestock level. We make the analysis in two subsections.

4.2.1 Marginal probability distribution for product k

In the sequel, first a priority policy with two products is analyzed. The transition rate diagram of a priority MTS queue for two-products is presented in Figure 4.1. Then analysis is generalized to n products. Let $\rho_k = \frac{\lambda_k}{\mu}$, and let $p_{i,j}$ denote the steady-state joint probability of state (i, j) in the two-product priority MTS queue, where i and j denote the shortfall levels for high and low priority products, respectively.

Lemma 4.2.1. *In priority MTS queue with two-products, the marginal probabilities for shortfall level of product 1 are expressed as*

$$p_i^1 = (\frac{\lambda_1}{\mu})^i(1 - \frac{\lambda_1}{\mu}), \quad i \geq 0, \tag{4.2.1}$$

and the marginal probabilities for shortfall level of product 2 are expressed as:

$$p_j^2 = \frac{1}{\rho_2} p_{0,(j+1)},$$

where

$$p_{0,j} = \rho_1 \sum_{l=0}^{j-1} (1 - \sum_{t=0}^{j-l-1} Pr_{0,t}^1) p_{0,l} + \rho_2 p_{0,j-1}, \; j > 1,$$

$$q_{0,j} = \frac{\lambda_1}{\Lambda} \sum_{t=0}^{j} q_{0,t} Pr_{0,j-t}^1 + \frac{\lambda_2}{\Lambda} q_{0,j-1} \qquad j > 0,$$

$$q_{0,0} = \frac{\Lambda - \sqrt{(\Lambda)^2 - 4\lambda_1 \mu}}{2\lambda_1},$$

and $p_{0,0} = (1 - \frac{\lambda_1+\lambda_2}{\mu})$*. In the equations,* $\Lambda = \lambda_1 + \lambda_2 + \mu$*.*

Proof. Note that the joint probability $p_{0,(i+1)}$ denotes the probability that in the two-product queue, the high priority product has a shortfall level of 0, and low priority product has a shortfall level of $i + 1$.

Derivation of joint and marginal probabilities in priority-queues are well-studied in the literature, see for instance Rozenshmidt (2008), Cidon and Sidi (1990) and the references therein. In the proof, an alternative derivation for the joint and marginal probabilities are provided. We use this alternative approach in Proposition 4.2.2. To obtain p_i^2, the marginal probability distribution for product 2, the following four steps are followed. First note that the following flow balance equations hold:

$$\begin{aligned}
(\lambda_1 + \lambda_2 + \mu) p_{i,j} &= \lambda_1 p_{(i-1),j} + \mu p_{(i+1),j} + \lambda_2 p_{i,(j-1)}, \qquad i > 0, j > 0 \\
(\lambda_1 + \lambda_2 + \mu) p_{0,j} &= \mu p_{0,(j+1)} + \lambda_2 p_{0,(j-1)}, \qquad j > 0 \\
(\lambda_1 + \lambda_2) p_{0,0} &= \mu p_{1,0} + \mu p_{0,1}
\end{aligned} \tag{4.2.2}$$

Step 1 If the $p_{i,j}$'s are added along diagonals (where $i + j = m$, for $i, j, m = 0, 1, 2, \cdots$, see Figure 4.1), the following equations are obtained:

$$(\lambda_1 + \lambda_2 + \mu)p_m = (\lambda_1 + \lambda_2)p_{m-1} + \mu p_{m+1}, \quad m = 1, 2, \cdots$$
$$(\lambda_1 + \lambda_2)p_0 = \mu p_1$$

$$(\lambda_1 + \lambda_2 + \mu)p_m = (\lambda_1 + \lambda_2)p_{m-1} + \mu p_{m+1}, \ m = 1, 2, \cdots$$

where $p_m = \sum_{i,j:i+j=m} p_{i,j}$. These are the well-known balance equations for $M/M/1$ queue with arrival rate $\lambda_1 + \lambda_2$ and service rate μ. Then p_m is obtained as:

$$p_m = (\frac{\lambda_1 + \lambda_2}{\mu})^m(1 - \frac{\lambda_1 + \lambda_2}{\mu}), \tag{4.2.3}$$

and therefore $p_0 = p_{0,0} = (1 - \frac{\lambda_1+\lambda_2}{\mu})$.

Step 2 Suppose the joint probabilities are added along the vertical lines (see Figure 4.1). Note $p^1_i = \sum_{j=0}^{\infty} p_{i,j}$, $i \in \mathbb{Z}^+$. Then,

$$\lambda_1 p^1_0 = \mu p^1_1, \quad i = 0,$$
$$(\lambda_1 + \mu)p^1_i = (\lambda_1)p^1_{(i-1)} + \mu p^1_{(i+1)}, \quad i > 0.$$

Again, these are the balance equations for $M/M/1$ queue. So,

$$p^1_i = (\frac{\lambda_1}{\mu})^i(1 - \frac{\lambda_1}{\mu}), \quad i \geq 0. \tag{4.2.4}$$

Step 3 For the low priority product, note $p^2_j = \sum_{i=0}^{\infty} p_{i,j}, j \in \mathbb{Z}^+$. If a horizontal cut is made between states $(\cdot, j)$ and $(\cdot, j + 1)$ in the flow diagram (see Figure 4.1), one observes that the following flow balance holds:

$$\lambda_2 p^2_j = \mu p_{0,j+1}, j > 0. \tag{4.2.5}$$

Step 4 In this step $p_{0,j}$ values are determined.

To determine $p_{0,j}$, first define C_i (Column i), as the set of states where the shortfall for product 1 is i,

$$C_i = \{(x_1, x_2) : x_1 = i, \ x_2 \in \mathbb{Z}^+\}, i = 0, 1, \ldots.$$

To determine $p_{0,j}$, the flow rates among the states $(x_1, x_2) \in C_0$ are analyzed. At any point in time, there is a flow-out from state $(0, j_1)$ with rate $(\lambda_1 + \lambda_2 + \mu)p_{0,(j+1)}$. Of the total flow-out rate, $\lambda_2 p_{0,(j+1)}$ is flowing into $(0, j_1 + 1)$, and $\mu p_{0,(j+1)}$ is flowing into $(0, j_1 - 1)$. These flows are within C_0. The remaining rate, $\lambda_1 p_{0,(j+1)}$ flows out of C_0. Since in steady-state, the probability that the process is in C_0 should not change, it should hold that due to state $(0, j_1)$, $\lambda_1 p_{0,j_1}$ must flow-into C_0 per unit time.

Define probability q_{j_1,j_2} as the probability that the process leaving C_0 from state $(0, j_1)$ enters into C_0 at state $(0, j_2)$. Thus, considering the states in C_0, flow rate from $(0, j_1)$ to $(0, j_2)$ is $\lambda_1 q_{j_1,j_2} p_{0,j_1}$. Denote $\lambda_1 q_{j_1,j_2}$ with λ_{j_1,j_2}. Then considering the states in C_0, the rate diagram can be obtained as in Figure 4.2(a).

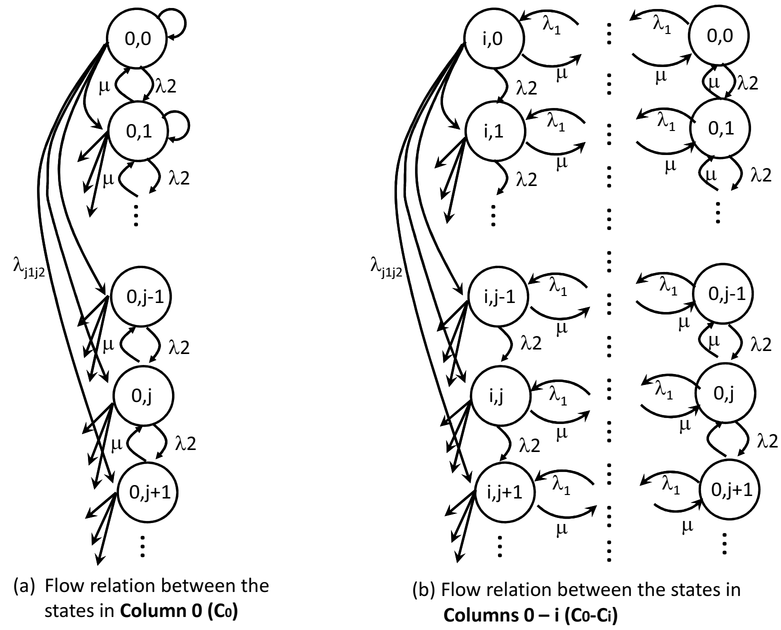


Figure 4.2: Flow rate relationship

Note that this flow rate relationship within the states in C_0 can be extended to any Column i (C_i). For example, for C_{i_1}, $\lambda_1 q_{j_1,j_2} p_{i_1,j_1}$ is the flow rate from state (i_1, j_1) into (i_1, j_2) given that the process leaves Column i_1 towards higher i (i,e., leaves C_{i_1} from the left, see Figure 4.2(b)).

In the following, we calculate the probabilities q_{j_1,j_2}. First define the probability $Pr^{\Delta}_{j1,j2}$, $\Delta = 1, 2, \ldots$ as the probability that the chain hits C_0 first time at state $(0, j_2)$, given that current state is $(\Delta, j_1), \Delta = 1, 2, \ldots$. Note that for $j_2 < j_1$, $Pr^{\Delta}_{j1,j2} = 0$.

We determine $Pr^1_{j_1,j_2}$, or equivalently, q_{j_1,j_2}. Start with $Pr^1_{0,0}$, which is the probability that given that the chain is in state $(1, 0)$ the first visit to C_0 is at $(0, 0)$. Let $\lambda_1 + \lambda_2 + \mu$ be denoted with Λ. Then,

$$Pr^1_{0,0} = \frac{\lambda_1}{\Lambda} Pr^2_{0,0} + \frac{\mu}{\Lambda}(1) + \frac{\lambda_2}{\Lambda}(0)$$

Note $Pr^2_{0,0} = (Pr^1_{0,0})^2$. The equation is second order and $Pr^1_{0,0}$ is obtained as:

$$Pr^1_{0,0} = \frac{\Lambda - \sqrt{(\Lambda)^2 - 4\lambda_1\mu}}{2\lambda_1}.$$

For the remaining states in C_0,

$$Pr^1_{0,j} \quad = \quad \frac{\lambda_1}{\Lambda} Pr^2_{0,j} + \frac{\lambda_2}{\Lambda} Pr^1_{1,j}, \qquad j > 0$$

Note that due to the repeating structure of the chain, $Pr^1_{j_1,j_2} = Pr^1_{j_1+u,j_2+u}$, $u = -j_1, \ldots, -1, 0, 1, 2, \ldots$. This implies for instance, for $j = 1, 2, \ldots$, $Pr^1_{j,j} = Pr^1_{0,0}$. Furthermore, it holds that $Pr^2_{0,j} = \sum_{t=0}^{j} Pr^1_{0,t} Pr^1_{t,j}$. Observing these, the following expression can be obtained for $Pr^1_{0,j}$:

$$Pr^1_{0,j} \quad = \quad \frac{\lambda_1}{\Lambda} \sum_{t=0}^{j} Pr^1_{0,t} Pr^1_{t,j} + \frac{\lambda_2}{\Lambda} Pr^1_{1,j}, \qquad j > 0$$

Observing the equivalences $Pr^1_{t,j} = Pr^1_{0,j-t}$ and $Pr^1_{1,j} = Pr^1_{0,j-1}$,

$$Pr^1_{0,j} = \frac{\lambda_1}{\Lambda} \sum_{t=0}^{j} Pr^1_{0,t} Pr^1_{0,j-t} + \frac{\lambda_2}{\Lambda} Pr^1_{0,j-1}, \qquad j > 0 \qquad (4.2.6)$$

It is possible to obtain $Pr^1_{0,j}$ values recursively. Given $\lambda_1 + \lambda_2 < 1$, the chain is recurrent which implies, if the process starts at state (1,j), it has to visit Column C_0. Thus $\sum_{j=0}^{\infty} Pr^1_{0,j} = 1$.

Note that for $j_2 \geq j_1$, $\lambda_{j_1,j_2} = \lambda_1 Pr^1_{0,j_2-j_1}$. In Figure 4.2(a), consider the cut between states $(0, j-1)$ and $(0, j)$. Local flow balance implies:

$$
\begin{aligned}
\mu p_{0,j} &= \sum_{l=0}^{j-1} p_{0,l} \sum_{t=0}^{\infty} \lambda_{l,j+t} + \lambda_2 p_{0,j-1} \\
&= \lambda_1 \sum_{l=0}^{j-1} p_{0,l} \sum_{t=0}^{\infty} q_{l,j+t} + \lambda_2 p_{0,j-1} \\
&= \lambda_1 \sum_{l=0}^{j-1} p_{0,l} \sum_{t=0}^{\infty} \lambda_1 Pr^1_{0,j+t-l} + \lambda_2 p_{0,j-1} \\
&= \lambda_1 \sum_{l=0}^{j-1} p_{0,l} (1 - \sum_{t=0}^{j-l-1} Pr^1_{0,t}) + \lambda_2 p_{0,j-1}, \quad j > 0.
\end{aligned}
$$

using $p_{0,0}$, $p_{0,j}$ can be obtained recursively. Note that $\sum_{j=0}^{\infty} p_{0,j} < 1$.Plugging in $q_{i,j}$ in place of $Pr^1_{i,j}$, we obtain:

$$
\mu p_{0,j} = \lambda_1 \sum_{l=0}^{j-1} p_{0,l} (1 - \sum_{t=0}^{j-l-1} q_{0,t}) + \lambda_2 p_{0,j-1}, \quad j > 0.
$$

Since from Eq. 4.2.5 $p^2_j = \frac{\mu}{\lambda_2} p_{0,j+1}$, the four steps complete the proof.

□

In the following, the analysis is extended to the n-product priority MTS queue. Suppose we are interested in the steady-state marginal probability of shortfall level for product k. Given p^k_i denotes the marginal probability that shortfall level of product k is i, let $\bar{\lambda}_1 = \sum_{\ell<k} \lambda_\ell$, and $\bar{\lambda}_2 = \lambda_k$. Also let $\bar{\rho}_1 = \frac{\bar{\lambda}_1}{\mu}$ and $\bar{\rho}_2 = \frac{\bar{\lambda}_2}{\mu}$. Finally, let $\bar{\Lambda} = \bar{\lambda}_1 + \bar{\lambda}_2 + \ldots + \bar{\lambda}_k + \mu$. Then, following Lemma 4.2.1 it is possible to obtain p^k_j as stated in the corollary below.

Corollary 4.2.1. *In priority MTS queue with n-products, the marginal probabilities for shortfall level of product k are expressed as:*

$$
p^k_j = \frac{1}{\bar{\rho}_2} r_{0,j+1},
$$

where

$$
r_j = \bar{\rho}_1 \sum_{l=0}^{j-1} r_{0,l} (1 - \sum_{t=0}^{j-l-1} \bar{q}_{0,t}) + \bar{\rho}_2 r_{0,j-1}, \; j > 1,
$$

$$
\bar{q}_{0,j} = \frac{\bar{\lambda}_1}{\bar{\Lambda}} \sum_{t=0}^{j} \bar{q}_{0,t} \bar{q}_{0,j-t} + \frac{\bar{\lambda}_2}{\bar{\Lambda}} \bar{q}_{0,j-1} \qquad j > 0,
$$

$$\bar{q}_{0,0} = \frac{\bar{\Lambda} - \sqrt{(\bar{\Lambda})^2 - 4\bar{\lambda}_1\mu}}{2\bar{\lambda}_1},$$

and $r_{0,0} = (1 - \frac{\bar{\lambda}_1+\bar{\lambda}_2}{\mu})$, *and* $\bar{\Lambda} = \sum_i \bar{\lambda}_i + \mu$.

Proof. First note that in the corollary $r_{0,j},\ j = 0, 1, 2, \ldots$ equivalently represents the joint probability $p_{0,j},\ j = 0, 1, 2, \ldots$ for the two-product priority queue where first product has an arrival rate $\bar{\lambda}_1$ and the second product has an arrival rate $\bar{\lambda}_2$. Note $r_{0,j}$ is expressed as,

$$r_{0,j} = \sum_{x_{k+1}} \cdots \sum_{x_n} p_{0,\ldots,0,x_k=j+1,x_{k+1},\ldots,x_n}.$$

Actually, the items that have lower priority than k does not impact the marginal probability distribution for item k and thus can be ignored. For the simplicity of the proof, in the remainder the items that have lower priorities than item k are ignored. The aim is now to obtain the marginal probability distribution of the lowest ranking item, k, in a k-product setting.

Note that the steady-state joint probability of state (i, j), in a two-product MTS queue with rates $\bar{\lambda}_1$ and $\bar{\lambda}_2$, is equal to the sum of joint probabilities in the $k-$product priority queue. Thus:

$$r_{i,j} = \sum_{x_1} \sum_{x_2} \cdots \sum_{\{x_{k-1}:\sum_l x_l = i\}} p_{x_1,x_2,\ldots,x_{k-1},j}.$$

For product k, first, the steady-state probability flow rates among the states $(0, \ldots, 0, i)$ and $(0, \ldots, 0, j)$ are obtained. Since the marginal probability distribution of the 2nd product in the two-product priority queue, will be equivalent to the marginal probability distribution of kth product in the k-product priority queue, the analysis follows exactly the same steps in Lemma 4.2.1. Then the following results hold for product k in a k-product priority queue:

$$p^k_j = \frac{1}{\bar{\rho}_2} r_{0,j+1},$$

where

$$r_{0,j} \;=\; \bar{\rho}_1 \sum_{l=0}^{j-1} r_{0,\ell}(1 - \sum_{t=0}^{j-l-1} \bar{q}_{0,t}) r_l + \bar{\rho}_2 r_{0,j-1},\ j > 1,$$

$$q_{0,j} = \frac{\bar{\lambda}_1}{\bar{\Lambda}} \sum_{t=0}^{j} \bar{q}_{0,t}\bar{q}_{0,j-t} + \frac{\bar{\lambda}_2}{\bar{\Lambda}} \bar{q}_{0,j-1} \qquad j > 0,$$

$$\bar{q}_{0,0} = \frac{\bar{\Lambda} - \sqrt{\bar{\Lambda}^2 - 4\bar{\lambda}_1\mu}}{2\bar{\lambda}_1},$$

and $r_{0,0} = (1 - \frac{\bar{\lambda}_1+\bar{\lambda}_2}{\mu})$,

where $\bar{\Lambda} = \sum_i \bar{\lambda}_i + \mu$.

Note that the same result holds for marginal probability distribution of product k in a n-product setting. □

4.2.2 Finding the optimal production idling policy

We observed that given the marginal probabilities, the cost incurred for each product can be calculated independent of other products. This implies, the basestock level for each product can be calculated independent of other products. Using the steady-state marginal probabilities, the total expected average cost incurred due to item k, TC_k, will be expressed below. In the two-product priority MTS-queue setting, we first determine the total expected average cost for item 1, and then for item 2. This is then generalized to n-product setting.

For product 1,

$$\begin{aligned} TC_1 &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}[(h_1)(S_1-i)^+ + (b_1)(S_1-i)^-]p_{i,j} \\ &= \sum_{i=0}^{\infty}[(h_1)(S_1-i)^+ + (b_1)(S_1-i)^-]\sum_{j=0}^{\infty}p_{i,j} \\ &= \sum_{i=0}^{\infty}[h_1(S_1-i)^+ + b_1(S_1-i)^-]\rho_1^i(1-\rho_1) \end{aligned}$$

where $\rho_1 = \frac{\lambda_1}{\mu}$.

For product 2,

$$\begin{aligned} TC_2 &= \sum_{j=0}^{\infty}[h_2(S_2-j)^+ + b_2(S_2-j)^-]p_j^2 \\ &= \sum_{j=0}^{\infty}[h_2(S_2-j)^+ + b_2(S_2-j)^-]\frac{1}{\rho_2}p_{0,(j+1)} \end{aligned}$$

where $\rho_2 = \frac{\lambda_2}{\mu}$.

In an n-product setting it is possible to express the expected total cost per unit time for product k by observing the analogy between the two-product and n-product priority queues. Thus, in the n-product setting, total expected average cost for item k is

$$TC_k = \sum_{j=0}^{\infty}[h_k(S_k - j)^+ + b_k(S_k - j)^-]p_j^k$$

where p_j^k is obtained as in Theorem 4.2.1.

We determine the optimal basestock levels, S_k, in Proposition 4.2.1. Before Proposition 4.2.1, introduce $P_j^k = \sum_{l=0}^{j} p_l^k$ as the cumulative probability that for product k shortfall level is lower than or equal to j. The basestock level for product k can be obtained considering the critical fractile implied by h_k and b_k.

Proposition 4.2.1. *The optimal basestock level, S_k, is defined as:*

$$S_k = \min_i \{P_i^k \geq \frac{b_k}{h_k + b_k}\}.$$

Proof. Proof simply follows from the property that $-TC_k$ is unimodal in S_k, for $k = 1, 2, .., n$. Basestock levels can be determined analogical to determining the order quantity in a newsvendor problem. □

Proposition 4.2.2. *In the two-product priority MTS-queue, an increase in λ_1 increases the optimal basestock level for both products.*

The proof consists of two parts. In the first part, we show an increase in λ_1 increases the optimal basestock level of product 1, i.e., S_1 increases with λ_1. In the second part, we show S_2 increases with λ_1. Intuitively, as λ_1 increases it is more likely that we observe higher amount of shortfall for both products, and thus optimal basestock levels increase as a response. Before the proof, we introduce the following lemma.

Lemma 4.2.2. *Consider an $M/M/1$ queue with arrival rate λ_1 and service rate μ. Let T denote the busy period. Consider a Poisson process with rate λ_2, $PP(\lambda_2)$. Let Z denote the number of arrivals from $PP(\lambda_2)$ during $(0, T]$. Then Z is stochastically increasing in λ_1.*

Proof. From Theorem 2 of Harris and Prabhu (1987), it holds that T is stochastically increasing in λ_1, and from Theorem 1.A.13 of Shaked and Shanthikumar (2007), it holds that Z is stochastically increasing in λ_1. □

Proof. (*(Proposition 4.2.2)*. **Part I. An increase in λ_1 increases S_1.** In the two product setting the marginal probabilities for the steady-state shortfall level of product 1 are equal to those under $M/M/1$ queue. For $t \geq 0$, the cumulative distribution function for steady-state product 1 shortfall level is $\sum_{\ell=0}^{j} p_\ell^1$. And the complement is $\sum_{\ell=j+1}^{\infty} p_k^1$. Since for all j, $\sum_{\ell=j+1}^{\infty} p_\ell^1$ increases with λ_1, optimal basestock is non-decreasing in λ_1. Note that S_1 is expressed as

$$S_1 = \min\{s : s \geq \frac{\ln \frac{h}{h+\ell}}{\ln \rho_1} - 1\},$$

which shows S_1 is non-decreasing in λ_1.

Part II. An increase in λ_1 increases S_2. The second part of the proof consists of the following steps. We first show that, given that product 1 is at zero shortfall level, the limiting shortfall level for product 2 (a conditional random variable) gets stochastically larger as λ_1 increases. Then, we show that this results in an increase in optimal basestock level S_2.

Specifically, the steps of the proof are as follows:

Step 1 For the Markov chain in Figure 4.2(a), we show that for a given state $(0, j_1)$, and for $t \geq j_1$, $j_1 = 0, 1, \ldots$, as λ_1 increases $\sum_{\ell=t}^{\infty} \lambda_{j_1,\ell}$ increases.

Step 2. Let $F(j) = \sum_{\ell=0}^{j} p_{0,\ell}$. Using Step 1, we show that for a given $j \geq 0$, $F(j)$ decreases in λ_1.

Step 3. Show that Step 2 implies an increase in λ_1 increases S_2.

Step 1. To show that for the chain in Figure 4.2(a), for a given state $(0, j_1)$ and for a given $t \geq j_1$, $\sum_{\ell=j_1}^{t} \lambda_{j_1,\ell}$ increases, we consider back the two-product priority queue. Consider the probability that "starting at state $(1, 0)$, first visit to states at C_0 is to the one of the states in $\{0, 1, \cdots, t - j_1\}$." This probability is simply $\sum_{\ell=0}^{t-j_1} q_{0,\ell}$. Let T denote the busy period of an $M/M/1$ queue with parameters λ_1 and μ. Note that for a given $t \geq 0$, $\sum_{\ell=0}^{t} q_{0,\ell}$ corresponds to the probability of t or fewer arrivals from a Poisson process with parameter λ_2 in the interval $(0, T]$ and from Lemma 4.2.2, $\sum_{\ell=0}^{t} q_{0,\ell}$ decreases with λ_1. This implies for $t \geq 0$, $\sum_{\ell=t}^{\infty} \lambda_{0,\ell}$ increases in λ_1.

Step 2. Consider the Markov chain defined by states in $\mathbf{C_0}$. A cut between states in $(0, j_1)$ and $(0, j_1+1)$ dissects C_0 into two sets of states $C_0^1(j_1)$ and $C_0^2(j_1)$ where $C_0^1(j_1) = \{(0,0), (0,1), \cdots, (0, j_1)\}$ and $C_0^2(j_1) = \{(0, j_1+1), (0, j_1+2), \cdots\}$. Note that rate of flow from $C_0^1(j_1)$ and $C_0^2(j_1)$ is

$$\lambda_1 \Big(\sum_{\ell_1=0}^{j_1} \sum_{\ell_2=j_1-\ell_1+1}^{\infty} q_{\ell_1,\ell_2} \Big) + \lambda_2 = \lambda_1 \Big(\sum_{\ell_1=1}^{j_1+1} \sum_{\ell_2=\ell_1}^{\infty} q_{0,\ell_2} \Big) + \lambda_2 \tag{4.2.7}$$

and flow from $C_0^2(j_1)$ to $C_0^1(j_1)$ is μ. Since for all possible cuts, at state j_1, flow from $C_0^2(j_1)$ to $C_0^1(j_1)$ increases in λ_1 (as shown in previous step), for any $j_1 \geq 0$, the ratio of fraction of time spent in $C_0^2(j_1)$ to fraction of time spent in $C_0^1(j_1)$ increases.

Step 3. Note for a given $t \geq 1$, if $\sum_{\ell=t}^{\infty} p_\ell^2$ increases (as a result of some effect), then optimal basestock level S_2 increases. To show that S_2 increases with λ_1, we just show that for a given t, $t \geq 1$, $\sum_{\ell=t}^{\infty} p_\ell^2$ increases with λ_1. Lemma 4.2.1 shows that limiting marginal probability for shortfall level j of product 2 is $p_j^2 = \frac{1}{\rho_2} p_{0,j+1}$. For $t \geq 1$, $\sum_{\ell=t}^{\infty} p_\ell^2 = \frac{1}{\rho_2} \sum_{\ell=t}^{\infty} p_{0,\ell+1}$. We show that as λ_1 increases, $\sum_{\ell=t}^{\infty} p_{0,\ell+1}$ increases, for $t \geq 1$.

Note that $p_j^2 = \frac{1}{\rho_2} p_{0,j+1}$, i.e., $\sum_{\ell=0}^{\infty} p_{0,\ell+1} = \rho_2$ and this holds for any λ_1. Considering partitions $C_0^1(0)$ and $C_0^2(0)$, an increase in λ_1 increases the fraction of time spent in $C_0^2(0)$ relative to the fraction of time spent in $C_0^1(0)$. Indeed, as λ_1 increases, "actual" fraction of time spent in $C_0^2(0)$, which is ρ_2, does not change, whereas actual fraction of time spent in $C_0^1(0)$ (which is $1 - \frac{\lambda_1+\lambda_2}{\mu}$) decreases. Thus, relatively fraction of time spent in $C_0^2(0)$ increases.

To show $\sum_{\ell=t}^{\infty} p_\ell^2$ increases in λ_1, we need to show that for $t \geq 1$, $\sum_{\ell=t}^{\infty} p_{0,\ell+1}$ increases. In other words, we need to show the "actual" time spent in $C_0^2(t)$ increases. Note that it is not straightforward to see whether actual time spent in $C_0^2(t)$ increases, since $p_{0,0}$ decreases with λ_1. For some $t \geq 1$, consider the two sets of states $C_0^1(t) \backslash (0,0)$ and $C_0^2(t)$. If state $(0,0)$ did not exist in the class at all, then following the same reasoning in Step 1, we would conclude that as λ_1 increases the fraction of time spent in $C_0^2(t)$ increases relative to fraction of time spent in $C_0^2(t)\backslash(0,0)$. Now, as λ_1 increases observe how flows out of $(0,0)$ into $\{C_0^1(t)\backslash(0,0) \cup C_0^2(t)\}$ affect the time spent in those states. For any $\ell_2 \geq 0$, $\sum_{\ell_1=0}^{\ell_2} q_{0,\ell_1}$ decreases in λ_1, whereas $\sum_{\ell_1=\ell_2+1}^{\infty} q_{0,\ell_1}^1$ increases in λ_1. This flow structure implies steady-state fraction of time spent in $C_0^2(t)$ relative to time spent in $C_0^1(t)\backslash(0,0)$ increases. Since fraction of time spent in $\{C_0^1(t)\backslash(0,0) \cup C_0^2(t)\}$ is constant at ρ_2, it should hold that for any $t \geq 1$, actual time spent in $C_0^2(t)$ increases with λ_1. Thus, for $t \geq 1$, $\sum_{\ell=t} \infty p_\ell^2$ increases. This completes the proof.

□

Corollary 4.2.2. *In the n-product priority MTS-queue, an increase in λ_k increases the optimal basestock level for products $k, k+1, \ldots, n$.*

Corollary 4.2.3. *Consider an n-product priority queue with identical products (i.e., $h_1 = h_2 = \cdots = h_n$, $b_1 = b_2 = \cdots = b_n$, $\lambda_1 = \lambda_2 = \cdots = \lambda_n$). The optimal basestock level for product k is higher than that of products $1, 2, \ldots, k-1$.*

4.3 Determining the production scheduling and substitution policy

To obtain the scheduling and substitution policy, the cost under priority-queue is improved by applying one step of the policy iteration algorithm. We first describe the generic policy iteration algorithm. The algorithm is described for discrete-time Markov Decision process under average reward criteria. The chain under consideration is a unichain.

Policy iteration algorithm

Step 0 Choose a stationary policy R. For $t = 0$ let the policy be denoted with R_t.

Step 1 Compute $g(R_t)$ (average cost per stage), $v(x, R_t)$ (relative value of state x).

Step 2 For each state $x \in S$, determine an action a_x that yields the minimum in

$$\min_{a \in A(x)} \{c(x,a) - g(R_t) + \sum_{y \in S} p(y|x,a) v(y, R_t)\}. \tag{4.3.1}$$

In Eq. 4.3.1 $A(x)$ denotes the set of actions available in state x, $p_{xy}(a)$ is the transition probability from state x to y under action a.

Step 3 Construct the new stationary policy R_{t+1}, by choosing action a_x for state x. If $R_t = R_{t+1}$ stop, the optimal policy is R_t. Otherwise let $t \leftarrow t+1$. Go to Step 1.

In this algorithm, Step 2 corresponds to the policy-improvement step. For our problem (defined in Eq. 3.1.1), we start with an initial policy (priority MTS queue policy) and apply the policy improvement step. For our problem, the policy-improvement step consists of the following decision for production idling and scheduling,

$$\min\{0, \min_{i \in N}\{\mu_i(v(x+e_i) - v(x))\}\}.$$

The expression states, if there exist one or more products with $v(x+e_i)-v(x) < 0$, then production is preferred to idling, and producing the product with lowest $\mu_i(v(x+e_i)-v(x))$ decreases the cost. Under the assumption of identical service rate μ, production will be scheduled for the product with minimum negative $v(x+e_i) - v(x)$. For the substitution decision (of product j with product i) policy improvement implies,

$$\min\{0, \min_{j \neq i}\{\mu_{ij}(v(x - e_i + e_j) - v(x) + c_{ij})\}\}.$$

The expression states that, considering all product combinations (given that the substitute product has positive on-hand inventory level), for each product i if there exist products that satisfy $v(x - e_i + e_j) - v(x) + c_{ij} < 0$, then product i must substitute one of those products. Under the identical substitution rate assumption, $\mu_{ij} = \mu_S$ the product that gives the lowest $v(x-e_i+e_j)-v(x)+c_{ij}$ will be substituted by i.

Step 1 of the policy iteration algorithm determines $g(R_t)$ and $v(x, R_t)$ under policy R_t by solving the system of functional equations.

For our problem, for $t = 0$ the starting Policy, R_0, is the priority MTS-queue policy and to apply the policy-improvement step, the relative values (bias differences) are required. For the n-product problem, as n increases, determining $g(R_0)$ and $v(x, R_0)$ by solving the system of equations is increasingly difficult. Therefore an alternative approach is proposed to determine the value differences. In the following, first the approach is described for the two-product problem, and then the approach is extended to the general n-product setting.

4.3.1 Value differences for the two-product problem

We first explicitly write down the functional equations for the priority MTS-queue policy. Let $\Lambda = \lambda_1 + \lambda_2 + \mu$.

$$\begin{aligned} v(x_1, x_2) + \frac{g}{\Lambda} \quad &= \frac{1}{\Lambda}\Big(\sum_{k=1}^{2} c_k(x_k) + \lambda_1 v(x_1 + 1, x_2) + \lambda_2 v(x_1, x_2 + 1) \\ &+\mu \mathbf{O} v(x_1, x_2)\Big), \quad x_1 \geq 0, x_2 \geq 0 \end{aligned} \tag{4.3.2}$$

where

$$c_k(x_k) = (h_k)(S_k - x_k)^+ + (b_k)(S_k - x_k)^-, \tag{4.3.3}$$

$$\mathbf{O} v(x_1, x_2) = \begin{cases} v(x_1 - 1, x_2) & \text{if } x_1 > 0 \\ v(x_1, x_2 - 1) & \text{if } x_1 = 0, x_2 > 0 \\ v(x_1, x_2) & \text{if } x_1 = 0, x_2 = 0 \end{cases} \tag{4.3.4}$$

is the priority operator for production.

Note that each product incurs holding and backordering cost independent of the other. Thus, the equation system can be decomposed, with respect to the products, into two equation systems as follows:

$$\begin{aligned} v_k(x_1, x_2) + \frac{g_k}{\Lambda} \quad &= \frac{1}{\Lambda}\{h_k(x) + \lambda_1 v_k(x_1 + 1, x_2) + \lambda_2 v_k(x_1, x_2 + 1) \\ &+\mu \mathbf{O} v_k(x_1, x_2)\}, \quad x_1 \geq 0, \ x_2 \geq 0, \ k = 1, 2. \end{aligned}$$

Note, $v(x_1, x_2) = v_1(x_1, x_2) + v_2(x_1, x_2)$. This implies, the bias value difference of two states can be expressed as sum of bias value differences of the states with respect to each product. For instance, for a given $x_1 \geq 0$, and $x_2 > 0$, the value difference $v(x_1+1, x_2-1) - v(x_1, x_2)$ can be expressed as,

$$v(x_1+1, x_2-1) - v(x_1, x_2) \quad = \sum_{k=1}^{2} \Big(v_k(x_1+1, x_2-1) - v_k(x_1, x_2) \Big)$$

In the following, we propose an approach to determine the bias value difference of two states. We first define the following functions:

$$g_1(x_1) = \begin{cases} v_1(x_1-1, x_{21}) - v_1(x_1, x_{22}) & \text{if } x_1 > 0, x_{21} \in \mathbb{Z}, x_{22} \in \mathbb{Z} \\ 0 & \text{if } x_1 = 0, \end{cases}$$

$$g_2^1(x_1, x_2) = \begin{cases} v_2(x_1, x_2-1) - v_2(x_1, x_2) & \text{if } x_2 > 0 \\ 0 & \text{if } x_2 = 0, \end{cases}$$

$$g_2^2(x_1, x_2) = \begin{cases} v_2(x_1-1, x_2) - v_2(x_1, x_2) & \text{if } x_1 > 0, x_2 \geq 0 \\ v_2(x_1, x_2-1) - v_2(x_1, x_2) & \text{if } x_1 = 0, x_2 > 0 \\ v_2(x_1, x_2) - v_2(x_1, x_2) & \text{if } x_1 = 0, x_2 = 0 \end{cases}$$

$$g_2^3(x_1, x_2) = \begin{cases} v_2(x_1-1, x_2+1) - v_2(x_1, x_2) & \text{if } x_1 > 0, x_2 \geq 0 \\ v_2(x_1, x_2) - v_2(x_1, x_2) & \text{if } x_1 = 0. \end{cases}$$

Note that the value of x_2 does not have an impact on $g_1(x_1)$, since product 1 is the higher priority product and the cost incurred by product 1 is independent of product 2.

Suppose we would like to obtain the contribution of product $i,\ i = 1, 2$ to the bias value differences of states (x_1-1, x_2+1) and (x_1, x_2). Two processes are considered where process 1 starts at state (x_1-1, x_2+1), and process 2 starts at state (x_1, x_2). First the two processes are coupled in that the same event sequences are assumed to take place. The difference in the cost of product i, under the two processes under each sample path gives the difference in bias values for that sample path. Taking expectation over all possible sample paths gives the contribution of product i to the bias difference.

For derivation of $g_1(x_1),\ g_2^1(x_1, x_2),\ g_2^2(x_1, x_2),\ g_2^3(x_1, x_2)$ see Appendix C. Note $g_1(x_1)$,$g_2^1(x_1, x_2)$, $g_2^2(x_1, x_2)$, $g_2^3(x_1, x_2)$ are functions whereas g in Eq. 4.3.2 is a scalar that denotes the average reward.

The functions, $g_1(x_1)$,$g_2^1(x_1,x_2)$, $g_2^2(x_1,x_2)$, $g_2^3(x_1,x_2)$, are defined only for two-product setting. Using these functions any bias value difference can be calculated. For instance, given $x_1 > 0,\ x_2 > 0$:

$$\begin{aligned} v(x_1+1,x_2-1) - v(x_1,x_2) \quad &= \sum_{k=1}^{2}\Big(v_k(x_1+1,x_2-1) - v_k(x_1,x_2)\Big) \\ &= g_1(x_1) + g_2^3(x_1,x_2) \end{aligned}$$

We use these relations when deriving the value differences for the n-product setting below.

4.3.2 Value differences for the n-product problem

For the n-product priority MTS-queue, let $\mathbf{x}$ denote the state vector. The system of functional equations are expressed as follows,

$$v(\mathbf{x}) + \frac{g}{\Lambda} = \frac{1}{\Lambda}\big(\sum_{k=1}^{n} c_k(x_k) + \sum_{k=1}^{n}\lambda_k v(\mathbf{x}+e_i) + \mu\mathbf{O}v(\mathbf{x})\big),\ \mathbf{x}\geq 0$$

where $\Lambda = \sum_{i=1}^{n}\lambda_i + \mu$, is the uniformization rate, $c_k(x_k)$ defined in Eq. 4.3.3,e_i is the unit vector with 1 at the i'th entry, and,

$$\mathbf{O}v(\mathbf{x}) = \begin{cases} v(x_1-1,x_2,..,x_n) & \text{if } x_1 > 0 \\ v(x_1,x_2,...,x_k-1,..,x_n) & \text{if } x_1 = 0, ..x_{k-1} = 0, x_k > 0,\ k \in N \\ v(\mathbf{x}) & \text{if } x_1 = 0,..,x_n = 0 \end{cases}$$

is the production priority operator. The functional equations can be decomposed, with respect to the products, into n groups of functional equations as follows.

$$v_k(\mathbf{x}) + \frac{g_k}{\Lambda} \quad = \quad \frac{1}{\Lambda}\Big(h_k(x_k) + \sum_{i=1}^{N}\lambda_i v_k(\mathbf{x}+e_i) + \mu\mathbf{O}v_k(\mathbf{x})\Big),\ \mathbf{x}\geq 0.$$

Then,

$$v(\mathbf{x}) = \sum_{k=1}^{n} v_k(\mathbf{x}). \tag{4.3.5}$$

For $x_i > 0$, let $f_{ij}(\mathbf{x})$ be defined as,

$$f_{ij}(\mathbf{x}) = v(\mathbf{x} - \mathbf{e_i} + \mathbf{e_j}) - v(\mathbf{x}).$$

The following observation is made.

Observation 4.3.1. *Given $i < j$,*

$$\begin{aligned}
f_{ji}(\mathbf{x}) &= f_{ji}(x_1, x_2, \ldots, x_i, \ldots, x_j, \ldots, x_N) \\
&= v(x_1, x_2, \ldots, x_i + 1, \ldots, x_j - 1, \ldots, x_N \\
&\quad -v(x_1, x_2, \ldots, x_i, \ldots, x_j, \ldots, x_N) \\
&= -(v(x_1, x_2, \ldots, x_i, \ldots, x_j, \ldots, x_N) \\
&\quad -v(x_1, x_2, \ldots, x_i + 1, \ldots, x_j - 1, \ldots, x_N)) \\
&= -f_{ij}(x_1, x_2, \ldots, x_i + 1, \ldots, x_j - 1, \ldots, x_N) \\
&= -f_{ij}(\mathbf{x} - \mathbf{e_j} + \mathbf{e_i})
\end{aligned}$$

4.3.2.1 Determining the value differences, $f_{ij}(\mathbf{x})$

Note, from Eq. 4.3.5, it is possible to write the bias value difference for states $(\mathbf{x} - \mathbf{e_i} + \mathbf{e_j})$ and $\mathbf{x}$ as the sum of individual value difference of each product

$$f_{ij}(\mathbf{x}) = \sum_{k=1}^{n} f_{ij}^k(\mathbf{x}),$$

where $f_{ij}^k(\mathbf{x})$ denotes the value difference due to product k,

$$f_{ij}^k(\mathbf{x}) = v_k(\mathbf{x} - \mathbf{e_i} + \mathbf{e_j}) - v_k(\mathbf{x})$$

To determine $f_{ij}(\mathbf{x})$, we determine individual bias value differences, $f_{ij}^k(\mathbf{x})$. The contribution of product k to the bias value difference, $f_{ij}^k(\mathbf{x})$, is determined as follows. Two processes are considered where process 1 starts at state $\mathbf{x} - \mathbf{e_i} + \mathbf{e_j}$, and process 2 starts at state $\mathbf{x}$. To obtain $f_{ij}^k(\mathbf{x})$, first the two processes are coupled in that the same event sequences are assumed to take place. The difference in the cost of product k under the two processes under each sample path gives the difference in bias values for that sample path. Taking expectation over all possible sample paths gives $f_{ij}^k(\mathbf{x})$. This is the same approach used to obtain the functions $g_1(x_1)$, $g_2^1(x_1,x_2), g_2^2(x_1,x_2), g_2^3(x_1,x_2)$. Actually, in the following we see that $f_{ij}^k(\mathbf{x})$ are expressed through those functions. Construction of $f_{ij}^k(\mathbf{x})$ may change depending on i, j and k. There are 5 possible cases, each of which is explained below. We derive the differences only for $i<j$. For $i>j$, we use the equivalence $f_{ij}(\mathbf{x}) = -f_{ji}(\mathbf{x}-\mathbf{e_i}+\mathbf{e_j})$(see Observation 4.3.1).

1. For $k<i$, $f_{ij}^k(\mathbf{x}) = 0$, since products i and j have lower priority than product k. The cost incurred due to product k is the same under both processes.

2. For $k=i$, $f_{ij}^k(\mathbf{x})$ $(= f_{ij}^i(\mathbf{x}))$ can equivalently be expressed as $v_i(x_1,x_2,\ldots,x_i-1,x_{i+1},\ldots,x_j,\ldots,x_n) - v_i(x_1,x_2,\ldots,x_i,x'_{i+1},\ldots,x'_j,\ldots,x'_n)$. Note in the expression, for $l>i$ it may possibly hold that $x'_l \neq x_l$. The reason is as follows. Of the two processes considered, suppose process 1 starts at state $\mathbf{x}-e_i$, and process 2 starts at state $\mathbf{x}$. When calculating the cost difference due to product i, it is sufficient to consider the costs incurred due to product i during the time until process 2 visits the state of 0 shortfall for product i. From that point on, the costs incurred by product i are identical under the two processes. Thus, the initial shortfall level of product j is irrelevant when calculating $f_{ij}^i(\mathbf{x})$. Observe that,

$$f_{ij}^i(\mathbf{x}) = \begin{cases} g_2^1(\sum_{l=1}^{i-1} x_l, x_i) & i>1 \\ g_1(x_1) & i=1. \end{cases}$$

   In g_2^1, all products with higher priority than product i are aggregated and defined as product 1, and product k is defined as product 2. Thus shortfall of "product 1" is defined as $\sum_{l=1}^{k-1} x_l$ and arrival rate of "product 1" is defined as $\sum_{l=1}^{k-1} \lambda_l$.

3. For $i<k<j$, to find $f_{ij}^k(\mathbf{x})$, assume process 1 starts at state $(x_1,x_2,..,x_i-1,..,x_k,..,x_j+1,..x_n)$ and process 2 starts at state $(x_1,x_2,..,x_i,..,x_k,..,x_j,..x_n)$. Then,

$$f_{ij}^k(\mathbf{x}) = g_2^2(\sum_{l=1}^{k-1} x_l, x_k)$$

4. For $k = j$,

$$f_{ij}^{j}(\mathbf{x}) = g_2^3(\sum_{l=1}^{j-1} x_l, x_k).$$

5. For $k > j$, to find $f_{ij}^{j}(\mathbf{x})$, assume process 1 starts at state $(x_1, x_2, .., x_i - 1, .., x_j{+}1, .., x_k, ..x_n)$ and process 2 starts at state $(x_1, x_2, .., x_i, .., x_j, .., x_k, ..x_n)$. Note the cost incurred due to product k is the same under process 1 and process 2, since products i and j have higher priority than k. Then, $f_{ij}^{k}(\mathbf{x}) = 0$.

4.3.2.2 Using the bias value differences, $f_{ij}(\mathbf{x})$, in the policy iteration

Given that the bias value differences are obtained, the final step is to use the differences in the policy improvement step of the policy iteration algorithm. We start with priority MTS-queue and using the bias value differences of the states, construct a new production scheduling and production substitution policy.

1. The policy-improvement step consists of the following decision for production idling and scheduling,

$$\min\{0, \min_{i \in N}\{\mu(v(x + e_i) - v(x))\}\}.$$

The expression states if there exist one or more products with $v(x + e_i) - v(x) < 0$, then production is preferred to idling, and producing the product with lowest $(v(x + e_i) - v(x))$ decreases the cost. Thus, production will be scheduled for the product with minimum negative $v(x + e_i) - v(x)$.

2. The policy-improvement step consists of the following decision for the substitution (of product j with product i),

$$\min\{0, \min_{j \neq i}\{\mu_{ij}(v(x - e_i + e_j) - v(x) + c_{ij})\}\}.$$

The expression states that, considering all product combinations (given that the substitute product has positive on-hand inventory level), for each product i if there exist products that satisfy $v(x - e_i + e_j) - v(x) + c_{ij} < 0$, then product i must substitute one of those products.

Note that through the policy improvement step, production scheduling and substitution policies that results in lower cost are obtained. However, production idling policy does not deviate from that of the priority MTS-queue. In other words, the policy improvement step proposes exactly the same idling levels as the priority MTS-queue.

To sum up, the Prio-DH policy is obtained as follows. For a priority MTS-queue policy, under a given priority order, optimal basestock levels are obtained considering the marginal probabilities and the critical fractile. Then, the bias value differences are calculated and through policy-improvement step of policy iteration algorithm, the differences are used in finding the production scheduling and substitution decisions that decrease the cost. The new policy is called Prio-DH.

4.4 Performance of the policy iteration step (numerical results)

Prio-DH improves the priority policy, to assess its performance the cost under Prio-DH is compared with that under the optimal policy (obtained by Markov decision process). Obviously, the order of the products in the priority policy has an effect on the cost under Prio-DH. In this section when comparing Prio-DH with optimal cost, we also try to obtain insights on the effect of the priority order. Then in § 4.5 we propose an algorithm to determine a "good" priority order.

Optimal solution requires a solution of a Markov decision process. Although the state-space can be truncated (i.e., the infinite state system is approximated by a finite state system), and MDP can be solved numerically, as the number of products increases a solution is harder to get due to the curse of dimensionality. We compare the Prio-DH with optimal policy in the two-product setting. For Product 1 and Product 2, all combinations of holding cost $h_k \in \{0.2, 0.05\}$ and $b_k \in \{2, 0.5\}$ for $k = 1, 2$ are evaluated. This makes a total of 16 instances. Both non-identical arrival rates ($\lambda_1 = 0.2, \lambda_2 = 0.6$) and identical arrival rates ($\lambda_1 = \lambda_2 = 0.4$) are considered, in order to understand how arrival rates affect the performances. Thus in total 32 instances are used in the analysis. Results are obtained under a substitution rate of $\mu_{ij} = \mu_s = 5$, and for production rate of $\mu = 1$. The cost of substitution is assumed identical for both products $c_{ij} = c = \{1, 5, \infty\}$. When $c = \infty$, substitution does not take place. For the state space, a dimension of 50×50 is used, and the value iteration algorithm for MDP stops after (at most) 1500 iterations.

To evaluate the cost under Prio-DH, either value iteration can be used for the specified policy or under the given policy, steady-state probabilities can be obtained. We prefer the second approach.

When evaluating the performance of Prio-DH, we evaluate both $1 \prec 2$ and $2 \prec 1$. Policy iteration step helps construct the switching curve (i.e., the production schedule) and the substitution curve. However, it does not have an impact on the hedging point (basestock level) suggested by the priority MTS-queue policy. Veatch and Wein (1996) observes that determining the hedging point correctly significantly improves the performance of a production policy. We also compare the hedging points obtained by the priority policies and the optimal policy. This helps to understand how well the priority policy approximates the idling levels.

In the following, we compare costs under Priority MTS-queue policy (shortly, **Prio**), Prio-DH policy and the optimal policy (**Opt**). In Table 4.1 and Table 4.2 the expected average cost values are presented under non-identical arrival rates $\lambda_1 = 0.2,\ \lambda_2 = 0.6$, and identical arrival rates $\lambda_1 = 0.4,\ \lambda_2 = 0.4$, respectively.

In the following, we present the observations based on the performances of Prio, Prio-DH and Opt. Before the observations, we present two additional tables, Table 4.4 and Table 4.3. In these tables the performance of Prio-DH are compared with Prio and Opt, under non-identical and identical arrival rates. In each table, the performances are compared under $c = \infty, \mu_S = 5$, $c = 5, \mu_S = 5$, and $c = 1, \mu = 5$. For Prio the priority order that gives the lower cost is considered. And Prio-DH is determined by applying one step of the policy iteration algorithm on the selected Prio. In the tables, under column Prio, and Opt, the values show $\frac{Cost_{Prio} - Cost_{PrioDH}}{Cost_{PrioDH}} 100\%$, and $\frac{Cost_{PrioDH} - Cost_{Opt}}{Cost_{PrioDH}} 100\%$.

Table 4.1: Performance of Prio-DH compared to Prio under non-identical λ

#	λ_1	λ_2	h1	h2	b1	b2	Cost					Base-stock		
							Prio		Prio-DH		OPT	Prio and Prio-DH		OPT
							1<2	2<1	1<2	2<1		1<2	2<1	
1	0.2	0.6	0.2	0.2	2	2	2.38	2.68	2.20	2.49	2.15	(3,7)	(1,10)	(7,4)
2	0.2	0.6	0.2	0.2	2	0.5	1.35	2.23	1.27	1.93	1.22	(2,4)	(1,5)	(7,2)
3	0.2	0.6	0.2	0.2	0.5	2	2.24	1.78	1.84	1.65	1.33	(2,4)	(0,10)	(3,4)
4	0.2	0.6	0.2	0.2	0.5	0.5	1.22	1.33	1.21	1.20	1.14	(1,4)	(0,5)	(3,2)
5	0.2	0.6	0.2	0.05	2	2	1.08	2.12	1.00	1.96	0.99	(1,15)	(1,16)	(7,7)
6	0.2	0.6	0.2	0.05	2	0.5	0.79	1.99	0.73	1.72	0.71	(1,9)	(1,10)	(7,4)
7	0.2	0.6	0.2	0.05	0.5	2	0.95	1.21	0.84	1.10	0.70	(0,11)	(0,16)	(3,7)
8	0.2	0.6	0.2	0.05	0.5	0.5	0.65	1.09	0.65	1.00	0.64	(0,10)	(0,10)	(3,4)
9	0.2	0.6	0.05	0.2	2	2	2.22	1.64	2.01	1.52	1.16	(14,2)	(2,10)	(13,4)
10	0.2	0.6	0.05	0.2	2	0.5	1.20	1.19	1.09	1.04	0.82	(9,1)	(2,5)	(13,2)
11	0.2	0.6	0.05	0.2	0.5	2	2.18	1.37	1.75	1.24	0.90	(9,2)	(1,10)	(7,4)
12	0.2	0.6	0.05	0.2	0.5	0.5	1.16	0.92	1.07	0.82	0.79	(8,1)	(1,5)	(7,2)
13	0.2	0.6	0.05	0.05	2	2	0.93	1.07	0.87	1.02	0.83	(5,11)	(2,16)	(13,7)
14	0.2	0.6	0.05	0.05	2	0.5	0.64	0.94	0.59	0.87	0.57	(3,8)	(2,10)	(13,4)
15	0.2	0.6	0.05	0.05	0.5	2	0.89	0.80	0.77	0.73	0.60	(3,8)	(1,16)	(7,7)
16	0.2	0.6	0.05	0.05	0.5	0.5	0.59	0.67	0.56	0.63	0.54	(3,7)	(1,10)	(7,4)

In the table h_i and b_i are the holding cost and backorder cost of product i. $i < j$ means, product i is given higher priority over product j. Prio is the cost of priority policy for given order. Prio-DH is heuristic policy cost for given product order. Opt is optimal policy cost. Base-stocks are hedging points of given policies. $\lambda_1 = 0.2, \lambda_2 = 0.6, c_S = 1, \mu_S = 5$

Table 4.2: Performance of Prio-DH compared to Prio under identical λ

#	h1	h2	b1	b2	Cost					Base-stock		
					Prio		Prio-DH		OPT	Prio and Prio-DH		OPT
					1<2	2<1	1<2	2<1		1<2	2<1	
1	0.2	0.2	2	2	2.54	2.54	2.32	2.32	2.16	(2,9)	(9,2)	(5,5)
2-3	0.2	0.2	2	0.5	1.53	2.29	1.45	1.88	1.27	(2,4)	(9,1)	(3,3)
4	0.2	0.2	0.5	0.5	1.29	1.29	1.18	1.18	1.15	(1,4)	(4,1)	(3,2)
5-9	0.2	0.05	2	2	1.30	2.24	1.20	2.04	1.07	(2,15)	(9,4)	(1,15)
6-11	0.2	0.05	2	0.5	1.01	2.16	0.94	1.76	0.80	(2,9)	(9,2)	(1,10)
7-10	0.2	0.05	0.5	2	1.06	1.23	0.94	1.09	0.75	(1,15)	(4,4)	(1,10)
8-12	0.2	0.05	0.5	0.5	0.76	1.16	0.72	1.02	0.71	(1,9)	(4,2)	(1,9)
13	0.05	0.05	2	2	1.00	1.00	0.94	0.94	0.83	(4,15)	(15,4)	(8,8)
14-15	0.05	0.05	2	0.5	0.71	0.93	0.65	0.82	0.58	(4,9)	(15,2)	(6,5)
16	0.05	0.05	0.5	0.5	0.63	0.63	0.59	0.59	0.55	(2,9)	(9,2)	(5,5)

In the table h_i and b_i are the holding cost and backorder cost of product i. $i < j$ means, product i is given higher priority over product j. Prio is the cost of priority policy for given order. Prio-DH is heuristic policy cost for given product order. Opt is optimal policy cost. Base-stocks are hedging points of given policies. $\lambda_1 = 0.2, \lambda_2 = 0.6, c = 1, \mu_S = 5$

The performances of Prio and Prio-DH are affected by the priority order selected under Prio. The parameters λ_i, h_i and b_i affect which priority order works better. Thus, we first present observations on how the priority order is affected by the parameters and then compare the performances of policies Prio, Prio-DH and Opt.

Observations

Observation 1 We observe that if $\lambda_i < \lambda_j$, a priority order of $i \prec j$ works better (given that all other parameter values are the same).

If λ_i is lower that λ_j, then product i changes its state less frequently compared to product j. Giving a higher priority to product i increases the speed at which the shortfall decreases. Such a policy is expected to result in lower cost. In Table 4.1, for the instances $h_1 = h_2$ and $b_1 = b_2$ ($\#1, 4, 13, 16$) arrival rate determines the best priority order.

Observation 2 If $h_i > h_j$ then it is more likely that $i \prec j$ gives lower cost under priority MTS-queue especially if the arrival rate is also lower for product 1, given that other parameters assume the same values for the products.

If unit holding cost is higher for a product, then lower basestock is kept for the product, and it is more likely that the product will have negative net inventory in steady state. Thus, giving higher priority to the product with higher holding cost will help to clean up the backorders in a faster rate. See that for the instances 5,8,9,12 in Table 4.1 and 4.2, if the product with higher unit holding cost is given a higher priority, then total cost is lower. On the other hand, if a product has high holding cost and high arrival rate, then it depends on the dominance of either factor. For instance, although $\lambda_1 < \lambda_2$ in Table 4.1 for the instances $\#9, 10, 11, 12$, $2 \prec 1$ results in lower cost due to $h_2 < h_1$.

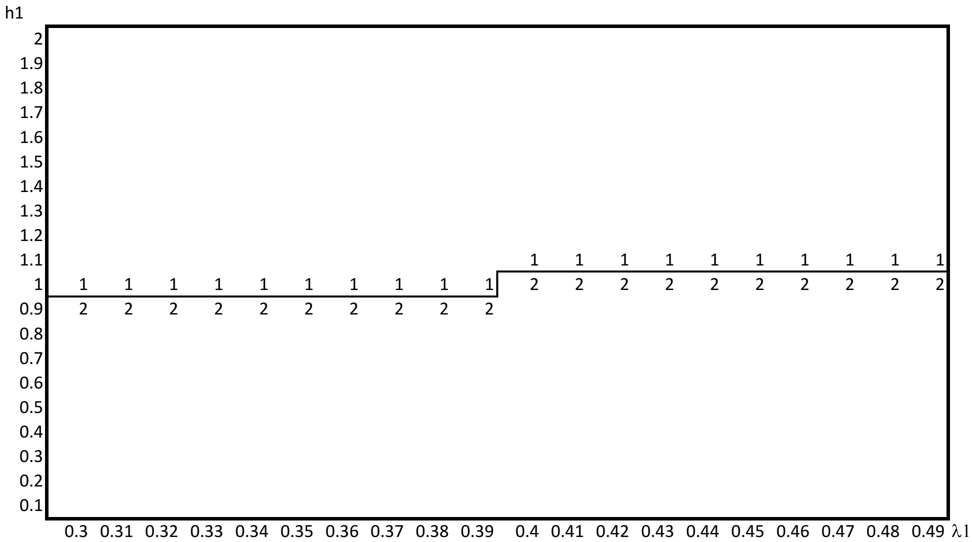
Observation 3 If $b_i > b_j$, then it is more likely that $i \prec j$, given that other parameters assume the same values for the products, however there might be some counter examples.

Table 4.3: Comparison of Prio, Prio-DH and Opt under $\lambda_1 = 0.2,\ \lambda_2 = 0.6$

							no subs		subs with c=5,$\mu_S = 5$		subs with c=1,$\mu_S = 5$	
							% difference Prio-DH					
#	λ_1	λ_2	h_1	h_2	b_1	b_2	Prio	Opt	Prio	Opt	Prio	Opt
1	0.2	0.6	0.2	0.2	2	2	5.26%	-3.11%	8.21%	-1.96%	5.48%	-3.35%
2	0.2	0.6	0.2	0.2	2	0.5	3.85%	-2.50%	6.77%	-4.07%	3.98%	-2.60%
3	0.2	0.6	0.2	0.2	0.5	2	4.73%	-6.14%	7.80%	-19.55%	7.14%	-10.31%
4	0.2	0.6	0.2	0.2	0.5	0.5	0.15%	-5.60%	0.75%	-5.64%	0.15%	-5.76%
5	0.2	0.6	0.2	0.05	2	2	2.87%	-0.60%	8.72%	-0.74%	3.20%	-0.72%
6	0.2	0.6	0.2	0.05	2	0.5	1.84%	-0.20%	8.58%	-2.50%	2.12%	-0.36%
7	0.2	0.6	0.2	0.05	0.5	2	11.81%	-7.89%	13.38%	-16.36%	11.86%	-11.11%
8	0.2	0.6	0.2	0.05	0.5	0.5	0.02%	-0.04%	1.06%	-0.45%	0.02%	-0.15%
9	0.2	0.6	0.05	0.2	2	2	0.30%	-6.06%	7.56%	-23.99%	5.74%	-10.02%
10	0.2	0.6	0.05	0.2	2	0.5	3.00%	-21.05%	14.00%	-21.88%	4.27%	-20.57%
11	0.2	0.6	0.05	0.2	0.5	2	1.79%	-1.32%	10.16%	-27.35%	8.30%	-7.84%
12	0.2	0.6	0.05	0.2	0.5	0.5	2.25%	-2.23%	11.61%	-3.44%	3.13%	-2.03%
13	0.2	0.6	0.05	0.05	2	2	5.90%	-4.32%	7.35%	-4.05%	6.06%	-4.47%
14	0.2	0.6	0.05	0.05	2	0.5	5.88%	-3.05%	7.95%	-4.10%	5.98%	-3.12%
15	0.2	0.6	0.05	0.05	0.5	2	4.55%	-11.95%	9.54%	-18.11%	7.95%	-12.77%
16	0.2	0.6	0.05	0.05	0.5	0.5	5.36%	-3.02%	5.76%	-3.22%	5.38%	-3.13%

Table 4.4: Comparison of Prio, Prio-DH and Opt under $\lambda_1 = \lambda_2 = 0.4$

					no subs		subs with c=5,$\mu_S = 5$		subs with c=1,$\mu_S = 5$	
					% difference Prio-DH					
#	h_1	h_2	b_1	b_2	Prio	Opt	Prio	Opt	Prio	Opt
1	0.2	0.2	2	2	5.23%	-9.24%	9.43%	-6.93%	6.26%	-8.60%
2	0.2	0.2	2	0.5	2.85%	-6.49%	6.06%	-12.10%	3.75%	-6.72%
4	0.2	0.2	0.5	0.5	7.12%	-3.79%	8.54%	-3.03%	7.15%	-3.85%
5	0.2	0.05	2	2	1.95%	-2.79%	8.35%	-10.94%	5.70%	-2.66%
6	0.2	0.05	2	0.5	1.14%	-1.13%	7.93%	-14.43%	4.75%	-1.81%
7	0.2	0.05	0.5	2	9.23%	-17.07%	12.20%	-20.35%	9.47%	-17.35%
8	0.2	0.05	0.5	0.5	3.35%	-0.51%	5.62%	-2.03%	3.50%	-0.44%
13	0.05	0.05	2	2	4.84%	-11.98%	7.31%	-10.76%	5.48%	-11.62%
14	0.05	0.05	2	0.5	6.60%	-7.89%	9.84%	-9.86%	7.32%	-7.95%
16	0.05	0.05	0.5	0.5	5.54%	-8.97%	7.37%	-7.73%	5.67%	-8.94%



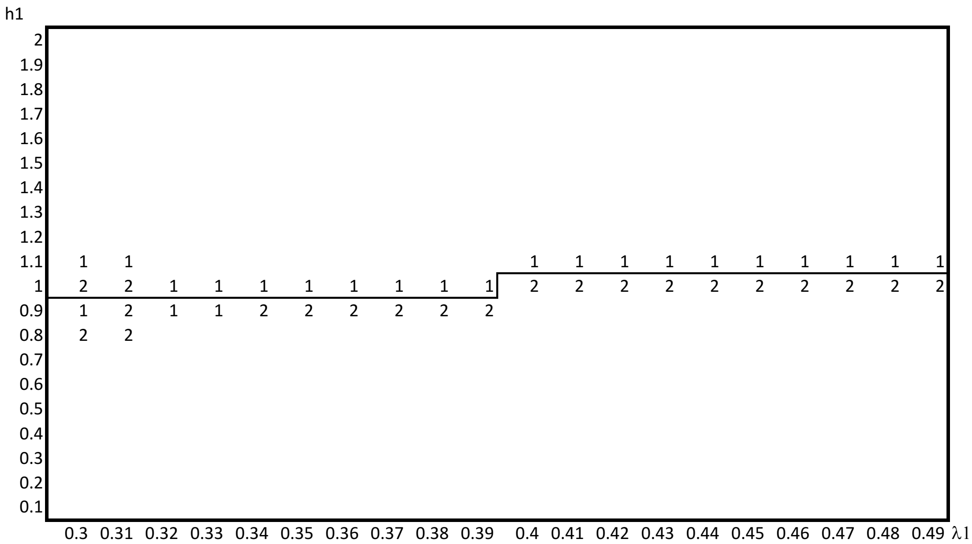


Figure 4.3: Optimal priority as a function of λ_1 and h_1 under parameters $b_1 = b_2 = 10, \lambda_2 = 0.4, h_2 = 1$. Above for Prio and Below for Prio-DH.

If $b_i > b_j$ then, lowering down the time spent in the backorder state for product i is likely to lower down the cost. One way to ensure this is to increase the basestock level for product i, and another is to assign a higher priority to product i. In the literature, it is shown that an optimal production scheduling policy for the general n-product MTS-queue is to produce for the product with higher unit backorder cost, whenever that product falls to backorderDe Vericourt et al. (2002). This policy is named as $c\mu$ rule. If higher priority is assigned to product i, then actions taken in the states where product i has backorders, coincides with the $c\mu$ rule. Thus, it is likely that the cost would be lower under such a priority scheme. In Table 4.5, instances $\#2, 3, 14, 15$ support this conjecture.

Finally, we observe that when unit holding cost and backorder cost orderings are not aligned (and furthermore when the arrival rate ordering is not aligned with those two), the performance of the priority policy compared to the optimal policy deteriorates. Contrast for example, in Table 4.4 instance 6 with instance 7. Also in Table 4.5 instance 6 with instance 10.

Observation 4 In Tables 4.4 and Table 4.5, under *no substitution* the difference between the costs under Prio and Prio-DH is due to the improved product scheduling policy. On the other hand, under $c = 1, \mu_S = 5$ and $c = 5, \mu_S = 5$ settings, the difference is due to both the improved production scheduling and substitution policies. We observe that cost under Prio-DH is lowest under $c = 1, \mu_S = 5$ and highest under no substitution, as expected.

Comparing Prio with Prio-DH, looking at the setting where arrival rates are non-identical and where $c = 1, \mu_S = 5$, we observe that the % improvement changes between $0.75 - 14.0\%$ with an average value of 8.1%. The average % improvement decreases to 5.0% under $c = 5$ and to 3.7% under no substitution setting. We further observe that average % improvements under the identical arrival rates are 8.26% for $c = 1$, 5.91% for $c = 5$ and 4.79% for no-substitution setting. These values are slightly larger than those under non-identical arrival rates.

Under the non-identical arrival rates, when substitution is taking place

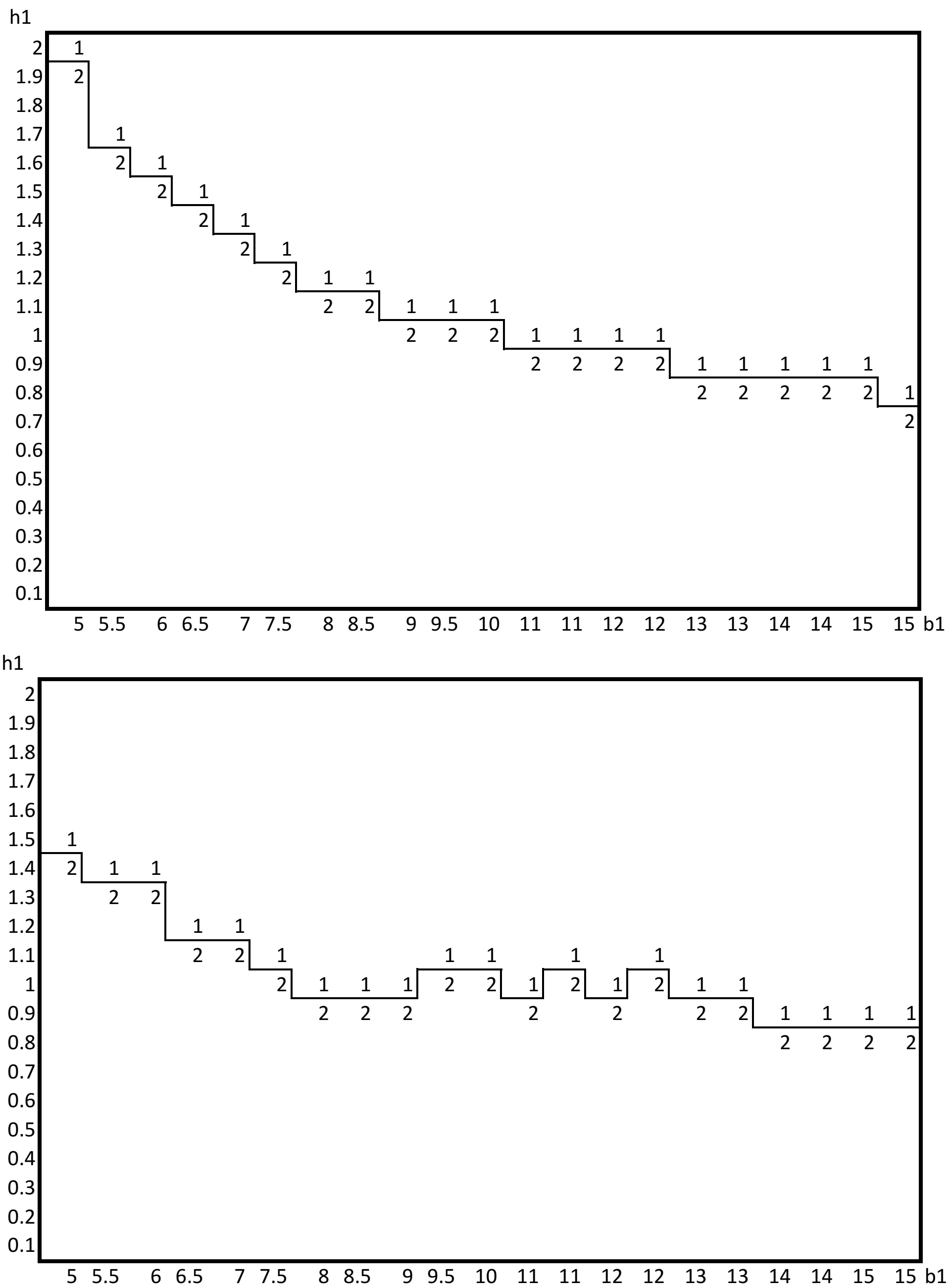


Figure 4.4: Optimal priority as a function of b_1 and h_1 under parameters $b_2 = 10, \lambda_1 = \lambda_2 = 0.4, h_2 = 1$. Above for Prio and Below for Prio-DH.

(i.e, when $c = 1$ or $c = 5$) the highest % improvements are obtained when $h_1 < h_2$ and $\lambda_1 < \lambda_2$. These are the settings, where holding cost and arrival rates imply opposite priority orderings. Under identical arrival rates we observe a similar pattern when $h_i < h_j$ and $b_i > b_j$, i.e., when the unit cost of holding and backordering imply opposite orderings.

In these settings, the misalignment between the priority orderings implied by different parameters lead to a deterioration of the performance of Prio compared to the optimal policy. Prio-DH is able to dampen the deterioration by improving the scheduling and substitution decisions.

For a given substitution cost, the settings where the improvement is lowest are when backordering costs are low and are identical for both products.

Observation 5 Looking at the difference between Prio-DH and Opt, we observe that when % improvement of Prio-DH is high, then % difference between Prio-DH and Opt is likely to be high. And the difference is high when the arrival rates, unit holding and unit backordering costs imply opposite priority ordering. When the ordering implied by the parameter are aligned, then it is likely that both Prio and Prio-DH operate close to optimal policy.

Observation 6 For the two-product case, under priority policy, the high priority product has lower basestock level, and low priority product has higher basestock level than those under the optimal (dynamic) policy. However, under certain instances this relation is not observed (see Table 4.2, instances $\#10, 12$).

When the holding cost of a product is high, the optimal policy tends to keep the inventory for the product with low holding cost, and meet the demand for both products through the lower cost product. This results in a very low basestock level for the product with high unit holding cost and a high basestock for the product with low unit holding cost. In those cases, the basestock levels under optimal policy and priority policy are aligned (see Tables 4.1 and 4.2), and thus the cost values are close. However, for the instances where basestock levels have completely different patterns (eg. Table 4.1 instances #9, 10, 11, 12 Table 4.2 instances #1, 13, 14) the difference between Prio and Opt is high.

Observation 7 There are a number of instances that do not follow any of the above discussed patterns. We also note that, in practice it is unlikely that a product has $h_1 < h_2$ and $b_1 > b_2$. Since unit backordering costs and unit holding costs are usually proportional to the "value" of the product. Furthermore, it is expected that more 'valuable" products face lower demand compared to staple products. Thus, assuming $\lambda_1 < \lambda_2$ and $h_1 > h_2$ might be reasonable.

Observation 8 In Tables 4.4 and Table 4.5, when we look at the difference between opt and Prio-DH, best case is $-\%0.04$ and worst case is $-\%27.35$. In 10 out of 96 instances, this difference is worse than %16 and in 38 out of 96 instances, this difference is better than %5. For the difference between Prio and Prio-DH, best improvement is %12.20 and worst improvement is %0.02. In 27 out of 96 instances, Prio-DH makes more than %7 percent improvement over Prio and in 13 out of 96 instance, it makes less than %3 percent improvement.

4.5 Determining the priority order

As discussed in § 4.2 the cost under the priority-queue policy could be lowered by selecting a "good" priority order of the products. Each product is characterized by arrival rate, holding cost and backordering cost. Consider the special case where products are all identical except for their backordering cost. The lowest priority product is the product that gets the lowest attention, ie. most likely to spend time incurring holding or backordering cost. Thus, intuitively it is more beneficial to give the product with highest backordering cost the highest priority. As discussed in the previous section, similar intuition holds for the case when a product has significantly lower arrival rate compared to the other, or when a product has considerably higher unit holding cost compared to the other.

However, there are settings where a "good" priority order cannot be determined following the intuition. This might be due to the fact that different parameters imply opposite priority orders. Or it might be the priority order does not simply follow the intuition. We present the example below to show the latter case.

Example Consider two products, each with arrival rate 0.1. Service rate is $\mu = 1$, holding and backorder costs are $h_1 = 1$, $h_2 = 1$, $b_1 = 10$ and $b_2 = 5$. Note that backordering cost of product 2 is higher. We compare the two settings where two different priorities are assigned to these products in Table 4.5.

Table 4.5: Comparison of priority orderings when unit backorder costs are different

Ordering	Optimal Basestock for Product 1	Optimal Basestock for Product 2	Cost due to Product 1	Cost due to Product 2	Total cost due to products
$1 \prec 2$	1	0	1.01	0.69	1.7
$2 \prec 1$	1	0	1.06	0.55	1.61

In this two-product example intuitively first product must be given higher priority, since $b_1 > b_2$ and the products are identical otherwise (see Figure 4.4). However, note that $1 \prec 2$ results in 5.6% higher cost than $2 \prec 1$. Moving from priority 1 to priority 2, note that the increase in cost of Product 2 is lower than the increase in Product 1 (5% versus 25%). This is interesting, since the backordering cost of Product 1 is higher. However, since basestock level for Product 1 is 1, and Product 2 is 0, higher basestock level results in lower increase in the total cost for Product 1.

In the following, a procedure to determine a "good" priority order will be discussed. Although after the one-step improvement, Prio-DH may perform better when it takes a non-optimal priority ordering as a base, the results in Table 4.1 and Table 4.2 imply that in a majority of the settings, starting with a good priority ordering results in better performance for Prio-DH. The complexity of finding the optimal policy is $O(n!)$, whereas the proposed procedure(which possibly yields non-optimal orders) has complexity $O(n^2)$, as will be shown below. The procedure varies for the identical arrival rate, and non-identical arrival rate cases. Each case is analyzed separately.

Identical arrival rates, $\lambda_k = \lambda$

Assume the products are ordered as $1 \prec \cdots \prec n$. In Figure 4.5, $i \prec j$. Suppose the priorities of products i and j are interchanged. Since the arrival rates are the same for all products, the cost of other products will not be affected by the interchange. The only cost change stems from the change in cost of product i and j. Thus only the change in cost of product i and j is calculated, to obtain the new total cost under the priority MTS-queue system.

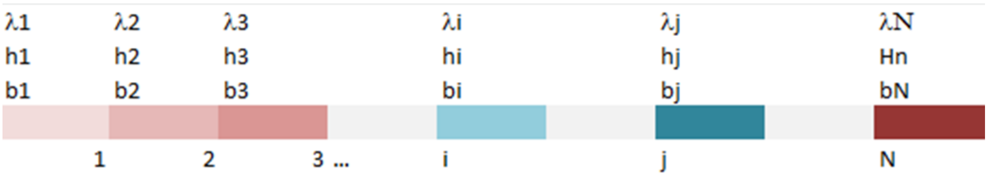


Figure 4.5: Product order under priority-queue.

Through a small numerical analysis, we study the impact of the priority of a product on the cost change after interchange. Suppose $n = 9$, $\lambda = 0.1$, $\mu = 1$, product i has $h_i = 1, b_i = 5$, and product j has $h_j = 1$, $b_j = 10$. Obviously, product j generates higher cost than product i for a given priority. Furthermore, since the shortfall level of a product with lower priority is stochastically larger than that of a product with higher priority, for the same product the magnitude of cost increases as the priority decreases. This means product i must have lower priority than j under the optimal order. For the analysis, we consider all possible priorities (i.e., ranks) for product i and product j. There are $\binom{9}{2}$ possible comparisons. Figure 4.6 shows that for a given priority rank for product i, the highest cost change (in absolute terms) takes place when product j has the lowest priority before the interchange. Furthermore, given that j has the lowest priority, higher cost change occurs when product i has higher priority. This small numerical analysis strengthened the intuition that, if a product is likely to generate a higher cost than others, then it should be given a higher priority.

We can formulate the problem of finding a good priority order under identical arrival rates as a (deterministic) job scheduling problem. The cost that a product

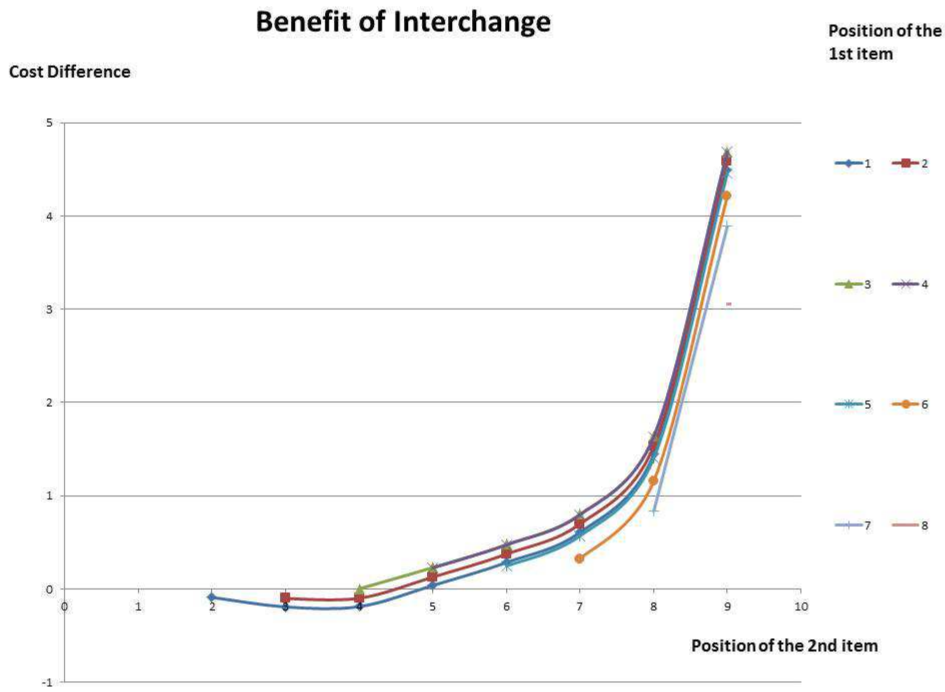


Figure 4.6: Cost of interchange $\lambda_i = \lambda_j = 0.1, h_i = h_j = 1, b_i = 5, b_j = 10$.

generates under a given priority (i.e., rank) is independent of the priority order of other products. However, the cost of each product changes with its assigned priority. Thus, if the cost values (TC_k) are regarded as the processing times, the job scheduling problem assumes the property that there do not exist any set-up times for production, but the position of each job affects its processing time. Furthermore, the later the job is scheduled, the higher its processing time. The objective is to find a sequence that minimize the makespan.

The proposed procedure for determining a priority order consists of the following steps.

Step 0. U is the set of products that are not assigned a priority, and P is the set of priority orders that are not assigned to a product. Initially $U = N$, $P = \{1, 2, \cdots, n\}$. Here, $p = 1$ denotes the highest priority.

Step 1. Select the lowest priority in P, call p. Note p also denotes the cardinality of P. The product that will be assigned to the p^{th} order will be determined.

Step 2. Calculate for all $k \in U$,

$$TC_k = \sum_{j=0}^{\infty} [h_k(S_k - j)^+ + b_k(S_k - j)^-] \frac{1}{\rho_k} p_{0,(j+1)}.$$

Here, TC_k is the cost incurred by product k when it has the priority order of p. In the equation, $\rho_k = \frac{\lambda_k}{\mu}$ and $p_{0,(j+1)}$'s are calculated using the same analysis under two product priority case, letting first product's arrival rate $\bar{\lambda}_1 = \sum_{l \in U \setminus \{k\}} \lambda_l$, and the second product's arrival rate $\bar{\lambda}_2 = \lambda_k$, and service rate μ. Note TC_k is calculated under optimal S_k.

Step 3. Find the product that has the lowest cost: $m = \arg\min_k TC_k$. Set order p as the priority order of product m.

The procedure has a complexity of $O(n^2)$ since, for the lowest priority position n comparisons are made, for the previous priority order $(n-1)$ comparisons are made, and so on.

For the non-identical arrival rate products, the independence property of TC_k from the priority order of other products does not hold anymore. Thus a modification is applied to the procedure.

Non-identical arrival rates

In this case, in Step 3 of the procedure (when the costs generated by the candidate products for a given priority order are compared) the comparisons will be flawed due to different arrival rates the products possess. The procedure is modified to mitigate the impact of arrival rate differences on the costs. Correction factor for product k (CF_k) is applied to cost TC_k to make the comparisons fairer.

Step 3 of the above procedure is modified as follows.

Step 3. Find the product that has the lowest *corrected* cost: $m = \arg\min_k \frac{TC_k}{CF_k}$. Set order p as the priority order of product m.

Three correction factors are considered.

1. (**ArC**) The first CF_k is set as λ_k. By this way the comparisons are made

over cost per customer for each product. However, this is a naive approach.

2. (**EsC**) Another approach is to normalize the cost by the expected number of shortfall for each product. Let x_k denote the random variable for the shortfall level of product k in steady state. The second CF_k is determined by $E[x_k] = \sum_{j=0}^{\infty} j \frac{1}{\rho_k} p_{0,(j+1)}$, where ρ_k and $p_{0,(j+1)}$ defined accordingly. For product k, the higher the arrival rate the higher the expected steady state shortfall. Assuming that the cost generated is somehow "proportional" to the expected shortfall, the correction factor enables to strip-out the impact of the differences in arrival rate on the cost. The cost comparison is in a way made based on the cost parameters, as in the identical arrival rate case.

3. (**StC**) The third correction factor considered normalizes the cost by the variance of the shortfall. This approach assumes that the arrival rate affects how the shortfall fluctuates over the expected value, and this fluctuation is the main cause of the cost difference. This approach assumes the expected shortfall is taken care of by adjusting the basestock level. Such an approach resembles the cost obtained under the newsvendor problem. Given that the only cost parameters are the overstocking and understocking cost, the total cost is due to the risk of overstocking and understocking. Expected shortfall does not have an affect on the cost, since it is taken care of by the order quantity. In the third approach CF_k is determined as $\sigma(x_k) = \sqrt{\sum_{j=0}^{\infty} (j - E[x_k])^2 \frac{1}{\rho_k} p_{0,(j+1)}}$.

In the following, the performance of the proposed procedures for determining the priority order are assessed. For the non-identical arrival rate setting, we denote the first approach with ArC, second approach with EsC and the third approach with StC.

4.6 Performance of the proposed procedures for determining the priority order (numerical results)

The performance of the proposed procedures are assessed through numerical analysis. For the numerical study, the number of products is taken as $n = 4$, and both identical and non-identical arrival rate cases are analyzed. Cost parameters are randomly generated for each product, $h_k \sim U[0.2, 1]$, $b_k \sim U[5, 20]$. Substitution costs are generated from U[5,20]. Substitution rate is assumed to be $\mu_=5$. Production rate is $\mu = 1$ for each product. The total number of instances generated is 100. For the identical arrival rate setting, $\lambda = 0.16$, and for the non-identical arrival rate setting $\sum_k \lambda_k = 0.64$. In this setting, for each instance, we generate a uniform random variable U_k for each product, such that $U_k \sim U[0, 1]$, and set $\lambda_k = 0.64\frac{U_k}{\sum_l^4 U_l}$.

4.6.1 Identical arrival rates

For identical arrivals case, we compare the performances of the priority order determined by the proposed procedure (Heur) with backorder-based priority order (Bo) and with the optimal priority order. In backorder-based priority order, the higher backorder product is given the higher priority. The optimal priority order is found by exhaustive search (by comparing all 24 orderings). The results are presented in Table 4.6. We also apply dynamic heuristic to priority policies obtained under Heur and Opt, and compare the improved costs.

When priority orders under Bo and Heur are compared, it is observed that the cost obtained under the priority order of Heur is considerably lower. In the median, the difference between the costs under the optimal order and heuristic order is zero. This means in more than % 50 of cases, the heuristic order finds the optimal priority order. On the average, cost under priority order of Heur policy is only 0.22% higher than that of the optimal priority order. Cost under priority order of Bo policy is on the other hand 6.94% higher than the cost under optimal order.

Note that, determining the optimal priority order not necessarily guarantees

Table 4.6: Comparison of heuristics for determining the priority order: identical λ

	% diff optimal order		% diff Prio-DH	
	Heur	Bo	Heur-DH	Bo-DH
average	0.22	6.94	0.46	4.75
maximum	2.17	39.22	10.59	33.10
minimum	0.00	0.00	-10.82	-8.82
median	0.00	5.24	0.00	2.75
10 percentile	0.00	0.00	-0.01	-1.56
25 percentile	0.00	1.07	0.00	0.00
75 percentile	0.16	10.16	0.00	8.77
90 percentile	0.97	15.67	4.98	13.06
stdev	0.47	7.09	3.04	7.09

the lowest cost after one-step of policy iteration. A non-optimal initial priority order might result in lower cost. When the cost performance of priority order obtained by Heur is compared with that of optimal order after policy iteration (in Table 4.6 the performance of Prio-DH obtained under optimal priority order is compared with that of Heur-DH and Bo-DH) on the average difference in costs is %0.46. Results in Table 4.6 indicate that cost under Heur-DH could be as low as 10.82% lower than Prio-DH, or could be as high as 10.59% higher. Under Heur-DH, for almost 75% of the instances, the cost is either lower or the same as the cost under Prio-DH. Even, for the Bo-DH, for 25% of the instances, the cost is either lower or the same as the cost under Prio-DH. However, on the average $Bo - DH$ results in 4.75% higher cost compared to Prio-DH.

4.6.2 Non-identical arrival rate

For the unequal arrival case, the heuristic policy is compared under three correction factors, arrival rate correction (ArC), expected shortfall correction (EsC), and deviation correction (StC). As a naive approach we also order the priority with respect to arrival rates only (Ar), in that the lower arrival rate product is given higher priority. Results are given in Table 4.7.

In this parameter set 100 randomly generated instances are considered. Holding cost of the products are generated from uniform distribution, U[0.2,1], and backorder costs are generated from U[5, 20]. Substitution costs are generated from U[5,20]. Substitution rate is assumed to be $\mu_{=}5$. Production rate is $\mu = 1$ for each product. Arrivals rates are generated as follows. Total rate is generated from U[0.2, 0.6]. To obtain the individual rates, four uniform random variables are generated, $U_i \sim U[0,1]$. To obtain λ_i, the total arrival rate (say Λ) is multiplied with $\frac{U_i}{\sum_i U_i}$.

The performances of the proposed procedures are listed in Table 4.7. The results show that StC gives priority orders close to the optimal priority order with a 3.56% cost difference between StC based priority order and the optimal priority order.The performances of Ar and ArC, are not very good. However, as a naive and a very easy ordering principle, arrival rate order work not so bad.

After policy iteration (DH) is applied, we see that optimal priority order and EsC perform almost the same. A hypothesis test is conducted for the paired differences of 100 instances (paired t-test). We test the hypothesis that "whether the mean of best Prio-DH is the same with other heuristics". The results of the test (presented in Table 4.8) indicate that at significance level (0.01), the null hypothesis is rejected for all heuristic results except StC.

Selecting the priority MTS-queue policy based on the best priority order is likely to improve the performance of Prio-DH. However, when finding the best priority ordering is computationally burdensome, StC based procedure could be used for determining the priority order.

Table 4.7: Comparison of heuristics for determining the priority order: non-identical λ

	% diff from Opt				% diff from Opt-DH			
	Ar	ARC	EsC	StC	Ar-dh	ArC-dh	EsC-dh	StC-dh
average	7.00%	7.09%	6.35%	3.56%	6.13%	6.11%	5.41%	2.75%
maximum	24.04%	24.04%	24.04%	13.84%	42.95%	42.95%	42.95%	34.79%
minimum	0.00%	0.00%	0.00%	0.00%	-14.17%	-14.17%	-14.17%	-14.17%
median	5.20%	5.20%	5.01%	2.70%	3.62%	3.70%	1.70%	0.35%
10 percentile	0.02%	0.73%	0.23%	0.00%	-4.75%	-5.45%	-4.75%	-7.91%
25 percentile	2.16%	2.16%	1.86%	0.90%	-0.23%	-0.35%	-0.50%	-0.59%
75 percentile	9.99%	9.99%	8.08%	5.30%	12.96%	12.96%	12.12%	6.81%
90 percentile	17.08%	17.08%	15.67%	7.98%	20.34%	20.21%	19.77%	15.77%

Table 4.8: T-values of comparisons

	Ar	ARC	EsC	StC
t-values	4.58	4.55	4.18	2.30
P-values	0.0000	0.0000	0.0001	0.0237

CHAPTER 5

POLICIES FOR THE SINGLE-ECHELON N-PRODUCT SYSTEM: OTHER HEURISTIC POLICIES

In this chapter five other heuristic policies will be introduced. We also present two benchmark policies that serve as an upper bound and a lower upper bound. The following policies will be described in this chapter:

1. Longest-Queue based Hedging policy with Service-Time Look Ahead (LQ-STLA)
2. Longest-Queue based Hedging policy with Service and Substitution Time Look Ahead (LQ-SSTLA)
3. Longest-Queue based Hedging policy with Busy-Time Look Ahead (LQ-BTLA)
4. Service Time Look Ahead with Substitution based on Inventory Level (SILEVEL)
5. Service Time Look Ahead with Substitution based on Inventory Cost (SICOST)

Benchmark policies:

1. First-Come-First-Served Policy (Upper Bound)
2. Policy under immediate substitution with no cost (Lower Bound)

5.1 Longest-Queue based Hedging policy with Service-Time Look Ahead (LQ-STLA)

This heuristic is adopted from Veatch and Wein (1996) and modified to incorporate the substitution decisions. It has two parts, the first part discusses identifying a hedging point (idling levels, or basestock levels), and the second part discusses production scheduling and substitution decisions.

Identifying the iding point

The proposed method for identifying the hedging point is originally suggested by Zipkin (1995). For the sake of completeness, the steps (as presented in Veatch and Wein (1996)) are listed here:

1. Assume that an LQ curve is optimal for production scheduling. An LQ curve is a policy that implies upon a production completion, the next production will be done for the item with highest shortfall.
2. Approximate the steady-state shortfall levels mean and variance.
3. Adjust for non-identical products.
4. Fit a geometric distribution to the shortfall level (with mean and variance approximated)
5. Treat each product independently for determining the basestock level.

In Step 2, the variance for steady-state shortfall level for product k is approximated by

$$\sigma_k^2 = \frac{\rho/(1-\rho)^2 + (K-1)\rho[1+\rho+(1-2\alpha_k)\rho^2 + (1-2\alpha_k)^2\rho^3]}{K^2}$$

where $\rho = \sum_k \rho_k = \sum_k \lambda_k/\mu$ and $\alpha_k = \rho_k/\rho$. Expected value of steady-state shortfall level, x_k is approximated by

$$E[x_k] = \rho_k/(1-\rho)$$

In Step 4, to fit a distribution to the steady state shortfall level of product k, a geometric distribution shifted a units to right is used. The reason is demand distribution in an $M/M/1$ queue is geometric, and in the n-product problem under LQ production scheme, shortfall distribution of each product can be approximated similarly. The probability mass function is, $f_x(x) = pq^{x-a-1},\ x = a+1, a+2, \ldots$ with mean $E[x_k]$ and variance $Var(x_k)$. So

$$\begin{aligned} q &= 1 - \frac{\sqrt{4\sigma_k^2 + 1} - 1}{2\sigma_k^2} \\ a &= E[x_k] - \frac{1}{1-q} \end{aligned}$$

This distribution is an input to the product k problem with threshold control. So optimal base-stock for x_k is

$$S_k = \lfloor ln[h_k/(h_k + b_k)/ln[q]] \rfloor + a$$

Identifying the production scheduling and substitution policy

Under the optimal policy (determined by the Markov decision process), when a scheduling (or substitution) decision is taken, the aim is to minimize over k: $\mu(v(x + e_k) - v(x)) = \mu\Delta_k v(x)$, which is the "expected cost change when producing product k". The bias contribution of a given decision at a given state can be approximated myopically by the expected cost rate of the next state, $c(x)$, or can be approximated with a look ahead by the expected cost rate after one service time neglecting the substitutions that occur. This is equivalent to redefining the stage by the time between two production completions.

When the stage is redefined as such, the next state upon production completion is not necessarily $x + e_k$, since all demand arrivals during the stage must be considered. Product k's contribution to bias $v_k(x)$ is approximated by,

$$v_k(x) \approx E[c_k(x_k + D_k(T))] \tag{5.1.1}$$

where x_k is the shortfall level, $c_k(x_k) = h_k(S_k - x_k)^+ + b_k(S_k - x_k)^-$, T is the service time (a random variable) and $D_k(t)$ is the number of demand for product k that arrive in time [0,t].

Let $F_k(x)$ denote the distribution function of $D_k(S)$. Note $D_k(S)$ is a geometric random variable with $p = 1 - \frac{\rho_k}{1+\rho_k}$. Letting $q = 1 - p$, $F_k(x) = q^x p$ and furthermore,

$$v_k(x+1) - v_k(x) \approx E[c_k(x_k + 1 + D_k(S))] \quad (5.1.2)$$

$$= -b_k(1 - F_k(x)) + h_k F_k(x) \quad (5.1.3)$$

$$= -bq^{x+1} + h(1 - q^{x+1}), x \geq 0 \quad (5.1.4)$$

For $x < 0$

$$v_k(x+1) - v_k(x) \approx E[c_k(x_k + 1 + D_k(S))] - E[c_k(x_k + D_k(S))]$$

$$= -b_k$$

So we are able to calculate value function approximation.Note $v(x+e_i)-v(x) = v_i(x_i+1) - v_i(x_i)$.Under the look ahead approximation, production order at state x is determined considering $\arg\min_{k\in\{1,..,K\}}(v(x+e_k)-v(x))$. Substitution control for product k inventory is $\arg\min_{\{i:x_i>S_i\}} \mu_S(v(x - e_k + e_i) - v(x))$.

Once the policy is determined, then total cost per unit time is obtained by obtaining the steady state joint distribution from flow balance equations, and then using this steady state probabilities to calculate total-cost.

5.2 Longest-Queue based Hedging policy with Service and Substitution Time Look Ahead (LQ-SSTLA)

The same approach as in the previous section will be adopted. Under service and substitution time look ahead version, the expected shortfall level is estimated after the minimum of service and substitution time. This is equivalent to setting the parameter q of the geometric distribution in Eq. 5.1.4 as $\frac{\lambda_k}{\lambda_k+\mu+k\mu_S}$.

5.3 Longest-Queue based Hedging policy with Busy Time Look Ahead (LQ-BTLA)

The same approach as in the previous section will be adopted. However, now the look-ahead time duration is extended to the busy time. This implies, the expected shortfall level is estimated after the busy time for that product. This is equivalent to setting the parameter q of the geometric distribution in Eq. 5.1.4 as $\frac{\sum_k \lambda_k}{\mu}$.

5.4 Longest-Queue based Hedging policy with Substitutions based on Inventory Cost (SICOST)

In this policy, base-stock levels are calculated as in the LQ-STLA policy. For production scheduling, if one or more products are backordered, then the product with the highest backorder cost is produced. If all the products have non-negative inventory levels, then the product with lowest holding cost is produced (up to its base-stock level). Among the backordered products, the product with highest backorder cost is substituted by a product which has a positive inventory. Among those products that have positive inventory, the one with highest inventory holding cost is selected.

5.5 Longest-Queue based Hedging policy with Substitutions based on Inventory Level (SILEVEL)

This policy is the same policy as SICOST except for the substitution decisions. Among the backordered products, the product with highest backorder cost is substituted by a product which has a positive inventory. Among those products that have positive inventory, the one with highest inventory level is selected.

5.6 Benchmark policies

We propose two benchmark policies (an upper bound and a lower bound) to obtain an understanding of the performance of the heuristic policies. For the benchmark policies the cost under each can be expressed in closed-form. Thus dimensionality does not affect the computational time.

5.6.1 FCFS policy (Upper bound)

We can use any feasible policy as an upper bound on the optimal multiproduct inventory-backorder model. To determine an upper bound, we assume no substitutions (or the substitution cost is infinity) and FCFS production scheduling. In Buzacott and Shanthikumar (1993), the authors provide the results for $M/M/1$ multiclass backlogged demand FCFS model. Customers of different types arrive according to mutually independent poisson processes with rate λ_i for product i. There is a single production facility, that operates with i.i.d. exponential production times with mean $1/\mu$. Here for each type i customer arrival, a request for i type product is placed. These requests are produced in the server in time order. Let $X_i(t)$ be the number of type i jobs at time t (i.e., shortfall level for product i), in the system. Let $\mathbf{X}(t) = (X_i(t), i = 1, ..., r)$. In this model, $N(t) = \sum_k X_k(t)$, the total shortfall from base-stock follows a Markov process, it has a stationary distribution which can be obtained from the analysis of an $M/M/1$ queue with arrival rate $\sum_k \lambda_k$ and service rate μ. That is, given z denotes the total shortfall amount,

$$p(z) = (1-\rho)\rho^z, \quad z = 0, 1, 2, \ldots.$$

Because the arrival processes of the jobs are mutually independent and Poisson, the stationary joint probability distribution $p(\mathbf{x}) = \lim_{t\to\infty} P(\mathbf{X}(t) = \mathbf{x})$ can be obtained from $\mathbf{p}$ by a multinomial thinning with probabilities $(\lambda_i/\lambda), i = 1, 2, .., n$.

$$\begin{aligned} P(X_1 = x_1, X_2 = x_2, .., X_n = x_n | N = z) &= \binom{z}{x_1, x_2, ..., x_n} \Pi_{i=1}^n (\frac{\lambda_i}{\lambda})^{x_i}, \\ z &= \sum_{i=1}^n x_i, \end{aligned}$$

then

$$p(\mathbf{x}) = \binom{z}{x_1, x_2, ..., x_n} (1-\rho) \Pi_{i=1}^n \rho_i^{x_i}$$

and the marginal distribution of X_i, $P(X_i = x_i) = p_i(x_i),\ i = 1, 2 \ldots, n$ is

$$p_i(x_i) = (1-\bar{\rho}_i)\bar{\rho}_i^{x_i}, \quad x_i \in \mathbb{Z}^+$$

where

$$\bar{\rho}_i = \lambda_i/(\mu - \sum_{j \neq i} \lambda_j)$$

Given this marginal distribution, the problem can be treated as if it is a single product problem with threshold control. So optimal base-stock for x_k is

$$S_k = \lfloor ln[h_k/(h_k + b_k)/ln[\bar{\rho}_k]] \rfloor$$

Then total cost for optimal FCFS policy is

$$TC_{FCFS} = \sum_{i=1}^{n} \sum_{x=1}^{\infty} c_i(x) p_i(x)$$

where $c_i(x_i) = h_i(S_i - x_i)^+ + b_i(S_i - x_i)^-,\ i = 1, 2, \ldots, n.$

5.6.2 Policy with immediate substitution and no substitution cost (Lower bound)

A lower bound for the problem can be obtained under the assumption that substitutions take no time and no cost. In this case the optimal operating policy is to hold inventory only on the product with lowest holding cost, and to substitute other products through the inventory of the lowest cost product. Let the products be ordered with respect to their backordering costs $b_1 > b_2 > \ldots > b_n$. Let the product with lowest holding cost be denoted with index ℓ. When demand arrives for products other than ℓ, the decision to be made is whether to meet the demand from stock through substitution or to backorder. If demand arrives for product ℓ, then the decision to be made is to meet the demand from the stock if stock is available, or otherwise to backorder. Note that since substitutions take no time or cost, it is as if there exists a single generic product demanded by n customer classes, where each class i is differentiated with respect to its arrival rate λ_i and backorder cost b_i. This problem is very similar to the problem studied earlier in the literature by Ha (1997a) and De Vericourt et al. (2002). The description of the earlier problem is as follows. There are n demand classes, and these classes are differentiated by their backordering cost and arrival rates. There exists a single product demanded by all customer classes, and a single inventory is kept for which there is a holding cost of h.

The demand for product i follows an independent Poisson process with rate λ_i. When demand arrives it is either met immediately from stock, or backordered (this may happen even if stock exists). The problem is to determine the rationing levels for each demand class. De Vericourt et al. (2002) obtains the following closed-form expression for cost per unit time.

Let z_k correspond to the inventory rationing threshold for customer class k. Let g_k denote the average cost of the system under the assumption that only products with indices $1, 2, \ldots, k$ are present. Let ρ_k denote the cumulative traffic intensity $\rho_k = \frac{\sum_i^k \lambda_i}{\mu}$. Initialize $z_1 = g_1 = \rho_0 = b_{n+1} = 0$. Construct the sequences z_k, g_k, and ρ_k as follows. For $k = 1, 2 \ldots, n$:

$$
\begin{aligned}
\rho_k &= \rho_{k-1} + \frac{\lambda_k}{\mu}, \\
z_{k+1} &= z_k + \lfloor \frac{\ln \frac{\rho_k (h + b_{k+1})}{\rho_k (h + b_k) + (1 - \rho_k)(g_k - (h + b_k) z_k)}}{\rho_k} \rfloor, \\
g_{k+1} &= (z_{k+1} - \frac{\rho_k}{1 - \rho_k})(h + b_{k+1}) + (g_k - (z_k - \frac{\rho_k}{1 - \rho_k})(h + b_k))\rho_k^{z_{k+1} - z_k}.
\end{aligned}
$$

Stop when the optimal rationing vector $z = (z_1, ..., z_K)$ and $g^* = g_{n+1}$ are obtained.

The inventory rationing levels z_k in this problem corresponds to the substitution threshold level in our problem. As an analog to our problem, we obtain z and g^* by taking $h = \min_{k=1,..n} h_k$. The problem studied and our problem differ in that, in our problem threshold level for product ℓ must be zero. On the other hand in the problem above only the product with highest unit backorder cost has a threshold level of zero. This difference in definition leads to the cost obtained under the inventory rationing problem to be a lower bound for our problem. Furthermore, the inventory rationing problem is optimal under the assumption of zero substitution cost and immediate substitutions. This further increases the gap between the cost values of the inventory rationing problem and the substitution problem. The optimal cost described above will be lower than the cost under the original problem. Thus a lower bound is obtained.

CHAPTER 6

COMPARISON OF THE PERFORMANCES OF THE HEURISTICS VIA NUMERICAL ANALYSIS

To assess the performance of the heuristics, we make comparisons under various parameter settings. The comparisons are made for the settings with $n = 2$, $n = 3$, and $n = 4$ products. For $n = 2$ we are able to compare the performance of the heuristics with that of the optimal policy. For $n = 3$ and $n = 4$, due to the curse of dimensionality to obtain the optimal policy is difficult. Even evaluation of a policy takes considerable amount of time. For these settings we compare the heuristics with each other only. We also look at the performances under various substitution schemes. One expects that a heuristic does not outperform others under all possible instances. The aim is to identify the settings under which each heuristic performs well.

6.1 Determining the state space for policy evaluation

To evaluate a policy, the corresponding transition rate matrix is constructed. The state space under consideration has n-dimensions where n is the number of products. The iterative method of successive overrelaxation is used for solving the flow balance equations Tijms (1994).

When determining the size of the state space the aim is come up with a size where small changes in size does not affect the expected average gain. For our problem each policy class has its own state space requirements. In priority MTS-queue based policies, assigning the same state space size to all dimensions

(products) is not meaningful, since the probability that there exists x or more backorders for product k considerably varies with the priority of the product. This probability also changes with the traffic intensity. Thus, we do not use a constant truncation point for the state space. We dynamically assign state space truncation points for each product based on the parameters.

The marginal probabilities under priority MTS-queue can be calculated up to the desired precision. These marginal probabilities are used for (dynamic) state space truncation for priority MTS-queue based heuristics. The procedure to determine the state space truncation is as follows. A precision level of ϵ is selected. For each product, the state that captures the cost under the priority MTS-queue with respect to the given precision level ϵ is calculated. Let the truncation shortfall point for product k be denoted with $n_k(\epsilon)$ where $n_k(\epsilon)$ satisfies:

$$\begin{aligned} TC_k^{n_k} &= \sum_{j=0}^{n_k}[h_k(S_k-j)^+ + b_k(S_k-j)^-]\frac{1}{\rho_k}p^k_{0,(j+1)} \\ n_k(\epsilon) &= \liminf\{n_k : \frac{TC-TC_k^{n_k}}{TC} < \epsilon\} \end{aligned}$$

For the priority MTS-queue policy, truncation based on the above calculation gives a cost figure in desired precision. For priority based heuristics there is no guarantee that truncation points determined as above will give the desired precision, however we adopt this procedure. We took several runs until the desired precision level is obtained.

For the LQ-based policies, the truncation levels are calculated similar to those of priority based policies. We do not have the exact joint or the marginal probabilities, but we have a geometric approximation given in Veatch and Wein (1996) which is used for hedging point determination. The truncation shortfall point for product k is set as $n_k(\epsilon)$ where $n_k(\epsilon)$ satisfies:

$$\begin{aligned} TC_k &= \sum_{j=a+1}^{\infty}[h_k(S_k-j)^+ + b_k(S_k-j)^-]\frac{1}{p}q^{j-a-1} \\ TC_k^n &= \sum_{j=a+1}^{n}[h_k(S_k-j)^+ + b_k(S_k-j)^-]\frac{1}{p}q^{j-a-1} \\ n(\epsilon) &= \liminf\{n : \frac{TC-TC_k^n}{TC} < \epsilon\} \end{aligned}$$

Where p and a values are as determined in § 5. This gives a rough approximation of the truncation level for LQ-based policies. These are used as relative truncation levels for each product, then we increase the truncation levels by some safety level, until we get a desired accuracy.

6.2 Notation for the Heuristics Compared

In this chapter the following notation is used to denote the heuristics.

Prio: Prio denotes the policy under the priority order that gives the minimum cost, before any one-step improvement is applied. In other words Prio is the policy evaluated under best priority order among the priority policies.

Prio-DH : The dynamic policy obtained after one-step improvement algorithm. In Prio-DH policy Prio policy under optimal ordering is taken as the base policy to apply the improvement. Note Prio-DH always gives lower cost than Prio.

Prio-DH-no-subs: While determining the dynamic policy after one-step improvement algorithm, substitution decisions are not allowed, only production scheduling decisions are allowed. Prio is used as the base policy. Note Prio-DH-no-subs always results in higher cost than Prio-DH. In this chapter Prio-DH denotes the policy under two-way substitution.

Prio-DH-one-way: While determining the dynamic policy after one-step improvement algorithm, substitution decisions are determined such that only the higher priority product can substitute a lower priority product. Prio is used as the base policy. Note Prio-DH-one-way results in higher cost than Prio-DH, lower cost than Prio-DH-no-subs.

Prio-DH-two-way: The Prio-DH policy.

LQ-STLA: This heuristic is adopted from Veatch and Wein (1996). It has been modified to incorporate the substitution decisions. It is a myopic policy considering only a single service time look-ahead.

LQ-STLA-no-subs: Defined similar to Prio-DH-no-subs, adopted to LQ-STLA. Note LQ-STLA-no-subs always results in higher cost than LQ-STLA.

LQ-STLA-one-way: Defined similar to Prio-DH-one-way, adopted to LQ-STLA. Note LQ-STLA-one-way results in higher cost than LQ-STLA, lower cost than LQ-STLA-no-subs.

LQ-STLA-two-way: The LQ-STLA policy.

LQ-SSTLA: This policy is similar to LQ-STLA, except that instead of service-time look ahead, the look ahead duration is the minimum of service and substitution time. Thus, the policy is more myopic compared to LQ-STLA. Hedging points are the same with LQ-STLA.

LQ-BTLA: This policy is similar to LQ-STLA, except that instead of service-time look ahead, the look ahead duration is extended to the busy time. Hedging points are the same with LQ-STLA.

SICOST: In this policy, production is done on the product with the highest backorder cost from the products that are backordered. If all the products have non-negative inventory levels, then the product with lowest holding cost is produced (up to its base-stock level). Hedging points are the same as LQ-STLA. Substitutions are done from the highest inventory cost product to the product with highest backorder cost among the backordered products.

SILEVEL: This policy is the same as SICOST except for the substitution decisions. In this policy, production is done on the product with the highest backorder cost from the products that are backordered. If all the products have non-negative inventory levels, then the product with lowest holding cost is produced (up to its base-stock level). Hedging points are the same as LQ-STLA. Substitutions are done from highest positive inventory level product to the product with highest backorder cost among the backordered products.

Upper Bound (UB): The FCFS policy with multiple product queues under no substitution is considered.

Lower Bound (LB): It is assumed there are no substitution costs and substitutions take no time. A single inventory is kept on the lowest unit holding cost product, substitutions are triggered by the rationing levels.

6.3 Summary of the Findings

In this chapter we determine Prio-DH based on the lowest cost Prio policy (we do not use the ordering heuristics). We test different parameters: holding, backorder and substitution cost tuples in order to identify the settings under which each heuristic performs well. Three sets of parameters are considered: Set 1, Set 2 and Set 3. Parameter Set 1 is a comprehensive random set which cover 100 instances. Parameter Set 2 is a group of select instances that helps understand the impact of a single parameter on the performance. Parameter Set 3 is a parameter set resembles the set considered by Perez and Zipkin (1997) where products are hierarchical, they can be ordered, with respect to both holding costs and backorder costs.

The following are the findings of the numerical study/

1. We test parameter Set 1 under light, medium-low and medium-high arrival rates. We would like to note that because of the state space limitation, heavy traffic results possibly may not reflect a true performance comparison of the heuristics. In Set 1, we observe in very light traffic there is not much discrepancy between different Prio-DH, LQ-STLA, SICOST and SILEVEL heuristics, however when we increase the arrival rate a bit, still in light traffic, we observe Prio-DH performs better than LQ-STLA in many of the instances. As we increase the traffic to medium-low and medium-high, there is a deterioration of performance for Prio-DH compared to LQ-STLA. In the specific instances that favor Prio-DH (i.e when products are hierarchical), it is observed that under high traffic, as the arrival rate increases, the performance of Prio-DH over LQ-STLA may increase.

2. In Set 2, it is observed that as the dispersion of the parameters increases (i.e., as the ranking among the products gets more apparent) so does the performance of Prio-DH.

3. In Set 3, our observations support that, in more hierarchical product sets, Prio-DH performs better, while for when products are symmetrical, LQ-

STLA excels. We observe that, as products become more differentiated, the substitution benefit for all policies increases. As in Set-1, in Set-3 we observe above very light traffic for arrival rate, there is a region where Prio-DH works extremely well which is explained with its resemblance to Lower Bound(LB) policy, for these parameter sets.

4. Considering Set 1, Set 2 and Set 3, in the summary, the priority heuristic (Prio-DH) performs well compared to LQ class of policies when:

   (a) The traffic intensity is low ($\sum \lambda_i < 0.2$),

   (b) The holding and backordering cost structure apparently implies a ranking among the items (e.g. $h_i > h_j$ and $b_i > b_j$) and traffic intensity is high ($\sum \lambda_i > 0.8$),

   (c) Backordering costs are identical but holding costs are not,

   (d) Finally, when unit holding and backordering costs are ordered, under low or medium traffic intensity, for every arrival rate considered there exists a ratio $\frac{h}{b}$ under which Prio-DH significantly outperforms other policies. Those are the cases where basestock levels are low, and basestock level for the lowest holding cost product serves as the inventory reservoir.

   LQ-based heuristics (LQ-STLA, LQ-SSTLA, SICOST, SILEVEL) outperform Prio-DH when:

   (a) Traffic intensity is not low or not very high,

   (b) Items are close to each other in terms of the cost parameters.

   (c) Unit holding costs are the same whereas unit backordering costs are different.

5. A major finding is that under a policy, how the basestock levels are set is a dominating factor in the performance of the policy.

6. Under very light traffic in the case of symmetric parameters, LQ based policies do not necessarily perform better than Prio-DH.

7. Under light traffic, the value of substitution is higher than under high traffic for both LQ-STLA and Prio-DH. Interesting observation is when a parameter set promotes Prio-DH, addition of substitution makes Prio-DH work even better. Prio-DH benefit from substitutions more than LQ-based policies. For higher traffic, if a heuristic is performing well, it is due to the success in scheduling rather than substitution decisions.

8. When we compare the performances of LQ-based policies, LQ-STLA and LQ-BTLA performs close to each other. On the other hand, the more myopic LQ-SSTLA may or may not outperform LQ-STLA.

9. SICOST and SILEVEL perform close to each other. They are using the same scheduling decision, they differ for substitution decisions. As noted before, scheduling decisions are more important for a policy performance. There are instances that these heuristics performs better than all other heuristics, which is interesting when the simplicity of the scheduling rules are thought.

10. Upper bound proximity tells us whether scheduling and substitution decisions are helpful compared to completely unscheduled case. If the performance is close to Lower bound, we can deduce, it is close to optimal as well. Under light traffic, or when holding cost for all products are the same, there are instances where the heuristic policy works very close to LB, which suggests, in these instances, the heuristic policy can be very close to optimal solution.

11. Our best policy's performance which is interpreted as being closer to Lower bound policy result rather than upper bound, seems to depend on whether there is a structure to parameters, such as hierarchy or symmetry. In our complete random parameter set, Set 1, our best policy covers only %32 of UB-LB gap, however for our structured data Set 3, it covers %55 percent of the gap.

6.4 Analysis of the numerical results

In the following, we present our results for the 2-product and 4-product settings. Our aim is to identify the parameters under which each heuristic performs well (significantly outperforms the other) especially under $n-$product settings. We observe findings under the 3-product setting are similar (we present the discussion on 3-product setting in Appendix D), and thus present only the results for 4-product setting.

6.4.1 Analysis under 2-product setting : comparison with the optimal policy

We analyze the performance of the heuristics under $c_S = \{1, 5\}$, and $(\lambda_1, \lambda_2) = \{(0.2, 0.6), (0.4, 0.4)\}$. For Product 1 and Product 2, out of all combinations of holding cost $h_k \in \{0.2, 0.05\}$ and $b_k \in \{2, 0.5\}$ for $k = 1, 2$, we present the result for instances that are realistic. Specifically,

1- All instances with $h_1 > h_2$ and $b_1 < b_2$ are excluded.

2- For any two products, we assume often a more valuable product face less frequent demand, and its arrival rate is lower. This implies, among the instances considered if $h_1 > h_2$ or if $b_1 > b_2$ then $\lambda_1 < \lambda_2$.

Such considerations are aligned with the instances considered in Veatch and Wein (1996) and Perez and Zipkin (1997). We consider a total of 46 instances. Results are obtained under a substitution rate of $\mu_{ij} = \mu_s = 5$, and for production rate of $\mu = 1$. For the state space of the optimal policy, a dimension of 50×50 is used, and the value iteration algorithm for MDP stops after (at most) 1500 iterations.

In the tables, under the column named "Diff" we present the % difference between the optimal policy and the best performing heuristic policy. Results show that for many of the instances heuristic policies perform close to the optimal policy. Performances deteriorate under lower substitution cost. Under non-identical arrival rate performance of the heuristic policies are slightly better than

Table 6.1: Performance of Heuristics c=1 identical λ

h1	h2	b1	b2	Cost							% Diff. Best Heur. vs. Opt	Base-stock		
				OPT	Prio-DH	LQ-STLA	LQ-SSTLA	LQ-BTLA	SICOST	SILEVEL		OPT	Prio-DH	LQ-BASED
0.2	0.2	2	2	2.16	2.32	2.17	2.17	2.17	2.23	2.23	0.55%	(5,5)	(2,9)	(5,5)
0.2	0.2	2	0.5	1.27	1.45	1.42	1.42	1.40	1.46	1.46	10.27%	(3,3)	(2,4)	(5,2)
0.2	0.2	0.5	0.5	1.15	1.19	1.18	1.18	1.18	1.21	1.21	3.13%	(3,2)	(1,4)	(2,2)
0.2	0.05	2	2	1.07	1.20	1.37	1.41	1.38	1.55	1.55	12.29%	(1,15)	(2,15)	(5,8)
0.2	0.05	2	0.5	0.80	0.94	1.09	1.12	1.08	1.27	1.27	16.86%	(1,10)	(2,9)	(5,5)
0.2	0.05	0.5	0.5	0.71	0.72	0.80	0.82	0.80	0.85	0.85	2.08%	(1,9)	(1,9)	(2,5)
0.05	0.05	2	2	0.83	0.94	0.84	0.84	0.84	0.88	0.88	0.66%	(8,8)	(4,15)	(8,8)
0.05	0.05	2	0.5	0.58	0.65	0.62	0.63	0.62	0.67	0.67	6.46%	(6,6)	(4,9)	(8,5)
0.05	0.05	0.5	0.5	0.55	0.59	0.55	0.55	0.55	0.60	0.60	0.65%	(5,5)	(2,9)	(5,5)

Table 6.2: Performance of Heuristics c=1 non-identical λ

h1	h2	b1	b2	Cost							% Diff. Best Heur. vs. Opt	Base-stock		
				OPT	Prio-DH	LQ-STLA	LQ-SSTLA	LQ-BTLA	SICOST	SILEVEL		OPT	Prio-DH	LQ-BASED
0.2	0.2	2	2	2.15	2.21	2.17	2.18	2.18	2.22	2.22	0.99%	(3,7)	(1,10)	(3,8)
0.2	0.2	2	0.5	1.22	1.27	1.30	1.31	1.30	1.32	1.32	4.24%	(2,4)	(1,5)	(3,4)
0.2	0.2	0.5	0.5	1.14	1.21	1.15	1.15	1.15	1.18	1.18	0.78%	(1,4)	(0,5)	(1,4)
0.2	0.05	2	2	0.99	1.00	1.13	1.15	1.13	1.26	1.26	1.02%	(1,15)	(1,16)	(3,12)
0.2	0.05	2	0.5	0.71	0.73	0.86	0.87	0.85	0.98	0.98	2.57%	(1,10)	(1,10)	(3,8)
0.2	0.05	0.5	0.5	0.64	0.65	0.66	0.66	0.66	0.70	0.70	0.45%	(0,10)	(0,10)	(1,8)
0.05	0.05	2	2	0.83	0.87	0.84	0.84	0.84	0.89	0.89	0.72%	(5,11)	(2,16)	(4,12)
0.05	0.05	2	0.5	0.57	0.59	0.59	0.59	0.59	0.62	0.62	3.43%	(3,8)	(2,10)	(4,8)
0.05	0.05	0.5	0.5	0.54	0.56	0.55	0.55	0.55	0.59	0.59	1.27%	(3,7)	(1,10)	(3,8)

Table 6.3: Performance of Heuristics c=5 identical λ

h1	h2	b1	b2	Cost							% Diff. Best Heur. vs. Opt	Base-stock		
				OPT	Prio-DH	LQ-STLA	LQ-SSTLA	LQ-BTLA	SICOST	SILEVEL		OPT	Prio-DH	LQ-BASED
0.2	0.2	2	2	2.18	2.39	2.20	2.20	2.20	2.44	2.44	0.57%	(5,5)	(2,9)	(5,5)
0.2	0.2	2	0.5	1.38	1.48	1.44	1.44	1.41	1.64	1.64	2.00%	(4,2)	(2,4)	(5,2)
0.2	0.2	0.5	0.5	1.15	1.20	1.18	1.18	1.18	1.45	1.45	2.68%	(3,2)	(1,4)	(2,2)
0.2	0.05	2	2	1.20	1.23	1.42	1.57	1.42	1.66	1.66	2.71%	(2,14)	(2,15)	(5,8)
0.2	0.05	2	0.5	0.95	0.96	1.13	1.25	1.11	1.48	1.48	1.84%	(2,9)	(2,9)	(5,5)
0.2	0.05	0.5	0.5	0.73	0.74	0.80	0.82	0.80	1.03	1.03	0.34%	(1,9)	(1,9)	(2,5)
0.05	0.05	2	2	0.84	0.96	0.85	0.85	0.85	1.02	1.02	0.67%	(8,8)	(4,15)	(8,8)
0.05	0.05	2	0.5	0.61	0.66	0.62	0.64	0.62	0.78	0.78	2.21%	(7,5)	(4,9)	(8,5)
0.05	0.05	0.5	0.5	0.55	0.60	0.55	0.55	0.55	0.81	0.81	0.38%	(5,5)	(2,9)	(5,5)

Table 6.4: Performance of Heuristics c=5 identical λ

h1	h2	b1	b2	Cost							% Diff. Best Heur. vs. Opt	Base-stock		
				OPT	Prio-DH	LQ-STLA	LQ-SSTLA	LQ-BTLA	SICOST	SILEVEL		OPT	Prio-DH	LQ-BASED
0.2	0.2	2	2	2.18	2.26	2.20	2.21	2.20	2.39	2.39	1.16%	(3,7)	(1,10)	(3,8)
0.2	0.2	2	0.5	1.27	1.30	1.31	1.31	1.31	1.46	1.46	2.66%	(2,4)	(1,5)	(3,4)
0.2	0.2	0.5	0.5	1.15	1.22	1.15	1.15	1.15	1.40	1.40	0.32%	(1,4)	(0,5)	(1,4)
0.2	0.05	2	2	1.04	1.05	1.15	1.22	1.16	1.32	1.32	0.97%	(1,15)	(1,16)	(3,12)
0.2	0.05	2	0.5	0.77	0.77	0.87	0.93	0.87	1.12	1.12	0.36%	(1,10)	(1,10)	(3,8)
0.2	0.05	0.5	0.5	0.65	0.66	0.66	0.66	0.66	0.84	0.84	0.17%	(0,10)	(0,10)	(1,8)
0.05	0.05	2	2	0.84	0.88	0.85	0.85	0.85	1.03	1.03	1.10%	(4,12)	(2,16)	(4,12)
0.05	0.05	2	0.5	0.58	0.60	0.59	0.59	0.59	0.75	0.75	1.12%	(4,7)	(2,10)	(4,8)
0.05	0.05	0.5	0.5	0.55	0.57	0.55	0.55	0.55	0.75	0.75	0.82%	(3,7)	(1,10)	(3,8)

the performance under identical arrival rates. For identical and non-identical arrival rates the % difference between the optimal policy and the best performing heuristic policy changes between $0.23\% - 12\%$, and $0.17\% - 10.42\%$, respectively. For the 46 instances, under identical arrival rate, performance of the heuristic policy compared to the optimal policy under $c_S = 1$ and $c_S = 5$ are 4.45% and 1.17%, on the average, respectively. Under non-identical arrival rate, performance of the heuristic policy compared to the optimal policy under $c_S = 1$ and $c_S = 5$ are 1.72% and 0.96%, respectively.

6.4.2 Analysis under 4-product setting

6.4.2.1 Parameter Set 1

In this parameter set 100 randomly generated instances are considered. Holding cost of the products are generated from uniform distribution, U[0.2,1], and backorder costs are generated from U[5, 20]. Substitution costs are generated from U[5,20]. Substitution rate is assumed to be $\mu_S = 5$. Production rate is $\mu = 1$ for each product.

Arrivals rates are generated as follows. Three level of traffic intensities are considered: light, medium-low and medium-high. For light traffic, total arrival rate is obtained from U[0,0.2], for medium-low traffic total rate is generated from U[0.2, 0.6], and for medium-high traffic intensity total rate is generated form U[0.6,0.8]. From the total arrival rate, individual rates are obtained as follows. Four uniform random variables are generated, $U_i \sim U[0,1]$. To obtain λ_i, the total arrival rate (say Λ) is multiplied with $\frac{U_i}{\sum_i U_i}$.

The details of random instance generation is as follows. For each instance $C(n = 4, 2)$ many substitution cost values, $n = 4$ many holding cost and backorder cost values are generated. To determine the total arrival rate, a random number R, distributed U[0,1], is generated. Under light traffic, medium-low traffic, and medium-high traffic, this number indicates a total arrival rate of $(R)0.2$, $0.2 + (R)(0.6 - 0.2)$, and $0.6 + (R)(0.8 - 0.6)$, respectively. Then using the random number U_i, arrival rates for each product under light, medium, and heavy traffic are obtained. Note that for a given instance, the same random number U_i is used for all traffic intensities.

The performance of the heuristics for each traffic intensity are presented in Table 6.5-Table 6.9.

Set 1- Light traffic ($\sum \lambda_k \in [0, 0.2]$)

In the Table 6.5, the average, standard deviation and percentiles of the 100 instances are given. The percentile information is provided in order to give information about how the performance is distributed. When total arrival rate is between $[0, 0.2]$ it is observed that Prio-DH performs better than LQ-STLA around %50 percent of time, and in 95% of the instances it works as good as LQ-STLA. For light traffic, very low basestocks are kept, in many cases maximum basestock level is 1 for any product. Products are backordered most of the time, and whenever there are backorders, the (conjectured) optimal policy is to produce the product with highest unit backorder cost product. Here, the priority policy resembles the (conjectured) optimal policy. The other heuristics, LQ-SSTLA performs close to LQ-STLA, and SICOST and SILEVEL perform 6% worse than LQ-STLA on the average. SICOST and SILEVEL heuristics never outperform LQ-STLA.

We compare the performance of the policies through statistical tests. It is observed that cost values under each policy come from Normal distribution. We conducted a paired t-test on mean of the difference, to compare the performances. Results show at 0.05 significance the performances are as follows:

$$Prio{-}DH \prec LQ{-}STLA \equiv LQ{-}BTLA \equiv LQ{-}SSTLA \prec SICOST \prec SILEVEL$$

Finally, comparing with the UB and LB, it is observed that the difference between UB and LQ-STLA and Prio-DH is around 7% on the average. This is a relatively small value which might be due to the low rate of substitution under the heuristic policies. Note that under UB policy, no substitution and no production scheduling takes place. According to Table 6.5, under light traffic, in

Table 6.5: The comparison of the heuristics under light traffic intensity (in percentages)

	PRIO	PRIODH	LQ-SSTLA	LQ-BTLA	SICOST	SILEVEL	UB	LB
average	4.44	-0.43	0.04	0.01	6.36	6.31	6.82	-44.59
stdev	7.24	1.34	0.27	0.04	10.16	10.26	8.49	21.19
min	-8.33	-9.68	-0.54	-0.01	0.00	0.00	0.00	-72.20
10 perc	0.00	-1.29	0.00	0.00	0.00	0.00	0.27	-64.47
25 perc	0.00	-0.25	0.00	0.00	0.00	0.00	0.78	-60.63
50 perc	0.99	0.00	0.00	0.00	0.81	0.53	3.19	-51.79
75 perc	6.08	0.00	0.00	0.00	10.27	10.07	9.94	-38.67
90 perc	13.54	0.00	0.09	0.00	19.34	17.37	17.93	0.00
max	39.30	1.16	1.73	0.28	56.24	56.24	45.39	0.00

25% of the time, the results are very close to LQ-STLA. So in one fourth of the instances, we are not getting any benefits from doing scheduling, or alternatively scheduling is not done.

The LB we consider is at least 40% loose in 75% percent of the instances. However for 10% of instances, it gives very tight results. Those are actually the instances where inventory rationing is not made due to the zero basestock level. Furthermore, in those instances UB difference is also $\leq 1\%$.

The value and impact of substitution

In this part the performances of Prio-DH and LQ-STLA are compared under different substitution schemes, no substitution, one-way and two-way substitution. Comparison results are presented in Table 6.6. Values in each column are obtained as follows. For Column 2, and Column 3, first for each of the 100 instances $\frac{Cost_{PrioDH-nosubs}-Cost_{LQSTLA-nosubs}}{COST_{LQSTLA-nosubs}}$ and $\frac{Cost_{PrioDH-oneway}-Cost_{LQSTLA-oneway}}{COST_{LQSTLA-oneway}}$ values are calculated, respectively. Then the performance measures are obtained. For Column 4 and Column 5, first for each of the 100 instances $\frac{Cost_{PrioDH}-Cost_{PrioDH-nosubs}}{COST_{PrioDH-nosubs}}$ and $\frac{Cost_{LQSTLA}-Cost_{LQSTLA-nosubs}}{COST_{LQSTLA-nosubs}}$ values are calculated, respectively. Then the performance measures are obtained.

Table 6.6: Comparison of Prio-DH and LQ-STLA under substitution types under light traffic (in percentages)

	Prio-DH over LQ-STLA		Value of (two-way) subs under Prio-DH	Value of (two-way) subs under LQ-STLA
	No-subs	One-way		
average	0.03	0.12	4.11	3.67
stdev	0.30	0.64	5.77	5.69
min	-2.10	-0.95	0.00	0.00
10 perc	0.00	0.00	0.00	0.00
25 perc	0.00	0.00	0.00	0.00
50 perc	0.00	0.00	1.38	0.50
75 perc	0.01	0.00	6.05	4.93
90 perc	0.27	0.19	12.84	11.73
max	0.99	4.39	28.21	28.21

In light traffic, when no substitution and one-way substitution are considered, LQ-STLA and Prio-DH performs equally well statistically. Since under two-way substitution Prio-DH performs better than LQ-STLA, this shows Prio-DH uses the substitutions more effectively than LQ-STLA. From Column 3 in Table 6.6 it is observed that under light traffic, when the substitutions are allowed only one way, Prio-DH loses its competitive edge.

Column 5 in Table 6.6 gives us information on value of substitution under LQ-STLA. Its results are similar the value of substitution under Prio-DH. It shows that in some instances ($<$ 10% of the instances), the substitution results in more than %10 cost decrease in light traffic. Under one-way substitution, the cost decrease drops from around $3-4\%$ to 0.2% on the average (not reported in Table 6.6). Finally, we note that comparing substitution rates under two-way substitution, under LQ-SSTLA, SICOST and SILEVEL, the rate is at least twice as much compared to LQ-STLA and Prio-DH. Since these heuristics performs worse, this implies there is an excessive substitution action taken by these policies. Indeed in SILEVEL, whenever a product has positive inventory and the other has backorders, a substitution action is enforced.

Set 1- Medium-Low traffic ($\sum \lambda_k \in [0.2, 0.6]$)

In Table 6.7, the average, standard deviation and percentiles of the 100 instances are given. The percentile information is provided in order to give information about how the performance is distributed.

When the arrival intensity is increased to medium-low [0.2,0.6]. In this setting, in 40% of the instances Prio-DH performs better, and in 60%, LQ-STLA outperforms. On the average Prio-DH results in 1.62% higher cost. Base=stock levels are higher, optimal policy work differently than priority policy, which can explain the deterioration of Prio-DH's performance. There is also a deterioration in the performance of the ordering heuristic, compared to the light traffic case. (All results for Prio and Prio-DH are given under optimal priority ordering).

For medium-low traffic the performance of the LQ-SSTLA deteriorates compared to that under light traffic, but the performances of SICOST and SILEVEL stays relatively the same. It is observed that due to the myopic behavior, LQ-SSTLA results in four times as much substitutions compared to LQ-STLA. Substitution rate for SICOST and SILEVEL are twice as much on the average. We conducted

Table 6.7: The comparison of the heuristics under medium-low traffic intensity (in percentages)

	PRIO	PRIODH	LQ-SSTLA	LQ-BTLA	SICOST	SILEVEL	UB	LB
average	15.17	1.62	1.20	0.05	7.91	6.87	18.33	-48.08
stdev	6.71	4.91	1.81	0.24	6.89	6.61	6.50	11.69
min	1.99	-15.55	-3.47	-0.53	0.58	-0.10	4.36	-72.89
10 perc	6.85	-2.47	-0.34	-0.09	2.64	1.63	10.31	-62.62
25 perc	10.13	-0.19	0.00	0.00	4.03	3.48	12.82	-55.25
50 perc	15.43	0.83	0.83	0.00	6.65	5.87	18.21	-49.86
75 perc	20.01	3.44	2.04	0.00	9.03	8.28	23.20	-39.55
90 perc	23.11	7.73	3.49	0.27	14.40	14.08	27.59	-32.64
max	36.82	15.94	7.58	1.03	47.07	44.96	31.26	-20.37

paired t-tests (cost values come from Normal distribution) to compare the performances of the policies. Comparison results reveal that under medium-low traffic

$$LQ{-}STLA \equiv LQ{-}BTLA \prec LQ{-}SSTLA \equiv Prio{-}DH \prec SILEVEL \prec SICOST$$

with 0.05 significance. As Table 6.7 shows, both Prio-DH and LQ-STLA operates far away from the UB(10%) and LB (50%) policies.

The value and impact of substitution

When the policies Prio-DH and LQ-STLA are compared under other substitution policies under medium traffic intensity, it is observed that when fewer substitutions are allowed, Prio-DH performs worse than LQ-STLA. For LQ-STLA the average value of substitution is 6%, while for Prio-DH it is 7.5%. This value can go up to 20% for Prio-DH, and 14% for LQ-STLA.

Table 6.8: Comparison of Prio-DH and LQ-STLA under substitution types under medium-low traffic (in percentages)

	Prio-DH over LQ-STLA		Value of (two-way) subs under Prio-DH	Value of (two-way) Subs under LQ-STLA
	No-subs	One-way		
average	3.42	4.20	7.56	5.89
stdev	3.49	4.43	3.49	2.82
min	-2.97	-2.61	1.30	0.15
10 perc	0.28	0.23	3.38	2.39
25 perc	1.14	1.12	5.31	3.97
50 perc	2.63	2.47	6.92	5.81
75 perc	4.86	6.82	9.26	7.63
90 perc	8.00	10.04	11.87	9.38
max	17.14	23.64	19.50	13.86

Set 1- Medium-High traffic ($\sum \lambda_k \in [0.6, 0.8]$)

In medium-high traffic, performance of Prio-DH over LQ-STLA further deteriorates. In only about %5 of the instances, it works better than LQ-STLA. In conclusion, when traffic intensity is $\in [0.2, 0.8]$, and when holding and backorder costs are not in proportion, LQ-STLA is a better policy to use. We compared

Table 6.9: The comparison of the heuristics under medium-high traffic intensity (in percentages)

	PRIO	PRIODH	LQ-SSTLA	LQ-BTLA	SICOST	SILEVEL	UB	LB
average	25.48	9.28	4.32	0.02	9.02	8.56	28.07	-34.38
stdev	7.08	5.93	2.59	0.28	3.20	3.05	4.78	13.12
min	8.49	-3.14	-0.50	-0.87	0.28	0.35	17.31	-61.31
10 perc	15.71	0.54	1.16	-0.33	4.79	5.05	22.27	-50.19
25 perc	20.43	5.86	2.43	-0.12	6.94	6.68	24.08	-44.41
50 perc	25.58	9.77	3.93	0.00	8.86	8.73	28.84	-34.49
75 perc	31.36	12.83	5.86	0.14	11.37	10.29	31.55	-25.60
90 perc	33.81	16.92	7.79	0.39	13.45	12.64	33.37	-16.56
max	39.46	23.02	15.08	0.75	17.18	16.96	39.07	-6.56

Table 6.10: Comparison of Prio-DH and LQ-STLA under substitution types under medium-high traffic (in percentages)

	Prio-DH - LQ-STLA		Value of (two-way) Subs under Prio	Value of (two-way) Subs under LQ-STLA
	No-subs	One-way		
average	10.44	11.69	4.98	3.97
stdev	5.64	5.95	2.59	1.70
min	-0.94	-1.22	0.31	0.11
10 perc	2.34	3.23	2.21	2.08
25 perc	6.96	7.66	2.85	2.74
50 perc	10.58	11.82	4.67	3.92
75 perc	14.29	15.40	6.38	5.07
90 perc	17.54	19.79	8.21	6.20
max	23.77	24.27	11.90	8.29

the performances of the policies using statistical tests. Under medium-high traffic, cost values show that LQ-STLA, LQ-BTLA, LQ-SSTLA, SILEVEL and SICOST follow Normal distribution, whereas cost values under Prio and Prio-DH not necessarily follow Normal distribution. We make the comparisons separately. For LQ class of policies we conducted paired t-tests with 0.05 significance. Results reveal that the performances can be ordered as follows:

$$LQ - STLA \equiv LQ - BTLA \prec LQ - SSTLA \prec SILEVEL \prec SICOST$$

To compare the performance of Prio-DH and LQ-STLA, first for each instance the difference in costs are recorded. Then the hypothesis of whether significant difference exists is tested via non-parametric tests. results reveal that LQ-STLA outperforms Prio-DH at 0.05 significance.

Under medium-high traffic, it is observed that UB is on the average 40% higher than the cost under LQ-STLA, 17% higher than the cost under Prio-DH. Substitution rate under Prio-DH is three times higher than the rate under LQ-STLA. This implies, under medium-high traffic, Prio-DH fails to do the scheduling effectively compared to LQ-STLA. The result shows that, the effect of the scheduling decision is closer to the “no scheduling” policy of the UB. Since LQ-STLA results in lower cost than Prio-DH, the gap between the LB is lower under LQ-STLA, which is 37%.

Value of substitution under LQ-STLA reaches up to 8% while under Prio-DH it can reach up to 12%. On the average under LQ-STLA the value of substitution is 4% and under Prio-DH it is 5%.

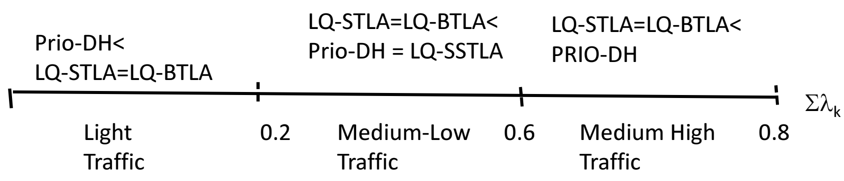


Figure 6.1: Graph of Heuristic Performance based on traffic intensity.

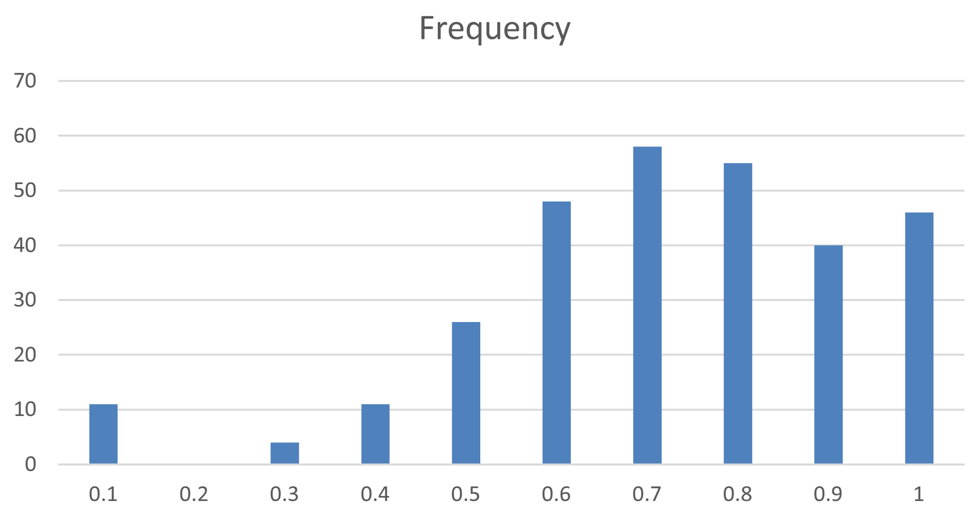


Figure 6.2: How much coverage our best policy gives based on difference of UB-LB.

In Figure 6.2, we look at the frequency diagram of the place of our best policy on UB-LB difference, scaled to 1. Hence higher values means, it is closer to UB, while lower values the best policy is closer to UB. When parameters of the system is randomly selected, average place is around 0.68. Although this means, our results are closer UB, this is partly due to, LB is a loose bound.

6.4.2.2 Parameter Set 2

In § 6.4.2.1 it is observed that under medium traffic LQ-STLA outperform Prio-DH for more of the instances, whereas under light traffic Prio-DH give lower cost for more of the cases. We conjecture that if the parameters favor a natural priority ordering, then Prio-DH may result in lower cost under medium or high traffic. In this section several more parameter settings are analyzed, to identify the conditions under which each heuristic performs well. In this part, only the results on comparison of the heuristics are presented, discussion on value of substitution is omitted. We present the results for the 4-product setting. The results for 3-product setting presented in Appendix D.

1. **Group 1: Identical h, varying** b The first group of parameters consist of the following setting. Unit holding cost are identical among the products and takes values $h \in \{1, 3, 5, 10\}$. Backordering cost has an increasing order, $b_1 \in \{5, 6, 7\}$, $b_2 = 2b_1$, $b_3 = 3b_1$, and $b_4 = 4b_1$. Arrival rate is identical for all products, and set to $\lambda_i = 0.2$. A total of 12 instances are studied. Priority order is taken as $4 \prec 3 \prec 2 \prec 1$.

Table 6.11: The comparison of the heuristics for Set 3 Group 1 (in percentages)

Instance	h	b_1	Prio-DH	LQ-SSTLA	LQ-BTLA	SICOST	SILEVEL	UB	LB
1	1	5	5.72	-0.46	0.17	1.06	2.19	38.64	-5.81
2		6	5.63	-0.47	0.22	1.02	2.23	36.80	-6.28
3		7	9.11	-0.63	0.02	0.83	2.09	39.37	-4.46
4	3	5	4.89	-0.65	0.00	-0.10	0.66	47.99	-4.44
5		6	5.69	-0.58	0.34	0.24	1.15	44.88	-5.28
6		7	1.48	-0.80	0.12	-0.12	0.66	41.85	-6.72
7	5	5	4.20	-0.30	0.00	0.07	0.74	54.86	-3.09
8		6	4.47	-0.64	0.40	-0.16	0.49	50.27	-4.14
9		7	1.51	-0.97	0.00	-0.48	0.24	45.93	-6.21
10	10	5	20.35	-0.18	0.00	-0.11	0.25	59.51	-1.92
11		6	8.96	-0.22	0.16	0.00	0.56	56.15	-1.78
12		7	4.24	-0.33	0.00	-0.19	0.33	56.94	-1.79

It is observed (in Table 6.11) that, under ordered backorder costs, when h (or $\frac{h}{b}$) is low or high, LQ-STLA performs better than Prio-DH.

If $\frac{h}{b}$ is low, then the stock levels are high. Prio-DH have a tendency to increase the gap between the stock levels of the products if $\frac{h}{b}$ ratio is low. On the other hand, LQ-STLA results in a more balanced stock level. Lowest priority product does not necessarily have the highest basestock in contrast to Prio-DH approach, since low backorder cost leads to a low basestock level.

When $\frac{h}{b}$ is high, stock levels are low (either 0 or 1). In that case, LQ-STLA keeps stock for high unit backorder cost products and no stock for low backorder products. Keeping stock helps avoid falling to backorder and facilitates substitution. Prio-DH does just the opposite: increases basestock level for low backorder cost products, and keeps no stock for high priority products. Under high holding cost, keeping unnecessary stock results in high cost. The stock levels are more balanced and are higher under LQ-STLA, and it is thus a more preferable approach.

Finally, under this setting the other heuristics (LQ-SSTLA, SICOST, SILEVEL) perform either very close, or outperform LQ-STLA. All heuristics, except Prio-DH, perform close to LB.

2. **Group 2: Identical h, varying b, changing λ.** In this group, $h = 3$, $b_1 = 5$, $b_2 = 10$, $b_3 = 15$, $b_4 = 20$. Arrival rates are identical and take the following values, $\lambda_i = \{0.16, 0.18, 0.20, 0.21, 0.22\}$.

Table 6.12: The comparison of the heuristics for Set 3 Group 2 (in percentages)

Instance	Arrival rate	Prio-DH	LQ-SSTLA	LQ-BTLA	SICOST	SILEVEL	UB	LB
1	0.16	1.4	-0.6	0	-0.5	0.5	47	-4.7
2	0.18	2.9	-0.7	0	-0.3	0.6	47.8	-5.9
3	0.2	4.9	-0.6	0	-0.1	0.7	48	-4.4
4	0.21	1.4	-0.7	0.2	-0.2	0.4	45.9	-4.9
5	0.22	-2.7	-0.3	0	0.4	0.9	42.6	-7.4

When the arrival rate is medium (0.16-0.18), the server can supply the shortfall of the product from designated level, without falling into backorder. This results in a decreased requirement to Prio-DH, as every request to server is served immediately. Since the basestock levels under LQ-STLA are adequately calculated considering the low queue length, all capacity can be allocated to the shortfall. This results in better performance for LQ-STLA (see Table 6.12)

When the arrival rates increase, however, the policy needs to consider which product to serve first, as the number of shortfall of products during a service gets larger. For this scenario, Prio-DH works better than LQ-STLA. As traffic rate increases, it is observed that Prio-DH performs closer to the lower bound. Therefore, under high traffic the optimal policy has the following structure: the stocks are pooled as if a single product and the stock is rationed with respect to the backorder cost. As traffic rate increases, Prio-DH sets the basestock levels that give way to such a policy. The lowest priority product has a considerably high basestock level compared to other products, and as a result substitution rate is much higher compared to LQ-STLA. Prio-DH "pools the stocks" and thus benefits from decreased holding cost. Since LQ-STLA has a tendency to treat the products equally, a basestock structure that facilitates pooling cannot be observed.

This pattern is observed for the 4-product setting. In 3-product setting we observe Prio-DH no necessarily outperforms LQ-STLA under high traffic (when $\lambda_k = 0.27, \sum \lambda_k = 0.81$).

All other heuristics perform close to LQ-STLA, under all arrival rates. Under high traffic, the substitution rates of the heuristics are close, whereas rate under Prio-DH is considerably higher.

3. **Group 3: Dispersed *b*.** In this setting it is assumed that $h = 3$, $\lambda_i = 0.2$. We take $b_1 = 12.5 - 3X,\ b_2 = 12.5 - X,\ b_3 = 12.5 + X, b_4 = 12.5 + 3X$ (keeping holding cost and arrival rate constant) in which the X factor defines the dispersion of backorders. The dispersion factor is selected from $X \in \{2, 2.25, 2.5, 2.75, 3\}$. A total of 5 instances are analyzed. By this way, the hierarchy among the items is made stronger.

   First of all, SICOST and LQ-SSTLA both outperform LQ-STLA (in Table 6.13). This implies the optimal scheduling policy is as defined in those

Table 6.13: The comparison of the heuristics for Set 3 Group 3 (in percentages)

Instance	Dispersion (X)	Prio-DH	LQ-SSTLA	LQ-BTLA	SICOST	SILEVEL	UB	LB
1	2	4.74	-0.80	0.13	-0.38	0.32	42.28	-2.14
2	2.25	4.94	-0.64	0.17	-0.16	0.57	45.18	-2.85
3	2.5	4.89	-0.65	0.00	-0.10	0.66	47.99	-4.44
4	2.75	1.23	-0.64	0.49	-0.02	0.77	50.99	-6.10
5	3	3.19	-0.60	0.37	-0.10	0.66	56.03	-6.51

heuristics, especially when the dispersion is high. Furthermore, increased substitution (SICOST has a higher rate of substitution compared to LQ-STLA)improves the performance under this setting.

For this system, when the system is in back-order there is an inherent priority order for processing items, but also when the system is holding inventory, all the inventory costs are the same, for this reason, it tries to allocate the scheduling decisions as a longest queue. So there is a mixed effect, but when we look at base-stocks of the systems, Prio-DH allocates less inventory to high- backorder product because it has higher priority, but this should not be the case, the higher backorder product should get higher base-stock as well, when other parameters are the same. For this reason, LQ-STLA performs better than Prio-DH.

Also another observation is, for Prio-DH (and LQ-STLA), as dispersion increases, the base-stock levels decrease. So with lower inventory, the production scheduling handles the cost well. Finally, under identical h and high λ (Groups 1,2 and 3) performance of the best policy is close to the lower bound. This implies performances are close to that of the optimal policy.

4. **Group 4: Identical b, varying h.** In this setting it is assumed that $h_1 = 5$, $h_2 = 3$, $h_3 = 2$, $h_4 = 1$, whereas unit backorder values are identical for all products. Unit backorder cost takes values $b \in \{15, 18, 21\}$. We consider $\lambda_k = \{0.2, 0.22\}$. The priority among the items is determined by the unit holding cost only.

Table 6.14: The comparison of the heuristics for Set 3 Group 4 (in percentages)

Instance	b	λ	Prio-DH	LQ-SSTLA	LQ-BTLA	SICOST	SILEVEL	UB	LB
1	15	0.2	-13.78	-2.36	1.68	8.49	8.63	38.51	-37.84
2	18		-14.67	-3.42	1.15	6.11	6.34	32.98	-40.97
3	21		-3.28	-3.97	1.07	10.99	11.15	31.81	-42.55
4	15	0.22	-19.99	-2.61	2.33	11.86	11.78	29.81	-36.3
5	18		-14.12	-3.25	1.83	16.44	16.47	29.79	-37.36
6	21		-11.66	-3.08	2.04	16.07	16.12	31.4	-37.92

When the products are ordered with respect to their holding cost, it is observed that Prio-DH outperforms LQ-STLA (see Table 6.14). The performance of Prio-DH slightly increases as the traffic intensity increases.

In this case, it is also observed that LQ-SSTLA outperforms LQ-STLA, whereas cost under SILEVEL and SICOST are much higher compared to LQ-STLA. The substitution rate under SICOST and SILEVEL are much lower compared to LQ-STLA.

Furthermore, all policies perform far away from the LB, and relatively closer to the UB. This may support the conclusion that the optimal policy does not necessarily pool the stock when holding cost are different, but rather the allocation of the capacity is done effectively at the onset.

In LQ-STLA basestock levels among the products differ only by 1, whereas in Prio-DH the difference between the basestocks is much higher. Substitution rates are close for the both policies. Increased basestock levels for the low holding cost products under Prio-DH helps to make the capacity allocation more effectively.

These observations are also valid for the 3-product setting.

5. **Group 5: Clustered products.** In this setting the products are clustered with respect to (h, b) values. It is assumed that a product is either in Cluster 1: $h = 5,\ b = 20$ or in Cluster 2: $h = 1,\ b = 10$. Cluster 2 is assigned the lower priority in Prio-DH.

   In Table 6.15 arrival rate is set to $\lambda_i = \{0.2\}$ and in Table 6.16 arrival rates are set to $\lambda_i = \{0.10,\ 0.11, \cdots, 0.18\}$. In both tables, to denote an instance, the number of products in cluster 1 is denoted with Nb-C1, and the number of products in Cluster 2 is denoted with Nb-C2.

Table 6.15: The comparison of the heuristics for Set 3 Group 5a when $\lambda_i = 0.2$ (in percentages)

Instance	(Nb-C1, Nb-C2)	Prio-DH	LQ-SSTLA	LQ-BTLA	SICOST	SILEVEL	UB	LB
1	(1,3)	7.30	-5.89	2.39	19.11	18.65	45.15	-25.35
2	(2,2)	0.47	-4.43	2.04	19.61	19.75	48.91	-42.21
3	(3,1)	-12.54	-1.79	0.9	11.33	11.52	41.16	-54.44

It is observed from Table 6.15 that as the number of products in Cluster 2 ($h = 1,\ b = 10$) increases the performance of Prio-DH compared to LQ-STLA deteriorates. If there are several products in Cluster 2, Prio-DH assigns one of them the lowest priority and sets the highest basestock to only to this product. In fact, this high basestock should be placed evenly for all the Cluster 2 members. LQ-STLA does this better, resulting in better performance.

On the other hand, when the number of products in Cluster 1 is higher the relative performance of Prio-DH improves. Since low basestock is assigned to priority products under Prio-DH, when the number of products in Cluster 1 is high, the basestock levels are low and evenly distributed, whereas the basestock levels of the products in Cluster 2 are much higher. Prio-DH does not assign the same basestock to two identical products due to its nature. LQ-STLA indeed assigns the same basestock level. However, in LQ-STLA the basestock levels differ very little among the products in Cluster 1 and Cluster 2, since only the ratio of $\frac{h}{b}$ is taken into consideration and not the actual values of h and b. Thus, Prio-DH does a better job in terms of assigning the basestock levels by inflating the basestock level of the product with low holding cost. As a result Prio-DH performs better when the number of low priority products decrease.

Table 6.16: The comparison of the heuristics for Set 3 Group 5b under varied λ_i (in percentages)

Instance	λ	Prio-DH	LQ-SSTLA	LQ-BTLA	SICOST	SILEVEL	UB	LB
1	0.10	0.79	0.00	0.00	0.00	0.00	91.32	-46.94
2	0.11	0.77	0.00	0.00	0.00	0.00	84.10	-52.43
3	0.12	0.71	0.00	0.00	0.00	0.00	77.26	-55.43
4	0.13	-22.12	0.00	0.00	0.00	0.00	61.33	-46.99
5	0.14	-30.28	0.00	0.00	6.53	7.06	33.07	-47.51
6	0.15	-23.47	0.00	0.00	6.64	7.22	38.26	-47.91
7	0.16	-4.49	0.00	0.00	6.44	7.04	42.56	-55.09
8	0.17	1.53	-0.50	0.84	8.96	9.39	54.74	-52.70
9	0.18	-2.79	-0.41	4.75	8.13	8.55	59.68	-48.83
10	0.19	-9.47	-0.80	3.32	6.19	6.58	61.29	-45.60
11	0.20	-12.54	-1.79	0.90	11.33	11.52	41.16	-47.91

In Table 6.16, the case where the number of products in Cluster 2 is 1, $(Nb-C1, Nb-C2) = (3, 1)$ is scrutinized. Only the arrival rate is varied.

When the arrival rate is low, the Prio-DH policy and LQ-STLA give similar results. It is observed that for low arrival rates ($\lambda = 0.10,\ 0.11,\ 0.12$), the basestock levels are the same for both policies. Basestock levels are zero for Cluster 1 products, and 1 for the Cluster 2 product. In both policies, when a shortfall occurs for one of the products, the server immediately replenishes the products before any other arrival. Hence the service order schedules cannot be differentiated.

When arrival rates increase slightly ($\lambda = 0.13,\ 0.14,\ 0.15$) Prio-DH prevails over LQ-STLA. It is observed that for LQ-STLA basestock levels are identical (or very close) among the products, whereas Prio-DH assigns higher basestock for low priority products.

As traffic intensity changes from low to high, it is observed that the performances of the two policies change. The relative performances are affected by the changes in the basestock levels, however Prio-DH seems to outperform LQ-STLA for a variety of arrival rates.

6. **Group 6: Identical products, dispersed arrival rates**. The final

group of instances aims to measure the joint impact of the traffic intensity and the dispersion in arrival rates and among the products.

It is assumed that $\lambda \in \{0.06, 0.07, 0.08, 0.085, 0.09\}$, and $\lambda_1 = \lambda$, $\lambda_2 = 2\lambda$, $\lambda_3 = 3\lambda$, $\lambda_4 = 4\lambda$. In the setting, $(h, b) = (1, 20)$ and $(5, 20)$. We expect when there is no dispersion, the LQ-STLA will prevail because of the optimal policy's longest queue structure. However, in the experiment set, we see that there is no simplistic relationship between arrival rate dispersion and performance differential of heuristics.

When the $\frac{h}{b}$ ratio is low, the increase in the arrival rate dispersion affects the performance mixedly. The performance of LQ-STLA may increase or decrease with λ, but it outperforms Prio-DH by 10%. When the $\frac{h}{b}$ ratio is high, with dispersion increase, the performance of Prio-DH gets better, but there are still contradicting results. In this case LQ-STLA outperforms by 0.2%.

Finally, under $\frac{h}{b} = \frac{5}{20}$, $\sum_i \lambda_i$ is kept as constant at 0.7 and the dispersion among λ_i is increased. The performances are close (around 1%) but a trend cannot be observed. Other heuristics, LQ-SSTLA, SICOST, and SILEVEL perform very close to LQ-STLA and the difference between the LB is around $2 - 3\%$. Note that when arrival rates are not dispersed, the difference between the lower bound increases to much higher values. This means, when arrival rates are dispersed, for identical products LQ-STLA performs close to the optimal policy.

6.4.2.3 Parameter Set 3

In this section, it is assumed that the products are ordered with respect to their "values". Ordered product values imply the unit holding cost and unit backordering costs are also ordered. We compare the policies under several instances. Each instance corresponds to a tuple of unit holding and unit backordering costs for all products and the arrival rates. In this set of runs we assume all four products have the same arrival rate. Each instance is constructed as follows: unit holding cost of the lowest priority product (say, product 4) is assumed $h_4 = 0.25$. A holding cost factor $h*$ is determined. Unit holding cost of product 3, product 2 and product 1 are set as $h_3 = \frac{h_4}{h^*}$, $h_2 = \frac{h_3}{h^*}$, $h_1 = \frac{h_2}{h^*}$, respectively. We let h^* take the following values; $h* = \{0,3; 0,5; 0.7; 0.9; 1\}$, a total of 5 multipliers. In each instance, for each product, the holding backorder ratio (h/b) is set at a constant value. We let h/b take the following values $\frac{h}{b} = \{10; 20; 50; 80\}$, a total of 4 ratios. In this parameter set, the substitution rate is taken as $\mu_S = 5$ and substitution cost is taken identical for all products as $c_s = 1$. Production rate is taken as $\mu = 1$. For the products, the arrival rates considered are $\lambda_k \in \{0.04, 0.07, 0.1, 0.13, 0.16, 0.18, 0.21, 0.22\}$, a total of 8 arrival rates. So we test the problem on 5*4*8=160 instances. Based on the numerical results in the previous sets, comparison of the performances within the class of LQ-based policies indicate that the performance of LQ-STLA is the same as that of LQ-BTLA and slightly dominates that of LQ-SSTLA. Performances of SICOST and SILEVEL are dominated by LQ-STLA, LQ-BTLA and LQ-SSTLA in almost all settings. Thus in this section we only compare LQ-STLA and Prio-DH.

Table 6.17 present the results on the performance of Prio-DH in contrast to LQ-STLA. The figures in the table show the difference in %:

$$\frac{Cost_{Prio-DH} - Cost_{LQ-STLA}}{Cost_{LQ-STLA}} 100\%$$

The priority heuristic is evaluated under the optimal priority ordering, which is many times in line with the unit holding cost ordering.

The following observations are made:

1. Under medium or high traffic ($\sum \lambda > 0.5$), when hierarchy among the

Table 6.17: Performance of Prio- over LQ-STLA(in percentages)

λ	$b/h \setminus h*$	0.3	0.5	0.7	0.9	1
0.04	10	0.00	0.00	0.00	0.00	0.00
	20	-68.77	-63.33	-55.81	-47.80	-32.34
	50	0.01	0.18	0.31	0.03	-0.08
	80	0.01	0.00	0.28	0.03	-0.08
0.07	10	-74.45	-61.87	-48.26	-36.37	-24.89
	20	1.78	1.69	0.41	-0.11	-0.45
	50	0.00	2.48	0.98	0.25	-0.41
	80	-0.26	2.70	10.29	23.44	29.48
0.1	10	7.16	1.62	1.17	0.88	-0.59
	20	1.06	3.75	3.29	-0.31	-1.11
	50	-2.30	2.61	14.12	29.20	27.36
	80	-27.54	-21.65	-13.54	-7.93	-8.63
0.13	10	3.19	11.23	8.78	10.16	11.57
	20	-2.90	0.87	13.07	16.08	12.45
	50	-20.60	-12.88	-5.44	4.20	6.85
	80	11.39	16.47	23.87	30.38	32.53
0.16	10	-3.92	4.49	8.48	8.81	9.25
	20	-15.55	-9.56	3.01	14.92	19.43
	50	-12.23	-10.17	-2.40	6.27	10.69
	80	-13.91	-5.75	7.09	21.03	24.43
0.18	10	-22.01	-14.15	0.34	10.98	11.54
	20	-12.28	-7.20	5.39	18.42	21.15
	50	-10.52	-2.30	10.79	24.49	30.17
	80	-19.90	-9.82	5.78	19.72	25.14
0.21	10	-20.77	-12.69	4.32	15.45	13.09
	20	-33.67	-19.17	-0.88	17.08	16.60
	50	-32.81	-18.48	2.37	16.38	22.85
	80	-27.74	-13.24	7.09	22.13	27.22
0.22	10	-28.11	-16.40	3.52	18.21	12.97
	20	-27.73	-17.37	1.41	15.58	17.74
	50	-41.77	-24.78	0.47	18.19	21.84
	80	-43.53	-25.43	0.16	19.73	23.70

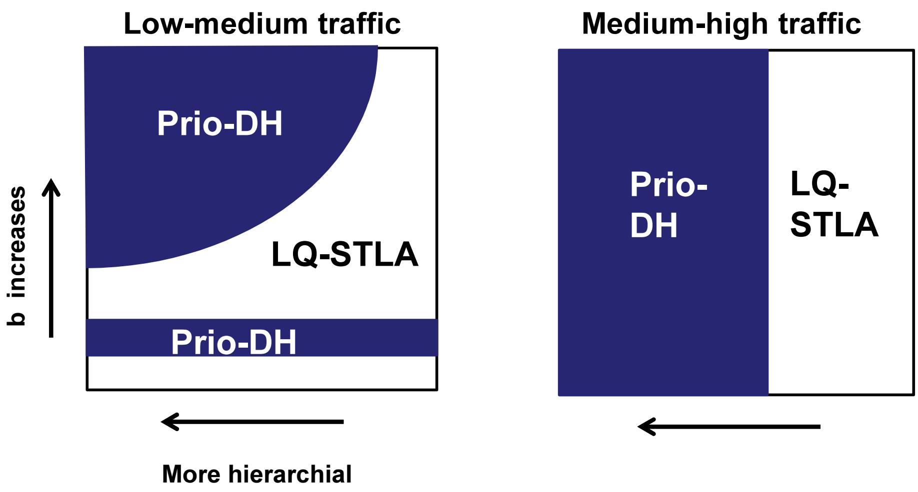


Figure 6.3: Performance of LQ-STLA versus Prio-DH under Set-2.

products is dominant ($h^* = 0.3, 0.5$) then Prio-DH outperforms LQ-STLA. In this case the performance difference between the two policies can be as high as 40%.

2. Under medium or high traffic ($\sum \lambda > 0.5$), when products are valued similarly, ($h^* = 0.7, 0.9, 1$) then LQ-STLA outperforms Prio-DH. When this is the case, the performance difference between the two policies can be as high as 20%.

3. Under low or very light traffic ($\sum \lambda < 0.5$), under some specific instances Prio-DH outperforms LQ-STLA, whereas in others LQ-STLA outperforms Prio-DH.

   Under low traffic, when the instances for which Prio-DH performs very well compared to LQ-STLA are examined in more detail, we encounter an interesting structure of the policy. Consider for example the following instances in Table 6.17: $\lambda_k = 0.04$, $\frac{b}{h} = 20$, $\lambda_k = 0.07$, $\frac{b}{h} = 10$, $\lambda_k = 0.1$, $\frac{b}{h} = 80$, $\lambda_k = 0.13$, $\frac{b}{h} = 50$, $h^* = 0.3, 0.5, 0.7$. For these instances the difference between the costs may be as high as 75%.

   In all of these instances, Prio-DH follows a policy structure that resembles the policy used to obtain the lower bound. The product with lowest priority (i.e., lowest holding cost) is identified. Compared to the other policies, a considerable amount of inventory is kept under the lowest priority product. Stock levels for higher priority products are lower, and whenever a backorder occurs, a substitution takes place. On the other hand, for a given $\frac{b}{h}$ ratio, LQ-STLA operates under the same inventory level for all products. Therefore it cannot benefit from substitution as much as Prio-DH.

The difference between the costs is most prominent when arrival rates are low. In those cases, the basestock levels are low as well. Thus, how the basestock levels are set considerably affect the performances of the policies. In Table Table 6.17, we observe that for a given arrival rate there exists a single $\frac{b}{h}$ under which Prio-DH remarkably outperforms LQ-STLA (for example for $\lambda_k = 0.04$, under $\frac{b}{h} = 20$, Prio-DH is 68% lower than LQ-STLA). Actually, rather than discrete values of $\frac{b}{h}$, if a continuous range is considered for $\frac{b}{h}$, we would observe that for each λ_k level, there are many values of $\frac{b}{h}$ under which a remarkable performance difference occur between Prio-DH and LQ-STLA. Those $\frac{b}{h}$ values correspond to the cases where a change in the basestock level takes place. In Figure 6.5 we present under light traffic ($\lambda_k = 0.12$) how $\frac{b}{h}$ affects the performance of Prio-DH with respect to LQ-STLA. For those instances where performance of Prio-DH is better than LQ-STLA, the basestock levels are low, and one or two products serve as the inventory repository under Prio-DH.

4. Finally, we observe that the benefit of substitution increases as the products are more differentiated (data not reported) under both Prio-DH and LQ-STLA. Under low traffic, for those instances where Prio-DH outperforms LQ-STLA the benefit of substitution under Prio-DH is very high. A more detailed discussion is presented in Appendix D.

In Figure 6.4, we look at the frequency diagram of the place of our best policy on UB-LB difference, scaled to 1. Hence higher values means, it is closer to UB, while lower values the best policy is closer to UB. When parameters of the system is randomly selected, average place is around 0.45. This means on average our results are closer to lower bound. When we compare this result with

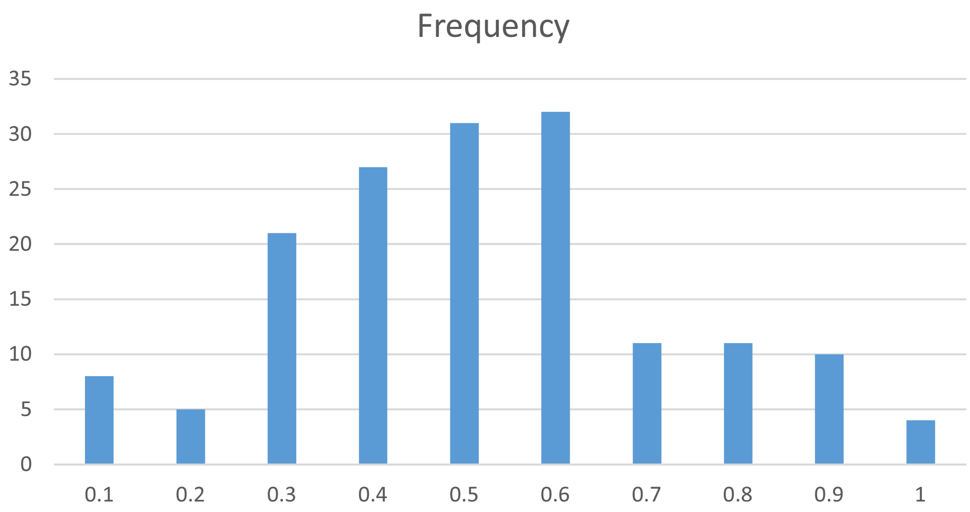


Figure 6.4: How much coverage our best policy gives based on difference of UB-LB.

Set-1 results, we can deduce that, when there is a structure of either hierarchy or symmetry, our policies works better compared to complete random parameters.

The findings under 3-product case are similar (see Appendix D).

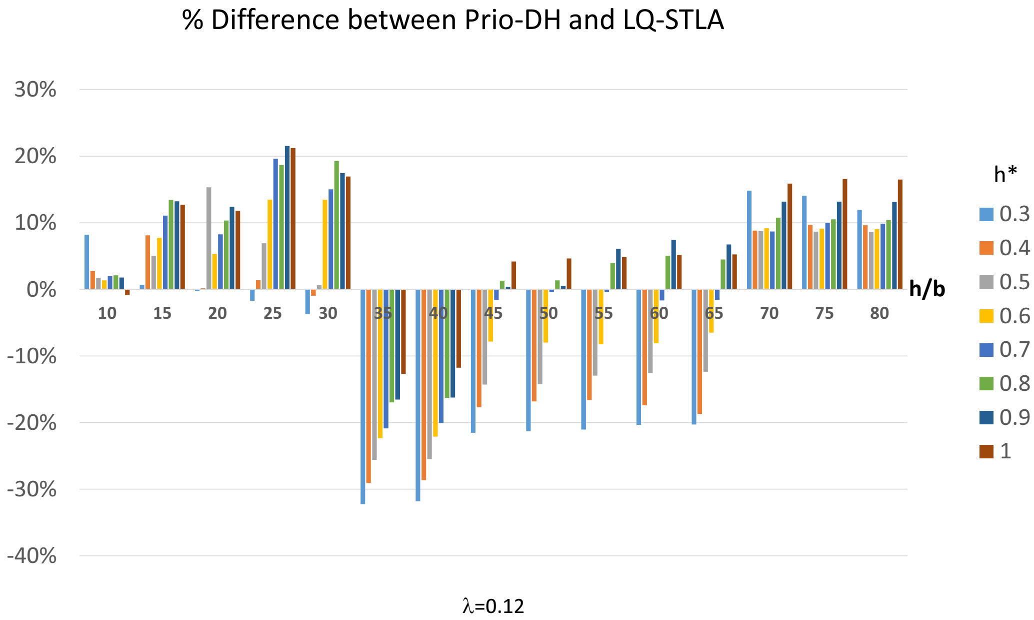


Figure 6.5: Performance of Prio-DH versus LQ-STLA as a function of $h\backslash b$, when products are hierarchical and $\lambda = 0.15$

CHAPTER 7

SELLER'S RESPONSE TO CUSTOMER-DRIVEN SUBSTITUTION

7.1 Introduction

In this second part of the thesis our aim is to address the manufacturer's substitution problem by incorporating the customer's substitution behavior. The first part of the thesis focuses only on construction and analysis of manufacturer-driven substitution policies. In manufacturer-driven substitution, the manufacturer takes the substitution decisions without taking the customer's response into consideration. The substitution decision may result in certain costs, but the customer accepts the substitute product for sure. Under customer-driven substitution on the other hand, the customer decisions on product selection are taken into consideration.

In this part, we aim to redefine the manufacturer's substitution problem by considering the manufacturer's response to customers' substitution behavior. We use the terms manufacturer and the seller interchangeably. We do not necessarily explicitly consider the production process, but rather only focus on how much the seller builds up inventory. Thus, here the seller might as well be a retailer.

To incorporate customer behavior into manufacturer's substitution decision-making process, it is assumed that the manufacturer does not explicitly offer a substitute product to the customer. Rather the manufacturer offers all products (in its product line) to the customer, but manipulates the customer's substitution behavior. Manufacturer may use three tools to affect the behavior: (i) number of products offered, (ii) pricing of the products, (iii) availability, i.e., stock levels of the products. We consider two types of substitution behavior. Customers may substitute a product with another one if price of the product is much higher compared to the other, or if a product they intend to purchase is out of stock. In our problem context, the first and last decisions (number of products, and stock levels) are proactive decisions, whereas the pricing decisions are used both proactively and reactively. Currently, in the thesis we consider only the pricing and stocking decisions.

In the study, it is assumed that the seller offers two products. As would be seen in § 7.4 analysis of more than two products is intractable. Insights on multiple products will be gained through the stylized two product setting. Future work includes the analysis for single product setting. Comparison of single and two-product settings enables us to understand the gain due to increased product variety, as a response to customer-driven substitution.

Our aim in this analysis is to understand the effect of substitution on the manufacturer decisions and on relevant performance measures. We first describe the problem in § 7.2. In the problem under consideration there exist two pricing periods. Before presenting the analysis for the complex two-period problem, we present a preliminary analysis for a single period problem in § 7.3. Then, we move onto two-period analysis in § 7.4. In Chapter 8, we present the specific research questions asked and aim to address those questions via a numerical study. The findings of the numerical study is also presented in this chapter.

7.2 Description of the problem

The manufacturer offers 2 products where demand for each product i is characterized as a linear function of the price for each product. Substitution is incorporated through the product's cross-price effect on the other product. In Table 7.1, we present the notation for this chapter.

Table 7.1: Table of notation for Chapter 8

Notation for all models

A_i	The deterministic component of the *base demand* for product i.
base demand	Demand when the price for both products is zero
β	Cross-price effect, substitution parameter, $0 \leq \beta \leq 1$.
ϵ_i	The random component of the base demand for product i. Here ϵ_i is "a constant" times a Bernoulli random variable
ϵ	vector of ϵ_is, a random vector
ϵ_i^h	Demand potential, the high value that ϵ_i takes (the low value is 0).
π_i	Probability of demand occurrence, probability that ϵ_i will take the high value. With probability $(1 - \pi_i)$, ϵ_i takes low value (0).
c	Unit cost for building stock for product i.
ω	denotes a realization of a random variable, for instance $\epsilon_i(\omega)$ is a realization of ϵ_i
I_i	Initial stock level for product i, a decision variable. $I_i \geq 0$.
s_i	Residual stock or inventory slack, the level beyond $A_i - (1 - \beta)p$, a decision variable for product i. Here $I_i = A_i - (1 - b)p + s_i$, s_i can take any real value as long as $I_i \geq 0$.

Table 7.2: Notation for single period n-product model

p	Uniform price (set for both products)
$\mathbf{s}$	Denotes the vector of residual stock variables for n products, $(s_1, \ldots, s_n)$

$E[R(p, \mathbf{s}, \epsilon)]$	Total expected profit
$E[R_k(p)]$	Total expected profit given $\frac{c}{\pi_k} \leq p < \frac{c}{\pi_{k+1}}$.
D_i	Demand for product i, a random variable.

Table 7.3: Notation for two period two-product model

p_1	Decision variable denoting the price set for both products in period 1. Products are priced uniformly at p_1.
p_2	Decision variable denoting the price set for the products in period 2. Products are priced uniformly.
K_i	Net inventory level for product i at the beginning of period 2.
$R_1(s_1, s_2, p_1, \epsilon)$	Profit obtained by the manufacturer in period 1. R_1 is a random variable.
$R_2(s_1, s_2, p_1, p_2, \epsilon)$	Profit obtained by the manufacturer in period 2. At the beginning of period 1, R_2 is a random variable. At the beginning of period 2, randomness resolves.
$R_2^*(s_1, s_2, p_1, \epsilon(\omega))$	Optimal profit obtained in period 2 for a given s_1, s_2, p_1 and for a realization of ϵ. (i.e., $R_2(s_1, s_2, p_1, p_2, \epsilon(\omega))$ under optimal p_2).
$R^2(K^1, K^2, p_1)$	Optimal profit obtained in period 2 for a given K_1, K_2, p_1. In R^2, $K^1 = \min\{K_1, K_2\}$, and $K^2 = \max\{K_1, K_2\}$.
$R(s_1, s_2, p_1, p_2, \epsilon)$	Total profit obtained by the manufacturer (in both periods). $R = R_1 + R_2$.
D_i^k	Demand for product i in period k

Demand has the following form:

$$D_i = A_i - p_i + \beta_{ij} p_j + \epsilon_i, \; i = 1, 2. \tag{7.2.1}$$

Here p_i is the price for product i, A_i is the deterministic part of the "base" demand for the product. Base demand is the demand when the prices of all products are zero, and β_{ij} is the substitution parameter.

An increase in the price of product j results in an increase in the demand of product i. A positive β_{ij} indicates that the products are substitutes while negativity indicates complementarity. We use a nonnegative β_{ij}, since we are considering substitutable products. Because the decrease in a product's demand due to its own price effect is assumed to be larger than the cross-price effect, β_{ij} is taken to be smaller than 1, i.e., $0 \leq \beta_{ij} < 1$. For analytical tractability it is further assumed that $\beta_{ij} = \beta_{ji} = \beta$.

In (7.2.1), ϵ_i is the random component of the base demand. The demand uncertainty is modelled such that the base demand for each product can be either in "High" status or in "Low" status. The probability of each event is known by the manufacturer. Specifically, the random part of the base demand is modeled as a function of a Bernoulli random variable as follows:

$$\epsilon_i = \begin{cases} \epsilon_i^h & \text{w.p. } \pi_i \\ 0 & \text{w.p. } 1 - \pi_i \end{cases}$$

The demand model expressed in (7.2.1) is the basic model under consideration. Such linear models are commonly used in operations management literature, see for instance Zhao and Atkins (2008); Tang and Yin (2007). In § 7.4, we will look at this demand model in more detail to better understand the dynamics of the customer behavior.

We assume, the following sequence of the events take place:

1. At the beginning (at time $t = 1$), the manufacturer decides on the quantity to stock for each of the products. For each unit stocked, a cost of c is

incurred. The products are substitutable and are assumed to belong to the same family. Thus, unit cost of replenishment, c, is assumed to be the same. The replenishment is completed before the demand occurs. The manufacturer also at this time decides on price of the products for the first period.

At the end of the period the demand is realized. Depending on the quantities stocked, there might be over-stock for some of the products and under-stock for others.

2. Based on the over-stock and under-stock quantities, in the second period(t=2) those customers whose demand are not met may find a chance to satisfy their demand, depending on the product availability. We model stockout based substitution through the demand overflow between the products.

   Such stockout based substitution models are considered for instance by Parlar (1988); Zhao and Atkins (2008); Lippman and McCardle (1997). However, in those models the regular demand and the overflow demand (due to stockout) are combined into the same period. We assume that stockout-based substitution takes place after the regular demand is realized, where the amount of substitution depends on the initial stock levels of both products.

   We assume that once the regular sales period (period 1) is over, the manufacturer may change the prices to affect the demand flow due to substitution. Note that change in prices may result in additional demand, due to those customers that did not make any purchase in the first period. It implies that substitutions can occur even if both products are overstock. When both products are overstock, stockout-based substitution does not take place, but only price-based substitutions take place.

Note that in the second decision period, the demand is deterministic, because the manufacturer at this time has the knowledge of the demand. In the second period no new demand arrival takes place, i.e, all demand arrives at the beginning of the first period. This is a common assumption in pricing literature, see for instance

Elmaghraby et al. (2008), Liu and van Ryzin (2008), Liu and Zhang (2013). The manufacturer tries to just allocate understock, overstock and possibly the additional demand by pricing and considering the substitution effects.

We call the period where the stock levels are decided and initial prices are set as period 1 and the period where price is updated and substitutions take place, as Period 2. We assume all the products have identical prices. This assumption can be justified by assuming these products are from the same product line, so that they need to be closely priced. In practice for instance, if the products are only differentiated by their colors, then usually they are set the same price.

Before analyzing the two-period problem, we first study a single period problem where price is not updated and stockout based substitutions are not allowed. In Appendix E, we also elaborate on a model where stock levels are set at the beginning of the period before the demand is realized, while price decisions are set after the demand is realized.

7.3 Single period model

In this setting, we assume stock level and price decisions are set at the beginning of the period before the demand is realized. There does not exist another chance for the manufacturer to change the prices. Thus, the only substitution type is price-based substitution. Manufacturer's objective is to maximize profit.

The sequence of events are as follows:

1. At the beginning, the manufacturer decides on how many units to order for each product. A cost c is incurred for each unit stocked.
2. The manufacturer also decides on the price for each product.
3. Demand is realized and profit is obtained.

In this section, we analyze the problem for the n product setting (since the complexity of the problem allows we analyze the problem for the most general case).

The manufacturer sells n products, indexed by $i \in N = \{1, \cdots, n\}$, and faces linear demand as a function of its own price, p_i, and the other products' prices. Let D_i denote the demand for product i, the demand curve is governed by the following function:

$$D_i = A_i - p + \sum_{j \neq i} \beta_{ij} p + \epsilon_i, \ i \in N \tag{7.3.1}$$

Again it is assumed that $\beta_{ij} = \beta_{ji}$, and that $0 < \beta_{ij} < 1$. Furthermore, $\sum_{j \neq i} \beta_{ij} < 1, \ \forall i$. Define $a_i = 1 - \sum_{j \neq i} \beta_{ij}$. (in the two product model we denote $\beta_{ij} = \beta_{ji}$ with β). Note $0 \leq a_i < 1$. Finally, it is assumed that the manufacturer sells positive quantities for all products under all realizations (ϵ_i^h or 0) i.e.,

$$p \leq \min_i \{\frac{A_i}{a_i}\} = \bar{p}.$$

At the beginning, the manufacturer determines the stock level I_i. It is assumed that $c < \bar{p}$, so that it is profitable to build up the stock.

Without loss of generality, an exchange of variable is made

$$I_i = A_i - a_i p + s_i, \ i \in N \tag{7.3.2}$$

Here, s_i is the extra stock held over base demand, in order to hedge against variability. Obviously, $0 \leq s_i \leq \epsilon_i^h$.

The profit function is the expected net revenue generated by the product demands, minus the cost of replenishment. Let $\mathbf{s}$ denote the vector of s_i, let ϵ denote the vector of $\epsilon_i s$.

$$E[R(p, \mathbf{s}, \epsilon)] = E[\sum_i \Big(p \min\{D_i, A_i - a_i p + s_i\} - c(A_i - a_i p + s_i) \Big)], \tag{7.3.3}$$

where D_i denotes the random demand. For simplicity, we omit ϵ in notation.

Explicitly,

$$\begin{aligned} E[R(p, \mathbf{s})] &= \sum_i \Big(p(\pi_i(A_i - a_i p + s_i) + (1 - \pi_i)(A_i - a_i p)) - c(A_i - a_i p + s_i) \Big) \\ \text{s. to } 0 \leq \ & s_i \leq \epsilon_i^h \\ c \leq \ & p \leq \bar{p} \end{aligned}$$

For a given p, the problem is separable for each product i, define

$$E[r_i(p,s_i)] = (p-c)(A_i - a_i p) + (p\pi_i - c)s_i, \quad i \in N.$$

Note $E[R(p,s)] = \sum_i E[r_i(p,s_i)]$. Then, if $p\pi_i - c < 0$, $s_i^* = 0$, and if $p\pi_i - c > 0$, $s_i^* = \epsilon_i^h$.

To determine p, let $L(p)$ be the set of products that satisfy $\frac{c}{\pi_i} < p$, and $U(p)$ be the set of products that satisfy $\frac{c}{\pi_i} \geq p$. Note for a given p, $L(p) \cup U(p) = N$. Note that for $i \in L(p)$, $s_i^* = \epsilon_i^h$ and $i \in U(p)$, $s_i^* = 0$.

As p increases from c to $\bar{p}$, $L(p)$ expands from $\emptyset$ to T, where $T = \{i : \frac{c}{\pi_i} < \bar{p}\}$.

For a given p, let

$$E[R(p)] = \max_{\mathbf{s}} E[R(p,\mathbf{s})] = E[R(p,\mathbf{s}^*(p))].$$

It is possible to express $E[R(p)]$ as,

$$\begin{aligned} E[R(p)] &= \sum_{i \in L(p)} \Big(p(\pi_i(A_i - a_i p + \epsilon_i^h) + (1-\pi_i)(A_i - a_i p)) - c(A_i - a_i p + \epsilon_i^h) \Big) \\ &\quad + \sum_{i \in U(p)} \Big(p(\pi_i(A_i - a_i p) + (1-\pi_i)(A_i - a_i p)) - c(A_i - a_i p) \Big) \\ &= \sum_i (p-c)(A_i - a_i p) + p \sum_{i \in L(p)} \pi_i \epsilon_i^h + c \sum_{i \in L(p)} \epsilon_i^h \end{aligned}$$

Let the products be indexed such that $\pi_1 > \pi_2 > \cdots > \pi_n$. Then for $\frac{c}{\pi_k} \leq p \leq \frac{c}{\pi_{k+1}}$, $k = 0, 1, \ldots, n-1$, $E[R(p)]$ is defined by a different function (Define $\pi_0 = 1$). Let Region 0 be defined as $[c, \frac{c}{\pi_1}]$, Region k be defined $(\frac{c}{\pi_k}, \frac{c}{\pi_{k+1}}]$ for $k \in \{1, 2, ..., m-1\}$ and Region m be defined as $(\frac{c}{\pi_m}, \bar{p}]$. Due to upper limit on p, it is sufficient to consider only regions $\{1, 2, ..., m\}$ where m is the cardinality of T. Let $E[\hat{R}_k(p)]$ denote the $E[R(p)]$ function in Region k, $k \in \{0, 1, \cdots, m\}$. The optimal price in each region, p_k^*, is determined as follows.

Determining p_k^*.

Let P_k be defined as,

$$P_k = \frac{\sum A_i + \sum_{i \leq k} \pi_k b_k}{2 \sum a_i} + \frac{c}{2}, \qquad k \in \{0, 1, \cdots, n\}$$

Note that $P_0 < P_1 < \cdots < P_n$.

1. Region 0.

   For $c \leq p < \frac{c}{\pi_1}$, $E[\hat{R}_0(p)] = (p - c)(\sum A_i - \sum a_i p)$.

   Note that $\frac{d^2 E[\hat{R}_0(p)]}{dp^2} < 0$, and that $R_0(p)$ is concave in p.

   Under the first order condition (FOC),

   $$\frac{dE[\hat{R}_0(p)]}{dp} = 0.$$

   Note that P_0 satisfies the FOC. Then the p value that maximizes $R(p)$ in Region 0 is,

   $$p_0^* = \min\{\max\{c, P_0\}, \frac{c}{\pi_1}\},$$

   which means maximizer is either the point that satisfies FOC or the boundaries.

2. Region k, $k \in \{1, 2 \ldots, m - 1\}$.

   For $\frac{c}{\pi_k} \leq p < \frac{c}{\pi_{k+1}}$, $E[\hat{R}_k(p)] = (p - c)(\sum A_i - \sum a_i p) + p \sum_{i \leq k} \pi_i \epsilon_i^h - c \sum_{i \leq k} \epsilon_i^h$.

   Note that $R_k(p)$ is concave in p and that P_k satisfies FOC, the maximizer is either P_k or boundaries. Note P_k is maximizer if it is inside boundaries: $p_k^* = P_k$ if $\frac{c}{\pi_k} \leq P_k < \frac{c}{\pi_{k+1}}$. In the following, we show that if $p_k^* = \frac{c}{\pi_k}$, then $p^* \neq p_k^*$. In other words, optimal uniform price is never on the boundary.

3. Region m.

   If $P_m > \bar{p}$, then $p_m^* = \bar{p}$. Otherwise $p_m^* = \min\{P_m, \frac{c}{\pi_m}\}$, which means either FOC is satisfied, or boundary is maximizer.

Before determining p^*, further observations are made.

Observation 1 At $\frac{c}{\pi_k}$,

$$\lim_{p \to \frac{c}{\pi_k}^-} \frac{dE[R(p)]}{dp} = \lim_{p \to \frac{c}{\pi_k}} \frac{dE[\hat{R}_{k-1}(p)]}{dp}$$
$$= \sum A_i - \sum a_i p - (p-c) \sum a_i + \sum_{i \le k-1} \pi_i \epsilon_i^h$$
$$\lim_{p \to \frac{c}{\pi_k}^+} \frac{dE[R(p)]}{dp} = \lim_{p \to \frac{c}{\pi_k}} \frac{dE[\hat{R}_k(p)]}{dp}$$
$$= \sum A_i - \sum a_i p - (p-c) \sum a_i + \sum_{i \le k} \pi_i \epsilon_i^h.$$

Thus, at the boundary $\frac{c}{\pi_k}$, $\frac{dE[\hat{R}_{k-1}(p)]}{dp} < \frac{dE[\hat{R}_k(p)]}{dp}$.

Observation 2 Furthermore, at $\frac{c}{\pi_k}$,

$$\lim_{p \to \frac{c}{\pi_k}^-} E[R(p)] = (p-c)(\sum A_i - \sum a_i p) + p \sum_{i \le k-1} \pi_i \epsilon_i^h - c \sum_{i \le k-1} \epsilon_i^h$$
$$\lim_{p \to \frac{c}{\pi_k}^+} E[R(p)] = (p-c)(\sum A_i - \sum a_i p) + p \sum_{i \le k-1} \pi_i \epsilon_i^h - c \sum_{i \le k-1} \epsilon_i^h + p\pi_k b_k - cb_k$$
$$= \lim_{p \to \frac{c}{\pi_k}^-} E[R(p)].$$

This implies $R(p)$ is continuous in p. Note, from Observation 1 that for $p = \frac{c}{\pi_k}$, $\lim_{p \to \frac{c}{\pi_k}} \frac{dE[\hat{R}_{k-1}(p)]}{dp} < \lim_{p \to \frac{c}{\pi_k}} \frac{dE[\hat{R}_k(p)]}{dp}$. This implies that $p = \frac{c}{\pi_k}$ is not the maximizer for each k, hence boundaries corresponding to $p = \frac{c}{\pi_k}$, are excluded from solution set.

Procedure to determine p

1. For $k \in \{1, 2 \ldots, m-1\}$, compare P_k and $\frac{c}{\pi_k}$ and $\frac{c}{\pi_{k+1}}$. If for some k, $P_k < \frac{c}{\pi_k}$ or $P_k > \frac{c}{\pi_{k+1}}$, then optimal p is not in Region k. If for k=0, $p_0 > \frac{c}{\pi_1}$ and for k=m, $p_m < \frac{c}{\pi_m}$, then optimal p is not in 0 or m.

   Let $K = \{k : \frac{c}{\pi_k < p_k < \frac{c}{\pi_{k+1}}}\}$. If $p_0 > \frac{c}{\pi_1}$ or $p_m < \frac{c}{\pi_m}$, add to the set K, $\{0\}$ and $\{m\}$.

2. $$p^* = p_j^* \text{ where } j = \arg\min_{k \in K} E[\hat{R}_k(p_k^*)]$$

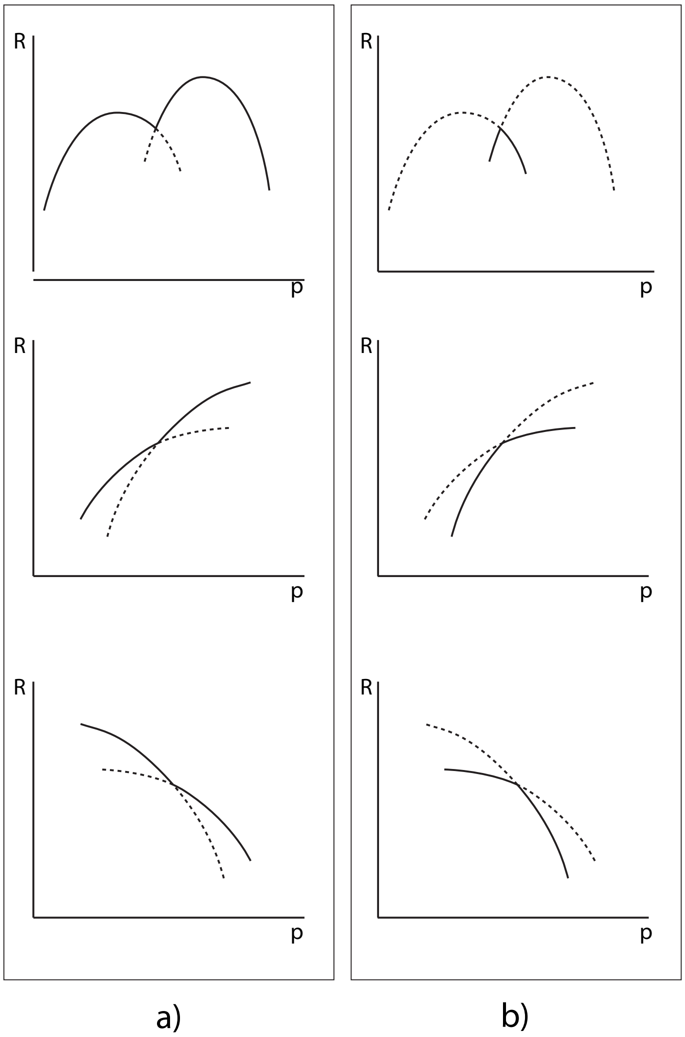


Figure 7.1: The function $E[R(p)]$ over two regions. (a) shows the possible cases whereas (b) shows the impossible cases.

Example. Consider a setting where there are 3 products with the following parameters. $A_1 = 10,\ A_2 = 13,\ A_3 = 14,\ \gamma_{12} = \gamma_{13} = \gamma_{23} = 0.15,\ \pi_1 = 0.8, \pi_2 = 0.5, \pi_3 = 0.2,\ b_1 = 2,\ b_2 = 5,\ b_3 = 6$. Cost of capacity is $c = 3$.

For this setting, $a_1 = 0.7,\ a_2 = 0.7$ and $a_3 = 0.7$. Note that $\frac{c}{\pi_i}$ is $3.75,\ 6,\ 15$ for products 1, 2 and 3, respectively. Finally, $P_0 = 10.31,\ P_1 = 10.69,\ P_2 = 11.28$ and $P_3 = 11.57$, and $\bar{p} = 14.286$.

Note for Region 0 and Region 1, $P_k > \frac{c}{\pi_{k+1}}$. Furthermore, for Region 3, $P_k < \frac{c}{\pi_k}$. This implies there is only one region to consider, Region 2. Optimal price is $p^* = P_2$.

Since $p^* = P_2$, capacity reservations for Product 1, 2 and 3 are $s_1 = 2,\ s_2 = 5,\ s_3 = 0$. Under these variable values the expected profit is 135.47.

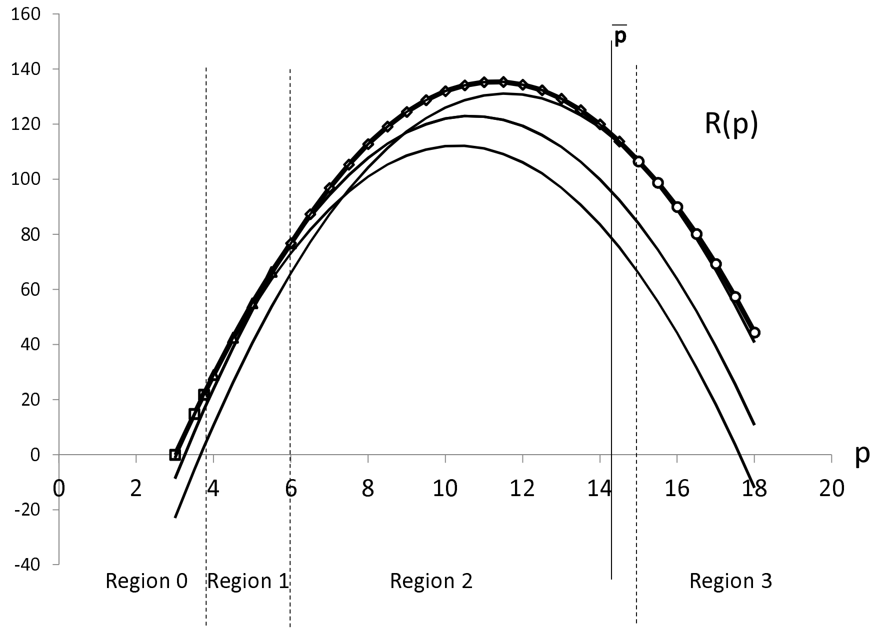


Figure 7.2: Determining the optimal p

7.4 Two period two-product problem

In this section, we assume there are two products, $i = 1, 2$. Before the demand is realized the manufacturer decides on the stock levels for both products, I_1, I_2 and the sales price, p_1. Demand is stochastic and stock levels and p_1 are determined considering the stochastic nature of the demand. After the demand is realized, it may occur that one of the products is overstock whereas the other is understock. Some of the customers, whose demand is not met, may now opt for the other product. This is termed as "stockout-based substitution". Before the substitution takes place the manufacturer may change the price of the overstock product in order to affect the substitution amount.

Other events that may occur by the end of the period is both products are understock after the demand is realized, or both products are overstock. When both products are understock, this means the initial price is set too low considering the actual valuations of the customers. In that case it is not possible to generate further sales by changing the price. If, on the other hand both products are overstock, the manufacturer may change the price to clear out the inventory. The new price both facilitates stock clearance and price-based substitution among the products.

Manufacturer's aim is to maximize the profit by deciding on the initial stock levels, the initial price, p_1, and the secondary price, p_2. Products are uniformly priced in both periods. Since there are two stages where the prices are set, we denote the problem as a two-period problem. We first determine optimal profit in the second period, and then study the first period.

We assume a linear form for the demand function. However now there is a need to define the demand function for Period 1 and Period 2. The second period demand should reflect the demand overflow due to stock-out based and price-based substitutions.

7.4.1 Demand functions

We first look at the underlying theory for the linear form of demand function, which is a commonly used functional form in the literature. In the literature, it is assumed that the linear demand function is an outcome of the consumers' utility maximization problem (see for instance Vives, 1994). More specifically, it is assumed that consumers with different tastes can be represented by a single "representative consumer." This representative consumer is assumed to gain utility by consuming a variety of products. How each product contributes to the utility of the "representative consumer" is defined by the structure of the utility function (Anderson et al., 1988).

Given a utility function of the representative consumer, we can derive the demand function by maximizing the surplus, where the surplus is defined as the *utility minus the total amount paid to the sellers of the products.*

In the following we first present how the demand function in (7.2.1) is obtained via surplus maximization for the single period problem. Then we propose a demand function for the two-period problem. Our approach in line with that of Atkins and Zhao (2009). However in Atkins and Zhao (2009) representative consumer framework is used to derive the demand function and spill rate for a single period problem. While we extend the approach to two-period problem.

Our representative consumer has a quadratic utility function expressed as follows:

$$U(\mathbf{q}) = \alpha^T \mathbf{q} - \frac{1}{2}\mathbf{q}^T C_N \mathbf{q}.$$

For our problem, there are two products, hence $N = 2$. The quantity demanded is represented by vector $\mathbf{q}$. Then, α is a size $N = 2$ vector, and C_N is $N \times N$ positive definite symmetric matrix. Suppose for the time being that no randomness exist in the coefficients. Define

$$\alpha_i = \frac{A_i + \beta A_j}{1 - \beta^2}, \qquad i, j \in \{1, 2\}, i \neq j,$$

$$C_N = \frac{1}{1-\beta^2}\begin{bmatrix} 1 & \beta \\ \beta & 1 \end{bmatrix}.$$

These two define the sensitivities of the utility function related to consumption.

The surplus of the representative consumer is $U(\mathbf{q}) - \mathbf{p}^T\mathbf{q}$, where $\mathbf{p}$ is the price vector. We get $\mathbf{q} = C_N^{-1}(\alpha - \mathbf{p})$ as the surplus maximizing quantity for the consumer. Note

$$C_N^{-1} = \begin{bmatrix} 1 & -\beta \\ -\beta & 1 \end{bmatrix}.$$

Thus $\mathbf{q} = C_N^{-1}(\alpha - \mathbf{p})$ is a linear function. Here $\mathbf{q}$ is simply the demand vector for the products. We simply obtain the demand function in Eq. 7.2.1.

Remarks

1. To include randomness in the base demand, α_i are defined as random variables as follows:

$$\alpha_i = \frac{A_i + \epsilon_i + \beta(A_j + \epsilon_j)}{1-\beta^2}, \qquad i, j \in \{1, 2\}, i \neq j.$$

2. As the structure of the utility function changes, so does the consumer's consumption behavior. For instance, it is possible to derive the discrete choice models from an appropriate utility function. Anderson et al. (1988) discusses the utility functions that lead to a logit choice model.

In the two-period model we define two demand functions, one for each period. Let the consumer's total demand quantity for product i be defined as $q_i = q_i^1 + q_i^2$. Here $q_i^t,\ t = 1, 2$ is the consumption quantity for product i in period t. If we write the intertemporal surplus of the consumer, we obtain

$$\begin{aligned} U(\mathbf{q}) - \mathbf{p}^T\mathbf{q} = {} & [\alpha_1 \;\; \alpha_2]\begin{bmatrix} q_1^1 + q_1^2 \\ q_2^1 + q_2^2 \end{bmatrix} - \frac{1}{2(1-\beta^2)}\begin{bmatrix} q_1^1 + q_1^2 \\ q_2^1 + q_2^2 \end{bmatrix}^T \begin{bmatrix} 1 & \beta \\ \beta & 1 \end{bmatrix}\begin{bmatrix} q_1^1 + q_1^2 \\ q_2^1 + q_2^2 \end{bmatrix} \\ & - p_1 q_1^1 - p_1 q_2^1 - p_2 q_1^2 - p_2 q_2^2 \end{aligned} \tag{7.4.1}$$

We assume the consumer is not strategic, i.e., the consumption decisions are taken myopically. This implies, in the first period the consumer demands the products as if there is no second period. This equivalently defines the first period demand function as:

$$
\begin{aligned}
D_i^1 &= A_i - p_1 + \beta p_1 + \epsilon_i, \;\; i,j = 1,2, i \neq j, \\
&= A_i - (1-\beta)p_1 + \epsilon_i, \;\; i,j = 1,2, i \neq j,
\end{aligned} \tag{7.4.2}
$$

D_i^1 is determined by taking derivative of the surplus function with respect to q_i^1 and ignoring q_i^2. Note the resulting demand function has the same functional form in (7.2.1).

In the second period, the consumer decides on $q_i^2, \; i = 1, 2$ to maximize second period surplus. The consumption quantity for the first period is already determined by the beginning of the second period. In the second period the consumer should maximize surplus with respect to q_1^2 and q_2^2. The consumption quantities for the first period will be treated as constants in the surplus function. Taking derivative of the surplus function in (7.4.1) with respect to q_i^2, FOC gives the surplus maximizing q_i^2 values, since the surplus is a positive definite quadratic function.

$$
\frac{\partial}{\partial q_i^2} U(\mathbf{q}) - p_2 = \alpha_i - \frac{1}{2(1-\beta^2)}[2(q_i^1 + q_i^2) + 2\beta(q_j^1 + q_j^2)] - p_2 = 0. \tag{7.4.3}
$$

Note that, in the second period, determining q_1^2 and q_2^2 is not as straightforward as in the first period. The reason is, the amount that will be sold over the periods by the seller is limited (by I_is), and the second period demand quantities should be determined taking the availability into consideration. In other words, the values q_i^2 depend on the initial stock level.

We analyze three different cases depending on the value of initial stock levels, I_i and realization of the demand in first period, $D_i^1(\omega)$ (we denote the realizations with ω). In each case, the demand function in the second period takes a different form.

Case(i) $D_i^1(\omega) \leq I_i,\ i = 1, 2$. When $I_i's$ are sufficiently high, when D_i^1 in (7.4.2) is plugged into the surplus function in (7.4.3), q_i^2 is obtained accordingly. Note that surplus maximizing q_i^2s are determined without taking into consideration the exact amount of remaining stock at the beginning of period 2. This is also the case when determining q_i^1s in period 1. The only information known by the consumer is the amount of demand satisfied in the first period for both products.

If we replace the values of D_i^1 (as defined in Eq. 7.4.2), the equations simplify to,

$$\alpha_i - \frac{1}{(1-\beta^2)}[\alpha_i - \beta\alpha_j - p_1 + \beta p_1 + q_i^2$$
$$+\beta\alpha_j - \beta^2\alpha_i - \beta p_1 + \beta^2 p_1 + \beta q_j^2] - p_2 = 0,\ i, j = 1, 2,\ i \neq j.$$

If we rearrange,

$$(1-\beta^2)(p_1 - p_2) = q_1^2 + \beta q_2^2$$
$$(1-\beta^2)(p_1 - p_2) = q_2^2 + \beta q_1^2.$$

If we simultaneously solve for q_1^2 and q_2^2, we obtain the demand function for period 2:

$$D_i^2 = (1-\beta)(p_1 - p_2),\ i = 1, 2.$$

Case(ii) $D_i^1(\omega) \leq I_i, D_j^1(\omega) \geq I_j,\ i, j \in \{1, 2\}, i \neq j$.

Without loss of generality, we take product i as product 1 and product j as product 2.

In period 1, satisfied quantity for product 1 is $q_1^1 = \alpha_1 - \beta\alpha_2 - p_1 + \beta p_1$, whereas for product 2, satisfied quantity is I_2. Hence, in (7.4.3) we plug I_2 in place of q_2^1. Furthermore, $q_2^2 = 0$ since there is no available stock for product 2 in the second period.

In order to maximize the surplus, take derivative with respect to q_1^2. From FOC we obtain:

$$\frac{\partial}{\partial q_1^2}U(\mathbf{q}) - p^2 = \alpha_i - \frac{1}{2(1-\beta^2)}[2(q_1^1 + q_1^2) + 2\beta I_2] - p_2 = 0.$$

Replacing q_1^1 with Eq. 7.4.2, the equations simplify to

$$\alpha_1 - \frac{1}{(1-\beta^2)}[\alpha_1 - \beta\alpha_2 - p_1 + \beta p_1 + q_1^2 + \beta I_2] - p_2 = 0.$$

If we solve for q_1^2,

$$q_1^2 = -\beta(\alpha_2 + \beta\alpha_1) - \beta I_2 + (p_1 - \beta p_1) - (1-\beta^2)p_2$$

If we plug in $A_2 - (1-\beta)p_1 + s_2$ in place of I_2, demand function for product 1 in the second period is obtained as

$$D_1^2 = \beta(\epsilon_2 - s_2) + (1-\beta^2)(p_1 - p_2). \qquad (7.4.4)$$

whereas $D_2^2 = 0$.

Case(iii) $D_i > I_i^1, D_j > I_j^1, i, j \in \{1, 2\}, i \neq j$

If both products are depleted by the beginning of period 2, then

$$D_i^2 = 0, i = 1, 2.$$

For an alternative explanation of the demand function see Appendix F

7.4.2 Determining the initial stock levels and prices

The manufacturer's problem is to determine the stock quantities for both products I_1 and I_2 (or, in terms of the residual stocks, s_1 and s_2), at the beginning of Period 1, and to determine the prices in Period 1 and Period 2, p_1 and p_2, that will maximize the expected profit. Let ϵ denote the vector of random variables $\epsilon = (\epsilon_1, \epsilon_2)$. Let $R(I_1, I_2, p_1, p_2, \epsilon)$ be the random variable denoting the profit over two periods:

$$\begin{aligned} R(I_1, I_2, p_1, p_2, \epsilon) &= p_1 \sum_{i=1}^{2} \min\{D_i^1, I_i\} - c(I_1 + I_2) \\ &+ p_2 \sum_{i=1}^{2} \min\{(I_i - D_i^1)^+, D_i^2\}, \qquad (7.4.5) \end{aligned}$$

where D_i^1 is a function of p_1 and ϵ_i as defined in (7.4.2), and D_i^2 is a function of $p_1,\ p_2,\ I_1,\ I_2,\ \epsilon_1,\ \epsilon_2$.

In Eq. 7.4.5, first two terms denote the profit in first period and the last term denotes the profit in the second period. Let the amount obtained in Period 1 be denoted with R_1, and the amount obtained in Period 2 be denoted with R_2. Then R_1 and R_2 are expressed as follows,

$$\begin{aligned} R_1(s_1, s_2, p_1, \epsilon) &= p_1 \sum_{i=1}^{2} (A_i - (1-\beta)p_1 + \min\{\epsilon_i, s_i\}) \\ &\quad - c(\sum_{i=1}^{2}(A_i - (1-\beta)p_1 + s_i)) \end{aligned} \tag{7.4.6}$$

$$R_2(s_1, s_2, p_1, p_2, \epsilon) = p_2 \sum_{i=1}^{2} (\min\{(s_i - \epsilon_i)^+, D_i^2\} \tag{7.4.7}$$

Let $R(I_1, I_2, p_1, p_2, (\epsilon_1(\omega), \epsilon_2(\omega))$ denote a realization of $R(I_1, I_2, p_1, p_2, \epsilon)$. The expected profit will be expressed as:

$$\begin{aligned} E[R] &= (1-\pi_1)(1-\pi_2)(R(I_1, I_2, p_1, p_2, (0,0))) \\ &+ \pi_1(1-\pi_2)(R(I_1, I_2, p_1, p_2, (\epsilon_1^h, 0))) \\ &+ (1-\pi_1)\pi_2(R(I_1, I_2, p_1, p_2, (0, \epsilon_2^h))) \\ &+ \pi_1\pi_2(R(I_1, I_2, p_1, p_2, (\epsilon_1^h, \epsilon_2^h))). \end{aligned}$$

Depending on the values of the decision variables, s_1, s_2, p_1, p_2, $E[R]$ takes different forms. It is not possible to determine the decision variables in closed-form. One approach is to exhaustively search for s_1, s_2, p_1, p_2 in the feasible region. The feasible region is constructed under the following constraints: Initial inventory level for product i, $A_i - (1-\beta)p_1 + s_i$, must be positive, price in the first period must ensure that demand in first period is (strictly) positive, without loss of optimality price in the first period must exceed cost per unit, $p_1 \geq c$, price in the second period must be non-negative. The second constraint is introduced for tractability purposes. Note that this constraint is not binding for the single product case, however it might be binding for multiple products. The feasible region is defined as $\{(s_1, s_2, p_1, p_2) : A_i - (1-\beta)p_1 + s_i \geq 0, A_i - (1-\beta)p_1 > 0,\ i = 1, 2, p_1 \geq c, p_2 \geq 0\}$.

Below, we propose an analytical approach to determine the optimal values of the decision variables. We first start with finding the optimal p_2 to maximize Period 2 profit for a given s_1, s_2, p_1. Then we determine optimal s_1, s_2, p_1 to maximize $E[R]$.

7.4.3 Determining optimal p_2 in Period 2

Let $R_2^*(s_1, s_2, p_1, \epsilon(\omega))$ denote the Period 2 profit function under optimal p_2 and under realization $\epsilon(\omega)$. We have a deterministic problem in Period 2, once the Period 2 beginning net inventory levels (K_1 and K_2) are revealed by the realization of the demands. Then

$$R_2^*(s_1, s_2, p_1, \epsilon(\omega)) = \max_{p_2} \sum_i \min((s_i - \epsilon_i(\omega))^+, D_i^2(\omega))p_2, \quad p_2 \geq 0.$$

Note Period 2 profit is independent of A_i. Thus, the profit function in the second period is symmetrical with respect to the net inventory levels of product 1 and product 2 at the beginning of Period 2.

In the following, we first make the analysis under the assumption that product 1 has lower net inventory level at the beginning of period 2. Then, we generalize as follows. Let the optimal period 2 profit obtained under given p_1, K_1, K_2 such that $K_1 \leq K_2$ be denoted with $R^2(K_1, K_2, p_1)$. Then, since in period 2 the problem is symmetric in terms of the products, a similar analysis would apply for the $K_2 \leq K_1$ case. Specifically, given that we have $R^2(K_1, K_2, p_1)$, it is possible to determine $R_2^*(s_1, s_2, p_1, \epsilon(\omega))$ based on the following relation:

$$R_2^*(s_1, s_2, p_1, \epsilon(\omega)) = R^2(\min(s_1 - \epsilon_1(\omega), s_2 - \epsilon_2(\omega)), \max(s_1 - \epsilon_1(\omega), s_2 - \epsilon_2(\omega)), p_1).$$

In the following we determine $R^2(K_1, K_2, p_1)$ for three cases: (i) $K_i \leq 0, i = 1, 2$, (ii) $K_1 \leq 0$, $K_2 > 0$, and (iii) $K_i > 0, i = 1, 2$. In all of the cases $K_1 \leq K_2$. We refer to Figure 7.3 in the analysis. The figure shows a partitioning of the space with respect to the values K_1 and K_2 can take. In the following, we show that for each partition, the form of the function R^2 changes. The partitioning is shown for both $K_1 \leq K_2$ and $K_1 \geq K_2$. Note in the figure that, the partitions are symmetrical around the $K_1 = K_2$ line.

(i) $K_i \leq 0, i = 1, 2$.

Let

$$d0 = \{(K_1, K_2) \in R \times R : K_i \leq 0\}.$$

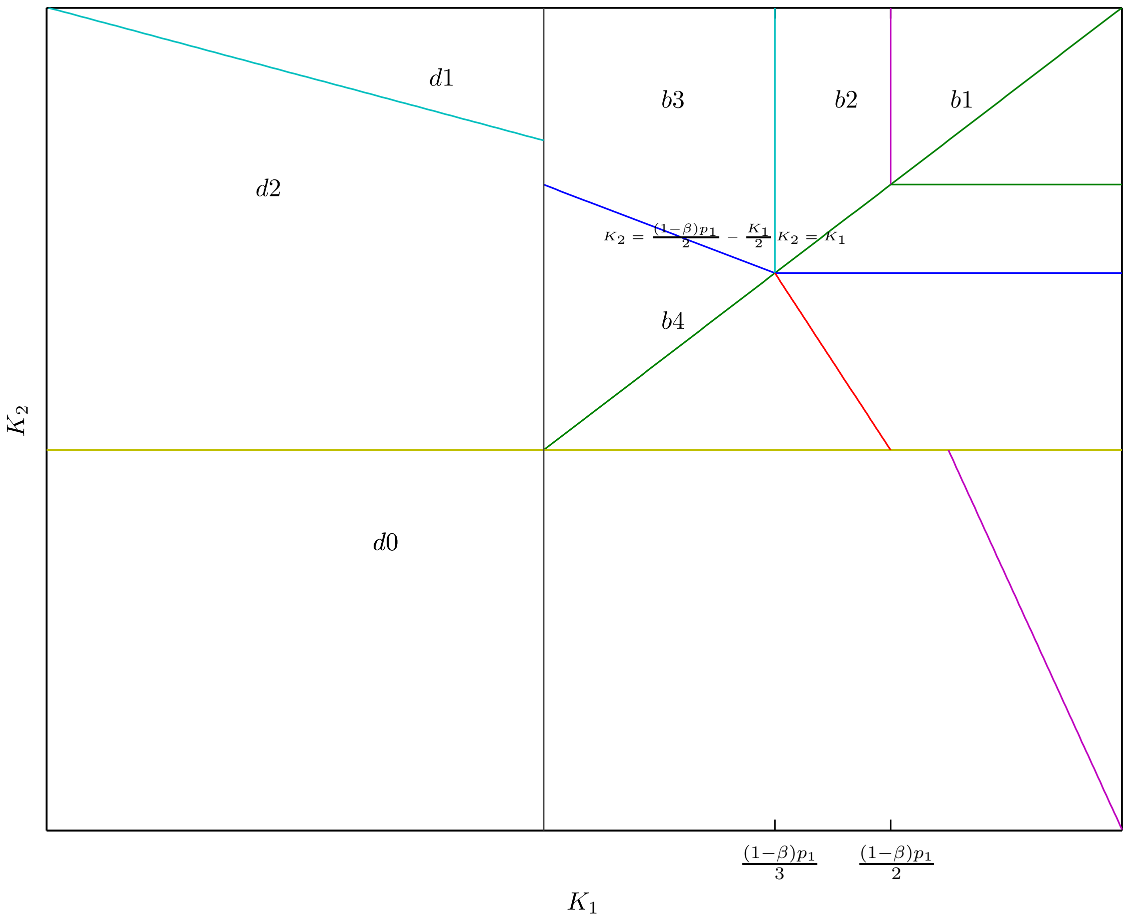


Figure 7.3: Partitioning of $R \times R$ with respect to the values K_1 and K_2 take

If there is no inventory left, i.e., for both products $K_i \leq 0$, then there will be no second period profit. Thus,

$$R^2(K_1, K_2) = 0, \ \ for \ \ (K_1, K_2) \in d0.$$

(ii) **If $K_1 \leq 0$ and $K_2 > 0$.**

In this case, there is inventory left for the product 2, $s_2 - \epsilon_2(\omega) = K_2 > 0$, whereas net inventory level for product 1 is non-negative, $s_1 - \epsilon_1(\omega) = K_1 < 0$. In other words, product 1 is understocked whereas product 2 is overstocked for at the beginning of period 2. We have defined the following demand functions for this case Eq. 7.4.4:

$$D_2^2 = -\beta K_1 + (1-\beta^2)(p_1 - p_2),$$
$$D_1^2 = 0.$$

Inventory to meet the total demand for product 2 is K_2. The problem will turn into a pricing problem in a single period, with deterministic demand D_2^2 (which is a function of p_2) and given stock level K_2.

Lemma 7.4.1. *Given $K_1 \leq 0 < K_2$,*

$$R^2(K_1, K_2, p_1) = \max_{p_2} \min\{-\beta K_1 + (1-\beta^2)(p_1 - p_2), K_2\} p_2,$$

is obtained as:

$$R^2(K_1, K_2, p_1) = \begin{cases} \frac{((1-\beta^2)p_1 - \beta K_1)^2}{4(1-\beta^2)}, & (K_1, K_2) \in d1, \\ K_2(-\frac{\beta K_1 + K_2}{1-\beta^2} + p_1), & (K_1, K_2) \in d2, \end{cases}$$

optimal p_2 is

$$p_2 = \begin{cases} \frac{((1-\beta^2)p_1 - \beta K_1)}{2(1-\beta^2)}, & (K_1, K_2) \in d1, \\ -\frac{\beta K_1 + K_2}{1-\beta^2} + p_1, & (K_1, K_2) \in d2, \end{cases}$$

where

$$d1 = \{(K_1, K_2) \in R \times R : \frac{-\beta K_1 + (1-\beta^2)p_1}{2} \leq K_2, K_1 \leq 0, K_2 > 0\},$$
$$d2 = \{(K_1, K_2) \in R \times R : K_2 < \frac{-\beta K_1 + (1-\beta^2)p_1}{2}, K_1 \leq 0, K_2 > 0\}.$$

Proof. See Figure 7.3 to see the partitions $d1$, d. The problem we would like to solve is :

$$R^2(K_1, K_2, p_1) = \max_{p_2} \min\{D_2^2, K_2\}p_2,$$

where $D_2^2 = -\beta K_1 + (1-\beta^2)(p_1 - p_2)$. Define functions $f_1(K_1, K_2, p_1, p_2)$ and $f_2(K_1, K_2, p_1, p_2)$ as $f_1 = K_2 p_2$ and $f_2 = D_2^2 p_2$. For a given K_1, K_2, p_1, $R^2(K_1, K_2, p_1) = \max_{p_2} \min\{f_1(p_2), f_2(p_2)\}$. We characterize the following points:

1. p_2 that maximizes f_2, $p_2^* = \frac{p_1}{2} - \frac{\beta K_1}{2(1-\beta^2)} \geq 0$.

2. p_2 at which f_1 and f_2 intersect. One intersection point is $p_2 = 0$. Denote the other with $\bar{p_2}$. Then $\bar{p_2} = p_1 - \frac{K_2 + \beta K_1}{1-\beta^2}$.

Then, either $p_2^* \geq \bar{p_2}$, or $p_2^* < \bar{p_2}$. If former, then $\min\{f_1, f_2\}$ is maximized at p_2^* and at p_2^*, $f_2 < f_1$. If latter, then $\min\{f_1, f_2\}$ is maximized at $\bar{p_2}$, and at $\bar{p_2}$, $f_1 = f_2$. Consider K_1 and K_2 values that make $p_2^* = \bar{p_2}$:

$$K_2 = (-\beta K_1 + (1-\beta^2)p_1)/2.$$

Given p_1, define d_1 and d_2 based on this separating plane. Plugging in the corresponding p_2^* and $\bar{p_2}$ values gives the expressions for R^2. □

(iii) **If** $K_i > 0, i = 1, 2$**.**

The problem will convert into a pricing problem with two products which have the same demand function.

$$D_i^2 = (1-\beta)(p_1 - p_2), \quad i = 1, 2,$$

with inventories K_i. Then,

$$R^2(K_1, K_2, p_1) = \max_{p_2}(\min(K_1, (1-\beta)(p_1 - p_2) + \min(K_2, (1-\beta)(p_1 - p_2))p_2,$$

if we solve this maximization problem

Lemma 7.4.2. *Given* $0 < K_1 \le K_2$,

$$R^2(K_1,K_2,p_1) = \begin{cases} (1-\beta)\frac{(p_1)^2}{2}, & (K_1,K_2)\in b1,\\ 2K_1(p_1-\frac{K_1}{(1-\beta)}), & (K_1,K_2)\in b2,\\ (1-\beta)(\frac{K_1}{2(1-\beta)}+\frac{p_1}{2})^2, & (K_1,K_2)\in b3,\\ (K_1+K_2)(p_1-\frac{K_2}{(1-\beta)}), & (K_1,K_2)\in b4,\end{cases}$$

where optimal p_2 *is*

$$p_2 = \begin{cases} \frac{p_1}{2}, & (K_1,K_2)\in b1,\\ p_1-\frac{K_1}{(1-\beta)}, & (K_1,K_2)\in b2,\\ \frac{K_1}{2(1-\beta)}+\frac{p_1}{2}, & (K_1,K_2)\in b3,\\ p_1-\frac{K_2}{(1-\beta)}, & (K_1,K_2)\in b4,\end{cases}$$

where

$$\begin{aligned} b1 &= \{(K_1,K_2)\in R\times R : \frac{(1-\beta)p_1}{2} \le K_1, 0<K_1\le K_2\},\\ b2 &= \{(K_1,K_2)\in R\times R : \frac{(1-\beta)p_1}{3} \le K_1 < \frac{(1-\beta)p_1}{2}, 0<K_1\le K_2\},\\ b3 &= \{(K_1,K_2)\in R\times R : K_1 < \frac{(1-\beta)p_1}{3}, K_2+\frac{K_1}{2} \ge \frac{(1-\beta)p_1}{2},\\ &\quad 0<K_1\le K_2\},\\ b4 &= \{(K_1,K_2)\in R\times R : K_2+\frac{K_1}{2} \le \frac{(1-\beta)p_1}{2}, 0<K_1\le K_2\}.\end{aligned}$$

Proof. See Figure 7.3 to see partitions $b1, b2, b3$ and $b4$. The problem we would like to solve is:

$$R^2(K_1,K_2,p_1) = \max_{p_2}\{\min\{K_1,(1-\beta)(p_1-p_2)\} + \min\{K_2,(1-\beta)(p_1-p_2)\}p_2\},$$

(Note that in R^2, $0 < K_1 \le K_2$)

Define functions $g_1(K_1,K_2,p_1,p_2)$, $g_2(K_1,K_2,p_1,p_2)$ and $g_3(K_1,K_2,p_1,p_2)$ as $g_1 = 2(1-\beta)(p_1-p_2)p_2$, $g_2 = K_1p_2 + (1-\beta)(p_1-p_2)p_2$ and $g_3 = (K_1 + K_2)p_2$. For a given K_1, K_2, p_1, $R^2(K_1,K_2,p_1) = \max_{p_2}\min\{g_1(p_2), g_2(p_2), g_3(p_2)\}$. Consider the p_2 values at which g_i and g_j, $i,j \in \{1,2,3\}, i \ne j$ intersect. Note that due to the structure of g_1, g_2, and g_3, each of the two functions intersect at two p_2 values only, one of which is 0. We characterize the following points:

1. p_2 that maximizes g_1: $p_2^1 = \frac{p_1}{2}$.

2. p_2 that maximizes g_2: $p_2^2 = \frac{p_1}{2} + \frac{K_1}{2(1-\beta)}$.

3. p_2 at which $g_1 = g_2$ (where $p_2 \neq 0$): $p_H = p_1 - \frac{K_1}{1-\beta}$.

4. p_2 at which $g_1 = g_3$ (where $p_2 \neq 0$): $p_M = p_1 - \frac{K_1+K_2}{2(1-\beta)}$.

5. p_2 at which $g_2 = g_3$ (where $p_2 \neq 0$): $p_L = p_1 - \frac{K_2}{1-\beta}$.

Note $p_2^2 > p_2^1$, and $p_L \leq p_M \leq p_H$.

To understand how g_1, g_2, and g_3 are positioned with respect to each other, we evaluate $\partial g_i/\partial p_2$ at $p_2 = 0$ for $i = 1, 2, 3$:

$$\partial g_1/\partial p_2|_{p_2=0} = 2(1-\beta)p_1$$

$$\partial g_2/\partial p_2|_{p_2=0} = K_1 + (1-\beta)p_1, \text{ and}$$

$$\partial g_3/\partial p_2|_{p_2=0} = K_1 + K_2.$$

Based on the derivatives, the following cases can be characterized:

1. If $(1-\beta)p_1 < K_2$, then $R^2 = \max_{p_2} \min\{g_1, g_2\}$. Note $p_L < 0$. Then optimal p_2, p_2^* is:

(i) If $p_H \leq p_2^1$, then $p_2^* = p_2^1$

(ii) If $p_2^1 \leq p_H \leq p_2^2$, then $p_2^* = p_H$.

(iii) If $p_2^2 \leq p_H$, then $p_2^* = p_2^2$.

(See Figure 7.4-a)

2. If $(1-\beta)p_1 \geq K_2$, then $0 \leq p_L \leq p_M \leq p_H$ and at $p_2 = 0^+$, $g_3 < g_2 < g_1$. Then optimal p_2 is:

(i) If $p_H \leq p_2^1$, then $p_2^* = p_2^1$.

(ii) If $p_2^1 < p_H \leq p_2^1$, then $p_2^* = p_2^2$.

(ii) If $p_2^2 < p_H$ and $p_L \leq p_2^2$, then $p_2^* = p_2^2$.

(iv) If $p_2^2 < p_L$, then $p_2^* = p_L$.

(See Figure 7.4-b)

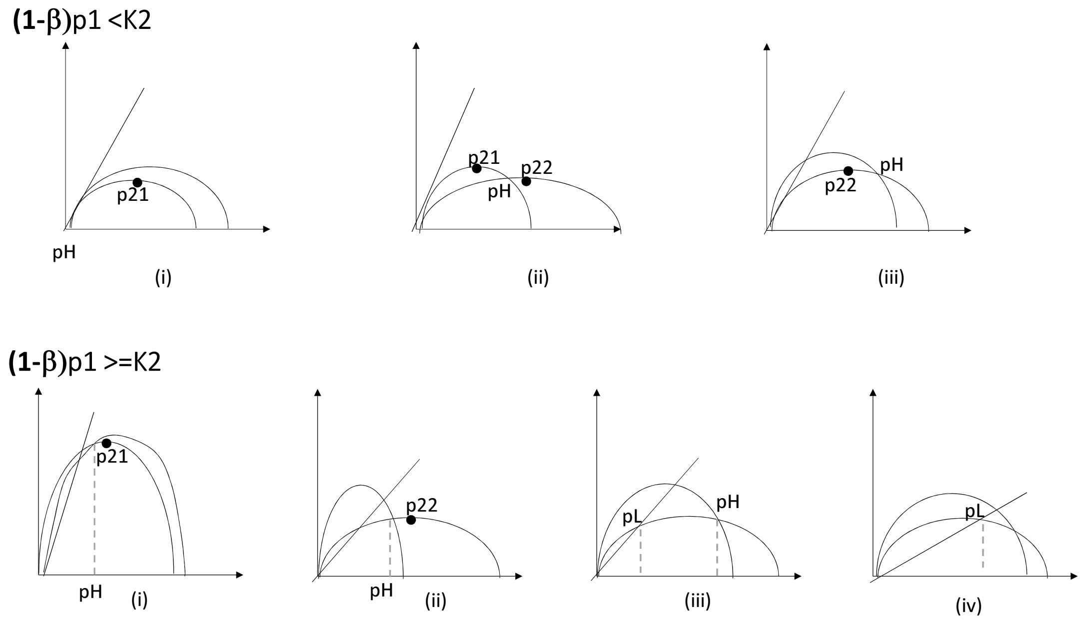


Figure 7.4: Functions g_1, g_2 and g_3. x-axis denotes p_2

In summary, depending on K_1, K_2 and p_1, R^2 takes one of the four following forms with p_2^* characterized as follows:

1. $p_H \leq p_2^1$, $p_2^* = p_2^1$, $R^2 = g_1(p_2^1)$.
2. $p_2^1 \leq p_H \leq p_2^2$, $p_2^* = p_H$, $R^2 = g_2(p_H)$.
3. $p_L \leq p_2^2 \leq p_H$, $p_2^* = p_2^2$, $R^2 = g_2(p_2^2)$.
4. $p_2^2 \leq p_L$, $p_2^* = p_L$, $R^2 = g_3(p_L)$.

Note these conditions correspond to partitions $b1, b2, b3$, and $b4$. □

Given $K_1 \leq K_2$, R^2 is characterized by 7 different functions, depending on the values of K_1 and K_2. In the following, for a given p_1 we discuss the continuity of R^2 on the region $R \times R$.

Discussion on Continuity of $R^2(K_1, K_2, p_1)$ with respect to K_1 and K_2, for a given p_1

In the following, considering partitions $b1, b2, b3, b4$ (valid for $0 < K_1 \leq K_2$) and $d0, d1, d2$ (valid for $K_1 < 0$), when (K_1, K_2) moving from one partition to the other, we check whether there is a discontinuity in R^2.

We conjecture that for $K_1 > 0$, R^2 is continuous, but, moving from $K_1 \leq 0$ to $K_1 > 0$, R^2 is not continuous. The left continuity (at $K_1 = 0$) of R^2 stems from the structure of the demand function. When only one product has overstock, it is as if possible to set two different prices in second stage (p_2^1, p_2^2). As discussed previously in § 7.4.1, for the understock product (say, product 1), the price p_2^1 depletes the unmet demand, whereas for the overstock product, the price p_2^2 maximizes the profit considering the demand for product 2 together with unmet and spill-over demand of product 1. The price p_2^1 is an implicit price, whereas p_2^2 is denoted with p_2.

We should point out that although different prices are set, there is no product sold at the depleted product's price, so in the second stage the uniformity of the price, which is realized, is intact under this scheme. However, when there is overstock for both products, a uniform price should be set for both of the products. This demand structure results in the discontinuity of R^2 at the border of these two cases (at $K_1 = 0$).

Lemma 7.4.3. *For a given p_1 (where p_1 takes values within the feasible region), and for $K_1 \leq K_2$,*

(i) When $K_2 \leq 0$, for a given K_2, $R^2(K_1, K_2, p_1)$ is continuous for all K_1,

(ii) When $K_2 > 0$, for a given K_2, $R^2(K_1, K_2, p_1)$ is continuous for all K_1 except at $K_1 = 0$, and is left continuous at $K_1 = 0$.

(iii) For a given K_1, $R^2(K_1, K_2, p_1)$ is continuous for all K_2.

Proof. $R^2(K_1, K_2, p_1)$ is a piecewise function. So in order to check continuity of the function we need to check the points where the definition of the function changes.

(i) **When** $K_2 \leq 0$**.** Since for R^2, it is defined $K_1 \leq K_2$, when $K_2 \leq 0$, $R^2(K_1, K_2) = 0$, hence it is continuous in K_1.

(ii) **When** $K_2 > 0$**.** As K_1 increases, depending on the value of K_2, definition of R^2 may change several times (actually as many as five times: from $d2$ to $d1$ to $b3$ to $b2$ to $b1$, or 4 times: from $d2$ to $b3$ to $b2$ to $b1$, or from $d2$ to $b4$ to $b3$ to $b2$, or twice: from $d2$ to $b4$).

We look at all the K_1 values at which definition of R^2 changes. Define set of partitions, $D = \{d1, d2\}$ and $B = \{b1, b2, b3, b4\}$. We first look at the continuity of R^2 at boundary of B and D and then within B and D.

1. *Continuity as K_1 moves from $d1$ to $b3$.* Equivalently, given $K_2 > \frac{(1-\beta^2)p_1}{2}$, continuity is checked at $K_1 = 0$.

$$\lim_{K_1 \to 0^-} R^2(K_1, K_2, p_1)_{|(K_1,K_2)\in d1} = (1-\beta^2)(\frac{p_1}{2})^2.$$

$$\begin{aligned}\lim_{K_1 \to 0^+} R^2(K_1, K_2, p_1)_{|(K_1,K_2)\in b3} &= \lim_{K_1 \to 0^+} (1-\beta)(\frac{K_1}{2(1-\beta)} + \frac{p_1}{2})^2, \\ &= (1-\beta)(\frac{p_1}{2})^2.\end{aligned}$$

Left limit is greater than the right limit.

2. *Continuity as K_1 moves from d2 to b3.* Given $\frac{(1-\beta)p_1}{2} < K_2 \leq \frac{(1-\beta^2)p_1}{2}$, continuity is checked at $K_1 = 0$.

$$\lim_{K_1 \to 0^-} R^2(K_1, K_2, p_1)_{|(K_1,K_2)\in d2} = K_2(-\frac{K_2}{1-\beta^2} + p_1)$$

$$\lim_{K_1 \to 0^+} R^2(K_1, K_2, p_1)_{|(K_1,K_2)\in b3} = (1-\beta)(\frac{p_1}{2})^2.$$

Left limit is greater than the right limit.

3. *Continuity as K_1 moves from d2 to b4.* Given $K_2 \leq \frac{(1-\beta)p_1}{2}$, continuity is checked at $K_1 = 0$.

$$\lim_{K_1 \to 0^-} R^2(K_1, K_2, p_1)_{|(K_1,K_2)\in d2} = K_2(p_1 - \frac{K_2}{1-\beta^2})$$

$$\begin{aligned}\lim_{K_1 \to 0^+} R^2(K_1, K_2)_{|(K_1,K_2)\in b4} &= \lim_{K_1 \to 0^+} (K_1 + K_2)(p_1 - \frac{K_2}{(1-\beta)}) \\ &= K_2(p_1 - \frac{K_2}{(1-\beta)}).\end{aligned}$$

Left limit is greater than the right limit.

We assume for a given K_2 and p_1, at $K_1 = 0$, $R^2 = R^2(0^-)$.

4. *Continuity as K_1 moves within $B = \{b1, b2, b3, b4\}$.* For $0 < K_1 \leq K_2$, for a given K_2, as K_1 increases, R^2 takes values in accordance with the partitions in B. The analysis in Lemma 7.4.2 implies R^2 is continuous in B.

5. *Continuity as K_1 moves within D.* For $K_1 < 0 < K_2$, for a given K_2, as K_1 increases, R^2 takes values in accordance with the partitions in D. The analysis in Lemma 7.4.1 implies R^2 is continuous in D.

(iii) **For a given K_1.** To check continuity with respect to K_2 for a given K_1, we look at continuity at $K_2 = 0$, and continuity within B and D.

1. *Continuity as K_2 moves from d0 to d2.*

$$\begin{aligned}\lim_{K_2 \to 0^-} R^2(K_1, K_2, p_1) &= 0, \\ \lim_{K_2 \to 0^+} R^2(K_1, K_2, p_1) &= \lim_{K_2 \to 0^+} K_2((\beta((1-\beta)p_1 - K_1) + p_1) - K_2) \\ &= 0.\end{aligned}$$

2. *Continuity as K_2 moves within B.* For $0 < K_1 \leq K_2$, for a given K_1, as K_2 increases, R^2 takes values in accordance with the partitions in B. The analysis in Lemma 7.4.2 implies R^2 is continuous in B.

3. *Continuity as K_2 moves within D.* For $K_1 < 0 < K_2$, for a given K_1, as K_2 increases, R^2 takes values in accordance with the partitions in D. The analysis in Lemma 7.4.1 implies R^2 is continuous in D. □

To complete the continuity analysis for all K_1 and K_2 values, we need to check the continuity of R^2 over the $K_1 = K_2$ line. If continuous, then can completely characterize the continuity of R_2^* over $R \times R$.Analysis in Lemmas 7.4.1 and 7.4.2 show that, $\lim_{K_2 \to K_1^+} R^2(K_1, K_2, p_1) = R^2(K_1, K_1, p_1)$. Thus R_2^* is continuous except at $K_1 = 0$ and $K_2 = 0$.

We close this part with the following results.

Property 7.4.1. *At the beginning of the second period, if both products have the same inventory level, then both products bring the same second period profit (which is possibly zero).*

Proof. Follows from the symmetrical structure of the second period demand function. □

Property 7.4.2. *At the beginning of second period, given that one product is overstock, it is more profitable to have the other product understock rather than to have it "slightly" overstock.*

Proof. Follows from the continuity analysis, since at $K_1 = 0$ left limit of R^2 is greater than the right limit. □

Total expected profit $E[R]$ is the sum of $E[R_1]$ and $E[R_2]$. Note in $E[R_2]$ there are four R_2 expressions (for each *low* and *high* combination). Each of the four terms has its corresponding K_1 and K_2 values. Each (K_1, K_2) can take values in partitions $d0, D$, or B, or in their symmetries. This makes a total of 13 partitions where (K_1, K_2) can take values in, for each of the four terms. Depending on s_1, s_2

and p_1, $E[R_2]$ can roughly take $13 \times 13 \times 13 \times 13 = 28561$ different forms in total. When determining s_1, s_2 and p_1 we need to consider the effect of a certain $(s_1, s2_, p_1)$ tuple on the form of $E[R_2]$.

7.4.4 Determining optimal I_1, I_2 and p_1 in Period 1

In this section, we determine the decision variables I_1, I_2, p_1 that maximizes the two-period profit function. Instead of the decision variables I_1 and I_2, we use the decisions variables that correspond to residual stock levels, s_1 and s_2. Lets define a partition P, on $R \times R$, $P = \{pp1, pp2, pp3, pp4, pn1, pn2, np1, np2, nn1\}$, where:

$$
\begin{aligned}
pp1 &= \{(s_1, s_2) \in R \times R : \epsilon_1^h < s_1, \epsilon_2^h < s_2\}, \\
pp2 &= \{(s_1, s_2) \in R \times R : \epsilon_1^h < s_1, 0 < s_2 \leq \epsilon_2^h\}, \\
pp3 &= \{(s_1, s_2) \in R \times R : 0 < s_1 \leq \epsilon_1^h, \epsilon_2^h < s_2\}, \\
pp4 &= \{(s_1, s_2) \in R \times R : 0 < s_1 \leq \epsilon_1^h, 0 < s_2 \leq \epsilon_2^h\}, \\
pn1 &= \{(s_1, s_2) \in R \times R : \epsilon_1^h < s_1, s_2 \leq 0\}, \\
pn2 &= \{(s_1, s_2) \in R \times R : 0 < s_1 \leq \epsilon_1^h, s_2 \leq 0\}, \\
np1 &= \{(s_1, s_2) \in R \times R : s_1 \leq 0, \epsilon_2^h < s_2\}, \\
np2 &= \{(s_1, s_2) \in R \times R : s_1 \leq 0, 0 < s_2 \leq \epsilon_2^h\}, \\
nn1 &= \{(s_1, s_2) \in R \times R : s_1 \leq 0, s_2 \leq 0\}.
\end{aligned}
$$

See Partition P in Figure 7.5. In the following $E[R(s_1, s_2, p_1, p_2, \epsilon)]$ is analyzed for each partition separately. Then optimal s_1, s_2, p_1 are determined as the variables that give the maximum of the $E[R]$'s obtained for each subset of the partition. $E[R(s_1, s_2, p_1, p_2, \epsilon)]$ is sum of $E[R_1(s_1, s_2, p_1, \epsilon)]$ and $E[R_2(s_1, s_2, p_1, p_2, \epsilon)]$. Since in Period 2, p_2^* is determined depending on the corresponding $\epsilon(\omega)$, $E[R]$ will be defined as sum of $E[R_1(s_1, s_2, p_1, \epsilon)]$ and $E[R_2^*(s_1, s_2, p_1, \epsilon)]$, and be defined as a function of s_1, s_2 and p_1 only.

We denote $E[R(s_1, s_2, p_1, p_2, \epsilon)]$ shortly with $E[R]$.

In the following, we describe how to determine s_1^*, s_2^* and p_1^* for nn1,np2,np1 and pp1. Obtaining the solution for other subsets follow similar lines. As we will se

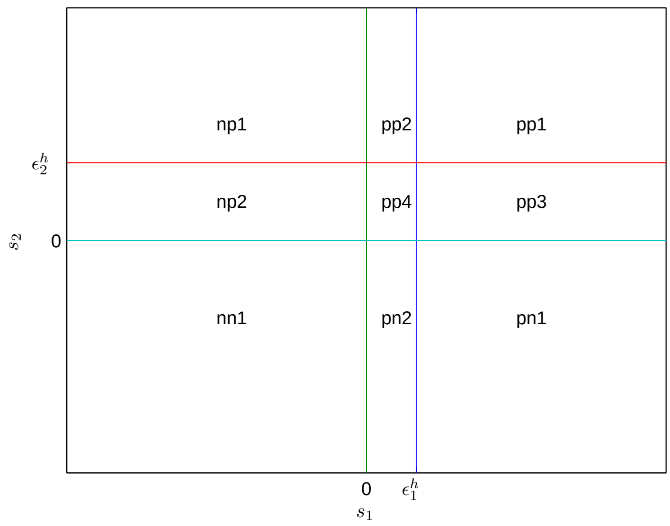


Figure 7.5: A partition of the space $R \times R$: P.

in the following, it is straightforward to determine the optimal values for nn1. Before describing the solution for subsets np2, np1 and pp1, a solution approach that will be used for analysis of the all the subsets will be introduced.

7.4.4.1 Region: nn1

In this partition, $s_1 \leq 0, s_2 \leq 0$. Thus, in Eq. 7.4.6 and Eq. 7.4.7, the terms $\min\{\epsilon_i, s_i\}$ are replaced with s_i. Since $s_i \leq \epsilon_i$, from Eq. 7.4.6 we observe that the second period profit is always zero. From Eq. 7.4.6, we end up with the following optimization problem:

$$\max E[R] = (A_1 + A_2 - 2(1-\beta)p_1 + s_1 + s_2)(p_1 - c),$$
$$s.t.$$
$$A_1 - (1-\beta)p_1 + s_1 \geq 0$$
$$A_2 - (1-\beta)p_1 + s_2 \geq 0$$
$$A_1 - (1-\beta)p_1 \geq 0$$
$$A_2 - (1-\beta)p_1 \geq 0$$
$$s_1 \leq 0$$
$$s_2 \leq 0$$
$$p_1 \geq c$$

Since p_1 is defined greater than c, first derivatives of $E[R]$ with respect to s_1 and s_2 are always positive. Hence, upper bound of (s_1, s_2) (which is $(0, 0)$ in region nn1) must be the profit maximizing values. After setting s_1 and s_2 to their maximizing values $(0, 0)$, we can find the maximizer p_1 for this region by using the first order conditions and bounds of p_1. Hence, we obtain $p_1^* = \max(\min(\frac{A_1+A_2}{2(1-\beta)}, \frac{A_1}{(1-\beta)}, \frac{A_2}{(1-\beta)}), c)$.

If $A_1 \leq A_2$, then $p^* = \max\{\frac{A_1}{(1-\beta)}, c)$. Note that $E[R]$ obtained under $(s_1, s_2, p_1) = (0, 0, \max\{\frac{A_1}{(1-\beta)}, c\})$ is a lower bound on the optimal profit.

7.4.4.2 A solution approach to find optimal solution within a subset of the Partition P

The problem of determining s_1, s_2 and p_1 values to maximize $E[R]$ within a subset of the partition P, can be formulated as finding the optimal solution of a quadratic programming problem under a set of linear constraints. In the following a solution approach is proposed for the problem.

Identifying KKT points of Quadratic Programming Problems

The structure of the total expected profit function in Eq. 7.3.3 together with the analysis in § 7.4.3 show that the functions to be maximized are all quadratic functions for each subset of partition. Furthermore, the elements of the partition P are all defined by linear inequalities. Thus, for each subset we have a "quadratic program". As a solution procedure we will devise an algorithm for identifying the KKT points of quadratic programs, but before doing this we need to give some preliminary information Bazaraa et al. (1999).

Quadratic problems have the following form:

$$\max f(x) = \frac{1}{2} x^T \ H \ x + g^T \ x$$
$$st.$$
$$Ax \leq b$$

where H is a symmetric matrix of $n \times n$ dimension. A is $m \times n$ matrix, x and g are column vectors both elements of $\mathbb{R}^n$. Let $N = \{1, 2, \ldots, n\}$ denote the set of indices of the decision variables, and $M = \{1, 2, \ldots, m\}$ denote the indices for the constraints. The symbol T indicates the transpose operator. Given a matrix B, Let B_i denote the i'th row of matrix B, B^j denote the j'th column of matrix B, B_{ij} is the ij^{th} element,$_{-i}B$ is the matrix formed by removing i'th row of matrix B, and ^{-i}B is the matrix formed by removing i'th column of matrix B.

Observe that are our problems will be of this type.

Lemma 7.4.4. *Consider the partial derivative of function $f(x)$ with respect to x_i, $i \in N$, $\frac{\partial f(x)}{\partial x_i}$. Let $[x_i]_{FOC}$ be the value that makes $\frac{\partial f(x)}{\partial x_i} = 0$. If x_i^* exists and bounded, then it is a linear function of $x_j s$, $j \in N \setminus i$.*

Proof. Wlog, we make the proof for $i = 1$.

If we take the derivative of $f(x)$ with respect to x_1:

$$\frac{\partial}{\partial x_1} f(x) = \frac{1}{2} \ H_1 \ x + \frac{1}{2} \ x^T \ H^1 + g_1.$$

x_1 setting this derivative 0 is, if $H_{11} \neq 0$:

$$[x_1]_{FOC} = \frac{1}{2 \ H_{11}} [^{-1}H_1 \ _{-1}x + [_{-1}x]^T \ _{-1}H^1 + g_1]$$

which shows such x_1 is a linear function of other variables. □

Property 7.4.3. *Suppose in the program constraint C_j holds at equality. Suppose furthermore that variable x_i has a non-zero coefficient in C_j. Then it is possible to express x_i as a linear function of other variables.*

Proof. Suppose

$$A_j\ x = b_j$$

If we write the x_i satisfying this constraint

$$[x_i]_{Cj} = \frac{1}{A_{ji}}[b_j -_{-1} [A_j]\ _{-1}x]$$

□

Definition 7.4.4 (Linear restriction of a quadratic program). We assume in the quadratic program, one of the variables is replaced with a linear function of other variables. Resulting program is named as a "linear restriction of a quadratic program."

Lemma 7.4.5. *When we look at a linear restriction of a quadratic program(QP), it is again QP, and it's optimal objective function value is equal or worse than original problem.*

Proof. Let QP be:

$$\max f(x) = \frac{1}{2}x^T\ H\ x + g^T\ x$$
$$st$$
$$A\ x \leq b.$$

Wlog, let x_1 be variable to be replaced with a linear function of other variables, i.e., we will take linear restriction along $x_1 = a^T\ _{-1}x + d$. Resulting objective function is

$$
\begin{aligned}
f(x) = & \frac{1}{2}[_{-1}x]^T \ {}^{-1}_{-1}H \ _{-1}x + [_{-1}g]^T \ _{-1}x \\
& \frac{1}{2}H^1_1[a^T \ _{-1}x + d]^T \ [a^T \ _{-1}x + d] \\
& \frac{1}{2}[_{-1}x]^T \ _{-1}H^1 \ [a^T \ _{-1}x + d] + \frac{1}{2} \ [a^T \ _{-1}x + d]^T \ {}^{-1}H_1 \ _{-1}x \\
& + g_1 \ [a^T \ _{-1}x + d] \\
& = \frac{1}{2}[_{-1}x]^T \ [{}^{-1}_{-1}H + H^1_1 \ a \ a^T + \ _{-1}H^1 \ a^T + a \ {}^{-1}H_1] \ _{-1}x \\
& + [[_{-1}g]^T + d \ a^T + \frac{d}{2}[_{-1}H^1]^T + \frac{d}{2}^{-1} H_1 + g_1 \ a^T] \ _{-1}x + \frac{1}{2}d^2 + g_1 d
\end{aligned}
$$

Hence objective function is quadratic. We also need to show new constraint set is linear.

$$
\begin{aligned}
& A \ x \leq b \\
& {}^{-1}A \ _{-1}x + A^1 \ [a^T \ _{-1}x + d] \leq b \\
& [{}^{-1}A + A^1 \ a^T] \ _{-1}x \leq b - d \ A^1
\end{aligned}
$$

We also showed, the constraint of the linear restriction is linear.

Finally, it is straightforward to show that, if $x_1 = a^T \ _{-1}x + d$ is redundant then the objective function value of the quadratic program would not change, otherwise it would decrease. □

Lemma 7.4.6. *Solution of the (original) quadratic program coincides with one of its linear restrictions on $x_i = [x_i]_{FOC}$ or $x_i = [x_i]_{Cj}$ $j \in \{1, ..., m\}$*

Proof. Wlog, we assume linear restriction is defined by $[x_1]_{FOC}$ or $[x_1]_{Cj}$. The KKT conditions of the quadratic program is

$$
\begin{aligned}
& u_0(Hx + g^T) + A^T \ u = 0 \\
& u^T(A_x - b) = 0 \\
& (u_0, u) \geq 0 \\
& (u_0, u) \neq 0.
\end{aligned}
$$

If at the optimal $u_T = 0$, then $u_0 \neq 0$ and $H \ x + g^T = 0$ which shows optimal solution satisfies $x_1 = [x_1]_{FOC}$.

If one of the $u_i \neq 0$ then associated $A_j x = b_j$ hence optimal satisfies $x_1 = [x_1]_{Cj}$

These two show that one of the restrictions satisfy the optimal solution of the original QP.

Note that from Lemma 7.4.5, QP with a linear restriction $[x_i]_{FOC}$ or $[x_1]_{Cj}$ has the same objective function value with the original QP. $\square$

With results from the previous lemmas, the following algorithm gives a subset of feasible solutions, of which one of them is the optimal solution of the quadratic program.

1. Create a linear restriction of the original QP with linear restrictions $x_1 = [x_1]_{FOC}$ and $x_1 = [x_1]_{Cj}$ $j \in \{1, ..., m\}$

2. Enlist the possible feasible solutions of the existing restriction problems with the algorithm.

3. Put the listed solutions in the x_1 formulas, to get the list of possible feasible solutions.

From the list of feasible solutions select the solution that gives the maximal result.

In the following for each element in P, we identify the corresponding QP. Then the optimal solution is determined following the approach described in § 7.4.4.2.

7.4.4.3 Region: np2 (Region: pn2)

The following QP is under consideration:

$$\begin{aligned}\max E[R] &= (A_1 + A_2 - 2(1-\beta)p_1 + s_1)(p_1 - c) + s_2(p_1\pi 2 - c) \\ &\quad + (1-\pi_2)((1-\pi_1)R^2(s_1, s_2, p_1) + \pi_1 R^2(s_1 - \epsilon_1^h, s_2, p_1)),\\ st.& \\ &A_1 - (1-\beta)p_1 \geq 0 \\ &A_2 - (1-\beta)p_1 \geq 0 \\ &s_1 \leq 0 \\ &0 < s_2 \leq \epsilon_2^h \\ &A_1 - (1-\beta)p_1 + s_1 \geq 0 \\ &p_1 \geq c\end{aligned}$$

Notice that condition $A_2 - (1-\beta)p_1 + s_2 \geq 0$ is implied by $0 < s_2$ and $A_2 - (1-\beta)p_1 \geq 0$, hence it drops. In order to find which function governs R_2, we need to look at the definition of R^2:

$$R^2(K_1, K_2, p_1) = \begin{cases} \frac{((1-\beta^2)p_1 - \beta K_1)^2}{4(1-\beta^2)}, & (K_1, K_2) \in d1, \\ K_2(-\frac{\beta K_1 + K_2}{1-\beta^2} + p_1), & (K_1, K_2) \in d2. \end{cases}$$

$(s_1, s_2) \in d_i$ and $(s_1 - \epsilon_1^h, s_2) \in d_j$ for $i, j = \{1, 2\}$ defines distinct formulation R^2 function.

Table 7.4: Possible values that (s_1, s_2) and $(s_1 - \epsilon_1^h, s_2)$ can take on the corresponding R^2 formulation

(s_1, s_2)	$(s_1 - \epsilon_1^h, s_2)$	Constraints	$R^2(s_1, s_2, p_1)$	$R^2(s_1 - \epsilon_1^h, s_2)$
$d1$	$d1$	$\frac{-\beta(s_1-\epsilon_1^h)+(1-\beta^2)p_1}{2} \leq s_2$	$\frac{((1-\beta^2)p_1-\beta s_1)^2}{4(1-\beta^2)}$	$\frac{((1-\beta^2)p_1-\beta(s_1-\epsilon_1^h))^2}{4(1-\beta^2)}$
$d1$	$d2$	$\frac{-\beta(s_1-\epsilon_1^h)+(1-\beta^2)p_1}{2} > s_2$ $\frac{-\beta(s_1)+(1-\beta^2)p_1}{2} \leq s_2$	$\frac{((1-\beta^2)p_1-\beta s_1)^2}{4(1-\beta^2)}$	$s_2(-\frac{\beta(s_1-\epsilon_1^h)+s_2}{1-\beta^2} + p_1)$
$d2$	$d2$	$\frac{-\beta s_1+(1-\beta^2)p_1}{2} > s_2$	$s_2(-\frac{\beta s_1+s_2}{1-\beta^2} + p_1)$	$s_2(-\frac{\beta(s_1-\epsilon_1^h)+s_2}{1-\beta^2} + p_1)$

$$R^2(K_1, K_2, p_1) = \begin{cases} \frac{((1-\beta^2)p_1 - \beta K_1)^2}{4(1-\beta^2)}, & (K_1, K_2) \in d1, \\ K_2(-\frac{\beta K_1 + K_2}{1-\beta^2} + p_1), & (K_1, K_2) \in d2, \end{cases}$$

Note in the table, if $(s_1, s_2) \in d_2$ then $(s_1 - \epsilon_1^h, s_2) \notin d_1$ since $\frac{\beta s_1 + (1-\beta^2)p_1}{2} > s_2$ and $\frac{\beta(s_1-\epsilon_1^h)+(1-\beta^2)p_1}{2} \leq s_2$ are contradicting inequalities.

We apply the algorithm (described in §7.4.4.2) for one of the subregions in np2: $np2 - (d1, d1)$.

7.4.4.4 Region: np2-(d1,d1)

The quadratic programming problem and the linear constraints under this subregion are as follows:

$$
\begin{aligned}
E[R] &= (A_1 + A_2 - 2(1-\beta)p_1 + s_1)(p_1 - c) + s_2(p_1\pi 2 - c) \\
&\quad + (1-\pi_2)(1-\pi_1)\frac{((1-\beta^2)p_1 - \beta s_1)^2}{4(1-\beta^2)} + \pi_1(\frac{((1-\beta^2)p_1 - \beta(s_1-\epsilon_1^h))^2}{4(1-\beta^2)})
\end{aligned}
$$

$(s.t)$

$$\frac{-\beta(s_1-\epsilon_1^h) + (1-\beta^2)p_1}{2} \leq s_2$$

$$A_1 - (1-\beta)p_1 \geq 0$$

$$A_2 - (1-\beta)p_1 \geq 0$$

$$s_1 \leq 0$$

$$0 < s_2 \leq \epsilon_2^h$$

$$A_1 - (1-\beta)p_1 + s_1 \geq 0$$

$$p_1 \geq c$$

We first identify candidate maximizer s_2 values (as a function of s_1 and p_1), then s_1 values (as a function of p_1) and finally p_1 values. Then, each candidate tuple (s_1, s_2, p_1) is checked for feasibility and optimality. Finally optimal (s_1, s_2, p_1) is identified for the corresponding subregion.

Note the first derivative of $E[R]$ with respect to s_2, is $(p_1\pi 2 - c)$. Depending on the value of p_1, either upper bound or lower bound of s_2 is a maximizer. So our maximizer candidates are

$$
\begin{aligned}
(s_2)_1 &= \epsilon_2^h, \\
(s_2)_2 &= \frac{-\beta(s_1-\epsilon_1^h) + (1-\beta^2)p_1}{2}.
\end{aligned}
$$

Now we embed $(s_2)_1$ into the original quadratic programming model to obtain the first linear restriction QP, namely, $Pr1$, and embed $(s_2)_2$ into the original quadratic programming model to obtain the second linear restriction QP, namely, $Pr2$.

$$
\begin{aligned}
Pr1 : E[R(s_1,(s_2)_1,p_1,\epsilon)] = {} & (A_1 + A_2 - 2(1-\beta)p_1 + s_1)(p_1 - c) \\
& + \epsilon_2^h(p_1\pi 2 - c) + (1-\pi_2)((1-\pi_1)\frac{((1-\beta^2)p_1 - \beta s_1)^2}{4(1-\beta^2)} \\
& + \pi_1\frac{((1-\beta^2)p_1 - \beta(s_1 - \epsilon_1^h))^2}{4(1-\beta^2)}) \\
(s.t) & \\
& \frac{-\beta(s_1 - \epsilon_1^h) + (1-\beta^2)p_1}{2} \le \epsilon_2^h \\
& A_1 - (1-\beta)p_1 \ge 0 \\
& A_2 - (1-\beta)p_1 \ge 0 \\
& s_1 \le 0 \\
& A_1 - (1-\beta)p_1 + s_1 \ge 0
\end{aligned}
$$

$$
\begin{aligned}
Pr2 : E[R(s_1,(s_2)_2,p_1,\epsilon)] = {} & (A_1 + A_2 - 2(1-\beta)p_1 + s_1)(p_1 - c) \\
& + \frac{-\beta(s_1 - \epsilon_1^h) + (1-\beta^2)p_1}{2}(p_1\pi 2 - c) \\
& + (1-\pi_2)(1-\pi_1)\frac{((1-\beta^2)p_1 - \beta s_1)^2}{4(1-\beta^2)} \\
& + \pi_1(\frac{((1-\beta^2)p_1 - \beta(s_1 - \epsilon_1^h))^2}{4(1-\beta^2)}) \\
(s.t) & \\
& \frac{-\beta(s_1 - \epsilon_1^h) + (1-\beta^2)p_1}{2} \le \epsilon_2^h \\
& A_1 - (1-\beta)p_1 \ge 0 \\
& A_2 - (1-\beta)p_1 \ge 0 \\
& s_1 \le 0 \\
& A_1 - (1-\beta)p_1 + s_1 \ge 0
\end{aligned}
$$

We now solve Pr1 and Pr2 with the algorithm. Since they have the identical feasible regions, solutions will be similar. Either lower or upper bound for the s_1, or FOC for s_1 is a maximizer. For Pr_i candidate maximizers are $(s_1)_i^j$. For $i = 1, 2$, there are $j = 4$ candidate maximizers.

$$
\begin{aligned}
(s_1)_i^1 &= 0 \\
(s_1)_i^2 &= -A + (1-\beta)p_1 \\
(s_1)_i^3 &= \frac{\beta\epsilon_1^h + (1-\beta^2)p_1 - 2\epsilon_2^h}{2\beta} \\
(s_1)_i^4 &= (s_1)_{FOC(i)}.
\end{aligned}
$$

If we calculate FOC for s_1:

$$
\begin{aligned}
(s_1)_1^4 = (s_1)_{FOC(1)} &= -\frac{(1+\frac{1}{2}\beta(1-\pi_2))p_1 - c}{\frac{1}{2}(1-\pi_2)\frac{\beta^2}{1-\beta^2}} \\
(s_1)_2^4 = (s_1)_{FOC(2)} &= -\frac{(1+\frac{1}{2}\beta(1-2\pi_2))p_1 - (1-\frac{1}{2}\beta)c}{\frac{1}{2}(1-\pi_2)\frac{\beta^2}{1-\beta^2}}
\end{aligned}
$$

For $(s_1, s_2) = ((s_1)_i^j, (s_2)_i)$, $i = 1, 2,\ j = 1, 2, 3, 4$, possible candidate p_1's will be determined(for each of the 8 cases). In the following, each candidate is denoted with ${}_k(p_1)_i^j$.

$$Pr1j : E[R((s_1)_1^j, (s_2)_1^j), p_1, \epsilon)] = (A_1 + A_2 + 2(1-\beta)p_1 + (s_1)_1^j)(p_1 - c) + \epsilon_2^h(p_1\pi 2 - c)$$
$$+ (1-\pi_2)((1-\pi_1)\frac{((1-\beta^2)p_1 - \beta(s_1)_1^j)^2}{4(1-\beta^2)}$$
$$+ \pi_1\frac{((1-\beta^2)p_1 - \beta((s_1)_1^j - \epsilon_1^h))^2}{4(1-\beta^2)})$$

$(s.t)$

$$\frac{-\beta(s_1 - \epsilon_1^h) + (1-\beta^2)p_1}{2} \leq \epsilon_2^h \quad \text{(C1)}$$
$$(s_1)_1^j \leq 0 \quad \text{(C2)}$$
$$A_1 - (1-\beta)p_1 + (s_1)_1^j \geq 0 \quad \text{(C3)}$$
$$A_1 - (1-\beta)p_1 \geq 0 \quad \text{(C4)}$$
$$A_2 - (1-\beta)p_1 \geq 0 \quad \text{(C5)}$$
$$p_1 \geq c \quad \text{(C6)}$$

$$Pr2j : E[R((s_1)_2^j, (s_2)_2^j, p_1, \epsilon)] = (A_1 + A_2 + 2(1-\beta)p_1 + (s_1)_2^j)(p_1 - c)$$
$$+ \frac{-\beta((s_1)_2^j - \epsilon_1^h) + (1-\beta^2)p_1}{2}(p_1\pi 2 - c)$$
$$+ (1-\pi_2)(1-\pi_1)\frac{((1-\beta^2)p_1 - \beta(s_1)_2^j)^2}{4(1-\beta^2)}$$
$$+ \pi_1(\frac{((1-\beta^2)p_1 - \beta((s_1)_2^j - \epsilon_1^h))^2}{4(1-\beta^2)})$$

$(s.t)$

$$\frac{-\beta((s_1)_2^j - \epsilon_1^h) + (1-\beta^2)p_1}{2} \leq \epsilon_2^h \quad \text{(C1)}$$
$$(s_1)_2^j \leq 0 \quad \text{(C2)}$$
$$A_1 - (1-\beta)p_1 + (s_1)_2^j \geq 0 \quad \text{(C3)}$$
$$A_1 - (1-\beta)p_1 \geq 0 \quad \text{(C4)}$$
$$A_2 - (1-\beta)p_1 \geq 0 \quad \text{(C5)}$$
$$p_1 \geq c \quad \text{(C6)}$$

$E[R((s_1)_i^j, (s_2)_i^j, p_1, \epsilon)]$ is only a function of p_1 in terms of decision variables. To determine the optimal p_1 that maximizes the objective function, we look at FOC and the boundaries.

$$ {}_1(p_1)_i^j = (p_1)_{FOC(i,j)}, $$

where $(p_1)_{FOC(i,j)}$ is the p_1 satisfying FOC for $E[R((s_1)_i^j, (s_2)_i^j, p_1, \epsilon)]$. Other possible values for p_1 related with np2 are

$$ {}_2(p_1)_i^j = c \tag{bC6} $$

$$ {}_3(p_1)_i^j = \frac{A_1}{1-\beta} \tag{bC4} $$

$$ {}_4(p_1)_i^j = \frac{A_2}{1-\beta} \tag{bC5} $$

Considering the constraint C1:

$$ {}_5(p_1)_i^1 = \frac{-\beta\epsilon_1^h + 2\epsilon_2^h}{2(1-\beta^2)} = (p_1)_{C1(i,1)} $$

$$ {}_5(p_1)_i^2 = \frac{\beta(-A_1 + (1-\beta)p_1 - \epsilon_1^h) + 2\epsilon_2^h}{2(1-\beta^2)} = (p_1)_{C1(i,2)} $$

$$ {}_5(p_1)_i^4 = (p_1)_{C1(i,4)} $$

where $(p_1)_{C1(i,j)}$ is the p_1 satisfying $C1(\frac{-\beta(s_1-\epsilon_1^h)+(1-\beta^2)p_1}{2} \leq \epsilon_2^h)$ for $(s_1, s_2) = ((s_1)_i^j, (s_2)_i^j)$. For j=3, condition is already met. If we calculate the last p_1

$$ {}_5(p_1)_1^4 = \frac{2\epsilon_2^h - \beta\epsilon_1^h + \frac{2c(1-\beta^2)}{(1-\pi_2)\beta}}{(1-\beta^2) + (1-\beta^2)\frac{2+\beta(1-\pi_2)}{\beta(1-\pi_2)}} $$

$$ {}_5(p_1)_2^4 = \frac{2\epsilon_2^h - \beta\epsilon_1^h + \frac{2c(1-\beta^2)(1-\beta)}{(1-\pi_2)\beta}}{(1-\beta^2) + (1-\beta^2)\frac{2+\beta(1-2\pi_2)}{\beta(1-\pi_2)}} $$

For Condition $C2(s_1 \leq 0)$, j=1, j=2 are already satisfied.

$$ {}_6(p_1)_i^3 = \frac{2\epsilon_2^h - \beta\epsilon_1^h}{2\beta(1-\beta^2)} = (p_1)_{C2(i,3)} $$

$$ {}_6(p_1)_i^4 = (p_1)_{C2(i,4)} $$

if we calculate the last p_1

$$
\begin{aligned}
{}_6(p_1)_1^4 &= \frac{c}{(1+\frac{1}{2}\beta(1-\pi_2))} \\
{}_6(p_1)_1^4 &= \frac{c(1-\beta)}{(1+\frac{1}{2}\beta(1-2\pi_2))}
\end{aligned}
$$

Since both of the p_1's are less than c, we will not consider these. For Constraint 3, $C3(A_1 - (1-\beta)p_1 + s_1 \geq 0)$. Note j=1, j=2 are already satisfied.

$$
\begin{aligned}
{}_7(p_1)_i^3 &= \frac{A_1 + \frac{\beta\epsilon_1^h - 2\epsilon_2^h}{2\beta}}{(1-\beta) - \frac{(1-\beta^2)}{2\beta}} = (p_1)_{C3(i,3)} \\
{}_7(p_1)_i^4 &= (p_1)_{C3(i,4)}
\end{aligned}
$$

if we calculate the last p_1

$$
\begin{aligned}
{}_7(p_1)_1^4 &= \frac{-A_1(\frac{1}{2}(1-\pi_2)\frac{\beta^2}{1-\beta^2}) + c}{(1-\beta)(\frac{1}{2}(1-\pi_2)\frac{\beta^2}{1-\beta^2}) + (1+\frac{1}{2}\beta(1-\pi_2))} \\
{}_7(p_1)_1^4 &= \frac{-A_1(\frac{1}{2}(1-\pi_2)\frac{\beta^2}{1-\beta^2}) + (1-\beta)c}{(1-\beta)(\frac{1}{2}(1-\pi_2)\frac{\beta^2}{1-\beta^2}) + (1+\frac{1}{2}\beta(1-2\pi_2))}
\end{aligned}
$$

Let ${}_k(s_1)_i^j$ denote $(s_1)_i^j|p_1 =_k$. Each $(s_1, s_2, p_1) = ({}_k(s_1)_i^j, {}_k(s_2)_i^j, {}_k(p_1)_i^j)$ which is feasible, is a candidate for optimal solution. We check whether a candidate is feasible, and then evaluate. Finally, the tuple that gives the maximum objective function value is identified. Note for instance, for the region $np2-d1d1$, there exist a total of 21 candidate tuples.

For the remaining elements of the set P, we only identify the QP. We do not pursue the solution, since the solution approach is analogical to that of $np2-d1d1$.

7.4.4.5 Region: np1 (Region: pn1)

The objective function of the QP under consideration is:

$$\begin{aligned}E[R] &= (A_1 + A_2 - 2(1-\beta)p_1 + s_1)(p_1 - c) + \epsilon_2^h p_1 \pi 2 - cs_2 \\ &+ \pi_2(\pi_1 R^2(s_1, s_2 - \epsilon_2^h, p_1) + (1-\pi_2)R^2(s_1 - \epsilon_1^h, s_2 - \epsilon_2^h, p_1)) \\ &+ (1-\pi_2)(\pi_1 R^2(s_1, s_2, p_1) + (1-\pi_2)R^2(s_1 - \epsilon_1^h, s_2, p_1)), \quad (s_1, s_2) \in np1\end{aligned}$$

Depending on s_1, s_2, and p_1 and the realization of ϵ, R_2 function changes. Specifically, relation between s_1 and s_2 defines for each realization of ϵ the form of R_2. There are 5 possible forms that

$$\begin{aligned}&\pi_2(\pi_1 R_2(s_1, s_2 - \epsilon_2^h) + (1-\pi_2)R_2(s_1 - \epsilon_1^h, s_2 - \epsilon_2^h)) \\ &+(1-\pi_2)(\pi_1 R_2(s_1, s_2) + (1-\pi_2)R_2(s_1 - \epsilon_1^h, s_2)),\end{aligned}$$

and accordingly $E[R]$ can take.

1. **Cond1-np1.** If

$$\frac{-\beta s_1 + (1-\beta^2)p_1}{2} \leq s_2 - \epsilon_2^h, \qquad (d1(s_1, s_2 - \epsilon_2^h))$$

   then $(s_1, s_2 - \epsilon_2^h) \in d1$, $(s_1 - \epsilon_1^h, s_2 - \epsilon_2^h) \in d1$, $(s_1, s_2) \in d1$ and $(s_1 - \epsilon_1^h, s_2) \in d1$. Then, considering the form that R_2 takes under each case

$$\begin{aligned}&\pi_2(\pi_1 R_2(s_1, s_2 - \epsilon_2^h) + (1-\pi_2)R_2(s_1 - \epsilon_1^h, s_2 - \epsilon_2^h)) \\ &+(1-\pi_2)(\pi_1 R_2(s_1, s_2) + (1-\pi_2)R_2(s_1 - \epsilon_1^h, s_2)),\end{aligned}$$

   and $E[R]$ is obtained accordingly.

2. **Cond2-np1.** If the conditions,

$$\frac{-\beta s_1 + (1-\beta^2)p_1}{2} > s_2 - \epsilon_2^h$$
$$\frac{-\beta(s_1 - \epsilon_1^h) + (1-\beta^2)p_1}{2} \leq s_2 - \epsilon_2^h,$$

hold, then $(s_1, s_2 - \epsilon_2^h) \in d2$, $(s_1 - \epsilon_1^h, s_2 - \epsilon_2^h) \in d1$, $(s_1, s_2) \in d1$ and $(s_1 - \epsilon_1^h, s_2) \in d1$. Then, considering the form that R_2 takes under each case $E[R]$ is obtained accordingly.

3. **Cond3-np1.** If the conditions,

$$\frac{-\beta(s_1 - \epsilon_1^h) + (1 - \beta^2)p_1}{2} > s_2 - \epsilon_2^h,$$
$$\frac{-\beta s_1 + (1 - \beta^2)p_1}{2} > s_2$$

hold, then $(s_1, s_2 - \epsilon_2^h) \in d2$, $(s_1 - \epsilon_1^h, s_2 - \epsilon_2^h) \in d2$, $(s_1, s_2) \in d1$ and $(s_1 - \epsilon_1^h, s_2) \in d1$. Then, considering the form that R_2 takes under each case $E[R]$ is obtained accordingly.

4. **Cond4-np1.** If the conditions,

$$\frac{-\beta s_1 + (1 - \beta^2)p_1}{2} > s_2,$$
$$\frac{-\beta(s_1 - \epsilon_1^h) + (1 - \beta^2)p_1}{2} \leq s_2$$

hold, then $(s_1, s_2 - \epsilon_2^h) \in d2$, $(s_1 - \epsilon_1^h, s_2 - \epsilon_2^h) \in d2$, $(s_1, s_2) \in d2$ and $(s_1 - \epsilon_1^h, s_2) \in d1$. Then, considering the form that R_2 takes under each case $E[R]$ is obtained accordingly.

5. **Cond5-np1.** Finally, if the condition,

$$\frac{-\beta(s_1 - \epsilon_1^h) + (1 - \beta^2)p_1}{2} > s_2$$

hold, then $(s_1, s_2 - \epsilon_2^h) \in d2$, $(s_1 - \epsilon_1^h, s_2 - \epsilon_2^h) \in d2$, $(s_1, s_2) \in d2$ and $(s_1 - \epsilon_1^h, s_2) \in d2$. Then, considering the form that R_2 takes under each case $E[R]$ is obtained accordingly.

In Table 7.5, for a given (s_1, s_2) and p_1, we present for each possible beginning period 2 inventory level, (K_1, K_2), which element of the partition $D = \{d1, d2\}$ is assumed. Note for a given (s_1, s_2), (K_1, K_2) can take the values $(s_1, s_2 - \epsilon_2^h)$, $(s_1 - \epsilon_1^h, s_2 - \epsilon_2^h)$, (s_1, s_2), and $(s_1 - \epsilon_1^h, s_2)$.

Table 7.5: Conditions and elements of the partition D that each (K_1, K_2) assumes

Condition	$(s_1, s_2 - \epsilon_2^h)$	$(s_1 - \epsilon_1^h, s_2 - \epsilon_2^h)$	(s_1, s_2)	$(s_1 - \epsilon_1^h, s_2)$
Cond1-np1	d1	d1	d1	d1
Cond2-np1	d2	d1	d1	d1
Cond3-np1	d2	d2	d1	d1
Cond4-np1	d2	d2	d2	d1
Cond5-np1	d2	d2	d2	d2

Note there are only 5 different forms that $E[R]$ can take in np1. Structure under pn1 is exactly the same except that indices must be switched. Once the objective function and the constraints are identified, a similar approach to np2 is followed.

7.4.4.6 Region: pp1

In this element of the partition P, expected profit function is defined as follows:

$$\begin{aligned} E[R] &= (A_1 + A_2 - 2(1-\beta)p_1)(p_1 - c) + (1-\pi_1)(1-\pi_2)R_2(s_1, s_2) \\ &+ \pi_1(1-\pi_2)R_2(s_1 - \epsilon_1^h, s_2, p_1)(1-\pi_1)\pi_2 R_2(s_1, s_2 - \epsilon_2^h) \\ &+ \pi_1\pi_2 R_2(s_1 - \epsilon_1^h, s_2 - \epsilon_2^h, p_1) - c(s_1 + s_2), (s_1, s_2) \in pp1 \end{aligned}$$

In order to find which function governs R_2, we first look at s_1 and s_2. We know that

$$R_2^*(s_1, s_2, p_1) = \begin{cases} R^2(s_1, s_2, p_1) & \text{if } s_1 < s_2 \\ R^2(s_2, s_1, p_1) & \text{if } s_1 \geq s_2 \end{cases}$$

Hence we make partition $A = \{a1, a2, a3, a4, a5\}$ where

$$a1 = \{(s_1, s_2) \in R \times R : s_1 < s_2 - \epsilon_2^h\}$$

$$a2 = \{(s_1, s_2) \in R \times R : s_1 - \epsilon_1^h \leq s_2 - \epsilon_2^h, s_1 \geq s_2 - \epsilon_2^h\}$$

$$a3 = \{(s_1, s_2) \in R \times R : s_1 < s_2, s_1 - \epsilon_1^h > s_2 - \epsilon_2^h\}$$

$$a4 = \{(s_1, s_2) \in R \times R : s_1 - \epsilon_1^h \leq s_2, s_1 \geq s_2\}$$

$$a5 = \{(s_1, s_2) \in R \times R : s_1 - \epsilon_1^h > s_2\}$$

The partition borders are selected in order to ensure symmetry which will ease the calculations. In fact, along the borders, there exist continuity.For these partitions the function R_2^* is defined as in the table below (in the table, we omit p_1 in R_2^* to ease the representation):

Area in pp1	$R_2^*(s_1-\epsilon_1^h, s_2)$	$R_2^*(s_1, s_2)$	$R_2^*(s_1-\epsilon_1^h, s_2-\epsilon_2^h)$	$R_2^*(s_1, s_2-\epsilon_2^h)$
a1	$R^2(s_1-\epsilon_1^h, s_2)$	$R^2(s_1, s_2)$	$R^2(s_1-\epsilon_1^h, s_2-\epsilon_2^h)$	$R^2(s_1, s_2-\epsilon_2^h)$
a2	$R^2(s_1-\epsilon_1^h, s_2)$	$R^2(s_1, s_2)$	$R^2(s_1-\epsilon_1^h, s_2-\epsilon_2^h)$	$R^2(s_2-\epsilon_2^h, s_1)$
a3	$R^2(s_1-\epsilon_1^h, s_2)$	$R^2(s_1, s_2)$	$R^2(s_2-\epsilon_2^h, s_1-\epsilon_1^h)$	$R^2(s_2-\epsilon_2^h, s_1)$
a4	$R^2(s_1-\epsilon_1^h, s_2)$	$R^2(s_2, s_1)$	$R^2(s_2-\epsilon_2^h, s_1-\epsilon_1^h)$	$R^2(s_2-\epsilon_2^h, s_1)$
a5	$R^2(s_2, s_1-\epsilon_1^h)$	$R^2(s_2, s_1)$	$R^2(s_2-\epsilon_2^h, s_1-\epsilon_1^h)$	$R^2(s_2-\epsilon_2^h, s_1)$

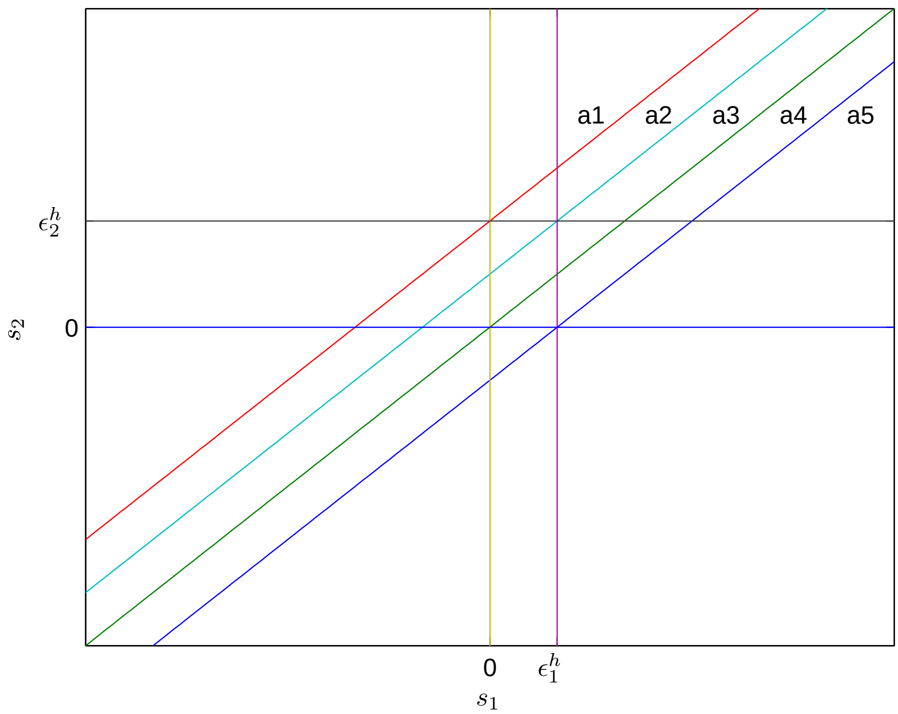


Figure 7.6: Partition P and A

The following lemmas state that regions a_1 and a_5 cannot include the optimal s_1 and s_2.

Lemma 7.4.7. *The directional derivative* R^2 *at every* $(K_1, K_2) \in \{(K_1, K_2) \in R \times R : K_1 \leq K_2\}$ *along the vector* $\boldsymbol{d} = (1, -1)$ *is positive.*

Proof. The function we will take derivative is

$$R^2(K_1, K_2, p_1) = \begin{cases} (1-\beta)\frac{(p_1)^2}{2} & (K_1, K_2) \in b1 \\ 2K_1(p_1 - \frac{K_1}{(1-\beta)}) & (K_1, K_2) \in b2 \\ (1-\beta)(\frac{K_1}{2(1-\beta)} + \frac{p_1}{2})^2 & (K_1, K_2) \in b3 \\ (K_1 + K_2)(p_1 - \frac{K_2}{(1-\beta)}) & (K_1, K_2) \in b4 \end{cases}$$

Then directional derivative along $\mathbf{d}$

$$R^2((K_1, K_2, p_1), \mathbf{d})' = \lim_{\lambda \to 0^+} \frac{R^2(K_1 + \lambda, K_2 - \lambda, p_1) - R^2(K_1, K_2, p_1)}{\lambda},$$

For area b1

$$\begin{aligned} R^2((K_1, K_2, p_1), \mathbf{d})' &= \lim_{\lambda \to 0^+} \frac{(1-\beta)\frac{(p_1)^2}{2} - (1-\beta)\frac{(p_1)^2}{2}}{\lambda} \\ &= 0, (K_1, K_2) \in b1 \end{aligned}$$

For area b2

$$\begin{aligned} R^2((K_1, K_2, p_1), \mathbf{d})' &= \lim_{\lambda \to 0^+} \frac{2(K_1 + \lambda)(p_1 - \frac{(K_1+\lambda)}{(1-\beta)}) - 2K_1(p_1 - \frac{K_1}{(1-\beta)})}{\lambda} \\ &= 2(p_1 - 2\frac{K_1}{(1-\beta)} > 0, (K_1, K_2) \in b2 : K1 \leq \frac{(1-\beta)p_1}{2} \end{aligned}$$

For area b3

$$\begin{aligned} R^2((K_1, K_2, p_1), \mathbf{d})' &= \lim_{\lambda \to 0^+} \frac{(1-\beta)(\frac{K_1+\lambda}{2(1-\beta)} + \frac{p_1}{2})^2 - (1-\beta)(\frac{K_1}{2(1-\beta)} + \frac{p_1}{2})^2}{\lambda} \\ &= \frac{1}{2}(\frac{K_1}{2(1-\beta)} + \frac{p_1}{2}) > 0, (K_1, K_2) \in b3 : K_1 \geq 0 \end{aligned}$$

For area b4

$$\begin{aligned} R^2((K_1, K_2, p_1), \mathbf{d})' &= \lim_{\lambda \to 0^+} \frac{(K_1 + K_2)(p_1 - \frac{K_2-\lambda}{(1-\beta)}) - (K_1 + K_2)(p_1 - \frac{K_2}{(1-\beta)})}{\lambda} \\ &= \frac{K_1 + K_2}{(1-\beta)} > 0, (K_1, K_2) \in b4 : K_1 \geq 0, K_2 \geq 0 \end{aligned}$$

□

Lemma 7.4.8. *In pp1, in area a1 and a5, $E[R]$ has improving direction $\mathbf{d} = (1, -1)$. Hence optimal point for $E[R]$ cannot be in the interior of a1 and a5. Since intersection of pp1 and a1 and a5 are defined as open sets, the optimal point is not included in a1 and a5 for pp1.*

Proof. For area a1,

$$\begin{aligned}E[R] &= (A_1 + A_2 - 2(1-\beta)p_1)(p_1 - c) + (1-\pi_1)(1-\pi_2)R^2(s_1, s_2, p_1)\\ &+ \pi_1(1-\pi_2)R^2(s_1 - \epsilon_1^h, s_2, p_1)(1-\pi_1)\pi_2 R^2(s_1, s_2 - \epsilon_2^h, p_1)\\ &+ \pi_1\pi_2 R^2(s_1 - \epsilon_1^h, s_2 - \epsilon_2^h, p_1) - c(s_1 + s_2)\end{aligned}$$

Hence directional derivative along $\mathbf{d}$

$$\begin{aligned}E[R'((s_1, s_2, p_1)], \mathbf{d}) &= (1-\pi_1)(1-\pi_2)R^2((s_1, s_2, p_1), \mathbf{d})'\\ &+ \pi_1(1-\pi_2)R^2((s_1 - \epsilon_1^h, s_2, p_1), \mathbf{d})'\\ &+ (1-\pi_1)\pi_2 R^2((s_1, s_2 - \epsilon_2^h, p_1), \mathbf{d})'\\ &+ \pi_1\pi_2 R^2(((s_1 - \epsilon_1^h, s_2 - \epsilon_2^h, p_1), \mathbf{d})'.\end{aligned}$$

From Lemma 7.4.7, we know each element is positive, hence $R'((s_1, s_2, p_1), \mathbf{d}) > 0$, which shows for interior of a1 $\mathbf{d}$ is an improving direction, which concludes there cannot be an optimal in the interior of a1. Ditto for a5. □

From Lemma 7.4.8, we conclude in order to find optimal in pp1, we need to consider only subregions a2, a3, and a4.

Region: pp1, Area:a2 (Area: a4)

$$a2 = \{(s_1, s_2) \in pp1 : s_1 - \epsilon_1^h \leq s_2 - \epsilon_2^h \leq s_1 \leq s_2\}$$

In this region and area, resulting expected profit function is:

$$\begin{aligned}E[R(s_1, s_2, p_1, \epsilon)] &= (A_1 + A_2 - 2(1-\beta) * p_1)(p_1 - c) + (1-\pi_1)(1-\pi_2)R^2(s_1, s_2)\\ &+ \pi_1(1-\pi_2)R^2(s_1 - \epsilon_1^h, s_2) + (1-\pi_1)\pi_2 R^2(s_2 - \epsilon_2^h, s_1)\\ &+ \pi_1\pi_2 R^2(s_1 - \epsilon_1^h, s_2 - \epsilon_2^h) - c(s_1 + s_2), \qquad (s_1, s_2) \in pp1 \bigcap a2\end{aligned}$$

Now we need find R^2's formulas based on partition B. In the following we list where the tuple $\{(s_1, \epsilon_1^h, s_2 - \epsilon_2^h), (s_2 - \epsilon_2^h, s_1), (s_1, s_2), (s_1 - \epsilon_1^h, s_2)\}$ falls within B.

- If $s_1 - \epsilon_1^h > \frac{(1-\beta)p_1}{2}$, then all other satisfy the inequality and thus tuple falls in (b_1, b_1, b_1, b_1).

- If $\frac{(1-\beta)p_1}{3} < s_1 - \epsilon_1^h < \frac{(1-\beta)p_1}{2}$, and $s_2 - \epsilon_2^h > \frac{(1-\beta)p_1}{2}$, then tuple falls in (b_2, b_1, b_1, b_2).
- If $\frac{(1-\beta)p_1}{3} < s_1 - \epsilon_1^h, s_2 - \epsilon_2^h < \frac{(1-\beta)p_1}{2}$, and $s_1 > \frac{(1-\beta)p_1}{2}$, then tuple falls in (b_2, b_2, b_1, b_2).
- If $s_1 - \epsilon_1^h < \frac{(1-\beta)p_1}{3}$ and $\frac{(1-\beta)p_1}{3} < s_2 - \epsilon_2^h < \frac{(1-\beta)p_1}{2}$ and $s_1 > \frac{(1-\beta)p_1}{2}$, then tuple falls in (b_4, b_2, b_1, b_3) or (b_3, b_2, b_1, b_3).

We list only some examples. Other cases are obtained following similar lines.

By interchanging s_1 and s_2, and ϵ_1^h and ϵ_2^h, the identified areas also hold for a4.

7.4.4.7 The remainder of the analysis

In the remainder of the analysis, for each remaining element of the partition P and its intersection with the partition $A = \{a_1, a_2, a_3, a_4, a_5\}$, we obtain all possible $E[R]$ functions that can be formed. Note that the form of $E[R]$ function changes with the possible values of the Period 2 beginning net inventory levels. Specifically, we consider the following regions,

- Region: pp1, Area: a3
- Region: pp2, Area: a1, a2
- Region: pp3, Area: a3, a4 (symmetrical to pp2-a2), a5 (symmetrical to pp2-a1)
- Region: pp4, Area: a2 ,a3, a4 (symmetrical to a2)

We list all possible combinations that (K_1, K_2) assume under each of these partitions, in Table G.1. We present the analysis in Appendix G.

CHAPTER 8

NUMERICAL STUDY FOR CONSUMER-DRIVEN SUBSTITUTION

To gain insights on the management of substitutable items, we analyze through a numerical study how decisions and performance measures are affected by the parameters. The numerical study consists of two parts. In the first part, we identify the performance measures of interest and observe how changes in parameters affect those measures. In the second part, we contrast the proposed models (single period model, two-period model under optimal inventory and pricing decisions, and two-period model with markdown-pricing as a late opportunity), and quantify the benefit of incorporating stock-out based substitution into inventory and pricing decisions.

8.1 The effect of the parameters on decisions and on the performance measures

We study the effect of parameters under two settings: symmetric products and asymmetric products. For the symmetric products case, all parameters are assumed identical. Specifically, demand bases are identical, $A_1 = A_2 = 100$, and the unit procurement cost, c, assumed values in $\{1, 5, 10\}$. The substitution parameter β is assumed in set $\{0, 0.2, 0.4, 0.6, 0.8\}$. Demand potential, ϵ_i^h is assumed in $\{10, 30, 70\}$ and probability of occurrence of this demand, π_i values are in $\{0.1, 0.5, 1\}$. We also look at the parameter settings where expected value of the demand noise, $E[\epsilon_i] = \epsilon_i^h \pi_i,\ i = 1, 2$, is constant while the variance of the demand changes. We set $\epsilon_i^h \pi_i = \{5, 10, 15, 20, 25, 30\}$, $i = 1, 2$, and $\pi_i = \{0.1, 0.2, 0.4, 0.6, 0.8\}$. In these settings, to keep $E[\epsilon_i]$ constant, as π_i value increases, ϵ_i^h is decreased. Note as π_i increases, the standard deviation of the ϵ_i decreases. For the symmetric products setting, a total of $3 \times 5 \times 3 \times 3 + 3 \times 5 \times 6 \times 6 = 675$ instances are run.

In the asymmetric products case, we fix $A_2 = 100$, while $A_1 = \{110, 120, 150, 200, 300\}$. Unit procurement cost is assumed as $c = 10$, and the substitution parameter β is assumed in set $\{0, 0.20.40.60.8\}$. Finally, $\epsilon_i^h \in \{0.1A_i, 0.3A_i, 0.7A_i\}$ and $\pi_i \in \{0.1, 0.5, 1\}$, $i = 1, 2$. For asymmetric products a total of $5 \times 5 \times 3^2 \times 3^2 = 2025$ instances are run.

The measures we consider are:

- Total expected profit, $\max_{p_1, s_1, s_2} E[R]$
- Period 1 price level p_1
- Inventory slack, s_i, $i = 1, 2$
- Initial stock levels, $I_i = A_i - (1 - \beta)p_1 + s_i$, $i = 1, 2$
- Quantity sold at initial price, $E[\min(D_i^1, I_i)]$, $i = 1, 2$, and quantity sold at second period $E[\min(D_i^2, s_i - \epsilon_i)]$, $i = 1, 2$
- Quantity overstocked at the end of first period $E[(I_i - D_i^1)^+]$, $i = 1, 2$, and at the end of the selling horizon $E[(s_i - \epsilon_i - D_i^2)^+]$, $i = 1, 2$ where $a^+ = max(0, a)$.
- Quantity understocked at the end of first period $E[(I_i - D_i^1)^-]$, $i = 1, 2$, and at the end of the selling period, $E[(s_i - \epsilon_i - D_i^2)^-]$, $i = 1, 2$, where $a^- = max(0, -a)$.

We do not include the expected quantity sold at markup price in the performance measures. The reason is, the numerical results show that second period price is almost always lower than the first period price (Only in 4 out of 2700 instances markup is observed with probability 0.01, i.e., when probability of occurrence of additional demand is very low and in the occasion of "high" demand realization for both products). Specifically, when one product is overstock whereas the other is understock, in the second period demand of both products are consolidated under the overstock product. If the overstock quantity is lower than the substitute quantity of the understock product, then it is preferable to increase the price in the second period. Numerical results show that in the second period p_2 is almost never set greater than p_1. Note, we are not able to infer this result from the analysis in Section 8.5.3.

8.1.1 Symmetrical products

We first analyze how the first period price (p_1) and the inventory slack (s) are affected by the changes in parameters. Then the effect of the parameters on sales, overstock and understock quantities are analyzed.

Effect of the parameters on the first period price, p_1

We make the following observations:

1. Price increases with c and with β, where the effect of the former is minor (when c increases 10 times, the increase in price is only around 1%). The reason for the minor increase is, the base demand A is much higher compared to unit cost. As β increases, cross-price effect increases, and this means the total demand for the two products are less likely to be lost with an increase in price. Thus, demand in effect is less sensitive to price, and the price increases with β. If the two products were owned by independent competing sellers, we would observe the opposite.
2. It is observed in a majority of the cases that as ϵ^h increases, the first period price p_1 increases. This is expected, since the increase in demand potential (ϵ^h) leads to an intention to reap additional revenue through increased

price. An increase in price results in a decrease in the net demand base (i.e., demand base eroded by the price). When c is low, or when cross-price effect (β) is high, the increase in price is more pronounced. However, under low β, high c and low π (e.g. when $\beta = 0$, $c = 10$ and $\pi = 0.1$) price does not change with ϵ_h. The stability in price is due to the fact that, all these three factors, lead to a more conservative price decision and thus change in ϵ_h is not sufficient to affect the price.

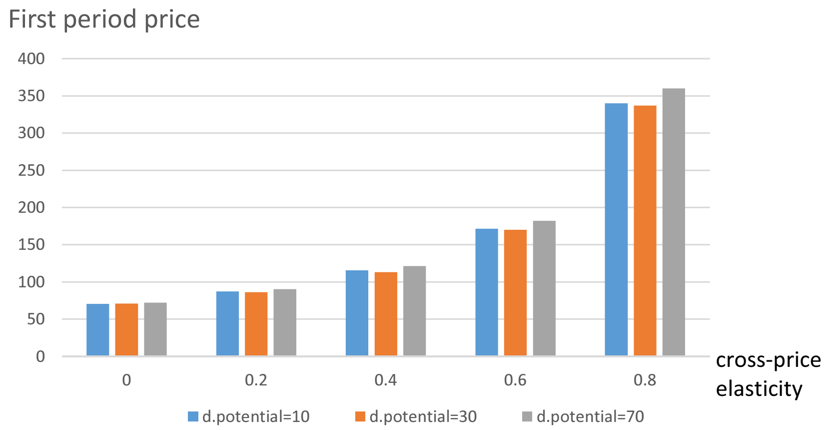
3. When the products have identical demand bases (A_i), but non-identical demand potentials (ϵ_h^i), we may observe a counter-intuitive effect of ϵ_h on the price. For instance there are cases where, when for one product ϵ_h is increased while for the other product ϵ_h is kept constant, the price may decrease. When the counter-intuitive instances are examined, it is observed that in those instances the response of inventory slack values (s_i) to a change in ϵ_h are also counter-intuitive, in that, an increase in ϵ_h decreases the inventory slack. The inventory slack not only decreases with ϵ_h, but also is set exactly equal to ϵ_h under the optimal policy.

   It is observed that there is one-to-one correspondence between the counter-intuitive response of inventory slack values and the price decrease. As ϵ_h increases, if inventory slack decreases, then this implies the optimal policy aims to leave one product overstock while the other out-of-stock.

   These instances are observed when at least for one of the products $\pi = 0.1$. For higher π values, the effect of ϵ_h on price follows the intuition (ie., price increases with ϵ_h) . the reason is, when π is low, leaving one product overstock while the other out-of-stock is a more profitable strategy. When π is high, on the other hand, the optimal policy prefers to leave the products overstock with the same second period inventory level for both.

Effect of the parameters on the inventory slacks, s_1 and s_2

The s_1 and s_2 values denote the "inventory slack" for product 1 and product 2, respectively. These values state how much to order above the deterministic part of the demand. The inventory slack is kept for the purpose of hedging against demand uncertainty, and to cover up for the second period demand due to price



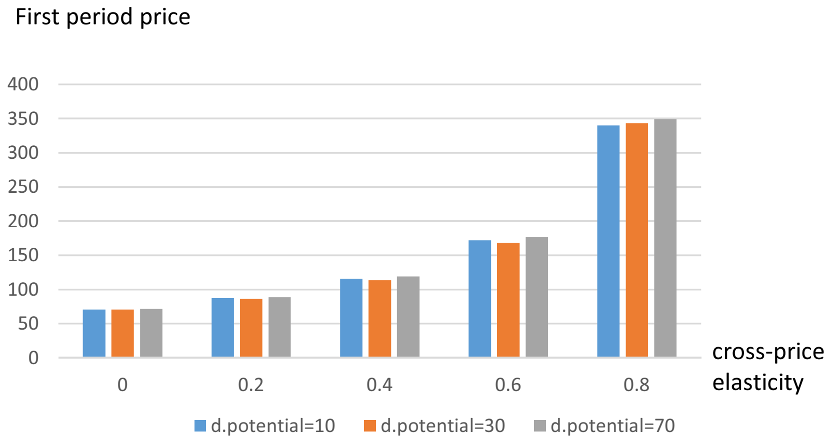


Figure 8.1: The effect of demand potential and cross-price elasticity on first period price. In the figure, $\pi_1 = \pi_2 = 0.1$. In (a) demand potentials are identical, whereas in (b) demand potential of product 1 is kept constant at 10 while demand potential for product 2 changes from 10 to 70.

decrease. In other words, even if there is no demand uncertainty, the quantity replenished might still exceed the first period demand. We make the following observations:

1. We observe that for the majority of the cases (when π_1 and π_2 are greater than 0.1) as the cross-price elasticity increase the inventory slack increases. As cross-price elasticity increases, substitution among the products increase and products' demands become less sensitive to the price. Thus the first period price increases. An increase in price means higher profit margin, i.e., lower implied overstock cost. Thus, inventory slack increases. An increase in ϵ_h also increases the inventory slack.

2. When the probability of occurrence of (potential) demand is low (when

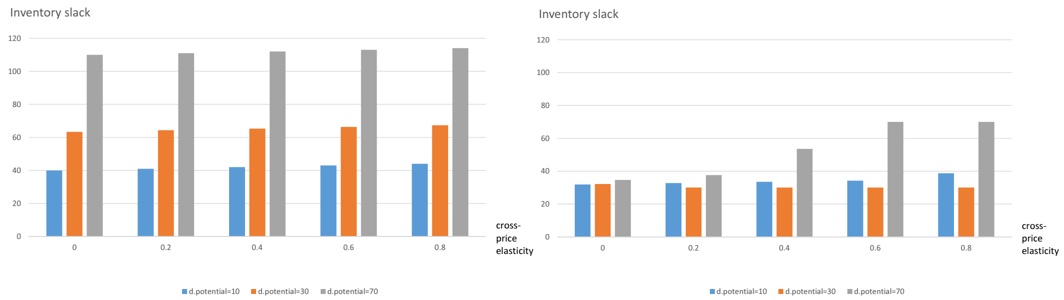


Figure 8.2: Effect of demand potential and cross-price elasticity on inventory slack for product 1. In (a) $\pi_1 = \pi_2 = 0.5$, and in (b) $\pi_1 = 0.1,\ \pi_2 = 0.5$.

either π_1 or $\pi_2 = 0.1$), then an increase in demand potential, ϵ_i^h, may actually decrease the inventory slack. See for instance in Figure 8.2-b, that when $\beta \in \{0.2, 0.4, 0.6\}$, as demand potential increases from 10 to 30, inventory slack decreases to 30. Interestingly, when inventory slack decreases with ϵ_i^h, then it is set exactly equal to ϵ_i^h under the optimal policy. The reason is, when the probability of occurrence of additional demand is low, then there is a relatively high probability that for one product additional demand occurs while for the other it does not. In that case, it is more profitable to leave one product overstock and the other out-of-stock (with zero backorders) at the end of period 1. As stated in Property 8.5.2, the manufacturer prefers this case over the case where both products are (slightly) over stock, since out-of-stock occasion gives the manufacturer flexibility when setting the markdown price of the overstock item. When both items are overstock, second-period markdown prices must be the same, and this deteriorates the profits. When the probability of occurrence of demand is low ($\pi_i = 0.1,\ i = 1, 2$), that s_i is "artificially" set equal to ϵ_i^h, governs around half of the cases. Furthermore, as ϵ_h increases it is more likely that this case is observed, since the benefit due to markdown price flexibility increases (when $\epsilon_h = 10$, s_i is always higher than ϵ_h, when $\epsilon_h = 30$ in 24% of the cases $s_i = \epsilon_h$ and when $\epsilon_h = 70$ in 56% of the cases $s_i = \epsilon_h$). The other factor that increases the likeliness of $s_i = \epsilon_h$ is the unit procurement cost, as unit procurement cost increases it is more likely that optimal decision is to set $s_i = \epsilon_h$. When $c = 1$, we do not observe $s_i = \epsilon_i^h$. This is because low c implies low overstock cost and thus s_i values are set more aggressively. Finally, we note that when $s_i = \epsilon_i^h$, it might hold that although products are symmetrical, s_i values may not be identical.

3. We also look at the response of inventory slack as expected value of additional demand, $\epsilon_h \pi_i$ increases. An increase in $\epsilon_h \pi_i$ results in an increase in inventory slack. The increase is more pronounced under low unit procurement cost.

   Under constant $\epsilon_h \pi$, as π is increased s_i decreases. Note that for a given expected (potential) demand, an increase in π implies a decrease in ϵ_h, which implies a decrease in variance of the demand potential. Thus, inventory slack decreases. The decrease is more pronounced under low unit procurement cost. We also observe that when expected potential demand is kept constant, under low π inventory slack is below ϵ_h and as π increases, inventory slack (although decreasing) exceeds ϵ_h.

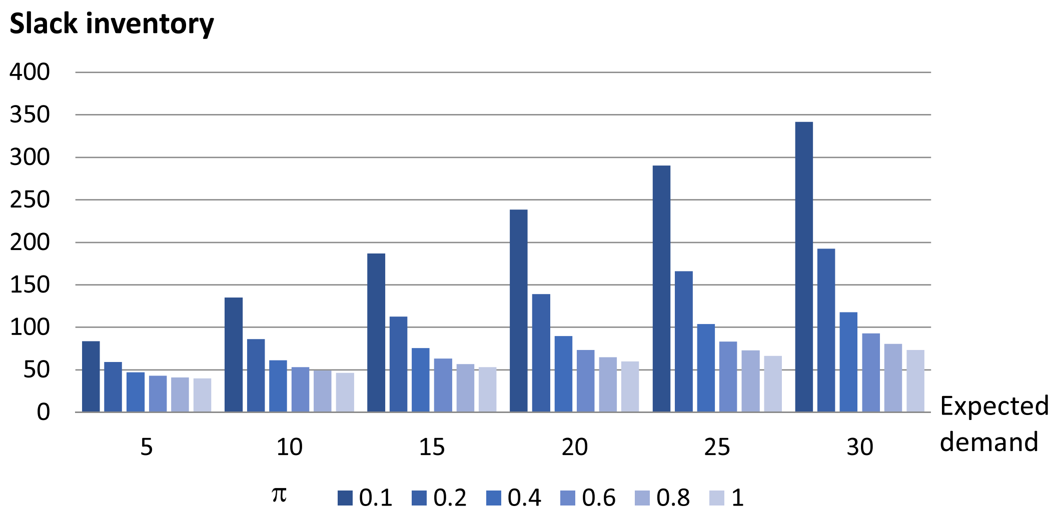


Figure 8.3: Effect of probability of demand potential and expected demand on slack inventory.

Effect of parameters on sales

We make the following observations:

1. When the products are identical, in a majority of the cases (approx. two-thirds of the cases), quantity sold at markdown price is almost equal to

quantity sold at initial price. We contrast this result with that under the stylized, linear demand, two-period, single product pricing problem (see Talluri and Van Ryzin (2006)) [1]. In the stylized problem, optimal first period and second period prices are $p_1 = \frac{2A+cb}{3b}$ and $p_2 = \frac{A+2cb}{3b}$, respectively. The corresponding sales quantities are $D_1 = D_2 = \frac{A-cb}{3}$.

Numerical analysis show that although there is uncertainty in demand, in a majority of the cases the decisions closely follow those of the stylized problem. In many of the remaining cases, quantity sold at markdown price is much lower, and those are the cases where either p_1 or $s_i,\ i = 1, 2$ are restricted by some factor. Specifically, when $\pi_i = 1,\ i = 1, 2$, and demand potential is high $\epsilon_i^h = 70$, sales in the second period is almost half of those in the first period. In those cases, the demand potential is so high that price cannot be set high enough to exploit the demand potential, since it is bounded by $\frac{A_i}{1-\beta},\ i = 1, 2$. Since the first period price is lower than optimal, this results in a decrease in second-period demand base, and thus sales in the second period drop. The other class of cases where sales drop in second period corresponds to those where π_i's are not necessarily high, but s_i is set exactly equal to ϵ_i^h with an effort to increase the probability of zero inventory for one product and positive for the other. In those cases, s_i values are asymmetrical as well. Finally, we observe that when demand potential is very high but probability of occurrence of that potential is low, s_i values are much lower compared to demand potential. This results in a slight increase in sales under markdown price compared to the sales under initial price.

2. As β increases, sales in first and second period increases, which is due to decreased sensitivity of the demand to price. An increase in probability of occurrence of demand potential also increases expected sales. However, an increase in demand potential may result in a decrease in expected sales for

[1] In the stylized two-period pricing problem, the linear demand functions for the first and the second period are as follows:$D_1 = A - bp_1$, $D_2 = b(p_1 - p_2)$. Given that the unit procurement cost is c, the two-period profit function is

$$\text{Profit} = p_1 D_1 + p_2 D_2 - c(D_1 + D_2) p_1(A - bp_1) + p_2 b(p_1 - p_2) - c(A - bp_2).$$

Solving for optimal $p_1,\ p_2$ gives $p_1 = \frac{2A+cb}{3b}$ and $p_2 = \frac{A+2cb}{3b}$. The corresponding sales quantities are $D_1 = D_2 = \frac{A-cb}{3}$.

both periods, when the probability of occurrence is low. This is because an increase in demand potential may result in a decrease in $s_i,\ i = 1, 2$ relative to $\epsilon_i^h,\ i = 1, 2$, and thus first period expected sales may decrease.

3. An increase in c decreases the expected sales in the second period. However, it may or may not decrease the sales in the first period, depending on the probability of occurrence of the demand. When probability is low, in a number of cases inventory slack are set equal to ϵ_i^h. This results in a decrease in inventory slack as c increases, while an increase in the deterministic part of the demand $(A_i - (1 - \beta)p_1)$. When expected demand potential $(\pi_i eps_i^h)$ is kept constant, an increase in ϵ_i^h does not affect the expected sales in any of the periods.

Effect of parameters on overstock and understock

Before the observations we can infer the following from the structure of the problem:

1. In both periods, for both products, the items might be overstock or understock. This is due to the uncertainty inherent in the demand. Specifically, if both items are overstock at the beginning of the second period, and if the overstock quantities are the same, then optimal markdown pricing policy is not to leave any unmet demand at the end of the selling horizon. Seller might prefer to leave some unsold items though. If overstock quantities at the beginning of second period are different, then markdown price could leave one item overstock whereas the other one understock. In that case, the understock amount would exceed the difference in overstock levels at the beginning of the first period. It is never optimal to leave both items understock at the end of the selling horizon.

2. If one of the items is overstock whereas the other is understock at the beginning of the second period, then the inventory of the overstock item will meet the demand of both products in the second period. Then, at the end of the selling horizon it might be preferable to leave some unsold stock, or totally meet the demand.

3. If both items are understock at the beginning of the second period, then the understock quantity will not change at the end of the selling horizon.

We now present the observations on quantities overstock and understock:

1. When the probability of occurrence of the demand potential is very high, overstock at the end of the second period is almost zero. When the probability of occurrence is low ($\pi_i = 0.1,\ i = 1, 2$), understock is observed either at both periods, or in the second period. As the demand potential gets higher while the probability of occurrence of such high potential is small, more conservative inventory slack is kept. As a result either understock is observed in both periods, or overstock in the first and understock in the second period is observed. Observing understock is more likely as c increases. This observation follows the intuition.

2. An increase in β increases both the expected sales and overstock quantities. Increase in β results in a decrease in price-sensitivity. This results in an increase in first period price, and implies an increase in profit margin, which results in an increase in inventory slack. Both first period and second period overstock increase (except for the cases where s_i are "artificially" set to ϵ_i^h).

3. When expected demand potential ($\pi_i \epsilon_i^h$) is kept constant while ϵ_i^h is increased, overstock quantities at the end of both periods increase.

8.1.2 Asymmetrical products

We make the following observations:

1. As demand base for product 1 increases (as A_1 increases), total expected profit increases. We observe that expected profit due to product 1 increases, whereas expected profit due to product 2 decreases. This is because, an increase in A_1 results in an increase in first period price p_1. For product 2 this means "a drift from the optimal price". For product 2,

the increase in price decreases the demand more than necessary, since its demand base is limited at $A_2 = 100$. For product 1 on the other hand, the increase in price is limited and demand for product 1 is more than necessary, since A_2 limits the first period price (see Figure 8.3). We also observe that as β increases, the average revenue increases. This is because as the products become more substitutable, the firm can increase its price without losing too much of the customer body.

Finally, an increase in A_1 increases the total expected profit in period 1, and generally increases the profit in period 2. However, it is observed that when probability of occurrence of demand is very low while A_1 is very high, an increase in A_1 may decrease the second period profit. When A_1 is increased "beyond a threshold value", the increase does not affect the price p_1 (since A_2 imposes a limit), but the inventory slack only. And if demand potential is high while the probability of demand occurrence is low, even s_i values do not increase with A_1. Then, an increase in A_1 may result in a decrease the stock levels at the end of the first period, and thus profit in the second period may decrease.

2. As A_1 increases, for product 1, both the amount of sold in period1 and period 2 increases For product 2 we observe that the amount sold in period 1 decreases while amount sold in period 2 increases. As stated before, as A_1 increases p_1 increases, and this explains the decrease of product 2 sales in period 1. For product 1, as A_1 gets higher, both price and demand gets higher. As a result, expected sales for product 1 in period 1 increases. For second period, as A_1 increases p_1 increases, which results in an increased "demand base" for second period. Hence, the amount sold in second period increases for both products. (see Figure 8.4 below)

3. We observe that inventory slack for product 1 and product 2 increase with A_1, so inventory slack is susceptible to demand base of either product. This can be explained through mechanisms of p_1, as there is more base demand, we can increase the price of the product, p_1. This results in higher demand values for the second period, and in order to compensate for that, both s_1 and s_2 increase.

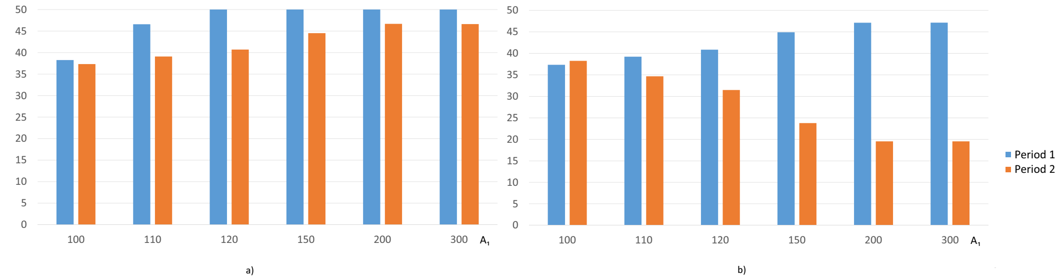


Figure 8.4: Amount of sold for both products. For (a) product 1 (b) product 2.

Under low π_1 (when $\pi_1 = 0.1$) and relatively low ϵ_1^h ($\epsilon_1^h = 0.3A_1$), s_1 is sensitive to β. As β increases, s_1 increases sharply, whereas when β is low, s_1 is conservative. As π_1 increases, sensitivity of s_1 to β decreases. When the probability of demand occurrence is low, whether uncertainty is hedged with high s_1 is related with the price. As β increases, p_1 increases, hence resulting in higher s_1 values. However when the probability of demand occurrence is 1, there is no randomness, resulting in robustness with respect to β when s_1 is considered. (see Figure 8.5)

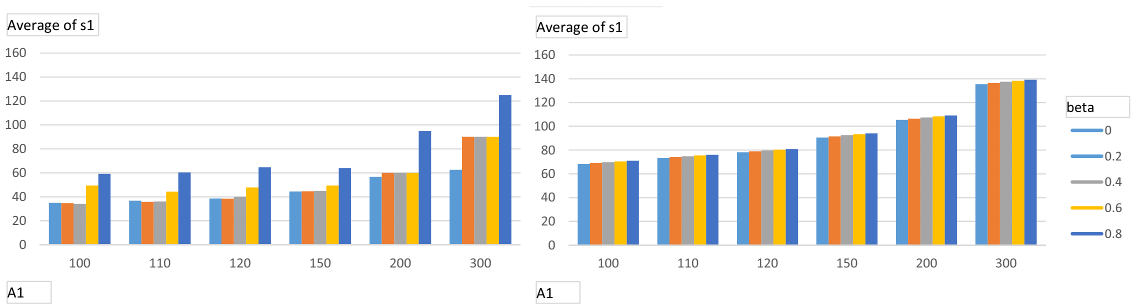


Figure 8.5: Average s_1 values according to $\pi_1 = 0.1$ and $\pi_1 = 1$

4. Expected quantity overstock at the end of first period is in range [56,317.88], whereas expected quantity understock is in range [0.066, 16.28]. At the end of the selling horizon, those ranges are [0028, 219.889] and [001, 16.28], respectively. Expected overstock quantity is much higher than expected understock quantity due to two reasons. First of all there is a second selling period and even in the absence of uncertainty, there would be overstock quantity. Secondly, price is much higher than unit procurement cost due to high values the demand base take (A_1 with respect to c). Thus, implied unit cost of understock is much higher compared to overstocking cost.

In 101 out of 2025 cases, in the first period expected understock is positive, and in 663 of 2025 cases in the second period expected understock is positive. Even if first period ends with overstock for both of the products, it may be preferable to leave one of the products understock at the end of the second period. Therefore observing understock at the end of the selling horizon is more likely than observing understock at the end of the first product. Product asymmetry seem to result in an increase in expected understock quantity. When products are symmetric, in 11 out of 45 cases expected second period understock quantity is positive, whereas that number when products are asymmetric, in 663 of 2025 cases we observe positive expected understock quantity.

8.2 The value of exploiting the stock-out-based substitution and the value of the optimal solution

Table 8.1: Model Summary

	First period	Second period	Period 1 price	Inventory decision	Period 2 price
Single period model	✓		before demand realization	before demand realization	
Two period model	✓	✓	before demand realization considering second period	before demand realization considering second period	based on remaining inventories
Two period model with late substitution	✓	✓	before demand realization as if there is one period	before demand realization as if there is one period	based on remaining inventories

In this part we compare the results obtained from single period scenario, two period model and two period model with late substitution. Define $Profit_{sp}$ as the total expected profit under the *single period model*, $Profit_{tp}$ as the total expected profit under the *two-period* model with late substitutions, and $Profit_{opt}$ as the total expected profit under the two-period optimal model. By the definition of the models, it holds that $Profit_{sp} < Profit_{tp} < Profit_{opt}$.

The results are obtained under the following parameter setting: we fix $A_2 =$

10, while $A_1 = \{10, 11, 12, 15, 20, 30\}$. Unit procurement cost is assumed as $c = 9$, and the substitution parameter β is assumed in $\{0, 0.2, 0.4, 0.6, 0.8\}$. Finally, $\epsilon_i^h \in \{0.1A_i, 0.3A_i, 0.7A_i\}$ and $\pi_i \in \{0.1, 0.5, 1\}$, $i = 1, 2$. A total of $6 \times 5 \times 3^2 \times 3^2 = 2430$ instances are run.

The following observations are made:

- The difference in profits of single-period and two-period optimal model gives the value of benefiting from stockout-based substitution in the presence of recourse pricing (second period pricing). We observe that as the base demand for product 1, A_1 increases, the "relative" benefit from stockout based substitution diminishes. As A_1 increases, the profit under both models increase and this results in a decrease in the "relative" benefit. We observe that under all β values except $\beta = 0$, the actual profit difference between single-period and two-period optimal models increase. When $\beta = 0$, and when A_1 is sufficiently high, further increase in A_1 brings limited benefit to two-period optimal model, since first period price hits a limit due to the restriction by A_2. Since first period price is bounded, a high portion of demand is met in the first period, which results in diminishing discrepancy between the models. As β increases, both the substitution event occurs at a higher rate, and the limit on the price is now less restrictive. Both effects results in increased the benefit of stockout based substitution.(see Figure 8.6)

  The difference between the profits under the two-period scenario with non-optimal solution, and under the single-period scenario gives *the benefit of a late markdown pricing*. We observe that for small to moderate β ($\beta \leq 0.4$), there is no benefit of markdown pricing due to stockout-based substitution or due to clearance. The reason is when β is small, initial price (p_1) is also small, which results in the slack inventories (s_i) to be set to 0. And when $s_i = 0$, there will be no overstock at the end of period 1, thus there is no inventory to meet the demand in period 2. The inventory slack increases with the probability of demand occurrence or with the demand potential. In either case, we observe benefit of a second period pricing is

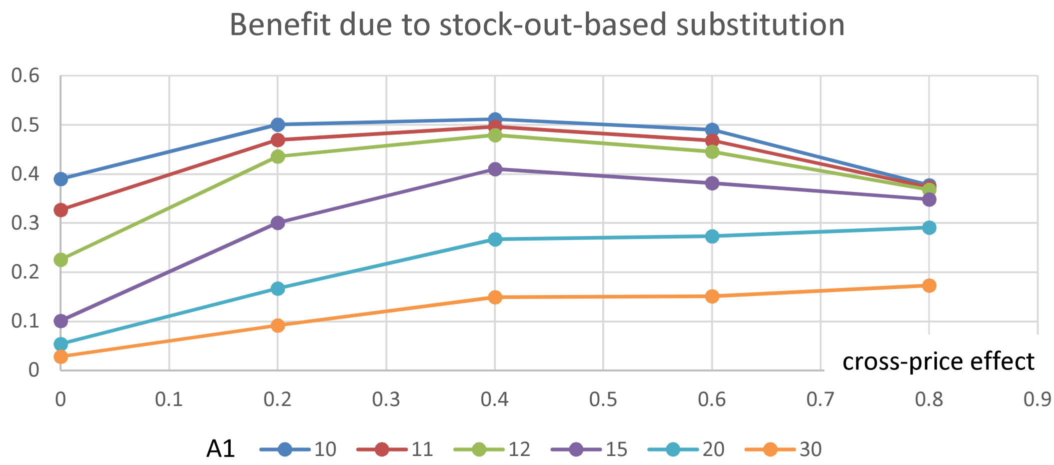


Figure 8.6: Instances are averaged out over all demand potentials and probability of demand occurrences.

high. When probability of occurrence of demand is moderate ($\pi_i = 0.5$) while demand potential is high ($\epsilon_i^h = 0.7A_i$) benefit of markdown pricing can be as high as 80%. However, when probability of demand occurrence is either very high or very low ($\pi_i = 0.1$ or 1) there does not exist any benefit of markdown pricing, given that inventory levels ignore the markdown pricing, (see Figure 8.7)

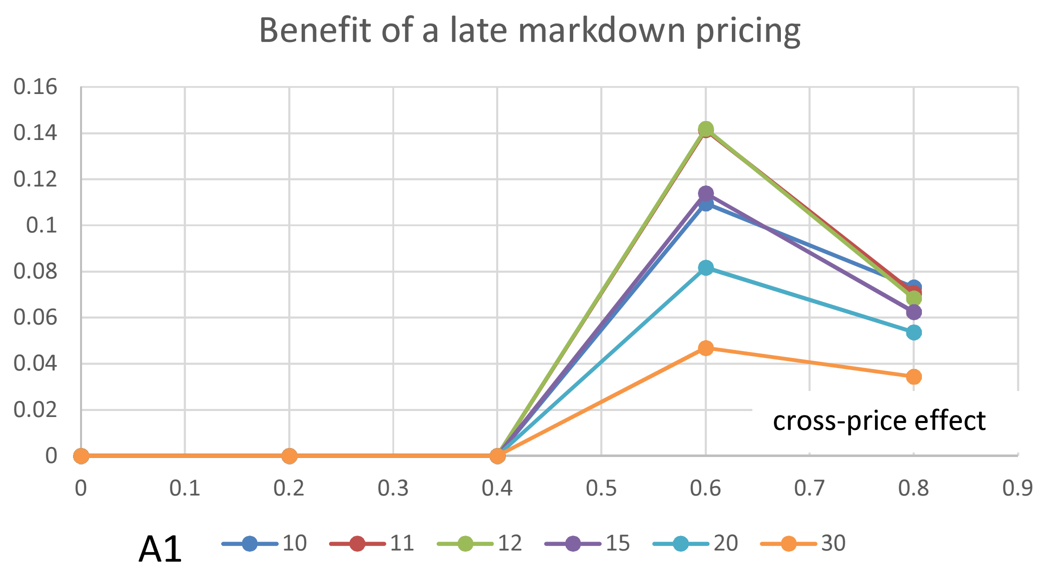


Figure 8.7: Instances are averaged out over all demand potentials and probability of demand occurrences.

- We observe that the first period price under the optimal policy in the two period scenario is always higher than the price under the single period scenario. In the two-period scenario, knowing that there will be a price update in the second period increases the price for the first period. We observe that in almost 50% of the cases, optimal prices are the same under the two-period and the single period scenarios. Asymmetry in the demand bases contribute to occurrence of identical prices. Specifically, when $A_1 = A_2 = 10$, in 55 (out of 405) cases the prices are identical, whereas when $A_1 = 30$, in 325 (out of 405) of the cases the prices are the same in both scenarios.

Inventory slack levels on the other hand are significantly different under the single-period and two-period models. In the single period model, for product i, inventory slack is either equal to 0 or ϵ_i^h. In the two period model on the other hand, considering the second period, other levels can possibly be optimal. We observe that, in the single period model inventory slack values are always lower than those of the two-period model. Furthermore, when $s_i < \epsilon_i^h$ in the two period model then $s_i = 0$ in the single period model, and when $s_i = \epsilon_i^h$ in the single period model, $s_i > \epsilon_i^h$ in the two-period model.

CHAPTER 9

CONCLUSIONS

Manufacturer may benefit from flexibility under uncertain demand. There are different kinds of flexibilities aiding manufacturer like operation, process and material handling. Substitution is an example of demand fulfillment flexibility and understanding how substitution can be used efficiently is a valuable tool in enhancing the profits. Our study contributes the literature in analysing the benefit of flexibility in context of substitution.

The aim of this study is to understand the manufacturer's production and stock management under manufacturer and customer-driven substitution. For manufacturer driven substitution, we study three different problems. In the first place, we investigate a two product model in which we analyze the optimal substitution and manufacturing policies, this gives us insight into how substitution flexibility can benefit our system and form of the optimal policies. In the second place, in n-product setting, we define the problem and we propose heuristics for better use of substitution options. Lastly, we look at a two-echelon problem in order to understand the benefit of WIP inventory, in this problem we analyze the substitution's behavior in the form of pooling resources under lower echelons. In manufacturer-driven substitution, customer's substitution behaviour is completely controlled by manufacturer. However, customer substitution has different dynamics based on customer's preference of the products. In customer driven substitution problem, our aim is to incorporate these dynamics into modeling.

Manufacturer driven substitution is studied under two product problem and manufacturer's joint production scheduling and substitution problem is studied in make-to-stock setting with a single manufacturing resource and the stocks for the products are kept separately. For this problem, we are able to show the optimal dynamic structure of the production and substitution policy. The production policy is hedging point switching curve type policy. Substitution policy is threshold type. With our numerical analysis, we are able to quantify the benefit of using dynamic policies over static policies and conditions that this benefit is higher. Dynamic policies fully use the state information when taking production and substitution decisions while static policies use partially. By comparing these, we look at the value of stock level information. Our observations confirm that substitution has higher benefits under static production policy in which we produce randomly until products inventories reach their basestocks rather than under dynamic production, especially under high traffic. The product asymmetry for arrival rates affects the benefit of substitution, this benefit decreases for static production for more asymmetric arrival rates, while it is in reverse direction for dynamic production. Furthermore, dynamic production is more beneficial under no substitution when compared to under substitution and the benefit of dynamic production is relatively high under moderate traffic.

Then, manufacturer driven substitution problem is studied under more general setting, when there are n-products. For the multi-product problem, determining the optimal policy in a reasonable time is not possible. Thus for this problem, we propose heuristics. Our heuristics can be classified under two main classes: one is priority based policies, and the other is longest-queue based policies. For analysis of priority heuristics, we first need an idling policy for this purpose we analyze the priority queue for a single server two-product system, by using this result we extend the problem to multi-product case. We show the optimal base-stock for a product will be higher if it is placed to have lower priority. In order to determine the priority order, several heuristics are proposed. We calculate the value function priority system for a multi-product production inventory system. And by applying one-step of policy improvement to this value function, we were able to devise a switching curve for priority based heuristic. For longest queue(LQ) based heuristics, idling policy of the system is calculated by a longest-queue approximation. For switching curve, we enhance on the service time look ahead policy which is proposed in the literature. We propose simpler substitution schemes like substitutions based on inventory cost or inventory level. Furthermore we propose a hybrid policy between priority based and longest queue based heuristics(Prio-LQ), and a randomized heuristic, where value function approximation is calculated based on randomized production scheme.

Performances of the proposed heuristics are assessed through a numerical study. In order to understand relative performance, we suggest one upper bound policy and one lower bound policy. By our numerical analysis, we observe when these policies perform better. For two-product setting, we compare the heuristics with optimal policy for symmetric and asymmetric arrival rates, % difference between the optimal policy is between 0.23%-12% and 0.17%-10.42%, respectively. When there is hierarchy in holding and back order costs, our priority based heuristics perform better and it gets better as dispersion of the hierarchy increases., however when there is symmetry for the products longest queue based policies performs better. Regardless of the cost structure, when arrival rate is just a little bit above light traffic, priority based heuristic works exceptionally well. For medium to high traffic, when the parameters are randomized, longest queue based policies perform better. When holding cost and backorder cost promotes Prio-DH, benefit of substitution is even better for Prio-DH than LQ-based heuristics. Production scheduling decisions are more important than substitution decisions for policy performance. For multi-product problem, it is not possible to compare the results with optimal policy, so comparing the results with upper bound(UB) and lower bound(LB) policies gives us a sense of performance of the policy. We observe that, when there is a structure in the parameter set as hierarchy or symmetry, our best-heuristic performs closer to LB.

Manufacturer's substitution problem is also affected by existence of upper echelons. In order to understand the problem, we formulate the two-echelon two-product problem with customer rejections and substitution. By numerical examples we show the benefit of WIP inventory. For this purpose, we compare

WIP system with no-WIP system. Existence of customer rejection limits the profitability due to substitution. In no-WIP, net benefit of using substitution is around %4, while this gain is %0.5 under WIP. WIP system controls the system costs by increasing the number of waiting customers, while no-WIP system does this by increasing customer rejections. Substitution can be seen as a form of pooling inventory at the customer level, our numerical examples also compare upper and lower echelon pooling strategies. Pooling at the lower echelon (substitutions) performs better than pooling at the higher echelon (using shared WIP inventory), when the arrival rates are asymmetrical, replenishment rates are equal, demand rate is high. So by numerical analysis, we show the effect of WIP inventory on substitution profitability, and the occasions when pooling in the form of substitutions is better than pooling in upper echelons by using shared inventory.

In customer-driven substitution problem, we want to incorporate customer side of the problem. We address the manufacturer's substitution problem by considering customer substitution behaviour. Consumer based substitution model investigate both price-based substitution behavior and stockout based substitution behaviour. Manufacturer can affect the customer substitution behaviour through pricing and initial stock levels of products. Two- period model is studied, in the first term substitution is incorporated to model by cross-price effect. In second period stockout based substitution is modeled by using Representative Consumer Theory framework. Before first period, pricing and inventory decisions are given, in second period, with remaining inventory, second term pricing decisions are given to affect customer's substitution behaviour and increasing the second term demand by markdown. We model the problem as abackward quadratic program and we give the structural properties of profit function. Furthermore constraint removal and feasibility analysis is done along with a solution procedure for the quadratic problem proposed.

In numerical study part, we analyze the effect of customer's substitution structure on the optimal ordering and pricing decisions. The properties of the profit functions are discussed. We include the effect of product asymmetry on the profit, the effect of base demand levels and product variability. In the case of asymmetry, there will be different capacity requirements, taking into consideration uniform pricing, contributions of the products will be different instead of single profit margin. The case where manufacture can not manipulate the customer-based substitution with a secondary pricing effort is considered, in this case the myopic models for single periods is considered. This give us the value of manufacturer's ability of planning ahead instead of making ad-hoc decisions only considering single period.

As a future work, a detailed analysis of priority order can be done. Our study has shown for the performance of the Priority Heuristic, selection of the priority order is very important. For our numerical analysis, we have used the best priority order. When the number of products increases selecting the best priority order becomes harder, trying every order constitutes a combinatorial problem. The heuristics we tried left much to be desired. Apart from increasing the heuristic performance, it is an interesting problem in its own right. Because priority production is easy to use for the businesses, as it requires minimal state information. However selecting the best-priority order is not an easy task to accomplish. To accomplish this, we need better and easier to calculate approximations of the cost for production inventory system for priority system.

For future work of customer driven driven substitution, extension to more than two periods can be done. Our analysis have shown that increasing the number of periods for exploitation of the same customer demand pool enhance the realized profits. Also extension to more than two-periods, in the case of demand occurrences between periods is welcome, in this case in addition to new demand, the remainder demand from previous periods are taken into consideration. Our model studied the single manufacturer problem, however decentralized, more than one-manufacturer problem can also be studied because it constitutes different dynamics than single manufacturer case. In our problem, consumers are myopic, and they did not consider the future occurrences of re-pricing. However in real-life consumers learn from the pricing patterns of the manufacturers. Incorporating this behaviour into consumer choice problem can have interesting implications, because in order to not teach consumers that steep markdowns can exist, we prefer not to make these discounts, although they are myopically optimal to employ.

BIBLIOGRAPHY

Ahiska, S. S. and Kurtul, E. (2014). Modeling and analysis of a product substitution strategy for a stochastic manufacturing/remanufacturing system. *Computers & Industrial Engineering*, 72:1–11.

Akcay, Y., Natarajan, H. P., and Xu, S. H. (2010). Joint dynamic pricing of multiple perishable products under consumer choice. *Management Science*, 56(8):1345–1361.

Akçay, Y. and Tan, B. (2008). On the benefits of assortment-based cooperation among independent producers. *Production and Operations Management*, 17(6):626–640.

Anderson, S. P., De Palma, A., and Thisse, J.-F. (1988). A representative consumer theory of the logit model. *International Economic Review*, pages 461–466.

Anderson, S. P., De Palma, A., and Thisse, J. F. (1992). *Discrete choice theory of product differentiation*. MIT press.

Anupindi, R., Dada, M., and Gupta, S. (1998). Estimation of consumer demand with stock-out based substitution: An application to vending machine products. *Marketing Science*, 17(4):406–423.

Archibald, T., Sassen, S., and Thomas, L. (1997). An optimal policy for a two depot inventory problem with stock transfer. *Management Science*, pages 173–183.

Atkins, D. and Zhao, X. (2009). On the modeling of demand spill for a stochastic demand system under competition. *Operations Research Letters*, 37(4):269–272.

Axsäter, S. (2003). A new decision rule for lateral transshipments in inventory systems. *Management Science*, pages 1168–1179.

Balakrishnan, A. and Brown, S. (1996). Process planning for aluminum tubes: An engineering-operations perspective. *Operations Research*, 44(1):7–20.

Balakrishnan, A. and Geunes, J. (2000). Requirements planning with substitutions: exploiting bill-of-materials flexibility in production planning. *Manufacturing & Service Operations Management*, 2(2):166–185.

Bassok, Y., Anupindi, R., and Akella, R. (1999). Single-period multiproduct inventory models with substitution. *Operations Research*, pages 632–642.

Bayındır, Z., Erkip, N., and Güllü, R. (2005). Assessing the benefits of remanufacturing option under one-way substitution. *Journal of the Operational Research Society*, 56(3):286–296.

Bayındır, Z. P., Erkip, N., and Güllü, R. (2007). Assessing the benefits of remanufacturing option under one-way substitution and capacity constraint. *Computers & Operations Research*, 34(2):487–514.

Bazaraa, M. S., Sherali, H. D., and Shetty, C. (1999). Linear complementary problem, and quadratic, separable, fractional, and geometric programming. *Nonlinear Programming: Theory and Algorithms*, pages 655–749.

Bish, E., Muriel, A., and Biller, S. (2005). Managing flexible capacity in a make-to-order environment. *Management Science*, 51(2):167–180.

Bish, E. and Suwandechochai, R. (2010). Optimal capacity for substitutable products under operational postponement. *European Journal of Operational Research*, 207(2):775–783.

Buzacott, J. and Shanthikumar, J. (1993). *Stochastic models of manufacturing systems*, volume 4. Prentice Hall Englewood Cliffs, NJ.

Carr, S. and Duenyas, I. (2000). Optimal admission control and sequencing in a make-to-stock/make-to-order production system. *Operations research*, 48(5):709–720.

Chod, J. and Rudi, N. (2005). Resource flexibility with responsive pricing. *Operations Research*, 53(3):532–548.

Cidon, I. and Sidi, M. (1990). Recursive computation of steady-state probabilities in priority queues. *Operations Research Letters*, 9(4):249–256.

Dawande, M., Gavirneni, S., Mu, Y., Sethi, S., and Sriskandarajah, C. (2010). On the interaction between demand substitution and production changeovers. *Manufacturing & Service Operations Management*, 12(4):682–691.

De Groote, X. (1994). The flexibility of production processes: a general framework. *Management Science*, 40(7):933–945.

De Vericourt, F., Karaesmen, F., and Dallery, Y. (2002). Optimal stock allocation for a capacitated supply system. *Management Science*, pages 1486–1501.

Dong, L., Kouvelis, P., and Tian, Z. (2009). Dynamic pricing and inventory control of substitute products. *Manufacturing & Service Operations Management*, 11(2):317–339.

Elmaghraby, W., Gülcü, A., and Keskinocak, P. (2008). Designing optimal preannounced markdowns in the presence of rational customers with multiunit demands. *Manufacturing & Service Operations Management*, 10(1):126–148.

Gallego, G., Katircioglu, K., and Ramachandran, B. (2006). Semiconductor inventory management with multiple grade parts and downgrading. *Production Planning & Control*, 17(7):689–700.

Gaur, V. and Honhon, D. (2006). Assortment planning and inventory decisions under a locational choice model. *Management Science*, 52(10):1528–1543.

Goyal, M. and Netessine, S. (2007). Strategic technology choice and capacity investment under demand uncertainty. *Management Science*, 53(2):192–207.

Goyal, M. and Netessine, S. (2011). Volume flexibility, product flexibility, or both: The role of demand correlation and product substitution. *Manufacturing & Service Operations Management*, 13(2):180–193.

Gruen, T. W., Corsten, D., and Bharadwaj, S. (2002). *Retail out-of-stocks: A worldwide examination of extent, causes and consumer responses.* Grocery Manufacturers of America Washington, DC.

Gumus, M., Kaminsky, P., and Mathur, S. (2011). The impact of product substitution and retail capacity on the timing and depth of price promotions: Theory and evidence. *Available at SSRN 1964325*.

Gupta, D. and Benjaafar, S. (2004). Make-to-order, make-to-stock, or delay product differentiation? a common framework for modeling and analysis. *IIE transactions*, 36(6):529–546.

Ha, A. (1997a). Inventory rationing in a make-to-stock production system with several demand classes and lost sales. *Management Science*, pages 1093–1103.

Ha, A. (1997b). Optimal dynamic scheduling policy for a make-to-stock production system. *Operations Research*, pages 42–53.

Harris, C. M. and Prabhu, N. U. (1987). Stochastic comparisons for single-server queues. *Naval Research Logistics (NRL)*, 34(4):555–567.

Hsu, V. N., Li, C.-L., and Xiao, W.-Q. (2005). Dynamic lot size problems with one-way product substitution. *IIE Transactions*, 37(3):201–215.

Inderfurth, K. (2004). Optimal policies in hybrid manufacturing/remanufacturing systems with product substitution. *International Journal of Production Economics*, 90(3):325–343.

Iravani, S., Kolfal, B., and Van Oyen, M. (2011). Capability flexibility: a decision support methodology for parallel service and manufacturing systems with flexible servers. *IIE Transactions*, 43(5):363–382.

Iravani, S. M., Kolfal, B., and Van Oyen, M. P. (2014). Process flexibility and inventory flexibility via product substitution. *Flexible Services and Manufacturing Journal*, 26(3):320–343.

Kök, A. and Fisher, M. (2007). Demand estimation and assortment optimization under substitution: Methodology and application. *Operations Research*, 55(6):1001–1021.

Kok, A., Fisher, M., and Vaidyanathan, R. (2008). Assortment planning: review of literature and industry practice, retail supply chain management. *Springer Science+ Business Media*, pages 99–153.

Lee, H. (1987). A multi-echelon inventory model for repairable items with emergency lateral transshipments. *Management Science*, pages 1302–1316.

Li, H., Graves, S., and Rosenfield, D. (2010). Optimal planning quantities for product transition. *Production and Operations Management*, 19(2):142–155.

Li, Q. and Ha, A. (2008). Reactive capacity and inventory competition under demand substitution. *IIE Transactions*, 40(8):707–717.

Li, Y., Chen, J., and Cai, X. (2006). Uncapacitated production planning with multiple product types, returned product remanufacturing, and demand substitution. *OR Spectrum*, 28(1):101–125.

Lippman, S. and McCardle, K. (1997). The competitive newsboy. *Operations Research*, 45(1):54–65.

Liu, Q. and van Ryzin, G. J. (2008). Strategic capacity rationing to induce early purchases. *Management Science*, 54(6):1115–1131.

Liu, Q. and Zhang, D. (2013). Dynamic pricing competition with strategic customers under vertical product differentiation. *Management Science*, 59(1):84–101.

Lus, B. and Muriel, A. (2009). Measuring the impact of increased product substitution on pricing and capacity decisions under linear demand models. *Production and Operations Management*, 18(1):95–113.

Maddah, B. and Bish, E. (2007). Joint pricing, assortment, and inventory decisions for a retailer's product line. *Naval Research Logistics (NRL)*, 54(3):315–330.

Mahajan, S. and van Ryzin, G. (2001). Stocking retail assortments under dynamic consumer substitution. *Operations Research*, 49(3):334–351.

Nagarajan, M. and Rajagopalan, S. (2008). Inventory models for substitutable products: Optimal policies and heuristics. *Management Science*, 54(8):1453–1466.

Netessine, S. and Rudi, N. (2003). Centralized and competitive inventory models with demand substitution. *Operations Research*, 51(2):329–335.

Parlar, M. (1988). Game theoretic analysis of the substitutable product inventory problem with random demands. *Naval Research Logistics (NRL)*, 35(3):397–409.

Perez, A. P. and Zipkin, P. (1997). Dynamic scheduling rules for a multiproduct make-to-stock queue. *Operations Research*, 45(6):919–930.

Pineyro, P. and Viera, O. (2010). The economic lot-sizing problem with remanufacturing and one-way substitution. *International Journal of Production Economics*, 124(2):482–488.

Rao, U. S., Swamithan, J. M., and Zhang, J. (2004). Multi-product inventory planning with downward substitution, stochastic demand and setup costs. *IIE Transactions*, 36(1):59–71.

Rozenshmidt, L. (2008). *On priority queues with impatient customers: Stationary and time-varying analysis*. PhD thesis, Technion–Israel Institute of Technology.

Ryzin, G. and Mahajan, S. (1999). On the relationship between inventory costs and variety benefits in retail assortments. *Management Science*, 45(11):1496–1509.

Schröder, K. (2012). Cannibalization on the german yoghurt market.

Sethi, A. and Sethi, S. (1990). Flexibility in manufacturing: a survey. *International journal of flexible manufacturing systems*, 2(4):289–328.

Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic orders*. Springer.

Sherbrooke, C. (1968). Metric: A multi-echelon technique for recoverable item control. *Operations Research*, pages 122–141.

Shumsky, R. and Zhang, F. (2009). Dynamic capacity management with substitution. *Operations research*, 57(3):671–684.

Stidham, S. and Weber, R. (1993). A survey of markov decision models for control of networks of queues. *Queueing Systems*, 13(1):291–314.

Suh, M. and Aydin, G. (2011). Dynamic pricing of substitutable products with limited inventories under logit demand. *IIE Transactions*, 43(5):323–331.

Talluri, K. T. and Van Ryzin, G. J. (2006). *The theory and practice of revenue management*, volume 68. Springer Science & Business Media.

Tang, C. S. and Yin, R. (2007). Joint ordering and pricing strategies for managing substitutable products. *Production and Operations Management*, 16(1):138–153.

Tanrisever, F., Morrice, D., and Morton, D. (2012). Managing capacity flexibility in make-to-order production environments. *European Journal of Operational Research*, 216(2):334–345.

Tijms, H. (1994). *Stochastic models: an algorithmic approach*. John Wiley & Sons New York.

Van Mieghem, J. A. and Dada, M. (1999). Price versus production postponement: Capacity and competition. *Management Science*, 45(12):1639–1649.

Van Ryzin, G. and Mahajan, S. (1999). On the relationship between inventory costs and variety benefits in retail assortments. *Management Science*, pages 1496–1509.

Veatch, M. and Wein, L. (1992). Monotone control of queueing networks. *Queueing Systems*, 12(3):391–408.

Veatch, M. and Wein, L. (1996). Scheduling a make-to-stock queue: Index policies and hedging points. *Operations Research*, pages 634–647.

Xu, H., Yao, D., and Zheng, S. (2011). Optimal control of replenishment and substitution in an inventory system with nonstationary batch demand. *Production and Operations Management*, 20(5):727–736.

Yang, J. and Qin, Z. (2007). Capacitated production control with virtual lateral transshipments. *Operations Research*, 55(6):1104–1119.

Zhao, H., Ryan, J., and Deshpande, V. (2008). Optimal dynamic production and inventory transshipment policies for a two-location make-to-stock system. *Operations research*, 56(2):400–410.

Zhao, X. (2008). Coordinating a supply chain system with retailers under both price and inventory *Production and Operations Management*, 17(5):532–542.

Zhao, X. and Atkins, D. (2008). Newsvendors under simultaneous price and inventory competition. *Manufacturing & Service Operations Management*, 10(3):539–546.

Zhao, X. and Atkins, D. (2009). Transshipment between competing retailers. *IIE Transactions*, 41(8):665–676.

Zheng, Y.-S. and Zipkin, P. (1990). A queueing model to analyze the value of centralized inventory information. *Operations Research*, 38(2):296–307.

Zipkin, P. H. (1995). Performance analysis of a multi-item production-inventory system under alternative policies. *Management Science*, 41(4):690–703.

APPENDIX A

PROOF FOR TWO-PRODUCT SINGLE ECHELON PROBLEM

Proof of Theorem 3.2.1 We characterize the structure of the optimal policy in two steps. In the first step we define a cost-to-go function for t periods, $V^t(x_1, x_2)$. Using induction we show that $V^t(x_1, x_2)$ satisfies certain conditions. Then, we show that for the problem under consideration, an optimal policy exists under the average reward criteria, and that since $V^t(x_1, x_2)$ satisfies certain conditions the policy can be characterized as in Theorem 3.2.1.

Define the t period cost-to-go function as,

$$\begin{aligned} V^t(x_1, x_2) = & \quad \frac{1}{\Lambda}(c(x) + \lambda_1 V^{t-1}(x_1 - 1, x_2) \\ & + \quad \lambda_2 V^{t-1}(x_1, x_2 - 1) \\ & + \quad \mu \mathbf{O}_1 V^{t-1}(x_1, x_2) + \mu_s \mathbf{O}_2 V^{t-1}(x_1, x_2) \end{aligned} \tag{A.0.1}$$

where $c(x) = \sum_{i=1}^{2} h_i x_i^+ + b_i x_i^-$, $\mathbf{O}_1$ and $\mathbf{O}_2$ as defined in Eq. 3.2.2.

Let F be the set of real-valued functions, where $f : (x_1, x_2) \to \mathbb{R}$ satisfies the following conditions.

$$f(x_1 + 1, x_2) - f(x_1, x_2) \geq f(x_1, x_2) - f(x_1 - 1, x_2) \tag{C1}$$

$$f(x_1, x_2 + 1) - f(x_1, x_2) \geq f(x_1 - 1, x_2 + 1) - f(x_1 - 1, x_2) \tag{C2}$$

$$f(x_1 + 1, x_2) - f(x_1, x_2 + 1) \geq f(x_1, x_2) - f(x_1 - 1, x_2 + 1) \tag{C3}$$

The following conditions are the respective counter parts of (C1)-(C3).

$$f(x_1, x_2+1) - f(x_1, x_2) \geq f(x_1, x_2) - f(x_1, x_2-1) \tag{C4}$$

$$f(x_1+1, x_2) - f(x_1, x_2) \geq f(x_1+1, x_2-1) - f(x_1, x_2-1) \tag{C5}$$

$$f(x_1, x_2+1) - f(x_1+1, x_2) \geq f(x_1, x_2) - f(x_1+1, x_2-1) \tag{C6}$$

If V^{t-1} satisfies the conditions C1-C6, then it is possible to infer the following for the decisions made at stage t. For a given x_2, if $V^{t-1}(x_1+1, x_2) - V^{t-1}(x_1, x_2) \leq 0$, then producing product 1 is preferred to no production. C1 implies when $V^{t-1}(x_1+1) - V^{t-1}(x_1) \leq 0$, $V^{t-1}(x_1) - V^{t-1}(x_1-1) \leq 0$ holds, which means if producing for product 1 is preferred to no production for (x_1, x_2), then producing for product 1 is also preferred for $(x_1 - 1, x_2)$. This implies at stage t, there exists an inventory level $S_1^t(x_2)$, below which producing for product 1 is preferred to no production and above which not producing is preferred to product 1. In a similar fashion, C4 implies there exists an $S_2^t(x_1)$.

From C5 it is possible to infer that $S_1^t(x_2)$ is decreasing in x_2. the condition states that if for a given x_2 producing product 1 is more preferable than not producing, then for $x_2 - 1$ producing product 1 is also preferable. This implies $S_1^t(x_2) \leq S_1^t(x_2 - 1)$. Conditions (C3) and (C6) together shows $K_1^t(x_2)$ exists and is increasing in x_2. Similarly $K_2^t(x_1)$ exists and is increasing in x_1. It is possible to infer from conditions C3) and (C6) also that one unit increase in x_2 leads to at most one unit decrease in $S_1^t(x_2)$. Same structure holds for $S_2^t(x_1)$.

Product substitution decisions are also characterized by conditions (C3) and (C6).

For the proof we introduce the following lemma at the beginning.

Lemma A.0.1. *If $V^{t-1} \in F$, then $V^t \in F$.*

Proof. In order prove $V^t \in F$, it is sufficient to show $\mathbf{O}_1 V^{t-1}$, $\mathbf{O}_2 V^{t-1}$, and $c(x)$ are in F. Note that conditions C2 and C3 imply C1. Furthermore, conditions C4-C6 are symmetric to C4-C6. Thus, it is sufficient to show $\mathbf{O}_1$, $\mathbf{O}_2$ and $c(x)$ satisfy conditions C2 and C3. For simplicity of notation we omit $t-1$ in V^{t-1}, and simply denote with V.

Condition C2

Under Production Operator, $\mathbf{O}_1$

Let

$$\begin{aligned} w^1(1,x_1,x_2) &= V(x_1+1,x_2) \\ w^1(0,x_1,x_2) &= V(x_1,x_2) \\ w^1(2,x_1,x_2) &= V(x_1,x_2+1), \end{aligned}$$

then $\mathbf{O}_1 V(x_1,x_2) = \min_{u\in\{0,1,2\}} w^1(u,x_1,x_2)$.

We show

$$\mathbf{O}_1 V(x_1,x_2+1) + \mathbf{O}_1 V(x_1-1,x_2) \geq \mathbf{O}_1 V(x_1,x_2) + \mathbf{O}_1 V(x_1-1,x_2+1)$$

Let $u_1, u_2 \in \{0,1,2\}$ denote the optimal production action taken in stage t under (x_1,x_2+1) and (x_1-1,x_2), respectively.

$$\mathbf{O}_1 V(x_1,x_2+1) = w^1(u_1,x_1,x_2+1), \quad \mathbf{O}_1 V(x_1-1,x_2) = w^1(u_2,x_1-1,x_2)$$

Since V satisfies conditions C1-C6, all possible values (u_1,u_2) can take are $u_1 \leq u_2$ and $(2,1)$. We make the analysis only for $u_1 = u_2$, other cases can be analyzed similarly.

$u_1 = u_2$

$$\begin{aligned} \mathbf{O}_1 V(x_1,x_2) + \mathbf{O}_1 V(x_1-1,x_2+1) &\leq w^1(u_1,x_1,x_2) + w^1(u_1,x_1-1,x_2+1) \\ &\leq w^1(u_1,x_1,x_2+1) + w^1(u_1,x_1-1,x_2) \\ &= \mathbf{O}_1 V(x_1,x_2+1) + \mathbf{O}_1 V(x_1-1,x_2). \end{aligned}$$

Under Substitution Operator, $\mathbf{O}_2$

Let

$$\begin{aligned} w^2(1,x_1,x_2) &= V(x_1-1,x_2+1) + c_{12} \\ w^2(0,x_1,x_2) &= V(x_1,x_2) \\ w^2(-1,x_1,x_2) &= V(x_1+1,x_2-1) + c_{21} \end{aligned}$$

Then $\mathbf{O}_2V(x_1,x_2) = \min_{u\in\{-1,0,1\}} w^2(u,x_1,x_2)$.

We show

$$\mathbf{O}_2V(x_1,x_2+1) + \mathbf{O}_2V(x_1-1,x_2) \geq \mathbf{O}_2V(x_1,x_2) + \mathbf{O}_2V(x_1-1,x_2+1)$$

Let u_1 and u_2 such that

$$\mathbf{O}_2V(x_1,x_2+1) = w^2(u_1,x_1,x_2+1), \quad \mathbf{O}_2V(x_1-1,x_2) = w^2(u_2,x_1-1,x_2).$$

Define

$$\Delta w^2(u,x_1,x_2) = w^2(u,x_1,x_2+1) - w^2(u,x_1,x_2)$$

Then from condition C6, we have

$$\Delta w^2(1,x_1,x_2) \geq \Delta_2 w^2(0,x_1,x_2) \geq \Delta_2 w^2(-1,x_1,x_2).$$

This implies w^2 is supermodular with respect to (u,x_2). Since V satisfies conditions C1-C6, all possible values (u_1,u_2) can take are $u_1 = u_2$, $u_1 = 0, u_2 \in \{-1,0,1\}$ and $u_2 = 0, u_1 \in \{-1,0,1\}$. We make the analysis only for $u_1 \geq u_2$, other cases can be analyzed similarly.

$u_1 \geq u_2$

$$\begin{aligned}
\mathbf{O}_2V(x_1,x_2) \quad + \quad & \mathbf{O}_2V(x_1-1,x_2+1) \leq w^2(u_1,x_1,x_2) + w^2(u_2,x_1-1,x_2+1) \\
\leq \quad & w^2(u_1,x_1,x_2) + w^2(u_2,x_1,x_2+1) \\
& +w^2(u_2,x_1-1,x_2) - w^2(u_2,x_1,x_2) \\
\leq \quad & w^2(u_1,x_1,x_2) + w^2(u_1,x_1,x_2+1) \\
& +w^2(u_2,x_1-1,x_2) - w^2(u_1,x_1,x_2) \\
= \quad & \mathbf{O}_2V(x_1,x_2+1) + \mathbf{O}_2V(x_1-1,x_2)
\end{aligned}$$

First inequality comes from definition of w^2, second from condition C2, and third from supermodularity of w^2 under (u,x_2).

Under $c(x)$

Condition C2 implies, $c(x_1,x_2+1) - c(x_1,x_2) \geq c(x_1-1,x_2+1) - c(x_1-1,x_2)$ should hold. It is trivial to show that $c(x)$ satisfies condition C2.

Condition C3

Under Replenishment operator, $\mathbf{O}_1$

We need to show

$$\mathbf{O}_1V(x_1+1,x_2)+\mathbf{O}_1V(x_1-1,x_2+1)\geq \mathbf{O}_1V(x_1,x_2)+\mathbf{O}_1V(x_1,x_2+1)$$

Let u_1 and u_2 be such that

$$\mathbf{O}_1V(x_1+1,x_2)=w^1(u_1,x_1+1,x_2),\quad \mathbf{O}_1V(x_1-1,x_2+1)=w^1(u_2,x_1-1,x_2+1).$$

Since V satisfies conditions C1-C6, all possible values (u_1,u_2) can take are $u_1=u_2=1$, $u_1=0,u_2\in\{0,1,2\}$ and $u_1=2,u_2\in\{0,1,2\}$. We make the analysis only for $u_1=u_2$, other cases can be analyzed similarly.

$u_1=u_2$

$$\begin{aligned}\mathbf{O}_1V(x_1,x_2)+\mathbf{O}_1V(x_1,x_2+1) &\leq w^1(u_1,x_1,x_2)+w^1(u_1,x_1,x_2+1)\\ &\leq w^1(u_1,x_1+1,x_2)+w^1(u_1,x_1-1,x_2+1)\\ &\leq \mathbf{O}_1V(x_1+1,x_2)+\mathbf{O}_1V(x_1-1,x_2+1),\end{aligned}$$

where the second inequality holds due to condition C3.

Under Substitution Operator, $\mathbf{O}_2$

We show

$$\mathbf{O}_2V(x_1+1,x_2)+\mathbf{O}_2V(x_1-1,x_2+1)\geq \mathbf{O}_2V(x_1,x_2)+\mathbf{O}_2V(x_1,x_2+1)$$

Let u_1 and u_2 be such that

$$\mathbf{O}_2V(x_1+1,x_2)=w^2(u_1,x_1+1,x_2),\quad \mathbf{O}_2V(x_1-1,x_2+1)=w^2(u_2,x_1-1,x_2+1).$$

Let

$$\Delta_1w^2(u,x_1,x_2)=w^2(u,x_1+1,x_2)-w^2(u,x_1,x_2).$$

Due to condition C2, it is possible to show that

$$\Delta_1w^2(1,x_1,x_2)\leq\Delta_1w^2(0,x_1,x_2)\leq\Delta_1w^2(-1,x_1,x_2)$$

Hence w^2 is submodular in (u, x_1). Since V satisfies conditions C1-C6, all possible values (u_1, u_2) can take are $u_1 = u_2$, $u_1 = 1, u_2 = 0$ and $u_1 = 0, u_2 = -1$. We make the analysis only for $u_1 = 1, u_2 = 0$, other cases can be analyzed similarly.

$u_1 = 1, u_2 = 0$

$$\begin{aligned}
\mathbf{O}_2V(x_1, x_2) \quad + \quad & \mathbf{O}_2V(x_1, x_2 + 1) \leq w^2(1, x_1, x_2) + w^2(0, x_1, x_2 + 1) \\
\leq \quad & V(x_1 - 1, x_2 + 1) + c_{12} + V(x_1, x_2 + 1) \\
& + w^2(1, x_1 + 1, x_2) + w^2(0, x_1 - 1, x_2 + 1) \\
= \quad & \mathbf{O}_2V(x_1 + 1, x_2) + \mathbf{O}_2V(x_1 - 1, x_2 + 1)
\end{aligned}$$

Under $c(x)$

Condition C3 implies, $c(x_1 + 1, x_2) - c(x_1, x_2 + 1) \geq c(x_1, x_2) - c(x_1 - 1, x_2 + 1)$ should hold. It is trivial to show that $c(x)$ satisfies condition C3.

□

Proof. (Theorem 3.2.1) Lemma A.0.1 show that $V^t \in F$ for $V^{t-1} \in F$. This implies as $t \to \infty$, V^t satisfies conditions C1-C6. Thus under the optimal policy there exists $S_1(x_2)$, where $S_1(x_2)$ is decreasing with $x_2(x_1)$. For one unit of increase in x_2, $S_1(x_2)$ decreases one unit. Furthermore, $K_1(x_2)$ and $T_1(x_2)$ are increasing in x_2. If $b_2 < b_1 + c_{12}$, then substitution does not take place when $x_1 \leq 0$.

For the policy to exist under the average reward criteria, and to have the same above-mentioned structure, one needs to show that the Markov decision process under consideration satisfies certain conditions.

The process under consideration is multi-chain, with countable state space and unbounded rewards. However, the cost structure is such that as number of products in the inventory, or in the backorder increases the cost rate increases linearly. This implies the number of states that the chain visits under the optimal policy must be bounded Stidham and Weber (1993). The proof of existence also follows similar lines with Ha (1997b). □

APPENDIX B

ANALYSIS OF A TWO-ECHELON TWO-PRODUCT MANUFACTURING SYSTEM WITH SUBSTITUTIONS AT THE LOWER ECHELON

Our problem in this chapter is to determine the optimal production, replenishment, substitution and customer rejection policy in a two echelon two product system. In the upper echelon, production resources are pooled, so the products share the limited capacity of a single processor. The output of the production at the upper echelon is a semi-finished product. This product is converted into the final product after another processing stage. It is possible to keep a stock of semi-finished product. This stock is common for both products. In the lower echelon each final product has its own inventory, and faces an exogenous and stochastic demand. Upper echelon inventory (work-in-process inventory, WIP) replenishes the lower echelon inventories, replenishment times are stochastic and takes place one replenishment at a time. We use the terms "replenishment of final product inventory" and "conversion of WIP into final product", interchangeably. Final products are substitutable. Substitutions involve some form of physical conversion of the products and thus take time and cost. Substitution between the products helps to pool the inventories at the lower echelon.

Upper echelon inventory helps the system utilize the limited production capacity better. Although having too much WIP inventory would increase the holding costs, it would decrease the backordering costs at the lower echelon, because demand for the final product can be met faster. In some systems upper echelon inventory is required because of the geographical reasons, for example the places of production and end-product inventories could be separated, and they might be a need for different location inventories for both echelons of the system.

A fundamental problem in production-inventory systems is pooling of resources and inventories. Pooling is grouping of resources for the common advantage of the system. Pooling in upper echelon can be done by sharing production capacity, and the WIP inventory following it. All the products use common WIP stock that is produced in the shared production facility. The replenishment time between the upper and lower echelons is the time required for processing the WIP to specialize into end-products. At the lower echelon, pooling can be done via substitution, products can be used for each other's place with some processing time and cost. This way the inventory of a product can be used by other products' customers.

B.1 Model Description

Our base model reflects a two-echelon system with two products (indexed by $i,\ i = 1, 2$), in which both production capacity and product inventories are pooled. In the upper echelon there is a single shared production facility, producing a single common WIP product. Production times are exponential with rate μ_p. At any point in time a decision as to whether to produce or not is made. In the lower echelon, the inventories of two different products are replenished from the upper echelon inventory one at a time. Replenishment times are exponential with rate μ_R, replenishment decisions are given on any point in time. The decision is whether or not to convert a WIP unit into a final product, and if so, into which final product.. System state is defined by the inventory levels of the final products, $x_1(t)$ and $x_2(t)$, and the inventory level of the WIP product at the upper echelon, $x_3(t)$. At the lower echelon, demand can be backordered, however at the upper echelon this is not allowed, i.e., if WIP stock is not available then conversion into final product is not initiated. At the lower echelon, demand arrives for the products and the manufacturer decides whether to accept or reject the demand. Finally, the manufacturer decides whether to convert one final product into the other. The substitution decisions are given at any point in time, provided that the final product "to substitute" has positive stock. When substitution takes place, inventory of the substituted final product increases by one, whereas inventory of substituting final product decreases by one. Substitution process time is an exponential random variable, with parameter μ_S. It costs c_s, to substitute one unit of final product. See Figure B.1 for the representation of the model.

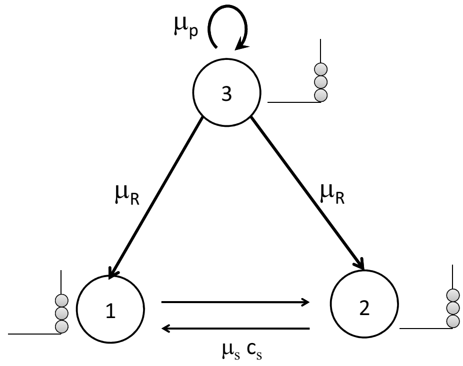


Figure B.1: Two-echelon System

In this manufacturing system, each accepted customer brings unit revenue R. There is a holding cost $h_i,\ i = 1,2$ and backorder cost $b_i,\ i = 1,2$ associated with each final product. For WIP inventory, the holding cost is denoted with h_3. Customer arrivals to product inventories are Poisson processes with rate $\lambda_i, i = 1,2$. We follow Lippman (1975) to uniformize the process by defining the uniformization constant $\tau = \frac{1}{\sum_{i\in\{1,2\}}(\lambda_i)+\mu_1+\mu}$. The optimality equation is given by,

$$
\begin{aligned}
V(x_1,x_2,x_3) + \frac{g}{\tau} &= \frac{1}{\tau}(-\sum_{i\in\{1,2,3\}} h_i(x_i) + \lambda_1 \mathbf{O}_1 V(x_1,x_2,x_3) \\
&+ \lambda_2 \mathbf{O}_2 V(x_1,x_2,x_3) + \mu_P \mathbf{O}_3 V(x_1,x_2,x_3) \\
&+ \mu_R I(x_3>0) \mathbf{O}_4 V(x_1,x_2,x_3) + \mu_R(1-I(x_3>0))V(x_1,x_2,x_3)) \\
&+ \mu_s \mathbf{O}_5 V(x_1,x_2,x_3)
\end{aligned}
$$

where $h_i(x_i) = h_i x_i^+ + b_i x_i^-$,$\mathbf{O}_1$ and $\mathbf{O}_2$ are customer acceptance rejection operators.

$$
\mathbf{O}_1 f(x_1,x_2,x_3) = \max \begin{cases} f(x_1,x_2,x_3) & a_1 = 0 \\ f(x_1-1,x_2,x_3)+R & a_1 = 1 \end{cases}
$$

$$\mathbf{O}_2 f(x_1,x_2,x_3) = \max \begin{cases} f(x_1,x_2,x_3) & a_2 = 0 \\ f(x_1,x_2-1,x_3)+R & a_2 = 1 \end{cases}$$

$\mathbf{O}_3$ is production in upper echelon operator.

$$\mathbf{O}_3 f(x_1,x_2,x_3) = \max \begin{cases} f(x_1,x_2,x_3) & a = 0 \\ f(x_1,x_2,x_3+1) & a = 1 \end{cases}$$

$\mathbf{O}_4$ is replenishment operator.

$$\mathbf{O}_4 f(x_1,x_2,x_3) = \max \begin{cases} f(x_1+1,x_2,x_3-1) & \bar{a} = 1 \\ f(x_1,x_2+1,x_3-1) & \bar{a} = 2 \\ f(x_1,x_2,x_3) & \bar{a} = 0 \end{cases}$$

$\mathbf{O}_5$ is substitution operator.

$$\mathbf{O}_5 f(x_1,x_2,x_3) = \max_{a_t=\{0,1,-1\}} f(x_1+a_t,x_2-a_t,x_3) - cx|a_t| \quad a_t = -1,0,1$$

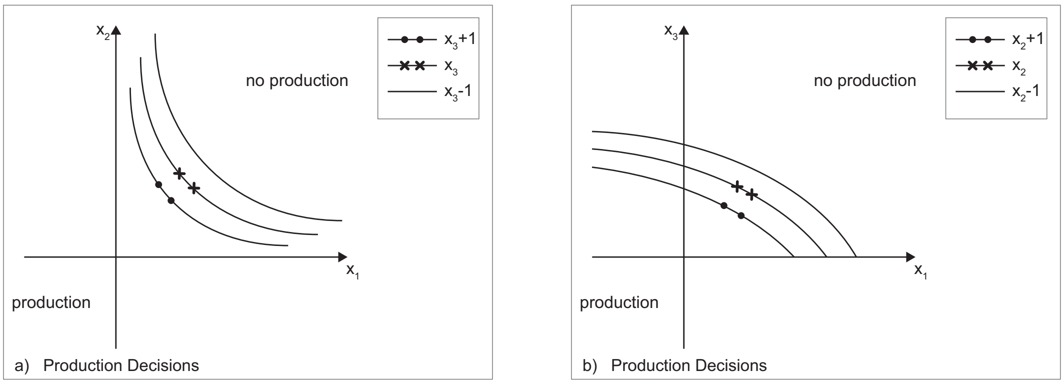


Figure B.2: Production policy

It is not possible to characterize the optimal policy (via induction approach). Numerical examples show certain characteristics of the optimal policy. Production policy on upper echelon is a base stock type policy, where the base-stock level depends on lower echelon inventory levels. As the inventory on lower echelon increases, the base stock level decreases. Replenishment decisions show hedging point-switching curve policy type where basestock levels and a production curve defines the production policy, their values depend on upper echelon

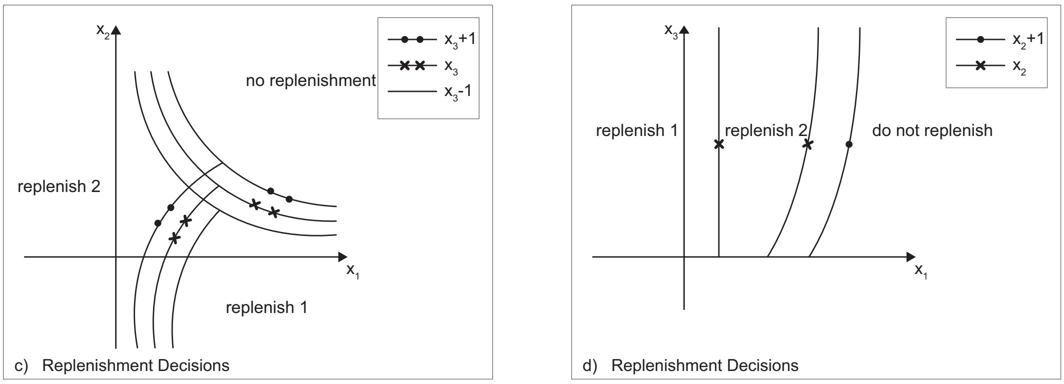


Figure B.3: Replenishment policy

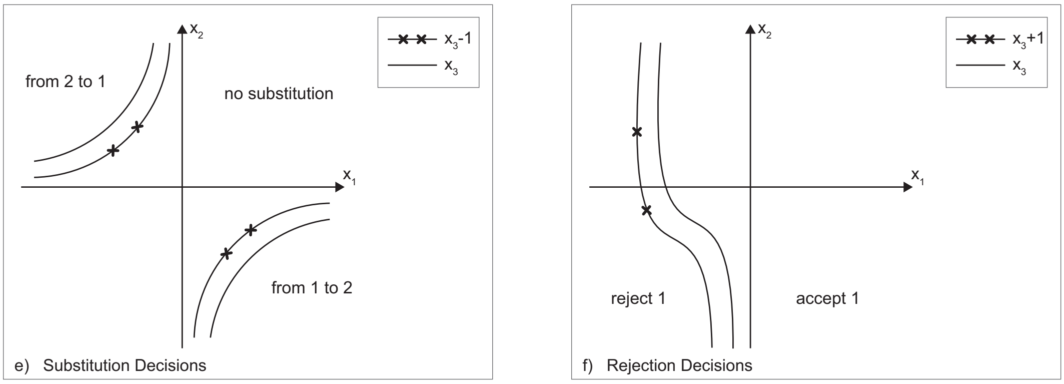


Figure B.4: Substitution and rejection policy

inventory level. When the substitution costs are high and takes a long time, the basestock for lower echelon decreases with the upper echelon inventory level as the system try hedge against backorders. However, this can change if substitutions are easy to make. Switching curve shifts as upper echelon inventory level increases. Policy substitutes fewer items as upper echelon inventory increases. At the lower echelon, more customers are accepted as upper echelon inventory level increases.

B.2 Numerical Analysis

To analyze the benefit of holding WIP inventory, and quantify the benefit of pooling at the upper vs. lower echelon, we conduct numerical experiments.

B.2.1 The benefit of upper echelon Inventory

To analyze the benefit of holding upper echelon inventory, we conduct numerical experiments. We compare two systems, in the first one optimal WIP inventory is kept at the upper echelon, in the second one zero inventory is kept at the upper echelon.. Parameters for customer arrival rate are $\lambda_1, \lambda_2 \in \{0.2, 0.4, 0.6, 0.8\}$, production rate is $\mu_p \in \{1.5\}$, replenishment rate of products is $\mu_R \in \{1\}$, substitution rate is $\mu_s \in \{4\}$, upper echelon inventory cost is $h_3 \in \{0.3, 0.5\}$, additional unit cost of inventory under lower echelon is $h_i \in \{0.2, 0.5, 1\}$ for i=1,2, the cost of substitution is $c_s = \{2, 5, 8\}$.

The following observations are made:

- Both in the presence or absence of WIP inventory (namely, WIP and no-WIP), if the lower echelon holding cost h increases, more substitutions take place. In both systems, the upper echelon holding cost h_3 affects the substitution decisions at the lower echelon similar to the lower echelon holding cost.
- As the customer arrival rate increases, the rate of substitution first increases then decreases in both WIP and no-WIP systems. In WIP system increasing the arrival rate increases the upper echelon inventory, and this results in reduced need for substitution. In no-WIP, increased customer arrival rate does not increase the lower echelon inventory, instead customer rejection is selected more frequently as an option. Because of this reason, under high demand rate the substitution rate is low. When there is no option for rejection, we expect the customer arrival rate to increase the substitution rate.
- Several performance measures are compared under no-WIP and WIP. In WIP system the number of waiting customers is higher. If the number of waiting customers considered as a surrogate service level, having inventory on upper echelon is unfavorable. This implies, in order to increase profitability, we give up from service level. On the other hand, in no-WIP customer rejection rates are higher. When we compare inventory levels of

both systems, we see that in WIP more inventory is kept. As the customer arrival rate increases, in the lower echelon, in WIP system inventory level increases while, in no-WIP system inventory level decreases.

- In the no-WIP system does not respond to increased customer arrival rate by increasing inventory levels since this decreases the profitability of the system. Instead customer rejection increases. In WIP both lower and upper echelon inventories increases as customer demand rate increases. However, when the lower echelon inventory holding cost is high, the increase in inventory levels is limited. In WIP system responds to increased customer demand rate by increasing especially, the upper echelon inventory, since upper echelon inventory cost is low. However, no-WIP responds to this situation by increasing customer rejection rate because inventory holding cost is relatively high. In general in no-WIP systems, the waiting times are higher and inventory levels are lower, but in this setup wait times do not get too high because of customer rejection. In fact in this setting, in WIP system, average number of waiting customers is higher. So for WIP system, both inventory holding costs and back-order costs are high. However, more customers are accepted in WIP, so the profit is higher than no-WIP even though costs are higher.

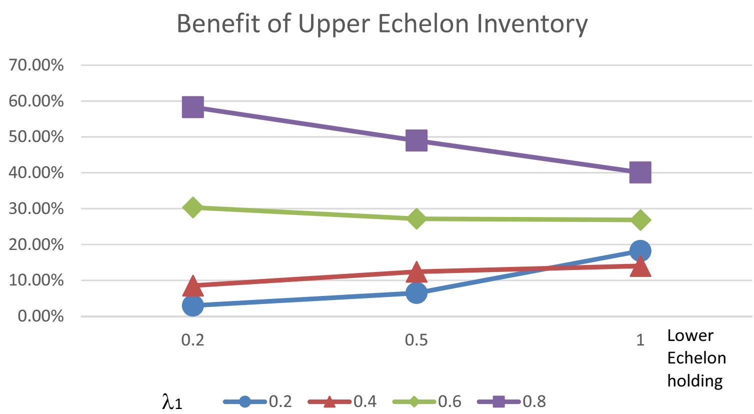


Figure B.5: Relative profit of MTS system compared to MTO system

- In Figure B.5, percent gain for using a WIP inventory in the upper echelon $(\frac{Profit_{MTS}-Profit_{MTO}}{Profit_{MTO}})$ is shown. The no-WIP system is always sub-optimal compared to the WIP system. Thus profit under WIP is always higher than that of no-WIP. Analysis shows that when the demand rate is high, the difference in profits of the two systems may be as high as %60. Under high arrival rate, the upper echelon's inventory effect is also high. The WIP system responds to increased arrival rate by increasing the upper echelon inventory since holding inventory in upper echelon is cheaper. Thus, net gain is high due to this flexibility.

- In no-WIP, an increase in the arrival rate merely affects the inventory level, while in WIP, both upper and lower echelon inventory levels increase. We also observe that an increase in h affects the two systems differently. In the WIP system when the demand rate is low, increase in holding cost parameter h do not affect the system profitability much. However when the demand rate is high, increase in h affects the profit profoundly. This is because under high arrival rate, inventory levels are also high. Thus, an increase in h decreases the net gain of keeping WIP stock in upper echelon. In low demand case, the relation is just the reverse.

- Lastly, we examine the profitability of substitution. The gain from substitution increases as lower echelon inventory holding cost increases. The profitability of substitution is higher under no-WIP. But in both systems this profitability is limited. In no-WIP, net gain of using substitution is around %4, while this gain is %0.5 under WIP. Under substitution, inventory levels are lower compared to no substitution case. Finally, the flexibility of customer rejection results in limited benefit of substitution.

B.2.2 Comparison of Upper and Lower Echelon Pooling Strategies

In this section, we compare the results on comparison of resource pooling at the upper and lower echelon. Resource pooling at upper echelon refers to sharing the same production capacity and upper echelon WIP inventory for the products. Resource pooling at the lower echelon is done via final product substitution. In Figure B.6, we present a representation of resource pooling at the upper and at the lower echelon. In Model 2, lower echelon pooling is done via substitution of products. In this model, at the upper echelon we have separate production facilities and separate WIP inventory is kept for each final product. In Model 3, resource pooling is done via production capacity sharing. In this model, we assume there is single upper echelon inventory and separate lower echelon inventories. There is no lower echelon pooling which means there is no substitution. Note that our base model, Model 1, is a hybrid model of these two pooling policies.

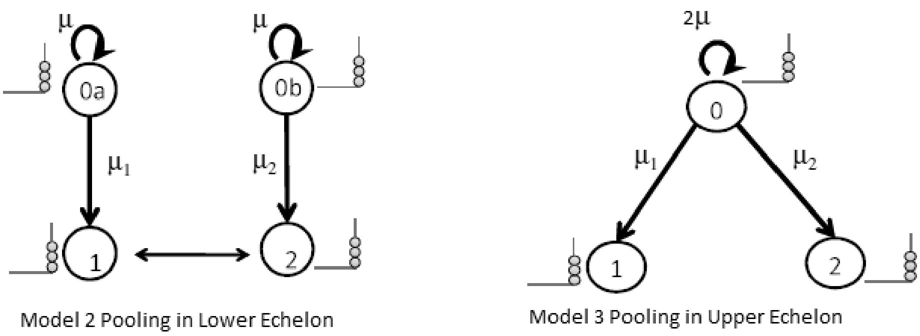


Figure B.6: Comparing the pooling types in Two-echelon Systems

In the analysis, production rate is assumed as $\mu_p \in \{1.5\}$, substitution rate is $\mu_T \in 4$, upper echelon inventory holding cost $h_3 \in \{0.3, 0.5\}$, lower echelon holding cost $h \in \{0.2, 0.5, 1\}$, unit cost of substitution is $c_s \in \{2\}$. Customer arrivals and replenishment are defined in three different ways:

- Symmetrical arrival rates(SASR case): $\lambda_1, \lambda_2 \in \{0.2, 0.4, 0.6, 0.8\}$ and $\mu_r = 1$.
- Asymmetrical arrival rates, but equal replenishment rates(AASR case): $(\lambda_1, \lambda_2) \in \{(0.12, 0.28), (0.24, 0.56), (0.36, 0.84), (0.48, 1.12)\}$ and $\mu_{r1} = \mu_{r2} = 1$
- Asymmetrical arrival rates, replenishment rates are proportional to the arrival rates(AAAR case):$(\lambda_1, \lambda_2) \in \{(0.12, 0.28), (0.24, 0.56), (0.36, 0.84), (0.48, 1.12)\}$ and $(\mu_{r1}, \mu_{r2}) = (0.6, 1.4)$

The following observations are made:

- According to the numerical analysis, pooling at the lower echelon (Model 2) is preferred over pooling at the higher echelon (Model 3), when the arrival rates are asymmetrical, replenishment rates are equal, demand rate is high and the holding cost at the lower echelon is low. When the replenishment rates are not proportional to the customer arrival rates, in order to respond to high demand rates, inventory levels need to be high. The product that does not have enough replenishment rate to satisfy the customer demand, selects the inventory holding option. In this case, pooling in lower echelon becomes more effective, substitution becomes a valuable tool for balancing the capacity. It is observed that when substitution is more profitable, then rate of substitution is also higher.

  Under very high or very low demand, fewer substitutions occur. In high demand, there is an option of customer rejection, which is used frequently, and substitution rate is low. When the lower echelon holding cost is high, pooling in upper echelon is more profitable. For high lower echelon holding cost, the system tends to hold more upper echelon inventory. Then, upper echelon pooling becomes more profitable. When there is an inclination to hold inventory at the lower echelon, pooling in lower echelon (for some reason) becomes more profitable.

- In the AAAR case, the profits are close under both pooling types. Especially, under high lower-echelon holding cost, and low arrival rate, pooling at the upper-echelon works better than pooling at the lower echelon.

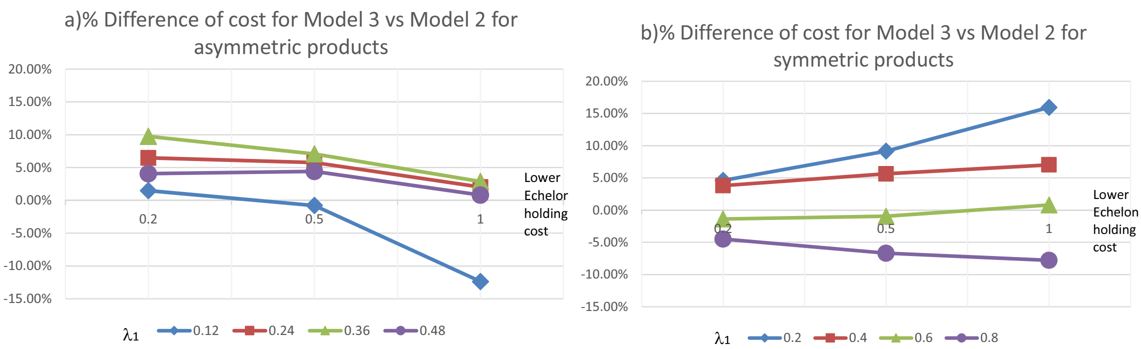


Figure B.7: % Difference for cost of Model 3 versus Model 2 a)AASR b)SASR

In Figure B.7, it is shown which pooling type is more effective under SASR case versus AASR case. According to the analysis, when the products are symmetrical, increase in demand rate decreases the net gain of using lower echelon pooling compared to upper echelon pooling. When the demand rate is low, the effect of lower echelon pooling is the highest. In the SASR case; when the customer demand rate is at the high end or low end, the benefit of using lower echelon pooling instead of upper echelon decreases.

- When final products are symmetrical (in the SASR case), in the case of high lower-echelon holding cost, pooling in lower echelon is more profitable. For asymmetrical products (in the AASR case), the relation is reverse. The reason is, symmetrical and asymmetrical products cases respond to high lower echelon holding cost differently. In SASR, the system reacts by decreasing the lower echelon inventory, while in AASR the systems reacts by shifting the inventory to upper echelon. Because the replenishment rates are not in proportion to the arrival rates in AASR, the capacity is not used effectively. In this case pooling in upper echelon becomes more valuable.

APPENDIX C

CALCULATION OF BIAS DIFFERENCES IN THE PRIORITY MTS-QUEUE

In the following, for the two-product priority MTS-queue system we propose an approach to determine the bias differences. Specifically, the following bias differences will be determined:

$$g_1(x_1) = \begin{cases} v_1(x_1-1, x_{21}) - v_1(x_1, x_{22}) & \text{if } x_1 > 0, x_{21} \in \mathbb{Z}, x_{22} \in \mathbb{Z} \\ 0 & \text{if } x_1 = 0, \end{cases}$$

$$g_2^1(x_1, x_2) = \begin{cases} v_2(x_1, x_2-1) - v_2(x_1, x_2) & \text{if } x_2 > 0 \\ 0 & \text{if } x_2 = 0, \end{cases}$$

$$g_2^2(x_1, x_2) = \begin{cases} v_2(x_1-1, x_2) - v_2(x_1, x_2) & \text{if } x_1 > 0, x_2 \geq 0 \\ v_2(x_1, x_2-1) - v_2(x_1, x_2) & \text{if } x_1 = 0, x_2 > 0 \\ v_2(x_1, x_2) - v_2(x_1, x_2) & \text{if } x_1 = 0, x_2 = 0 \end{cases}$$

$$g_2^3(x_1, x_2) = \begin{cases} v_2(x_1-1, x_2+1) - v_2(x_1, x_2) & \text{if } x_1 > 0, x_2 \geq 0 \\ v_2(x_1, x_2) - v_2(x_1, x_2) & \text{if } x_1 = 0 \end{cases}$$

In the following, in LemmasC.1.1, C.2.1,C.3.1,C.4.1, we determine $g_1(x_2)$, $g_2^1(x_1, x_2)$, $g_2^2(x_1, x_2)$ and $g_2^3(x_1, x_2)$, respectively.

Consider the Markov chain corresponding to priority MTS-queue. Consider two states (x_{11}, x_{21}) and (x_{12}, x_{22}). Let $V_k(x_{11}, x_{21}, t)$ and $V_k(x_{12}, x_{21}, t)$ denote the expected total cost incurred due to product k during interval $[0, t]$ given that the initial states at time $t = 0$ are (x_{11}, x_{21}) and (x_{12}, x_{22}), respectively. Then, the bias difference $v_k(x_{11}, x_{21}) - v_k(x_{12} - x_{22})$ is simply the difference in the expected costs (see Puterman, 1994):

$$v_k(x_{11}, x_{21}) - v_k(x_{21} - x_{22}) = \lim_{t\to\infty} V_k(x_{11}, x_{21}, t) - V_k(x_{21}, x_{22}, t)$$

To find $\lim_{t\to\infty} V_k(x_{11}, x_{21}, t) - V_k(x_{12}, x_{22}, t)$ one can take the following approach. Consider one sample path, i.e., a single realization of a sequence of events. Let the process that is characterized by the initial state (x_{11}, x_{21}) be named as Process 1, and the process characterized by the initial state (x_{12}, x_{21}) named as Process 2. If throughout the sample path at some point Process 1 and Process 2 coincide, in that there exists a time point t_1 at which the same state is visited by both processes, then from that time point on same cost is incurred under both processes. Thus, to determine the cost difference due to that sample path, it is sufficient to consider the cost difference until t_1. Taking expectation over all possible sample paths gives the expected cost difference $\lim_{t\to\infty} V_k(x_{11}, x_{21}, t) - V_k(x_{12}, x_{22}, t)$. To determine this expected cost difference, it is sufficient to consider the expected time until the two processes coincide and the cost difference accumulated until that time. In the following we further elaborate on this approach. Before doing so, note that to determine $g_1(x_1)$, $g_2^1(x_1, x_2)$, $g_2^2(x_1, x_2)$, $g_2^3(x_1, x_2)$ $g_2^3(x_1, x_2)$, we only consider the cost incurred due to product 1, where the cost consists of holding and backordering cost. Due the definition of these difference functions, it might be that the costs incurred due to product k become identical for both processes, before process 1 and process 2 totally coincide. We explain what this argument means in details in the following.

C.1 Determining $g_1(x_1)$

The definition of $g_1(x_1)$ is $v_1(x_1 - 1, x_{21}) - v_1(x_1 - x_{22})$. In this expression note that shortfall level for product 2, x_{21} and x_{22} does not affect the value of $g_1(x_1)$. In the following we describe how to determine $g_1(x_1)$. Let Process 1 be the (Markov) process that starts at the state $(x_1 - 1, x_{21})$, and Process 2 be the (Markov) process that starts at state (x_1, x_{22}). Soon it will be clear that the shortfall level due to Product 2, x_{21} and x_{22} are irrelevant in determining the expected cost difference due to product 1. Note $x_1 > 0$. Both processes are using priority policy, giving priority to product 1. For Process j ($j = 1, 2$), denote the state variable corresponding to product k ($k = 1, 2$) at time t with $X_k^j(t)$, and the scalar shortfall value of product k with x_k. The difference function, $g_1(x_1)$, can be expressed as follows:

$$\begin{aligned} g_1(x_1) = \; & E \; [\int_0^\infty \Big(h_1((S_1 - X_1^1(t))^+ - (S_1 - X_1^2(t))^+) \\ + \; & b_1((S_1 - X_1^1(t))^- - (S_1 - X_1^2(t))^-) \Big) \, dt | X_1^1(0) = x_1 - 1, X_1^2(0) = x_1], \end{aligned}$$

where $a^+ = max(0, a)$, $a^- = max(0, -a)$. Note, when deriving $g_1(x_1)$ product 2 shortfall level is irrelevant.

Consider both processes under a given sample path. Let the time point at which the shortfall level of product 1 under process 2 becomes zero be denoted with T_{x_1}. In other words, T_{x_1} is the first time $X_1^2(t) = 0$ given $X_1^2(0) = x_1$. Thus, T_{x_1} is the random variable denoting the time until product 1 shortfall is cleared under process 2. At time T_{x_1}, a production completion occurs. For process 2 the production completion decreases the shortfall for product 1 by one to zero, whereas for process 1 it decreases the shortfall for product 2 or if there is no product 2 shortfall, there will be no production (i.e., the production completion would be a fictitious one). Therefore, at T_{x_1}, $X_1^1(T_{x_1}) = 0$ and $X_1^2(T_{x_1}) = 0$ and for $t \geq T_{x_1}$ the evolution of shortfall for product 1 is identical for both processes (note evolution of shortfall of product 2 may not be identical under the two processes, however that is irrelevant). The costs incurred by product 1 under the two processes are also identical past T_{x_1}.

Hence, $g_1(x_1)$ can be expressed as,

$$\begin{aligned}
g_1(x_1) &= E[\int_0^{T_{x_1}} \Big(h_1((S_1 - X_1^1(t))^+ - (S_1 - X_1^2(t))^+) \\
&+ b_1((S_1 - X_1^1(t))^- - (S_1 - X_1^2(t))^-)\Big)\, dt | X_1^1(0) = x_1 - 1, X_1^2(0) = x_1] \\
&= E[\int_0^{T_{x_1}} \Big(h_1((S_1 - X_1^2(t) + 1)^+ - (S_1 - X_1^2(t))^+) \\
&+ b_1((S_1 - X_1^2(t) + 1)^- - (S_1 - X_1^2(t))^-)\Big)\, dt | X_1^2(0) = x_1] \\
&= E[\int_0^{T_{x_1}} (h_1 I_{\{S_1 - X_1^2(t) \geq 0\}} - b_1(1 - I_{\{S_1 - X_1^2(t) \geq 0\}}))\, dt | X_1^2(0) = x_1] \\
&= (h_1 + b_1) E[\int_0^{T_{x_1}} I_{\{S_1 - X_1^2(t) \geq 0\}}\, dt | X_1^2(0) = x_1] \\
&- b_1 E[\int_0^{T_{x_1}} 1\, dt | X_1^2(0) = x_1] \\
&= (h_1 + b_1) e_1(x_1) - b_1 E[T_{x_1}],
\end{aligned}$$

where $e_1(x_1)$ denotes $E[\int_0^{T_{x_1}} I_{\{S_1 - X_1^2(t) \geq 0\}}\, dt | X_1^2(0) = x_1]$. In other words, $e_1(x_1)$ is the expected total time process 2 spends in states with non-negative net inventory level for product 1, starting with a shortfall level of x_1 for product 1 at $t = 0$, until $t = T_{x_1}$.

In the following, we discuss how $E[T_{x_1}]$ and $e_1(x_1)$ are derived, respectively. Note $e_1(0) = 0$, since $T_{x_1} = 0$. Considering product 1, T_{x_1} is the time until all shortfall for product 1 is cleaned under process 2. Note that T_{x_1} is the busy time for an $M/M/1$ queue with λ_1 arrival rate and μ service rate given at the beginning x_1 customers are present in the system, since product 1 is the higher priority product. Hence,

$$E[T_{x_1}] = \frac{x_1}{\mu - \lambda_1}. \tag{C.1.1}$$

For derivation of $e_1(x_1)$ we introduce the following lemma.

Lemma C.1.1. *The difference function, $g_1(x_1)$, is expressed as:*

$$g_1(x_1) = -b_1 \frac{x_1}{\mu - \lambda_1} + (h_1 + b_1)(\frac{\min(S_1, x_1)}{\mu - \lambda_1} - \frac{1}{\mu - \lambda_1} \sum_{l=(S_1 - x_1)^+}^{S_1 - 1} (\frac{\lambda_1}{\mu})^{l+1}.$$

The function $e_1(x_1)$ defined on $x_1 \geq 0$ is expressed as follows:

$$e_1(x_1) = \frac{\min(S_1, x_1)}{\mu - \lambda_1} - \frac{1}{\mu - \lambda_1} \sum_{l=(S_1 - x_1)^+}^{S_1 - 1} (\frac{\lambda_1}{\mu})^{l+1}$$

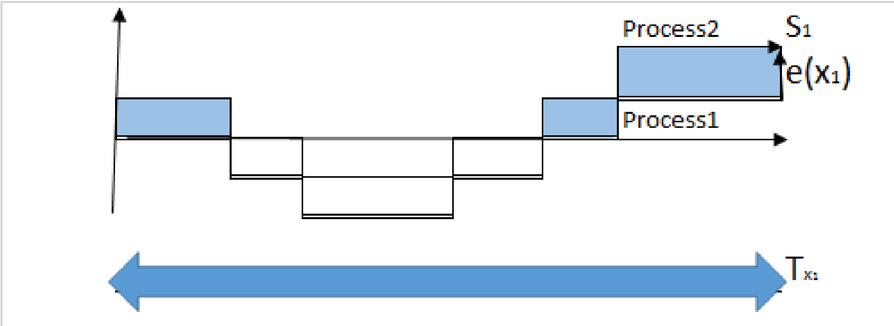


Figure C.1: Coupling based on single product problem-Veatch and Wein

Proof. We have already discussed why $g_1(x_1)$ is a function of $e_1(x_1)$. In the proof we show how to derive $e_1(x_1)$.

To derive $e_1(x_1)$, first define $\bar{e}_1(x_1)$ as the expected total time the process spends in states corresponding to non-negative inventory level for product 1, starting with state x_1 until the process visits state $x_1 - 1$. Hence $\bar{e}_1(x_1) = e_1(x_1) - e_1(x_1 - 1)$.

Note that, $e_1(x_1) = \sum_{l=1}^{x_1} \bar{e}_1(l) + e_1(0)$. By definition $e_1(0) = 0$. Three mutually exclusive cases are under consideration:

<u>For $x_1 > S_1$.</u> In this case, $\bar{e}_1(x_1) = 0$, since starting with x_1 the process is always in negative net inventory until it visits $x_1 - 1$.

<u>For $x_1 = S_1$.</u> The function $\bar{e}_1(x_1)$ can be expressed as:

$$\bar{e}_1(S_1) = \frac{1}{\lambda_1 + \mu} + \frac{\mu}{\lambda_1 + \mu}(0) + \frac{\lambda_1}{\lambda_1 + \mu}\bar{e}_1(S_1). \tag{C.1.2}$$

The equation follows from the first-step analysis. When product 1 has a shortfall level of S_1, the net inventory level of product 1 is 0. Expected time until next event is $\frac{1}{\lambda_1+\mu}$. If the next (relevant) event is a product 1 arrival, then expected time from the point of arrival of the customer until $S_1 - 1$ shortfalls is simply equal to $\bar{e}_1(S_1)$. The reason is, in the case of a customer arrival, state is $S_1 + 1$, and this corresponds to negative net inventory. From that point on, the process will not spend time at states that correspond to nonnegative net inventory, until it again visits state S_1.

If on the other hand, the next event is a service completion, then the state with shortfall level of $S_1 - 1$ have been reached.

From Eq. C.1.2, $\bar{e}_1(S_1) = \frac{1}{\mu}$.

For $0 < x_1 < S_1$. The function $\bar{e}_1(x_1)$ is expressed as:

$$\bar{e}_1(x_1) = \frac{1}{\lambda_1 + \mu} + \frac{\lambda_1}{\lambda_1 + \mu}(\bar{e}_1(x_1) + \bar{e}_1(x_1 + 1)) \qquad \text{for } 0 < x_1 < S_1, \quad \text{(C.1.3)}$$

where $\bar{e}_1(S_1) = \frac{1}{\mu}$. If the terms of the equation are arranged,

$$\mu \bar{e}_1(x_1) - \lambda_1 \bar{e}_1(x_1 + 1) = 1, \qquad 0 < x_1 < S_1.$$

This is a non-homogeneous recursive equation with the characteristic equation,

$$\mu - \lambda_1 r = 0$$

Multiply both sides with $(r - 1)$, and the characteristic equation of the homogeneous recursive equation is,

$$(\mu - \lambda_1 r)(r - 1) = 0$$

This equation has two roots one is $r = \frac{\mu}{\lambda_1}$, the other is $r = 1$. So, $\bar{e}_1(x_1)$ has following form:

$$\bar{e}_1(x_1) = C_0 1^{x_1} + C_1 (\frac{\mu}{\lambda_1})^{x_1}, \qquad 0 < x_1 \leq S_1. \qquad \text{(C.1.4)}$$

Note that the equation also holds for $x_1 = S_1$, because $\bar{e}_1(S_1) = \frac{1}{\mu}$ is of form C.1.3 given $\bar{e}_1(S_1 + 1) = 0$. If Eq. C.1.4 is placed in Eq. C.1.3, then for each x_1 satisfying $0 < x_1 < S_1$:

$$\mu C_0 + \mu C_1 (\frac{\mu}{\lambda_1})^{x_1} - \lambda_1 C_0 - \lambda_1 C_1 (\frac{\mu}{\lambda_1})^{x_1 + 1} = 1.$$

Then $C_0 = \frac{1}{\mu - \lambda_1}$. From $\bar{e}_1(S_1) = \frac{1}{\mu}$ and Eq. C.1.4, $C_1 = \frac{-\lambda_1}{(\mu - \lambda_1)\mu}(\frac{\mu}{\lambda_1})^{-S_1}$. Therefore,

$$\bar{e}_1(x_1) = \frac{1}{\mu - \lambda_1}(1 - (\frac{\lambda_1}{\mu})^{S_1 - x_1 + 1}), \qquad 0 < x_1 \leq S_1.$$

Since $e_1(x_1) = \sum_{\ell=1}^{x_1} \bar{e}_1(\ell) + e_1(0)$, $e_1(x_1)$ is obtained as follows,

$$e_1(x_1) = \begin{cases} \frac{x_1}{\mu - \lambda_1} - \frac{1}{\mu - \lambda_1} \sum_{l = S_1 - x_1}^{S_1 - 1} (\frac{\lambda_1}{\mu})^{l+1} & 0 < x_1 \leq S_1 \\ \frac{S_1}{\mu - \lambda_1} - \frac{1}{\mu - \lambda_1} \sum_{l=0}^{S_1 - 1} (\frac{\lambda_1}{\mu})^{l+1} & x_1 > S_1 \end{cases}$$

□

C.2 Determining $g_2^1(x_1, x_2)$

Next, we determine $g_2^1(x_1, x_2)$ in Lemma C.2.1. To derive $g_2^1(x_1, x_2)$, we again define two processes as follows. At time $t = 0$, process 1 starts at state $(x_1, x_2 - 1)$ and process 2 starts at state (x_1, x_2). By definition of $g_2^1(x_1, x_2)$, the cost difference incurred only due to product 2 is under consideration. Let T_{x_1,x_2} be for Process 2, the first passage time to state $(X_1^2(t), X_2^2(t)) = (0, 0)$, given that current state (at time $t = 0$) is (x_1, x_2).

For a given x_1 and x_2, in the following we first express $g_1^2(x_1, x_2)$.

$$\begin{aligned} g_2^1(x_1, x_2) &= E[\int_0^\infty \Big(h_2((S_2 - X_2^1(t))^+ - (S_2 - X_2^2(t))^+) \\ &+ b_2((S_2 - X_2^1(t))^- - (S_2 - X_2^2(t))^-)\Big)\, dt \\ &\quad |X_2^1(0) = x_2 - 1, X_1^1(0) = x_1, X_2^2(0) = x_2, X_1^2(0) = x_1] \end{aligned}$$

Note that, for a given sample path, the cost incurred due to product 2 will be identical under both processes once process 2 visits state $(X_1^2(t), X_2^2(t)) = (0, 0)$, which happens at time $t = T_{x_1,x_2}$. Thus,

$$\begin{aligned} g_2^1(x_1, x_2) &= E[\int_0^{T_{x_1,x_2}} \Big(h_2((S_2 - X_2^1(t))^+ - (S_2 - X_2^2(t))^+) \\ &\quad +b_2((S_2 - X_2^1(t))^- - (S_2 - X_2^2(t))^-)\Big)\, dt \\ &\quad |X_2^1(0) = x_2 - 1, X_1^1(0) = x_1, X_2^2(0) = x_2, X_1^2(0) = x_1] \\ &= E[\int_0^{T_{x_1,x_2}} (h_2((S_2 - X_2^2(t) + 1)^+ - (S_2 - X_2^2(t))^+) \\ &\quad +b_2((S_2 - X_2^2(t) + 1)^- - (S_2 - X_2^2(t))^-))\, dt|X_2^2(0) = x_2, X_1^2(0) = x_1] \\ &= E[\int_0^{T_{x_1,x_2}} (h_2 I_{\{S_1 - X_2^2(t) \geq 0\}} \\ &\quad -b_2(1 - I_{\{S_2 - X_2^2(t) \geq 0\}}))\, dt|X_2^2(0) = x_2, X_1^2(0) = x_1] \\ &= (h_2 + b_2)E[\int_0^{T_{x_1,x_2}} I_{\{S_1 - X_2^2(t) \geq 0\}}\, dt|X_2^2(0) = x_2, X_1^2(0) = x_1] \\ &\quad -b_2 E[\int_0^{T_{x_1,x_2}} 1\, dt|X_2^2(0) = x_2, X_1^2(0) = x_1] \\ &= (h_2 + b_2)e_2^1(x_1, x_2) - b_2 E[T_{x_1,x_2}], \end{aligned}$$

where $e_2^1(x_1, x_2)$ is the expected time process 2 spends in states with non-negative net inventory level for product 2 during the time interval $[0, T_{x_1,x_2}]$.

First observe that, in analogy to an $M/M/1$ queue with arrival rate $\lambda_1 + \lambda_2$ and service rate μ, and $x_1 + x_2$ customers waiting,

$$E[T_{x_1,x_2}] = \frac{x_1 + x_2}{\mu - \lambda_1 - \lambda_2}.$$

Also observe that, $e_2^1(0,0) = 0$ since $T_{0,0} = 0$. In the following lemma, we determine $e_2^1(x_1, x_2)$.

Lemma C.2.1. *For $x_1 \geq 0$, $0 < x_2$, the difference function, $g_2^1(x_1, x_2)$, is expressed as:*

$$g_2^1(x_1, x_2) = -b_2(\frac{x_1 + x_2}{\mu - \lambda_1 - \lambda_2}) + (h_2 + b_2)e_2^1(x_1, x_2)$$

The function $e_2^1(x_1, x_2)$, defined for $x_1 \geq 0$ and $x_2 > 0$, is expressed as follows:

$$e_2^1(x_1, x_2) = \begin{cases} C_0^{x_2} + \sum_{\ell=1}^{S_2 - x_2 + 1} C_\ell^{x_2}(x_1)^{\ell-1}(\alpha)^{x_1}, & 0 < x_2 \leq S_2 \\ e_2^1(0, S_2), & x_2 > S_2 \end{cases} \tag{C.2.1}$$

where,

(i)

$$e_2^1(0, x_2) = \sum_{\ell=1}^{x_2} \bar{e}_2^1(0, \ell) + e_2^1(0,0), \ x_2 = 1, 2, \ldots, S_2$$

In the equation, for a given x_2, $0 < x_2 \leq S_2$, $\bar{e}_2^1(0, x_2)$ can be computed recursively as:

$$\begin{aligned} \bar{e}_2^1(0, x_2) \quad = \quad & \frac{1}{\mu} + \frac{\lambda_1}{\mu}(\frac{S_2 - x_2 + 1}{\lambda_2} + \alpha \sum_{m=1}^{S_2 - x_2 + 1} C_m^{x_2} \\ & + \sum_{m=x_2+1}^{S_2} \bar{e}_2^1(0, m)) + \frac{\lambda_2}{\mu}\bar{e}_2^1(0, x_2 + 1) \\ & \textit{for } x_2 = 1, 2, \ldots, S_2, \end{aligned}$$

where $\bar{e}_2^1(0, S_2 + 1) = 0$ and $e_2^1(0,0) = 0$,

(ii)

$$
\begin{aligned}
C_0^{x_2} &= \frac{S_2 - x_2 + 1}{\lambda_2} + e_2^1(0, S_2), \;\; x_2 = 1, 2, \ldots, S_2 \\
C_1^{x_2} &= -\frac{S_2 - x_2 + 1}{\lambda_2} - e_2^1(0, S_2) + e_2^1(0, x_2), \;\; x_2 = 1, 2, \ldots, S_2 \\
C_\ell^{x_2} &= \frac{1}{(\ell - 1)(\mu - \lambda_1\alpha^2)} \Big(\sum_{m=\ell+1}^{S_2 - x_2 + 1} C_m^{x_2} \binom{m-1}{\ell-2} (\lambda_1\alpha^2 \\
&\quad + \mu(-1)^{m-\ell+1}) + \lambda_2 \alpha C_{\ell-1}^{x_2+1} \Big) \\
&\qquad\qquad \ell = 2, 3, \ldots, S_2 - x_2 + 1, \;\; x_2 = 1, 2, \ldots, S_2 - 1
\end{aligned}
$$

(iii) $\alpha = \frac{(\lambda_1+\lambda_2+\mu) - \sqrt{(\lambda_1+\lambda_2+\mu)^2 - 4(\lambda_1)\mu}}{2\lambda_1}$.

Proof. We have already discussed $g_1^2(x_1, x_2)$ as a function of $e_1^2(x_1, x_2)$. In the proof we show how to derive the expression for $e_1^2(x_1, x_2)$.

We do the proof in two parts. In the first part we show by induction that $e_2^1(x_1, x_2)$ is of the form Eq. C.2.1. In the second part we propose a method to determine the coefficients $C_\ell^{x_2}$, $x_2 = 1, 2, \ldots, S_2$ and $\ell = 0, 1, \ldots, S_2 - x_2 + 1$.

Part I We derive $e_2^1(x_1, x_2)$ for all possible values of x_1 and x_2 to show that Eq. C.2.1 holds.

<u>For $x_2 > S_2$.</u> Consider the time until process 2 visits the state with zero shortfall for product 1 and S_2 shortfall for product 2. During this time, the process will never visit a state with non-negative net inventory level for product 2. This implies, for $x_2 > S_2$, $e_2^1(x_1, x_2) = e_2^1(0, S_2)$.

<u>$0 < x_2 \leq S_2$.</u> To show by induction that Eq. C.2.1 holds, we first need to show that the equation holds at the boundary. Let the boundary be defined by $x_2 = S_2$. So, we first prove for $x_2 = S_2$ that $e_2^1(x_1, x_2)$ is of the form Eq. C.2.1. Then, using induction we do the proof for $0 < x_2 < S_2$.

<u>$x_2 = S_2$.</u> If we make the first step analysis for $e_2^1(x_1, S_2)$:

$$
\begin{aligned}
e_2^1(x_1, S_2) &= \frac{1}{\lambda_1 + \lambda_2 + \mu} + \frac{\lambda_1}{\lambda_1 + \lambda_2 + \mu} e_2^1(x_1 + 1, S_2) \\
&+ \frac{\lambda_2}{\lambda_1 + \lambda_2 + \mu} e_2^1(0, S_2) + \frac{\mu}{\lambda_1 + \lambda_2 + \mu} e_2^1(x_1 - 1, S_2) \;\; \text{for } x_1 > 0,
\end{aligned}
$$

Rearranging the terms:

$$\begin{aligned}(\lambda_1+\lambda_2+\mu)e_2^1(x_1,S_2)-(\lambda_1)e_2^1(x_1+1,S_2)&\\ -\mu e_2^1(x_1-1,S_2)&=1+\lambda_2 e_2^1(0,S_2)\\ &\qquad\text{for } x_1>0. \quad \text{(C.2.2)}\end{aligned}$$

This is a second-order non-homogenous linear difference equation. The characteristic equation is:

$$-(\lambda_1 r^2-(\lambda_1+\lambda_2+\mu)r+\mu)=0.$$

To obtain a homogeneous difference equation, multiply the characteristic equation with $(r-1)$,

$$-(r-1)(\lambda_1 r^2-(\lambda_1+\lambda_2+\mu)r+\mu)=0.$$

The roots of the characteristic equation is $r_1=1$ and $r_{\{2,3\}}=\frac{(\lambda_1+\lambda_2+\mu)\pm\sqrt{(\lambda_1+\lambda_2+\mu)^2-4(\lambda_1)\mu}}{2\lambda_1}$ such that $r_2<r_3$. Let $\alpha=r_2$, and $\alpha_+=r_3$. Note that $\alpha_+>1$ and $0<\alpha<1$. So solution of the difference equation is:

$$e_2^1(x_1,S_2)=C_0^{S_2}(1)^{x_1}+C_1^{S_2}(\alpha)^{x_1}+C_2^{S_2}(\alpha_+)^{x_1} \quad \text{for } x_1\geq 0. \quad \text{(C.2.3)}$$

Note that for $x_1=1$ the equation includes the term $e_2^1(0,S_2)$. Thus $e_2^1(0,S_2)$ must also follow the same equation form.

In the following, we show that $e_2^1(x_1,S_2)$ follows the equation form in Eq. C.2.1. This is equivalent to saying that $C_2^{S_2}=0$. Then, we show that $e_2^1(x_1,x_2),\ 0<x_2<S_2$ also follow the equation form in Eq. C.2.1.

To show that $e_2^1(x_1,S_2)$ is of form Eq. C.2.1, we need to show $C_2^{S_2}=0$. Observe that $\alpha_+>1$ which implies, if $C_2^{S_2}\neq 0$ then $e_2^1(x_1,S_2)$ must be diverging if we let x_1 go to ∞. However, note

$$\lim_{x_1\to\infty} e_2^1(x_1,S_2)=\frac{1}{\lambda_2}+e_2^1(0,S_2).$$

The expression states that, given that there is too much shortfall for product 1, the amount of time spent in non-negative states for product 2 will be equal to the time until the shortfall of product 2 becomes S_2+1, plus

$e_2^1(0,S_2)$. The reason is the server will not get chance to work on the second priority product when $x_1 \Rightarrow \infty$, and thus product 2 will definitely visit state S_2+1 after an expected time of $\frac{1}{\lambda_2}$.

By definition, $e_2^1(0,S_2) \leq E[T_{0,S_2}] = \frac{S_2}{\mu-\lambda_1-\lambda_2}$. Hence the term on the right is less than ∞. Which shows, the limit does not diverge. This is possible only if $C_2^{S_2}=0$. Therefore, for $x_2=S_2$, $e_2^1(x_1,S_2)$ is of the form in Eq. C.2.1.

$\underline{0 < x_2 < S_2.}$ Next, we show $e_2^1(x_1,x_2)$ is of the form Eq. C.2.1. We use induction to show this. Specifically, it is assumed that if $e_2^1(x_1,x_2+1)$ is of the form Eq. C.2.1, then $e_2^1(x_1,x_2)$ is also of the form Eq. C.2.1. Note that for the boundary value, $e_2^1(x_1,S_2)$ is already shown to be of the form Eq. C.2.1.

For $x_1>0$, following the first step analysis, $e_2^1(x_1,x_2)$ is expressed as:

$$\begin{aligned} e_2^1(x_1,x_2) &= \frac{1}{\lambda_1+\lambda_2+\mu}+\frac{\lambda_1}{\lambda_1+\lambda_2+\mu}e_2^1(x_1+1,x_2) \\ &+\frac{\mu}{\lambda_1+\lambda_2+\mu}e_2^1(x_1-1,x_2) \\ &+\frac{\lambda_2}{\lambda_1+\lambda_2+\mu}e_2^1(x_1,x_2+1) \quad \text{for } 0<x_2<S_2 \end{aligned} \tag{C.2.4}$$

Rearranging the terms in Eq. C.2.4,

$$\begin{aligned} (\lambda_1+\lambda_2+\mu)e_2^1(x_1,x_2)-(\lambda_1)e_2^1(x_1+1,x_2)-\mu e_2^1(x_1-1,x_2) = \\ 1+\lambda_2 e_2^1(x_1,x_2+1) \quad \text{for } 0<x_2<S_2. \end{aligned}$$

This is a second order non-homogeneous difference equation. Assume for x_2+1, Eq. C.2.1 holds. Then,

$$\begin{aligned} (\lambda_1+\lambda_2+\mu)e_2^1(x_1,x_2)-(\lambda_1)e_2^1(x_1+1,x_2)-\mu e_2^1(x_1-1,x_2) = \\ \lambda_2 C_0^{x_2+1}(1)^{x_1}+\lambda_2\sum_{l=1}^{S_2-x_2} C_\ell^{x_2+1}(x_1)^{\ell-1}(\alpha)^{x_1} \quad \text{for } 0<x_2<S_2 \end{aligned}$$

The characteristic equation of the second order non-homogeneous difference equation is

$$-(\lambda_1 r^2-(\lambda_1+\lambda_2+\mu)r+\mu)=0.$$

This characteristics equation has roots $r_1 = \alpha$, and $r_2 = \alpha_+$ (as determined in the previous characteristic equation). Converting the non-homogeneous equation into an homogeneous one, the characteristic equation becomes:

$$-(\lambda_1 r^2 - (\lambda_1 + \lambda_2 + \mu)r + \mu)(r - \alpha)^{S_2 - x_2}(r - 1) = 0$$

This equation has $S_2 - x_2 + 3$ roots. One of the roots is 1, the other root is α_+, and the remaining $S_2 - x_2 + 1$ of the roots has value α. So for $x_1 \geq 0$, $e_2^1(x_1, x_2)$ is of the form

$$\begin{aligned} e_2^1(x_1, x_2) &= C_0^{x_2}(1)^{x_1} + \sum_{\ell=1}^{S_2 - x_2 + 1} C_\ell^{x_2}(x_1)^{\ell-1}(\alpha)^{x_1} + C_{S_2 - x_2 + 2}^{x_2}(\alpha_+)^{x_1} \\ &\quad \text{for } 0 < x_2 < S_2 \end{aligned}$$

In the equation above $C_0^{x_2},\ C_1^{x_2}, \ldots, C_{S_2-x_2+2}^{x_2}$ are constants. (The number of constants are equal to the number of roots). The constant may take different values for different values of x_2.

To complete the proof that $e_2^1(x_1, x_2)$ is of form Eq. C.2.1, we show that the constant $C_{S_2+x_2+2}^{x_2}$ is equal to 0. To see why, consider $\lim_{x_1 \to \infty} e_2^1(x_1, x_2)$. Note that,

$$\lim_{x_1 \to \infty} e_2^1(x_1, x_2) = \frac{S_2 - x_2 + 1}{\lambda_2} + e_2^1(0, S_2), \quad 0 < x_2 < S_2 \qquad \text{(C.2.5)}$$

Since $0 \leq e_2^1(0, S_2) \leq E[T_{0,S_2}] = \frac{S_2}{\mu - \lambda_1 - \lambda_2}$, this limit exists. So $C_{S_2-x_2+2}^{x_2}$ must be 0 since otherwise $e_2^1(x_1, x_2)$ diverges as $x_1 \Rightarrow \infty$.

This completes the induction-based proof. We conclude that for $x_1 \geq 0,\ 0 < x_2 \leq S_2$, $e_2^1(x_1, x_2)$ is of the form in Eq. C.2.1.

Part II

Next step is to determine the coefficients $C_\ell^{x_2}$, $x_2 = 1, 2, \ldots, S_2$ and $\ell = 0, 1, \ldots, S_2 - x_2 + 1$ (it is already shown that $C_{S_2-x_2+2}^{x_2} = 0$). To determine these coefficients, a procedure is proposed. Before the procedure the following observations are made.

1. For calculation of $C_0^{x_2}$ and $C_1^{x_2}$, we look at the limit

$$\begin{aligned}\lim_{x_1\to\infty} e_2^1(x_1,x_2) &= \lim_{x_1\to\infty} C_0^{x_2} + \sum_{l=1}^{S_2-x_2+1} C_l^{x_2} x_1^{l-1}\alpha^{x_1} \\ &= C_0^{x_2} \\ &= \frac{S_2-x_2+1}{\lambda_2} + e_2^1(0,S_2),\ 0<x_2\le S_2. \quad \text{(C.2.6)}\end{aligned}$$

. Since

$$\begin{aligned}e_2^1(0,x_2) &= C_0^{x_2} + \sum_{l=1}^{S_2-x_2+1} C_l^{x_2} x_1^{l-1}|_{x_1=0} \\ &= C_0^{x_2} + C_1^{x_2}\end{aligned}$$

we deduce

$$C_1^{x_2} = -\frac{S_2-x_2+1}{\lambda_2} - e_2^1(0,S_2) + e_2^1(0,x_2) \quad \text{(C.2.7)}$$

.

Note $C_1^{S_2} = -\frac{1}{\lambda_2}$.

2. To determine $C_0^{x_2}$, $0 < x_2 \le S_2$, $e_2^1(0,S_2)$ needs to be determined. To determine $C_1^{x_2}$, $0 < x_2 \le S_2$, $e_2^1(0,x_2)$ needs to be determined.

3. Let $\bar{e}_2^1(x_1,x_2)$ denote the expected time spent under the nonnegative net inventory for product 2, until, product 2 shortfall decreases to x_2-1, starting with state (x_1,x_2). Note $\bar{e}_2^1(x_1,x_2) = e_2^1(x_1,x_2) - e_2^1(0,x_2-1)$ and thus

$$e_2^1(0,x_2) = \sum_{l=1}^{x_2} \bar{e}_2^1(0,l) + e_2^1(0,0),$$

where $e_2^1(0,0)=0$.

4. From first step analysis, for $0 < x_2 \le S_2$

$$\begin{aligned}\bar{e}_2^1(0,x_2) &= \frac{1}{\lambda_1+\lambda_2+\mu} + \frac{\lambda_1}{\lambda_1+\lambda_2+\mu}\bar{e}_2^1(1,x_2) \\ &+ \frac{\lambda_2}{\lambda_1+\lambda_2+\mu}(\bar{e}_2^1(0,x_2) + \bar{e}_2^1(0,x_2+1)). \quad \text{(C.2.8)}\end{aligned}$$

where $\bar{e}_2^1(0,S_2+1)) = 0$, since $e_2^1(0,S_2+1) = e_2^1(0,S_2)$. In Eq. C.2.8,

$\bar{e}_2^1(1,x_2)$ is determined as follows. From Eq. C.2.1

$$\begin{aligned}
\bar{e}_2^1(1,x_2) &= C_0^{x_2} + \alpha \sum_{t=1}^{S_2-x_2+1} C_t^{x_2} - e_2^1(0,x_2-1) \\
&= \frac{S_2-x_2+1}{\lambda_2} + e_2^1(0,S_2) + \alpha \sum_{t=1}^{S_2-x_2+1} C_t^{x_2} - e_2^1(0,x_2-1) \\
&= \frac{S_2-x_2+1}{\lambda_2} + \alpha \sum_{t=1}^{S_2-x_2+1} C_t^{x_2} + \sum_{t=x_2}^{S_2} \bar{e}_2^1(0,t). \qquad \text{(C.2.9)}
\end{aligned}$$

If we plug Eq. C.2.9 in Eq. C.2.8:

$$\begin{aligned}
\bar{e}_2^1(0,x_2) &= \frac{1}{\mu} + \frac{\lambda_1}{\mu}\Big(\frac{S_2-x_2+1}{\lambda_2} + \alpha \sum_{t=1}^{S_2-x_2+1} C_t^{x_2} \\
&\quad + \sum_{t=x_2+1}^{S_2} \bar{e}_2^1(0,t)\Big) + \frac{\lambda_2}{\mu}\bar{e}_2^1(0,x_2+1). \qquad \text{(C.2.10)}
\end{aligned}$$

Given $0 < x_2 \leq S_2$, to determine $\bar{e}_2^1(0,x_2)$, one needs to have determined all $\bar{e}_2^1(0,\ell),\ x_2+1 < \ell \leq S_2$. Furthermore, one needs coefficient values $C_1^{x_2}, C_2^{x_2}, \ldots, C_{S_2-x_2+1}^{x_2}$.

5. From first step analysis

$$\begin{aligned}
e_2^1(x_1,x_2) &= \frac{1}{\lambda_1+\lambda_2+\mu} + \frac{\lambda_1}{\lambda_1+\lambda_2+\mu} e_2^1(x_1+1,x_2) \\
&\quad + \frac{\mu}{\lambda_1+\lambda_2+\mu} e_2^1(x_1-1,x_2) \\
&\quad + \frac{\lambda_2}{\lambda_1+\lambda_2+\mu} e_2^1(x_1,x_2+1),
\end{aligned}$$

Given that $e_2^1(x_1,x_2)$ has the form in Eq. C.2.1, the equation will look like:

$$\begin{aligned}
C_0^{x_2} &+ \sum_{\ell=1}^{S_2-x_2+1} C_\ell^{x_2} x_1^{\ell-1}(\alpha)^{x_1} = \frac{1}{\lambda_1+\lambda_2+\mu} \\
&+ \frac{\lambda_1}{\lambda_1+\lambda_2+\mu}\Big(C_0^{x_2} + \sum_{\ell=1}^{S_2-x_2+1} C_\ell^{x_2}(x_1+1)^{\ell-1}(\alpha)^{x_1+1}\Big) \\
&+ \frac{\mu}{\lambda_1+\lambda_2+\mu}\Big(C_0^{x_2} + \sum_{\ell=1}^{S_2-x_2+1} C_\ell^{x_2}(x_1-1)^{\ell-1}(\alpha)^{x_1-1}\Big) \\
&+ \frac{\lambda_2}{\lambda_1+\lambda_2+\mu}\Big(C_0^{x_2+1} + \sum_{\ell=1}^{S_2-x_2} C_\ell^{x_2+1} x_1^{\ell-1}(\alpha)^{x_1}\Big)
\end{aligned}$$

To determine $C_\ell^{x_2},\ \ell = 2,3,\ldots,S_2-x_2+1,\ 0 < x_2 \leq S_2-1$, the coefficients of the terms $x_1^{\ell-1}\alpha^{x_1}$ on the RHS and LHS of the equation above are matched.

Specifically, to obtain $C_\ell^{x_2}$ for a given ℓ and x_2 (where $\ell = 2, 3, \ldots, S_2 - x_2 + 1,\ 0 < x_2 \leq S_2 - 1$) the coefficient of the terms $x_1^{\ell-2}\alpha^{x_1}$ on the LHS and RHS are matched. Then one obtains:

$$\begin{aligned}
C_{l-1}^{x_2}\alpha^{x_1} &= \frac{\lambda_1}{\lambda_1+\lambda_2+\mu}\Big(\sum_{t=l-1}^{S_2-x_2+1} C_t^{x_2}\binom{t-1}{l-2}\alpha^{x_1+1}\Big) \\
&+\frac{\mu}{\lambda_1+\lambda_2+\mu}\Big(\sum_{t=l-1}^{S_2-x_2+1} C_t^{x_2}\binom{t-1}{l-2}(-1)^{t-l+1}\alpha^{x_1-1}\Big) \\
&+\frac{\lambda_2}{\lambda_1+\lambda_2+\mu}(C_{l-1}^{x_2+1}\alpha^{x_1}), \\
&\text{for } l \in \{2, 3, ..., S_2 - x_2 + 1\}, 0 < x_2 \leq S_2 - 1 \quad \text{(C.2.11)}
\end{aligned}$$

In Eq. C.2.11 look at the terms with coefficient $C_{\ell-1}^{x_2}$ on both sides:

$$\begin{aligned}
C_{l-1}^{x_2}\alpha^{x_1} &= \frac{\lambda_1}{\lambda_1+\lambda_2+\mu}C_{\ell-1}^{x_2}\binom{\ell-2}{\ell-2}\alpha\alpha^{x_1}) \\
&+\frac{\mu}{\lambda_1+\lambda_2+\mu}C_{\ell-1}^{x_2}\binom{\ell-2}{\ell-2}(-1)^0\frac{1}{\alpha}\alpha^{x_1}) \quad \text{(C.2.12)}
\end{aligned}$$

Since it holds that $(\lambda_1 + \lambda_2 + \mu)\alpha = \lambda_1\alpha^2 + \mu$, all terms with $C_\ell^{x_2}$ vanish. Only the terms $C_{\ell-1}^{x_2+1}$, $C_\ell^{x_2}$, $C_{\ell+1}^{x_2}, \ldots, C_{S_2-x_2+1}^{x_2}$ remain. Taking the terms with $C_\ell^{x_2}$ to the LHS, one obtains:

$$\begin{aligned}
C_l^{x_2} &= \frac{1}{(l-1)(\mu-\lambda_1\alpha^2)}\Big(\sum_{m=l+1}^{S_2-x_2+1} C_m^{x_2}\binom{m-1}{l-2}(\lambda_1\alpha^2 \\
&+\mu(-1)^{m-l-1}) + \lambda_2 C_{l-1}^{x_2+1}\alpha) \\
&\ell = 2, 3, \ldots, S_2 - x_2 + 1,\ x_2 = 1, 2, \ldots, S_2 - 1. \quad \text{(C.2.13)}
\end{aligned}$$

Now we present the procedure to determine $C_\ell^{x_2},\ \ell = 0, 1, \ldots, S_2 - x_2 + 1,\ x_2 = 1, 2 \ldots, S_2$. The coefficients $C_\ell^{x_2}$ with $\ell > S_2 - x_2 + 1$ are already shown to be 0.

Procedure to determine $C_\ell^{x_2}$ in Eq. C.2.1

Step 0 Let $i = S_2$.

Step 1 Determine C_1^i. From Eq. C.2.7

$$C_1^{x_2} = -\frac{S_2 - x_2 + 1}{\lambda_2} - \sum_{l=i+1}^{S_2} \bar{e}_2^1(0, l). \quad \text{(C.2.14)}$$

Step 2 Determine $C_2^i, C_3^i, \ldots, C_{S_2-i+1}^i$ using Eq. C.2.11. First determine $C_{S_2-i+1}^i$ im terms of $C_{S_2-i}^{i+1}$, then determine $C_{S_2-i}^i$ in terms of $C_{S_2-i+1}^i$ and $C_{S_2-i}^{i+1}$, and so on.

Step 3 Find $\bar{e}_2^1(0, i)$ using Eq. C.2.10. Note $C_1^i, C_2^i, \ldots, C_{S_2-i+1}^i$ are already determined. Furthermore $\bar{e}_2^1(0, i+1), \bar{e}_2^1(0, i+2), \ldots, \bar{e}_2^1(0, S_2)$ are already determined.

Set $i \Leftarrow i - 1$. If $i = 0$ move to Step 4. Otherwise go to Step 1.

Step 4 By now all C_l^i, $i = 1, 2, \ldots, S_2$, $l = 1, 2, \ldots, S_2 - i + 1$ have been determined. Note $C_{S_2-i+2}^i$ are zero. Determine $C_0^{x_2}$, $0 < x_2 \leq S_2$ from Eq. C.2.6.

This completes the procedure to determine the coefficients in the Eq. C.2.1,and the proof.

□

C.3 Determining $g_2^2(x_1, x_2)$

Next, we determine $g_2^2(x_1, x_2) = v_2(x_1 - 1, x_2) - v_2(x_1, x_2)$, $x_1 > 0, x_2 \geq 0$ in Lemma C.3.1. To derive $g_2^2(x_1, x_2)$, we again define two processes as follows. At time $t = 0$, process 1 starts at state $(x_1 - 1, x_2)$ and process 2 starts at state (x_1, x_2). By definition of $g_2^2(x_1, x_2)$, the cost difference incurred only due to product 2 is under consideration. Let T_{x_1} be the first passage time for process 2 to the state with the shortfall level of 0 for product 1 given that current state (at time $t = 0$) has shortfall level of x_1 for product 1. Let T_{x_1,x_2} be for Process 2 the first passage time to the state with shortfall level of 0 for product 1 and product 2, given that current state (at time $t = 0$) has shortfall levels of x_1 and x_2 for products 1 and 2. Considering a sample path, note for $t \in [0, T_{x_1}]$ no cost difference exist due to product 2 under the two processes. The shortfall difference only exists in $t \in [T_{x_1}, T_{x_1,x_2}]$. Thus when determining $g_2^2(x_1, x_2)$ we consider the time interval $[T_{x_1}, T_{x_1,x_2}]$.

For a given $x_1 > 0$ and $x_2 \geq 0$, we first express $g_2^2(x_1, x_2)$ as follows.

$$\begin{aligned}
g_2^2(x_1,x_2) &= E[\int_0^\infty (h_2(((S_2 - X_2^1(t))^+ - (S_2 - X_2^2(t))^+) \\
&+ b_2((S_2 - X_2^1(t))^- - (S_2 - X_2^2(t))^-))\ dt \\
&\quad |X_2^1(0) = x_2, X_1^1(0) = x_1 - 1, X_2^2(0) = x_2, X_1^2(0) = x_1]
\end{aligned}$$

Since $X_2^1(t)$ and $X_2^2(t)$ are different only for $t \in [T_{x_1}, T_{x_1,x_2}]$, and since for $t \in [0, T_{x_1}]$ they assume the same value:

$$\begin{aligned}
g_2^2(x_1,x_2) &= E[\int_{T_{x_1}}^{T_{x_1,x_2}} (h_2(((S_2 - X_2^1(t))^+ - (S_2 - X_2^2(t))^+) \\
&+ b_2((S_2 - X_2^1(t))^- - (S_2 - X_2^2(t))^-))\ dt \\
&\quad |X_2^1(0) = x_2, X_1^1(0) = x_1 - 1, X_2^2(0) = x_2, X_1^2(0) = x_1] \\
&= E[\int_{T_{x_1}}^{T_{x_1,x_2}} (h_2(((S_2 - X_2^2(t) + 1)^+ - (S_2 - X_2^2(t))^+) \\
&+ b_2((S_2 - X_2^2(t) + 1)^- - (S_2 - X_2^2(t))^-))\ dt | X_2^2(0) = x_2, X_1^2(0) = x_1] \\
&= E[\int_{T_{x_1}}^{T_{x_1,x_2}} (h_2 I(S_2 - X_2^2(t) \geq 0) - b_2(1 - I(S_2 - X_2^2(t) \geq 0)))\ dt \\
&\quad |X_2^2(0) = x_2, X_1^2(0) = x_1] \\
&= (h_2 + b_2)E[\int_{T_{x_1}}^{T_{x_1,x_2}} I(S_1 - X_k^2(t) \geq 0)\ dt | X_2^2(0) = x_2, X_1^2(0) = x_1] \\
&\quad -b_2 E[\int_{T_{x_1}}^{T_{x_1,x_2}} 1\ dt | X_2^2(0) = x_2, X_1^2(0) = x_1] \\
&= -b_2 E[T_{x_1,x_2} - T_{x_1}] + (h_2 + b_2) e_2^2(x_1,x_2),
\end{aligned}$$

where $e_2^2(x_1,x_2)$ as the expected time process 2 visits the states with non-negative net inventory level for product 2 for $t \in [T_{x_1}, T_{x_1,x_2}]$ beginning at state (x_1,x_2). Observe that, $e_2^2(0,0) = 0$ since $T_{0,0} - T_0 = 0$. Furthermore $e_2^2(0,x_2) = e_2^1(0,x_2)$, where e_2^1 is as defined in the previous section. Finally, note

$$E[T_{x_1,x_2} - T_{x_1}] = \frac{x_1 + x_2}{\mu - \lambda_1 - \lambda_2} - \frac{x_1}{\mu - \lambda_1}.$$

In the following lemma, we determine $e_2^1(x_1,x_2)$.

Lemma C.3.1. *The difference function $g_2^2(x_1,x_2)$ defined for $0 < x_1, 0 \leq x_2$ is expressed as:*

$$g_2^2(x_1,x_2) = -b_2 E[T_{x_1,x_2} - T_{x_1}] + (h_2 + b_2) e_2^2(x_1,x_2).$$

The function $e_2^2(x_1, x_2)$ defined for $0 \le x_1,\ 0 \le x_2$ is expressed as follows:

1. *For $x_1 = 0$, $e_2^2(0, x_2) = e_2^1(0, x_2)$.*

2. *For $x_1 > 0$,*

$$e_2^2(x_1, x_2) = \begin{cases} C_0^{x_2} + \sum_{\ell=1}^{S_2-x_2+1} C_\ell^{x_2}(x_1)^{\ell-1}(\alpha)^{x_1}, & 0 \le x_2 \le S_2 \\ e_2^1(0, S_2), & x_2 > S_2 \end{cases} \quad \text{(C.3.1)}$$

where,

(i)

$$\begin{aligned} C_0^{x_2} &= e_2^1(0, S_2),\ x_2 = 0, 1, \ldots, S_2 \\ C_1^{x_2} &= -e_2^1(0, S_2) + e_2^1(0, x_2),\ x_2 = 0, 1, \ldots, S_2 \\ C_\ell^{x_2} &= \frac{1}{(\ell-1)(\mu - \lambda_1\alpha^2)}\Big(\sum_{m=\ell+1}^{S_2-x_2+1} C_m^{x_2}\binom{m-1}{\ell-2}(\lambda_1\alpha^2 \\ &\quad +\mu(-1)^{m-\ell+1}) + \lambda_2\alpha C_{\ell-1}^{x_2+1}\Big) \\ &\quad \ell = 2, 3, \ldots, S_2 - x_2 + 1,\ x_2 = 0, 1, 2, \ldots, S_2 - 1 \end{aligned}$$

(ii) $\alpha = \frac{(\lambda_1+\lambda_2+\mu)-\sqrt{(\lambda_1+\lambda_2+\mu)^2-4(\lambda_1)\mu}}{2\lambda_1}$.

Proof. We have already discussed $g_2^2(x_1, x_2)$ as a function of $e_2^2(x_1, x_2)$. In the proof we show how to derive the expression for $e_2^2(x_1, x_2)$.

1. For $x_1 = 0$, $e_2^2(0, x_2) = e_2^1(0, x_2)$ follows from the definition of the two functions. 2. For $x_1 > 0$, we do the proof in two parts. In the first part we show by induction that $e_2^2(x_1, x_2)$ is of the form Eq. C.3.1. In the second part, we propose a method to determine the coefficients $C_\ell^{x_2}$, $x_2 = 0, 1, 2, \ldots, S_2$ and $\ell = 0, 1, \ldots, S_2 - x_2 + 1$.

Part I

We derive $e_2^2(x_1, x_2)$ for all possible values of x_1 and x_2 to show that Eq. C.3.1 holds.

For $x_2 > S_2$. Consider the time until process 2 visits the state with zero shortfall for product 1 and S_2 shortfall for product 2. During this time, the process will never visit a state with non-negative net inventory level for product 2. This implies, for $x_2 > S_2$, $e_2^2(x_1, x_2) = e_2^1(0, S_2)$.

$0 \le x_2 \le S_2$. To show by induction that Eq. C.3.1 holds, we first need to show that the equation holds at the boundary. Let the boundary be defined by $x_2 = S_2$. So, we first prove for $x_2 = S_2$ that $e_2^2(x_1, x_2)$ is of the form Eq. C.3.1. Then, using induction we do the proof for $0 \le x_2 < S_2$.

$x_2 = S_2$. If we make the first step analysis for $e_2^2(x_1, S_2)$:

$$\begin{aligned} e_2^2(x_1, S_2) &= \frac{\lambda_1}{\lambda_1+\lambda_2+\mu} e_2^2(x_1+1, S_2) \\ &+ \frac{\lambda_2}{\lambda_1+\lambda_2+\mu} e_2^2(0, S_2) + \frac{\mu}{\lambda_1+\lambda_2+\mu} e_2^2(x_1-1, S_2) \quad \text{for } x_1 > 0, \end{aligned}$$

Rearranging the terms:

$$\begin{aligned} (\lambda_1+\lambda_2+\mu)e_2^2(x_1, S_2) - (\lambda_1)e_2^2(x_1+1, S_2) & \\ -\mu e_2^2(x_1-1, S_2) &= \lambda_2 e_2^1(0, S_2) \\ & \text{for } x_1 > 0. \end{aligned} \quad \text{(C.3.2)}$$

This is a second-order non-homogenous linear difference equation. The characteristic equation is:

$$-(\lambda_1 r^2 - (\lambda_1+\lambda_2+\mu)r + \mu) = 0.$$

To obtain a homogeneous difference equation, multiply the characteristic equation with $(r-1)$,

$$-(r-1)(\lambda_1 r^2 - (\lambda_1+\lambda_2+\mu)r + \mu) = 0.$$

The roots of the characteristic equation is $r_1 = 1$ and $r_{\{2,3\}} = \frac{(\lambda_1+\lambda_2+\mu)\pm\sqrt{(\lambda_1+\lambda_2+\mu)^2-4(\lambda_1)\mu}}{2\lambda_1}$ such that $r_2 < r_3$. Let $\alpha = r_2$, and $\alpha_+ = r_3$. Note that $\alpha_+ > 1$ and $0 < \alpha < 1$. So solution of the difference equation is:

$$e_2^2(x_1, S_2) = C_0^{S_2}(1)^{x_1} + C_1^{S_2}(\alpha)^{x_1} + C_2^{S_2}(\alpha_+)^{x_1} \quad \text{for } x_1 \ge 0. \quad \text{(C.3.3)}$$

Note $e_2^2(0, S_2)$ has the same equation form.

In the following, we show that $e_2^2(x_1, S_2)$ follows the equation form in Eq. C.3.1. This is equivalent to showing that $C_2^{S_2} = 0$. Afterwards, we show that $e_2^2(x_1, x_2)$, $0 \le x_2 < S_2$ also follows the equation form in Eq. C.3.1.

To show that $e_2^2(x_1, S_2)$ is of form Eq. C.3.1, we need to show $C_2^{S_2} = 0$. Observe that $\alpha_+ > 1$ which implies, if $C_2^{S_2} \neq 0$ then $e_2^1(x_1, S_2)$ must be diverging if we let x_1 go to ∞. However, note

$$\lim_{x_1 \to \infty} e_2^2(x_1, S_2) = e_2^1(0, S_2).$$

By definition, $e_2^1(0, S_2) \le E[T_{0,S_2}] = \frac{S_2}{\mu - \lambda_1 - \lambda_2}$. Hence the term on the right is less than ∞. Which shows, the limit does not diverge. This is possible only if $C_2^{S_2} = 0$. Therefore, for $x_2 = S_2$, $e_2^2(x_1, S_2)$ is of the form in Eq. C.3.1.

$\underline{0 \le x_2 < S_2.}$ Next, we show $e_2^2(x_1, x_2)$ is of the form Eq. C.3.1. We use induction to show this. Specifically, it is assumed that if $e_2^2(x_1, x_2 + 1)$ is of the form Eq. C.3.1, then $e_2^2(x_1, x_2)$ is also of the form Eq. C.3.1. Note that for the boundary value, $e_2^2(x_1, S_2)$ is already shown to be of the form Eq. C.3.1.

For $x_1 > 0$, following the first step analysis, $e_2^2(x_1, x_2)$ is expressed as:

$$\begin{aligned} e_2^2(x_1, x_2) &= \frac{\lambda_1}{\lambda_1 + \lambda_2 + \mu} e_2^2(x_1 + 1, x_2) + \frac{\mu}{\lambda_1 + \lambda_2 + \mu} e_2^2(x_1 - 1, x_2) \\ &\quad + \frac{\lambda_2}{\lambda_1 + \lambda_2 + \mu} e_2^2(x_1, x_2 + 1) \quad \text{for } 0 \le x_2 < S_2 \end{aligned} \tag{C.3.4}$$

Rearranging the terms in Eq. C.3.4,

$$\begin{aligned} (\lambda_1 + \lambda_2 + \mu) e_2^2(x_1, x_2) - (\lambda_1) e_2^2(x_1 + 1, x_2) - \mu e_2^2(x_1 - 1, x_2) = \\ \lambda_2 e_2^2(x_1, x_2 + 1) \quad \text{for } 0 \le x_2 < S_2. \end{aligned}$$

This is a second order non-homogeneous difference equation. Assume for $x_2 + 1$, Eq. C.3.1 holds. Then,

$$\begin{aligned} (\lambda_1 + \lambda_2 + \mu) e_2^2(x_1, x_2) - (\lambda_1) e_2^2(x_1 + 1, x_2) - \mu e_2^2(x_1 - 1, x_2) = \\ \lambda_2 C_0^{x_2+1} (1)^{x_1} + \lambda_2 \sum_{l=1}^{S_2 - x_2} C_\ell^{x_2+1} (x_1)^{\ell - 1} (\alpha)^{x_1} \quad \text{for } 0 \le x_2 < S_2 \end{aligned}$$

The characteristic equation of the second order non-homogeneous difference equation is

$$-(\lambda_1 r^2 - (\lambda_1 + \lambda_2 + \mu)r + \mu) = 0.$$

This characteristics equation has roots $r_1 = \alpha$, and $r_2 = \alpha_+$ (as determined in the previous characteristic equation). Converting the non-homogeneous equation into an homogeneous one, the characteristic equation becomes:

$$-(\lambda_1 r^2 - (\lambda_1 + \lambda_2 + \mu)r + \mu)(r - \alpha)^{S_2 - x_2}(r - 1) = 0$$

This equation has $S_2 - x_2 + 3$ roots. One of the roots is 1, the other root is α_+, and the remaining $S_2 - x_2 + 1$ of the roots has value α. So for $x_1 > 0$, $e_2^2(x_1, x_2)$ is of the form

$$\begin{aligned} e_2^2(x_1, x_2) \quad &= \quad C_0^{x_2}(1)^{x_1} + \sum_{\ell=1}^{S_2 - x_2 + 1} C_\ell^{x_2}(x_1)^{\ell-1}(\alpha)^{x_1} \\ &\quad + C_{S_2 - x_2 + 2}^{x_2}(\alpha_+)^{x_1} \quad \text{for } 0 \le x_2 < S_2 \end{aligned}$$

Note the same equation form also holds for $e_2^2(0, x_2) = 0$.

To complete the proof that $e_2^2(x_1, x_2)$ is of form Eq. C.3.1, we show that the constant $C_{S_2 + x_2 + 2}^{x_2}$ is equal to 0. To see why, consider $\lim_{x_1 \to \infty} e_2^2(x_1, x_2)$. Note that,

$$\lim_{x_1 \to \infty} e_2^2(x_1, x_2) = e_2^1(0, S_2), \quad 0 \le x_2 < S_2 \tag{C.3.5}$$

Since $0 \le e_2^1(0, S_2) \le E[T_{0,S_2}] = \frac{S_2}{\mu - \lambda_1 - \lambda_2}$, this limit exists. So $C_{S_2 - x_2 + 2}^{x_2}$ must be 0 since otherwise $e_2^2(x_1, x_2)$ diverges as $x_1 \Rightarrow \infty$.

This completes the induction-based proof. We conclude that for $x_1 \ge 0$, $0 \le x_2 \le S_2$, $e_2^2(x_1, x_2)$ is of the form in Eq. C.3.1.

Next step is to determine the coefficients $C_\ell^{x_2}$, $x_2 = 0, 1, 2, \ldots, S_2$ and $\ell = 0, 1, \ldots, S_2 - x_2 + 1$ (it is already shown that $C_{S_2 - x_2 + 2}^{x_2} = 0$). To determine these coefficients, a procedure is proposed. Before the procedure the following observations are made.

1. For calculation of $C_0^{x_2}$ and $C_1^{x_2}$, we look at the limit

$$\begin{aligned}\lim_{x_1\to\infty} e_2^2(x_1,x_2) &= \lim_{x_1\to\infty} C_0^{x_2} + \sum_{l=1}^{S_2-x_2+1} C_l^{x_2} x_1^{l-1}\alpha^{x_1}\\ &= C_0^{x_2}\\ &= e_2^1(0,S_2).\end{aligned}$$

The last equality holds from Eq. C.3.5. Since

$$\begin{aligned}e_2^2(0,x_2) &= e_2^1(0,x_2)\\ &= C_0^{x_2} + \sum_{l=1}^{S_2-x_2+1} C_l^{x_2} x_1^{l-1}\alpha^{x_1}|_{x_1=0}\\ &= C_0^{x_2} + C_1^{x_2},\end{aligned}$$

we deduce $C_1^{x_2} = e_2^1(0,x_2) - e_2^1(0,S_2),\ for\ 0 \le S_2$.

Given $e_2^1(x_1,x_2)$ are all determined, it is possible to obtain $C_0^{x_2}$ and $C_1^{x_2}$.

2. From the first step analysis

$$\begin{aligned}e_2^2(x_1,x_2) \;=\; & \frac{\lambda_1}{\lambda_1+\lambda_2+\mu} e_2^2(x_1+1,x_2)\\ &+\frac{\mu}{\lambda_1+\lambda_2+\mu} e_2^2(x_1-1,x_2)\\ &+\frac{\lambda_2}{\lambda_1+\lambda_2+\mu} e_2^2(x_1,x_2+1) \quad \text{for } 0<x_1, 0\le x_2<S_2,\end{aligned}$$

Given that $e_2^2(x_1,x_2)$ has the form in Eq. C.3.1, the equation will look like:

$$\begin{aligned}C_0^{x_2} + \sum_{l=1}^{S_2-x_2+1} C_l^{x_2} x_1^{l-1}\alpha^{x_1} \;=\; & \frac{\lambda_1}{\lambda_1+\lambda_2+\mu}(C_0^{x_2} + \sum_{l=1}^{S_2-x_2+1} C_l^{x_2}(x_1+1)^{l-1}\alpha^{x_1+1})\\ &+\frac{\mu}{\lambda_1+\lambda_2+\mu}(C_0^{x_2} + \sum_{l=1}^{S_2-x_2+1} C_l^{x_2}(x_1-1)^{l-1}\alpha^{x_1-1})\\ &+\frac{\lambda_2}{\lambda_1+\lambda_2+\mu}(C_0^{x_2+1} + \sum_{l=1}^{S_2-x_2} C_l^{x_2+1} x_1^{l-1}\alpha^{x_1})\\ &\text{for } 0<x_1, 0\le x_2<S_2\end{aligned}$$

To determine $C_\ell^{x_2},\ \ell = 2,3,\ldots,S_2-x_2+1,\ 0\le x_2\le S_2-1$, the coefficients of the terms $x_1^{\ell-1}\alpha^{x_1}$ on the RHS and LHS of the equation above are matched. Then one obtains:

$$
\begin{aligned}
C_{l-1}^{x_2}\alpha^{x_1} &= \frac{\lambda_1}{\lambda_1+\lambda_2+\mu}(\sum_{t=l-1}^{S_2-x_2+1} C_t^{x_2}\binom{t-1}{l-2}\alpha^{x_1+1}) \\
&+\frac{\mu}{\lambda_1+\lambda_2+\mu}(\sum_{t=l-1}^{S_2-x_2+1} C_t^{x_2}\binom{t-1}{l-2}(-1)^{t-l+1}\alpha^{x_1-1}) \\
&+\frac{\lambda_2}{\lambda_1+\lambda_2+\mu}(C_{l-1}^{x_2+1}\alpha^{x_1}), \\
&\quad \text{for } l \in \{2,3,...,S_2-x_2+1\}, 0 \le x_2 \le S_2-1 \quad \text{(C.3.6)}
\end{aligned}
$$

In Eq. C.3.6 look at the terms with coefficient $C_{\ell-1}^{x_2}$ on both sides:

$$
\begin{aligned}
C_{l-1}^{x_2}\alpha^{x_1} &= \frac{\lambda_1}{\lambda_1+\lambda_2+\mu}C_{\ell-1}^{x_2}\binom{\ell-2}{\ell-2}\alpha\alpha^{x_1}) \\
&+\frac{\mu}{\lambda_1+\lambda_2+\mu}C_{\ell-1}^{x_2}\binom{\ell-2}{\ell-2}(-1)^0\frac{1}{\alpha}\alpha^{x_1}) \qquad \text{(C.3.7)}
\end{aligned}
$$

Since it holds that $(\lambda_1+\lambda_2+\mu)\alpha = \lambda_1\alpha^2+\mu$, all terms with $C_\ell^{x_2}$ vanish. Only the terms $C_{\ell-1}^{x_2+1}$, $C_\ell^{x_2}$, $C_{\ell+1}^{x_2}, \ldots,$ $C_{S_2-x_2+1}^{x_2}$ remain. Taking the terms with $C_\ell^{x_2}$ to the LHS, one obtains:

$$
\begin{aligned}
C_l^{x_2} &= \frac{1}{(l-1)(\mu-\lambda_1\alpha^2)}(\sum_{m=l+1}^{S_2-x_2+1} C_m^{x_2}\binom{m-1}{l-2}(\lambda_1\alpha^2 \\
&+\mu(-1)^{m-l-1}) + \lambda_2 C_{l-1}^{x_2+1}\alpha) \\
&\le l \le S_2-x_2+1,\ 0 \le x_2 \le S_2-1. \qquad \text{(C.3.8)}
\end{aligned}
$$

Now we present the procedure to determine $C_\ell^{x_2},\ \ell = 0,1,\ldots,S_2-x_2+1,\ 0 \le x_2 \le S_2$. The coefficients $C_\ell^{x_2}$ with $\ell \ge S_2-x_2+2$ are already shown to be 0.

Procedure to determine $C_\ell^{x_2}$ in Eq. C.3.1

Step 1 For $i = 0,1,2\ldots,S_2$, $C_0^i = e_2^1(0,S_2)$, $\ C_1^i = e_2^1(0,i) - e_2^1(0,S_2)$.

Step 2 Set $i = S_2-1$.

Step 3 For $i = 0,1,2\ldots,S_2$, determine $C_2^i,\ C_3^i,\ldots,C_{S_2-i+1}^i$ using Eq. C.3.8. First determine $C_{S_2-i+1}^i$ in terms of $C_{S_2-i}^{i+1}$, then determine $C_{S_2-i}^i$ in terms of $C_{S_2-i+1}^i$ and $C_{S_2-i}^{i+1}$, and so on.

Step 4 Set $i \to i-1$. If $i = -1$ stop. All $C_l^i,\ i = 0,1,2,\ldots,S_2,\ l = 0,1,2,\ldots,S_2-i+1$ is determined. If $i \ge 0$, go to Step 3.

□

C.4 Determining $g_2^3(x_1, x_2)$

Next, we determine $g_2^3(x_1, x_2)$, $0 < x_1$, $0 \le x_2$ in Lemma C.4.1. To derive $g_2^3(x_1, x_2)$, we again define two processes as follows. At time $t = 0$, process 1 starts at state $(x_1 + 1, x_2 - 1)$ and process 2 starts at state (x_1, x_2). By definition of $g_2^3(x_1, x_2)$, the cost difference incurred only due to product 2 is under consideration. Note that under the two processes, the shortfall levels for product 2 are different only until the shortfall for product 1 reaches 0 under process 2. Thus, consider for process 2, the first passage time until state with 0 product 1 shortfall is reached. We have already defined this first passage time as T_{x_1}.

For a given x_1 and x_2, in the following we first express $g_2^3(x_1, x_2)$.

$$\begin{aligned}
g_2^3(x_1, x_2) &= E[\int_0^\infty (h_2(((S_2 - X_2^1(t))^+ - (S_2 - X_2^2(t))^+) \\
&+ b_2((S_2 - X_2^1(t))^- - (S_2 - X_2^2(t))^-)) \; dt \\
& \quad X_2^1(0) = x_2 + 1, X_1^1(0) = x_1 - 1, X_2^2(0) = x_2, X_1^2(0) = x_1]
\end{aligned}$$

$$\begin{aligned}
g_2^3(x_1, x_2) &= E[\int_0^{T_{x_1}} (h_2(((S_2 - X_2^1(t))^+ - (S_2 - X_2^2(t))^+) \\
&+ b_2((S_2 - X_2^1(t))^- - (S_2 - X_2^2(t))^-)) \; dt \\
& \quad |X_2^1(0) = x_2 + 1, X_1^1(0) = x_1 - 1, X_2^2(0) = x_2, X_1^2(0) = x_1] \\
&= E[\int_0^{T_{x_1}} (h_2(((S_2 - X_2^2(t) - 1)^+ - (S_2 - X_2^2(t))^+) \\
&+ b_2((S_2 - X_2^2(t) - 1)^- - (S_2 - X_2^2(t))^-)) \; dt | X_2^2(0) = x_2, X_1^2(0) = x_1] \\
&= E[\int_0^{T_{x_1}} (-h_2 I(S_2 - X_2^2(t) > 0) + b_2(1 - I(S_2 - X_2^2(t) > 0))) \; dt \\
& \quad |X_2^2(0) = x_2, X_1^2(0) = x_1] \\
&= -(h_2 + b_2) E[\int_0^{T_{x_1}} I(S_1 - X_k^2(t) > 0) \; dt | X_2^2(0) = x_2, X_1^2(0) = x_1] \\
& \quad + b_2 E[\int_0^{T_{x_1}} 1 \; dt | X_2^2(0) = x_2, X_1^2(0) = x_1] \\
&= b_2 E[T_{x_1}] - (h_2 + b_2) e_2^3(x_1, x_2),
\end{aligned}$$

where $e_2^3(x_1, x_2)$ is the expected total time process 2 spends in **positive** net inventory for product 2 until T_{x_1}, given initial state is (x_1, x_2). Note, as derived previously, $E[T_{x_1}] = \frac{x_1}{\mu - \lambda_1}$.

We introduce the following lemma.

Lemma C.4.1. *For $x_1 > 0, 0 \leq x_2$, the difference function $g_3^2(x_1, x_2)$ is expressed as:*

$$g_2^3(x_1, x_2) = b_2 \frac{x_1}{\mu - \lambda_1} - (h_2 + b_2) e_2^3(x_1, x_2),$$

The function $e_2^3(x_1, x_2)$ defined on $x_1 \geq 0$, $x_2 \geq 0$ is obtained as follows:

1. *For $x_1 = 0$, $e_2^3(0, x_2) = 0$.*

2. *For $x_1 > 0$,*

$$e_2^3(x_1, x_2) = \begin{cases} C_0^{x_2} + \sum_{l=1}^{S_2 - x_2} C_l^{x_2} x_1^{l-1} \alpha^{x_1}, & 0 \leq x_2 < S_2 \\ 0, & x_2 \geq S_2 \end{cases} \tag{C.4.1}$$

where,

(i)

$$\begin{aligned} C_0^{x_2} &= \frac{S_2 - x_2}{\lambda_2}, \; x_2 = 0, 1, \ldots, S_2 - 1 \\ C_1^{x_2} &= -\frac{S_2 - x_2}{\lambda_2}, \; x_2 = 0, 1, \ldots, S_2 - 1 \\ C_l^{x_2} &= \frac{1}{(l-1)(\mu - \lambda_1 \alpha^2)} \Big(\sum_{m=l+1}^{S_2 - x_2} C_m^{x_2} \binom{m-1}{l-2} (\lambda_1 \alpha^2 \\ &\quad + \mu(-1)^{m-l+1}) + \lambda_2 C_{l-1}^{x_2+1} \alpha \Big) \\ &\quad \ell = 2, 3 \ldots, S_2 - x_2, \; x_2 = 0, 1, \ldots, S_2 - 2 \end{aligned}$$

Proof. We have already discussed $g_2^3(x_1, x_2)$ as a function of $e_2^3(x_1, x_2)$. In the proof we show how to derive the expression for $e_2^3(x_1, x_2)$.

1. For $x_1 = 0$, $e_2^3(0, x_2) = 0$ since $T_{x_1=0} = 0$.

2. For $x_1 > 0$, we do the proof in two parts. In the first part we show by induction that $e_3^2(x_1, x_2)$ is of the form Eq. C.4.1. In the second part we propose a method to determine the coefficients $C_\ell^{x_2}$, $x_2 = 0, 1, 2, \ldots, S_2 - 1$ and $\ell = 0, 1, \ldots, S_2 - x_2$.

Part I

We derive $e_3^2(x_1, x_2)$ for all possible values of $x_1 > 0$ and $x_2 \geq 0$ to show that Eq. C.4.1 holds.

For $x_2 \geq S_2$. If the process starts with non-positive net inventory for product 2, the process will never reach positive inventory for product 2 until time T_{x_1}. This implies, for $x_2 \geq S_2$, $e_2^3(x_1, x_2) = 0$.

For $x_2 = S_2 - 1$. To show by induction that Eq. C.4.1 holds, we first need to show that the equation holds at the boundary. Let the boundary be defined by $x_2 = S_2 - 1$. So, we first prove for $x_2 = S_2 - 1$ that $e_2^3(x_1, x_2)$ is of the form Eq. C.4.1. Then, using induction we do the proof for $0 \leq x_2 \leq S_2 - 2$.

$x_2 = S_2 - 1$. If we make the first step analysis for $e_2^3(x_1, S_2 - 1)$:

$$\begin{aligned} e_2^3(x_1, S_2-1) &= \frac{1}{\lambda_1+\lambda_2+\mu} \\ &\quad + \frac{\lambda_1}{\lambda_1+\lambda_2+\mu} e_2^3(x_1+1, S_2-1) \\ &+ \frac{\lambda_2}{\lambda_1+\lambda_2+\mu}(0) + \frac{\mu}{\lambda_1+\lambda_2+\mu} e_2^3(x_1-1, S_2-1) \end{aligned}$$

Rearranging the terms:

$$(\lambda_1+\lambda_2+\mu)e_2^3(x_1, S_2-1) - (\lambda_1)e_2^3(x_1+1, S_2-1) - \mu e_2^3(x_1-1, S_2-1) = 1$$

This is a second-order non-homogenous linear difference equation. The characteristic equation is:

$$-(\lambda_1 r^2 - (\lambda_1+\lambda_2+\mu)r + \mu) = 0$$

To obtain a homogeneous recursive equation, we multiply the characteristic equation with $(r-1)$,

$$-(r-1)(\lambda_1 r^2 - (\lambda_1+\lambda_2+\mu)r + \mu) = 0.$$

The roots of the characteristic equation is $r_1 = 1$ and $r_{\{2,3\}} = \frac{(\lambda_1+\lambda_2+\mu) \pm \sqrt{(\lambda_1+\lambda_2+\mu)^2 - 4(\lambda_1)\mu}}{2\lambda_1}$ such that $r_2 < r_3$. Let $\alpha = r_2$, and $\alpha_+ = r_3$.

Note that $\alpha_+ > 1$ and $0 < \alpha < 1$. So solution of the difference equation is:

$$e_2^3(x_1, S_2 - 1) = C_0^{S_2-1}(1)^{x_1} + C_1^{S_2-1}(\alpha)^{x_1} + C_2^{S_2-1}(\alpha_+)^{x_1}$$

Note $e_2^3(0, S_2 - 1)$ has the same equation form.

In the following, we show that $e_2^3(x_1, S_2 - 1)$ follows the equation form in Eq. C.4.1. This is equivalent to showing that $C_2^{S_2-1} = 0$. Afterwards, we show that $e_2^3(x_1, x_2),\ 0 \leq x_2 \leq S_2 - 2$ also follows the equation form in Eq. C.4.1.

To show that $e_2^3(x_1, S_2 - 1)$ is of form Eq. C.4.1, we need to show $C_2^{S_2} = 0$. Observe that

$$\lim_{x_1 \to \infty} e_2^3(x_1, S_2 - 1) = \frac{1}{\lambda_2},$$

Since there is abundant product 1 shortfall, the expected time the process spends in positive inventory is equal to the expected time until a product 2 demand occurs. Since the limit is bounded, this implies $C_2^{S_2-1} = 0$. Therefore, $e_2^3(x_1, S_2 - 1)$ is of the form in Eq. C.4.1. □

APPENDIX D

NUMERICAL ANALYSIS FOR 3-PRODUCT SETTING

D.1 Analysis of the numerical results-3 product

D.1.1 Parameter Set 1-3 product

We claim, when holding costs and backorder costs are ordered, Priority-DH works better. In order to support this claim, the ordered holding and back-order tuples are tested. We test the problem on 3 product setting, each product has dedicated holding cost, back order cost and arrival rate. We call a setting defining a 3 product holding cost, backorder cost, and arrival rate tuple, a case. We test the problem for different cases. For each case, the holding cost of the first product is $h_1 = 0.25$, the other products holding costs are calculated by dividing previous product holding cost with a holding cost multiplier($h_{i+1} = h_i/h^*$). We test the following multipliers $h* = \{0,3;0,5;0.7;0.9;1\}$, a total of 5 multipliers. In each sample, for each product the holding backorder ratio is the same (h/b). We test the following ratios $h/b = \{10;20;50;80\}$, a total of 4 ratios. In this parameter set, the substitution rate is taken as $\tau = 5$ and substitution cost is taken as $c_s = 1$. Production rate is taken as $\mu = 1$. Finally, in terms of the arrival rates the products are taken as identical. The range of arrival rate values are $\lambda_i \in \{0.04, 0.07, 0.1, 0.13, 0.16, 0.18, 0.21, 0.22, 0.24, 0.25, 0.26, 0.27\}$, a total of 8 arrival rates. So we test the problem on 5*4*13=260 cases.

Performance of Prio-dh and LQ-STLA

We first compare the performance of Prio-DH with LQ-STLA in Table 6.17. The figures in the table show the difference in %:

$$\frac{Cost_{Prio-DH} - Cost_{LQ-STLA}}{Cost_{LQ-STLA}} 100\%$$

The priority heuristic is evaluated under the optimal priority ordering. When

Table D.1: Performance of Prio-DH over LQ-STLA-light traffic

λ	$b/h \setminus h*$	0.3	0.5	0.7	0.9	1
0.04	10	0.00%	0.00%	0.00%	0.00%	0.00%
	20	0.00%	0.00%	0.00%	0.00%	0.00%
	50	0.00%	0.29%	0.03%	0.02%	0.01%
	80	0.00%	0.00%	0.36%	0.03%	0.01%
0.07	10	-58.38%	-48.95%	-39.74%	-31.58%	-27.95%
	20	4.03%	0.01%	0.16%	0.02%	0.03%
	50	0.00%	0.00%	1.09%	0.16%	0.05%
	80	0.00%	0.01%	0.47%	0.23%	0.06%
0.1	10	1.09%	0.02%	0.32%	0.17%	0.08%
	20	2.81%	1.14%	0.80%	0.28%	0.09%
	50	-3.14%	6.10%	16.44%	21.27%	24.08%
	80	-31.15%	-23.61%	-17.29%	-12.54%	-10.51%
0.13	10	5.61%	0.04%	1.55%	0.50%	0.16%
	20	0.62%	5.02%	16.29%	11.71%	13.45%
	50	-27.30%	-20.27%	-13.74%	-8.93%	-6.75%
	80	8.34%	8.57%	10.07%	14.46%	15.77%
0.16	10	1.55%	9.56%	7.41%	10.01%	11.00%
	20	-24.03%	-17.90%	-9.95%	-10.52%	-6.94%
	50	12.21%	9.04%	9.34%	12.39%	14.47%
	80	5.12%	12.65%	18.76%	24.68%	24.79%
0.18	10	-2.98%	2.08%	4.69%	1.53%	1.05%
	20	-19.44%	-11.18%	-0.30%	7.92%	9.52%
	50	7.00%	8.81%	13.51%	16.74%	18.50%
	80	-10.74%	-2.69%	7.42%	12.95%	13.60%
0.21	10	-15.42%	-9.03%	1.27%	5.92%	6.81%
	20	-16.42%	-6.85%	2.61%	12.65%	13.34%
	50	-6.32%	1.73%	11.14%	17.21%	19.77%
	80	-19.68%	-9.33%	2.98%	12.39%	15.52%

holding multiplier is low, meaning the products differ to much, Prio-DH performs usually better than LQ-STLA. When multiplier is in 67% of the instances, Prio-DH perform at least as good as LQ-STLA

When the products are identical, (in the cases holding multiplier is 1), almost always LQ-STLA performs at least as good as Prio-DH. However there are instances that Prio-DH prevails. In these instances, arrival rates are low. We

Table D.2: Performance of Prio-DH over LQ-STLA-medium high traffic

λ	$b/h \setminus h*$	0.3	0.5	0.7	0.9	1
0.22	10	-14.69%	-5.86%	8.00%	13.73%	13.49%
	20	-8.41%	-6.44%	-0.41%	5.80%	8.50%
	50	-6.76%	3.06%	12.72%	18.52%	22.40%
	80	-17.53%	-5.99%	7.09%	18.19%	22.27%
0.24	10	-13.33%	-8.13%	4.91%	9.83%	9.64%
	20	-1.45%	1.51%	10.95%	18.19%	22.60%
	50	-15.10%	-4.20%	9.86%	18.44%	22.08%
	80	-10.65%	-3.34%	8.21%	16.01%	19.43%
0.25	10	-23.88%	-14.02%	0.07%	8.82%	11.13%
	20	-13.94%	-9.23%	2.62%	10.53%	15.39%
	50	-23.26%	-11.49%	3.74%	14.17%	19.52%
	80	-6.67%	0.54%	11.86%	21.87%	26.69%
0.26	10	-21.60%	-10.44%	3.45%	9.93%	10.83%
	20	-11.04%	-6.53%	5.30%	12.24%	15.05%
	50	-21.56%	-7.39%	7.11%	14.69%	19.02%
	80	-13.27%	-4.63%	9.87%	21.09%	25.19%
0.27	10	-28.99%	-14.78%	2.66%	12.74%	13.53%
	20	-20.23%	-10.45%	4.90%	12.32%	14.82%
	50	-12.67%	-3.54%	10.28%	19.02%	22.93%
	80	-19.69%	-7.34%	8.29%	18.27%	21.77%

need to keep a little inventory. LQ-STLA as it is symmetric, try to assign either 0 inventory to all products or 1 inventory, but Prio-DH manages to assign 1 inventory, to a few products, thus benefit from both inventory, while enjoying low inventory costs.

There are certain cases which Prio-DH works extremely well over LQ-STLA in low traffic. Usually in these sets the Prio-DH cost is very close to our LB approximation. The reason is its working principle in these cases is similar to LB approximation. It holds inventory in lowest holding cost product, and when there is backorder for other products from the lowest holding cost products inventory by substitution. Since substitution rate is high, it requires very little time, which makes Prio-DH perform close to lower bound.

In the cases holding cost multiplier is low, thus promoting the Prio-DH, Prio-DH performs better in light traffic and high traffic compared to medium traffic.

In very light traffic, in most cases, Prio-DH and LQ-STLA do not differ. In these instances for both heuristics, there is no need to hold inventory, and no inventory means no-substitution decisions as well. Hence they perform closely. However as we increase the traffic a little bit, but still can be considered as light traffic, Prio-DH performs better than LQ-STLA. The reason is in these instances, the inventory need is very scarce, but there is need for inventory, and Prio-DH assigns this scarce inventory to appropriate product, and manages the substitution decisions better. In very low traffic, production scheduling other than base-stock decision is not very important, because, usually there is no shortfall of inventory from base-stock, and when there is shortfall, it is located in one product usually, which is quickly filled. So base-stock decisions are important.

In high traffic for low holding multiplier cases, Prio-Dh performs better. If we ignore the substitution decision, from Wein's work, we know that in heavy traffic, the inventory required is held on lowest holding cost product, and the products that are backordered are processed in backorder order. This defines an inherent priority order for the cases we discuss. So we expect priority based heuristics, performs well on these cases, which our numerical results confirm.

It is hard to interpret how the holding cost backorder cost ratio affect the performance. When h/b ratio is high, the back order of the products are more apparent, which promotes prioritization better. Although there is this effect, we have to consider another as well. Consider the time, low priority product is backordered, while the others has inventory but they are still in shortfall from base-stock. When this ratio is very high, the backorder of the low priority product will be so high that optimal policy will try to supply this product, although we consider it low priority product which will interfere with the performance of the Prio-DH policy. As a result of these mixed effects, it is hard to say something about the performance conjecture based on holding cost backorder cost ratio.

About the holding- backorder cost ratio, when we consider single product problem, this ratio simply defines the optimal base-stock level for the product, in multi-product problem it gives us an idea about the inventory levels. When we look at Table D.1, we observe that, When a Prio-DH performs extremely well it corresponds to the same holding cost backorder cost ratio for a given arrival rate. This behavior is realized at certain base-stock levels which we explained earlier, this explains this phenomena.

The value and the impact of substitution Next, the value of substitution under LQ-STLA and Prio-DH is analyzed. By doing so: (1) The settings under which substitution is beneficial are identified, (2) The impact of scheduling decisions and substitution decisions on the performance of the heuristics are differentiated. The figures in Table D.3 and Table D.4 show:

$$\text{Value of Subs under LQ-STLA} = \frac{Cost_{LQ-STLA-nosubs} - Cost_{LQ-STLA}}{Cost_{LQ-STLA-nosubs}} 100\%$$

and

$$\text{Value of Subs under Prio-DH} = \frac{Cost_{Prio-DH-nosubs} - Cost_{Prio-DH}}{Cost_{Prio-DH-nosubs}} 100\%,$$

respectively. Prio-DH is evaluated under Prio-DH.

Table D.3: Value of subs for LQ-STLA

λ	$b/h \setminus h*$	0.3	0.5	0.7	0.9	1
0.04	10	0.00%	0.00%	0.00%	0.00%	0.00%
	20	0.00%	0.00%	0.00%	0.00%	0.00%
	50	4.63%	7.10%	7.09%	7.08%	7.07%
	80	7.37%	8.23%	10.96%	10.98%	10.99%
0.07	10	0.00%	0.00%	0.00%	0.00%	0.00%
	20	8.52%	8.46%	8.35%	8.21%	8.14%
	50	13.63%	14.71%	18.82%	18.90%	18.93%
	80	22.41%	20.28%	23.13%	26.74%	26.81%
0.1	10	8.43%	8.14%	7.75%	7.44%	7.15%
	20	10.31%	15.17%	15.09%	14.95%	14.87%
	50	24.23%	25.06%	29.10%	29.29%	29.34%
	80	5.92%	5.13%	4.57%	4.09%	3.82%
0.13	10	12.34%	12.01%	11.53%	11.01%	10.76%
	20	15.74%	15.94%	20.23%	20.09%	20.01%
	50	8.82%	7.44%	6.66%	5.99%	5.56%
	80	10.10%	9.84%	9.91%	10.18%	10.13%

Continued on next page

Table D.3 – continued from previous page

λ	$b/h \setminus h*$	0.3	0.5	0.7	0.9	1
0.16	10	8.95%	14.34%	13.77%	13.17%	12.89%
	20	8.77%	5.62%	5.55%	6.44%	6.29%
	50	16.42%	14.10%	13.09%	12.26%	11.70%
	80	14.17%	17.20%	17.12%	18.34%	18.41%
0.18	10	10.85%	14.41%	13.75%	13.12%	12.83%
	20	11.27%	7.87%	7.83%	9.05%	8.84%
	50	15.70%	14.52%	14.83%	15.61%	15.60%
	80	7.42%	5.76%	6.47%	6.53%	6.45%
0.21	10	9.65%	6.55%	7.31%	6.67%	6.38%
	20	14.47%	11.13%	10.56%	11.83%	11.51%
	50	8.70%	8.16%	7.92%	7.99%	7.90%
	80	7.38%	4.33%	3.95%	3.72%	3.67%
0.22	10	10.44%	7.21%	7.83%	7.12%	6.80%
	20	11.03%	5.48%	4.28%	4.56%	4.72%
	50	10.17%	9.43%	9.17%	9.25%	9.14%
	80	8.84%	5.71%	4.89%	5.03%	4.97%
0.24	10	11.00%	7.72%	7.93%	7.06%	6.70%
	20	12.07%	7.44%	5.86%	5.65%	6.10%
	50	8.92%	5.84%	5.55%	5.16%	5.09%
	80	8.80%	3.88%	3.16%	3.46%	3.12%
0.25	10	10.35%	5.63%	4.80%	3.81%	3.62%
	20	10.36%	3.90%	3.46%	2.82%	2.79%
	50	8.11%	4.22%	3.57%	2.98%	2.71%
	80	10.47%	5.18%	4.18%	4.32%	4.15%
0.26	10	10.91%	6.08%	5.02%	3.93%	3.71%
	20	11.59%	4.64%	4.08%	3.33%	3.27%
	50	9.28%	5.38%	4.40%	3.76%	3.44%
	80	8.71%	4.19%	3.02%	2.85%	2.57%

Continued on next page

Table D.3 – continued from previous page

λ	$b/h \setminus h*$	0.3	0.5	0.7	0.9	1
0.27	10	10.69%	4.93%	3.58%	2.56%	2.13%
	20	8.60%	3.61%	3.09%	2.12%	1.97%
	50	9.16%	4.77%	3.52%	2.79%	2.26%
	80	8.27%	2.92%	2.56%	2.19%	1.83%

Table D.4: Value of subs for Prio-DH

λ	$b/h \setminus h*$	0.3	0.5	0.7	0.9	1
0.04	10	0.00%	0.00%	0.00%	0.00%	0.00%
	20	0.00%	0.00%	0.00%	0.00%	0.00%
	50	4.67%	6.94%	7.48%	7.56%	7.35%
	80	7.43%	8.39%	11.23%	11.68%	11.40%
0.07	10	58.32%	48.81%	39.48%	31.20%	27.51%
	20	4.92%	8.67%	9.05%	9.31%	8.74%
	50	13.72%	15.08%	19.46%	20.57%	20.02%
	80	22.51%	20.76%	24.74%	28.78%	28.17%
0.1	10	7.57%	8.44%	8.89%	8.92%	8.04%
	20	7.90%	14.64%	16.39%	17.29%	16.25%
	50	26.38%	21.11%	19.41%	18.38%	16.50%
	80	31.24%	24.49%	19.46%	14.83%	12.82%
0.13	10	7.57%	12.47%	12.78%	12.85%	12.35%
	20	15.40%	13.64%	11.81%	17.21%	16.02%
	50	31.77%	25.47%	20.92%	15.89%	13.38%
	80	4.34%	6.71%	8.49%	6.77%	6.44%
0.16	10	7.90%	8.54%	13.02%	12.07%	10.89%
	20	20.10%	14.79%	11.05%	13.30%	10.53%
	50	5.73%	8.89%	10.38%	8.25%	6.93%
	80	9.74%	9.87%	8.73%	7.39%	8.36%

Continued on next page

Table D.4 – continued from previous page

λ	$b/h \setminus h*$	0.3	0.5	0.7	0.9	1
0.18	10	11.44%	11.29%	11.46%	13.67%	12.45%
	20	24.21%	17.57%	12.87%	9.07%	8.90%
	50	8.83%	9.06%	8.47%	8.33%	7.30%
	80	11.50%	9.25%	6.68%	5.95%	6.99%
0.21	10	14.72%	10.88%	9.13%	7.52%	7.98%
	20	29.31%	20.44%	17.08%	11.94%	11.71%
	50	9.47%	9.43%	7.47%	6.08%	6.40%
	80	16.60%	12.65%	8.45%	5.76%	5.37%
0.22	10	17.21%	11.90%	7.85%	6.37%	7.57%
	20	5.19%	6.38%	6.68%	5.75%	4.64%
	50	11.74%	9.21%	7.66%	6.28%	5.32%
	80	19.07%	14.30%	10.09%	6.64%	5.40%
0.24	10	18.38%	14.20%	9.60%	7.13%	7.08%
	20	5.14%	6.32%	5.14%	3.80%	1.97%
	50	14.95%	11.55%	7.46%	5.43%	4.57%
	80	4.09%	3.26%	4.17%	4.30%	3.75%
0.25	10	19.49%	13.63%	10.61%	8.14%	6.62%
	20	4.49%	6.20%	5.16%	3.78%	2.78%
	50	15.63%	12.91%	8.47%	5.80%	5.28%
	80	4.83%	4.37%	4.88%	3.47%	2.62%
0.26	10	21.14%	13.14%	10.02%	8.74%	7.22%
	20	5.14%	6.20%	5.45%	3.46%	3.76%
	50	17.81%	12.13%	7.92%	7.16%	6.20%
	80	3.92%	4.78%	3.97%	2.50%	2.31%
0.27	10	21.54%	13.20%	7.95%	5.25%	4.83%
	20	6.66%	6.20%	3.72%	2.37%	3.57%
	50	3.36%	3.23%	3.45%	2.23%	1.96%
	80	5.72%	3.58%	3.13%	2.22%	2.63%

For LQ-STLA and Prio-DH, when arrival rate is very low the average earnings due to substitutions is higher. The thing that is realized is: When the arrival rate is very low, there is no base-stock, which leads no substitutions and also result in value of substitutions being not high. If we increase the arrival rates a little bit, but still in low traffic region, there is very high earnings due to substitutions. If we increase the arrival rates further, the value of substitution decreases. When traffic intensity is higher, although the rate of substitution increases, scheduling becomes more important than the substitution decision. Therefore relatively, the contribution to cost savings is lower. In light traffic, substitution generates higher % savings because when the costs are low, the amount of saving that substitution generates is relatively large.

For high traffic the vale of substitutions tend to get smaller. In very high traffic, the ratio of the customers that are served from their own products to the customers who are served by substitutions decreases. However for high traffic, the cost difference created by substitutions is much lower. The observation is substitutions do better when the inventory is scarce, so it should be required to be used efficiently.

One might expect for high backorder holding cost ratio because, high backordered product is quickly supplied in the case of substitution, so that the savings due to substitution are always higher. On the contrary to intuition, we are not observing this. The reason can be, for high ratio systems, base stock levels get very high, so that the probability of getting backordered, thus being served by substitution drops. We observe, when the arrival rates are low, the savings due to substitutions increases as the products get similar, as the holding cost multiplier increases. However as the arrival rate increases, the savings due to substitution decreases with the holding cost multiplier. When the arrival rates are high, ability to satisfy the high back order cost product from low holding cost inventory by substitutions still causes the system still be less costly although apparent effect of arrival rate and the occurrence of substitution cost. But this effect is affected by substitution cost, if long term keeping of low cost inventory plus the substitution costs is higher than keeping high holding cost inventory, it will not be observed.

For Prio-DH, in the previous- subsection we observed there are instances, where Prio-DH performs extremely well over LQ-STLA. In such instances, value of substitutions is very high also. We had explained substitutions role on the mimicking the lower bound system. Since substitution have such effect on certain cases, savings due to substitutions are very high in these cases.

It is observed that the type of substitution policy affects the relative performance of Prio-DH and LQ-STLA. To understand the impact of the additional substitution decisions on the performance, we compare Prio-DH and LQ-STLA under (i) no-substitution, (ii) one-way substitution (where only high priority product can substitute a low product customer's need) and (iii) two-way substitution.

We calculate for Prio-DH vs. LQ-STLA: the measure $\frac{Cost_{Prio-DH} - Cost_{LQ-STLA}}{Cost_{LQ-STLA}} 100\%$ under no-substitution, one-way and two-way substitution, respectively for the 160 cases. This measure is compared for different substitution policies in Table D.4.

As Prio-DH utilizes the substitutions better than LQ-STLA, two way substitution results are usually better for Prio-DH over LQ-STLA comparison than the no-substitution results. However the savings due to substitutions from higher priority products, and from lower priority products change a lot.Our one-way substitution policy only takes the substitutions from higher priority products into consideration. Hence, when we look at oneway substitution and no substitution comparioson, in high traffic, this type of substitution helps less to the Prio-DH than LQ-STLA. However when arrival rates are low, this type of substitutions takes importance.

Table D.5: PrioDH Performance over LQ-STLA

Nosubs<twoway	33.33%
Nosubs<oneway	74.17%
one way<twoway	26.67%

In 33.33% of the cases performance of Prio-DH over LQ-STLA decreases, in the rest increases when two-way substitution policy is employed instead of no substitution policy. This tells us, Prio-DH benefits from two-way substitutions more than LQ-STLA compared to no substitution case hence prio-DH utilizes

two way substitutions better.

In 74.17% of the cases performance of Prio-DH over LQ-STLA decreases when one way substitution is allowed instead of no substitution. Hence allowing one way substitution benefits LQ-STLA more than Prio-DH. Prio-DH utilizes low product inventory for satisfying high priority products, when two way substitution is allowed. However when this is not allowed, Prio-DH performance deteriorates.

D.1.2 Parameter Set 2-3 product

The performances of the heuristics are compared under certain settings. In § 6.4.2.1 and § 6.4.2.2 it is observed that under medium traffic LQ-STLA is likely to outperform Prio-DH, whereas under light traffic Prio-DH is likely to give lower cost. When traffic intensity is relatively higher, depending on the parameters either LQ-STLA or Prio-DH might outperform the other. If the parameters favor a natural priority ordering, then Prio-DH is likely to result in lower cost. In this section several more parameter settings are analyzed, to identify the conditions under which each heuristic performs well. In this part, only the results on comparison of the heuristics are presented, discussion on value of substitution is omitted.

1. **Group 1: Identical h, varying b** The first group of parameters consist of the following setting. Unit holding cost are identical among the products and takes values $h \in \{1, 3, 5, 10\}$. Backordering cost has an increasing order, $b_1 \in \{5, 6, 7\}$, $b_2 = 2.5b_1$, and $b_3 = 4b_1$. Arrival rate is identical for all products, and set to $\lambda_i = 0.27$. A total of 12 instances are studied. Priority order is taken as $3 \prec 2 \prec 1$.

   It is observed (in Table D.6) that, under ordered backorder costs, when h (or $\frac{h}{b}$) is low or high, LQ-STLA performs better than Prio-DH.

   If $\frac{h}{b}$ is low, then the stock levels are high. Prio-DH have a tendency to increase the gap between the stock levels of the products if $\frac{h}{b}$ ratio is low. On the other hand, LQ-STLA results in a more balanced stock level.

Table D.6: The comparison of the heuristics for Set 2 Group 1 (in percentages)

Instance	h	b_1	Prio-DH	LQ-SSTLA	SICOST	SILEVEL	UB	LB
1	1	5	2.8%	-0.2%	1.3%	1.8%	28.2%	-7.0%
2		6	5.6%	-0.2%	1.4%	2.1%	30.7%	-4.8%
3		7	3.7%	-0.2%	1.6%	2.4%	28.9%	-5.4%
4	3	5	3.1%	-0.7%	-0.1%	0.2%	41.0%	-3.7%
5		6	2.2%	-0.6%	0.3%	0.7%	38.5%	-4.7%
6		7	-2.6%	-0.2%	0.7%	1.1%	30.2%	-8.7%
7	5	5	0.8%	-0.4%	0.2%	0.6%	43.3%	-4.8%
8		6	-0.6%	-0.6%	-0.1%	0.3%	37.4%	-6.7%
9		7	3.4%	-0.5%	0.0%	0.4%	39.7%	-4.9%
10	10	5	6.8%	-0.1%	0.0%	0.2%	54.5%	-1.3%
11		6	1.1%	-0.1%	0.0%	0.3%	51.3%	-1.5%
12		7	3.1%	-0.1%	0.1%	0.3%	51.7%	-1.4%

Lowest priority product does not necessarily have the highest basestock in contrast to Prio-DH approach, since low backorder cost leads to a low basestock level.

When $\frac{h}{b}$ is high, stock levels are low (either 0 or 1). In that case, LQ-STLA keeps stock for high unit backorder cost products and no stock for low backorder products. Keeping stock helps avoid falling to backorder and facilitates substitution. Prio-DH does just the opposite: increases basestock level for low backorder cost products, and keeps no stock for high priority products. Under high holding cost, keeping unnecessary stock results in high cost. The stock levels are more balanced and are higher under LQ-STLA, and it is thus a more preferable approach.

Finally, under this setting the other heuristics (LQ-SSTLA, SICOST, SILEVEL) perform either very close, or outperform LQ-STLA, especially under moderate $\frac{h}{b}$ ratio. All heuristics, except Prio-DH, perform close to LB.

2. **Group 2: Identical h, varying b, changing λ.** In this group, $h = 3$, $b_1 = 5$, $b_2 = 12.5$, $b_3 = 20$. Arrival rates are identical and take the following values, $\lambda_i = \{0.15, 0.18, 0.21, 0.24, 0.27\}$.

When there is very low traffic, the system does not need to hold inventory for all products, instead, it can quickly feed the backorders with substitu-

Table D.7: The comparison of the heuristics for Set 2 Group 2 (in percentages)

Instance	Arrival rate	Prio-DH	LQ-SSTLA	SICOST	SILEVEL	UB	LB
1	0.15	-3.3	0.0	0.0	0.5	30.5	-9.5
2	0.18	2.4	-1.0	-0.6	0.0	34.0	-5.7
3	0.21	1.4	-0.6	-0.3	0.1	38.1	-5.2
4	0.24	4.7	-0.9	-0.2	0.3	38.0	-3.8
5	0.27	3.1	-0.7	-0.1	0.2	41.0	-3.7

tions, and this can be explained by priority structure. Hence in very low traffic, Prio-DH works better.

When the arrival rate is medium (0.18-0.27), the server can supply the shortfall of the product from designated level, without falling into backorder. This results in a decreased requirement to Prio-DH, as every request to server is served immediately. Since the basestock levels under LQ-STLA are adequately calculated considering the low queue length, all capacity can be allocated to the shortfall. This results in better performance for LQ-STLA (see Table D.7)

All other heuristics perform close to LQ-STLA, under all arrival rates. Under high traffic, the substitution rates of the heuristics are close. For this instance, we can observe that heuristics perform closer to our LB, hence we can say that, they work well under this circumstance. SICOST outperform LQ-STLA (in Table D.7). This implies the optimal scheduling policy is as defined in this heuristic, especially when arrival rate is low.

3. **Group 3: Dispersed** b**.** In this setting it is assumed that $h = 3$, $\lambda_i = 0.27$. We take $b_1 = 12.5 - X$, $b_2 = 12.5 + X$, $b_3 = 12.5 + 3X$ (keeping holding cost and arrival rate constant) in which the X factor defines the dispersion of backorders. The dispersion factor is selected from $X \in \{2, 2.25, 2.5, 2.75, 3\}$. A total of 5 instances are analyzed. By this way, the priority among the items is made stronger.

   When the backorder dispersion is low, LQ-STLA works better than Prio-DH. In fact when there is no dispersion, when backorders are also equal, we

Table D.8: The comparison of the heuristics for Set 2 Group 3 (in percentages)

Instance	Dispersion (X)	Prio-DH	LQ-SSTLA	SICOST	SILEVEL	UB	LB
1	2	5.14	-0.29	0.51	0.88	27.00	-3.25
2	2.25	5.30	-0.29	0.56	0.95	28.36	-3.47
3	2.5	0.04	-0.38	0.52	0.93	29.64	-3.98
4	2.75	-0.20	-0.38	0.57	1.00	31.05	-5.16
5	3	1.54	-0.58	0.15	0.59	36.27	-3.70

expect that the optimal policy is longest queue type. When the dispersion is low, the policy is closer to the longest queue policy, hence, base-stock in LQ-STLA is calculated better. This results in better performance of LQ-STLA.

When the backorder dispersion is high, the server gives priority to higher backorder products, in order to decrease the cost. Prio-DH is already considering this, while LQ-STLA calculates the basestocks according to equal priorities. This results in closer performance of Prio-DH over LQ-STLA.

For this system, when the system is in back-order there is an inherent priority order for processing items, but also when the system is holding inventory, all the inventory costs are the same, for this reason, it tries to allocate the scheduling decisions as a longest queue. So there is a mixed effect, but when we look at base-stocks of the systems, Prio-DH allocates less inventory to high- backorder product because it has higher priority, but this should not be the case, the higher backorder product, should get higher base-stock as well, when other parameters are the same. For this reason, LQ-STLA performs better than Prio-DH.

Another interesting observation is, for both LQ-STLA and Prio-DH, as the dispersion of backorders increases, the cost of the system decreases. This is explained by high backorder product is immediately processed, its backorder magnitude is not that important, though, the backorder costs of the low-priority products are important, because they wait for processing, and when the dispersion is high, the waiting backorders have relatively less cost.

Also another observation is, for Prio-DH (and LQ-STLA), as dispersion increases, the base-stock levels decrease. So with lower inventory, the production scheduling handles the cost good.

4. **Group 4: Identical b, varying h.** In this setting it is assumed that $h_1 = 5$, $h_2 = 2.5$, $h_3 = 1$, $b \in \{15, 18, 21\}$, and $\lambda_i = \{0.25, 0.27\}$. The priority among the items is determined by the unit holding cost only.

Table D.9: The comparison of the heuristics for Set 2 Group 4 (in percentages)

Instance	b	λ	Prio-DH	LQ-SSTLA	SICOST	SILEVEL	UB	LB
1	15	0.25	-3.58	-2.28	6.73	6.78	35.87	-37.50
2	18		-3.12	-2.91	11.12	11.04	26.79	-42.48
3	21		-3.21	-5.18	9.27	9.19	29.94	-41.54
4	15	0.27	-7.50	-3.40	11.24	11.15	25.40	-38.58
5	18		-6.89	-4.06	10.74	10.66	31.88	-37.48
6	21		-10.11	-4.61	8.80	8.77	30.32	-39.50

When the products are ordered with respect to their holding cost, it is observed that Prio-DH outperforms LQ-STLA (see Table D.9). The performance of Prio-DH slightly increases as the traffic intensity increases.

In this case, it is also observed that LQ-SSTLA performs close to LQ-STLA. LQ-SSTLA is slightly better, whereas cost under SILEVEL and SICOST are much higher compared to LQ-STLA. The substitution rate under SICOST and SILEVEL is much lower compared to LQ-STLA.

Furthermore, all policies perform far away from the LB, and relatively closer to the UB. This may support the conclusion that the optimal policy does not necessarily pool the stock when holding cost are different, but rather the allocation of the capacity is done effectively at the onset.

In LQ-STLA basestock levels among the products differ only by 1, whereas in Prio-DH the difference between the basestocks is much higher. Substitution rates are close for the both policies. Increased basestock levels for the low holding cost products under Prio-DH helps to make the capacity allocation more effectively.

5. **Group 5: Clustered products.** In this setting the products are clustered with respect to $h-b$ values. It is assumed that a product is either in Cluster 1: $h = 5,\ b = 20$ or in Cluster 2: $h = 1,\ b = 10$. Cluster 2 is assigned the lower priority in Prio-DH.

   In Table D.10 arrival rates are set to $\lambda_i = \{0.10,\ 0.13,\ 0.16,\ 0.18,\ 0.21,\ 0.24,\ 0.27\}$. In the table, to denote an instance, the number of products in cluster 1 is denoted with Nb-C1, and the number of products in Cluster 2 is denoted with Nb-C2.

Table D.10: The comparison of the heuristics for Set 2 Group 5a when $\lambda_i = 0.2$ (in percentages)

(Nb-C1, Nb-C2)	λ	Prio-DH	LQ-SSTLA	SICOST	SILEVEL	UB	LB
(1,2)	0.1	0.94	0.00	0.01	4.35	103.39	-14.14
	0.13	2.12	0.01	0.02	3.89	114.93	-14.24
	0.16	-31.78	0.00	12.90	13.00	38.02	-42.94
	0.18	-24.28	0.01	12.36	12.48	43.42	-37.51
	0.21	7.77	-9.69	3.91	4.27	41.91	-34.84
	0.24	4.67	-6.77	3.62	3.91	48.78	-31.30
	0.27	2.57	-8.72	10.97	10.80	48.21	-27.50
(2,1)	0.1	-0.01	0.00	0.01	0.01	101.10	-34.02
	0.13	0.00	0.01	0.01	0.01	87.05	-42.98
	0.16	-23.77	0.00	6.94	7.52	28.46	-60.28
	0.18	-16.76	0.00	6.96	7.56	34.38	-55.89
	0.21	11.64	-5.53	3.25	3.94	43.52	-49.97
	0.24	-3.88	-3.71	2.55	3.07	47.23	-49.30
	0.27	-13.30	-4.99	8.50	8.90	50.54	-46.08

It is observed from Table 6.15 that for both cluster groups, when the arrival rate is a little bit above the light traffic, Prio-DH works exceptionally well, this can be explained by Prio-DH's ability to mimick the LB policy. On the other hand, when the number of products in Cluster 1 is higher the relative performance of Prio-DH is good for higher traffic. Since low basestock is assigned to priority products under Prio-DH, when the number of products in Cluster 1 is high, the basestock levels are low and evenly distributed, whereas the basestock levels of the products in Cluster 2 are much higher. Prio-DH does not assign the same basestock to two identical products due to its nature. LQ-STLA indeed assigns the same basestock level. However, in LQ-STLA the basestock levels differ very little among the products in Cluster 1 and Cluster 2, since only the ratio of $\frac{h}{b}$ is taken into consideration and not the actual values of h and b. Thus, Prio-DH does a better job in terms of assigning the basestock levels by inflating the basestock level of the product with low holding cost. As a result Prio-DH performs better when the number of low priority products decrease, in high traffic.

6. **Group 6: Identical products, dispersed arrival rates**. The final group of instances aims to measure the joint impact of the traffic intensity and the dispersion in arrival rates and among the products.

   It is assumed that $\lambda \in \{0.06, 0.07, 0.08, 0.085, 0.09, 0.1\}$, and $\lambda_1 = \lambda,\ \lambda_2 = 2.5\lambda$, $\lambda_3 = 5\lambda$. In the setting, $(h, b) = (1, 20)$ and $(5, 20)$. We expect when there is no dispersion, the LQ-STLA will prevail because of the optimal policy's longest queue structure. However, in the experiment set, we see that there is no simplistic relationship between arrival rate dispersion and performance differential of heuristics.

   When the $\frac{h}{b}$ ratio is low, the increase in the arrival rate dispersion affects the performance mixedly. The performance of LQ-STLA may increase or decrease with λ, but it outperforms Prio-DH by 10%. When the $\frac{h}{b}$ ratio is high, with dispersion increase, the performance of Prio-DH gets better, but there are still contradicting results. In this case LQ-STLA outperforms by 0.2%.

APPENDIX E

CONSUMER DRIVEN SUBSTITUTION: ADDITIONAL MODELS

Price set only once, after the demand is realized

In this scenario, we assume stock levels are set at the beginning of the period before the demand is realized, while price decisions are set after the demand is realized. Such pricing schemes are called as "responsive pricing" in the literature (Bish and Suwandechochai, 2010; Van Mieghem and Dada, 1999). There does not exist another chance for the manufacturer to change the prices. Thus, again the only substitution type is price-based substitution. Manufacturer's objective is to maximize profit. The sequence of events are as follows:

1. At the beginning, the manufacturer decides on how many units to order for each product. A cost c is incurred for each unit stocked.
2. Demand is realized.
3. Prices are set and profit is obtained.

This scenario is defined to contrast the benefit of responsive pricing with the benefit of markdown pricing and stockout-based substitution.

Price set only once, after the demand is realized

Table E.1: Notation for single period two-product model where price is set after demand realization

A	Deterministic component of the base demand which is identical for both products, $A_1 = A_2 = A$
I	Initial stock level, which is identical for both products, a decision variable. $I_1 = I_2 = I$.
p	Uniform price (set for both products)

This problem is not tractable under general assumptions (such as n products, non-identical A_i, etc.) We assume the manufacturer is selling two products, indexed by $i = \{1, 2\}$. The manufacturer faces a linear demand curve which is a function of price. Let D_i be the demand for product i, and p will be the price assigned to both of the products. We make the following further assumptions:

1. Base demand $A_i,\ i = 1, 2$ are identical for both products.
2. ϵ_i has the same distribution for both products.

   The ϵ_i is a Bernoulli random variable with

   $$\epsilon_i = \begin{cases} \epsilon^h, & \text{w.p. } \pi, \\ 0, & \text{w.p. } 1 - \pi. \end{cases}$$

3. For notational simplicity, we scale the coefficient $a = 1 - \beta$, A and ϵ^h by $(1 - \beta)$. Then the expected revenue must be multiplied by $(1 - \beta)$ (to be on the same scale with the expected profit in Model 1). In the following, to distinguish between the random variable ϵ_i and the realization ϵ^h, we denote ϵ^h with b.

Under these assumptions, the demand curve is governed the by following function:

$$D_i = A - p + \epsilon_i,\ i = 1, 2. \tag{E.0.1}$$

Here A is the intercept of the demand curve, and ϵ_i is the random part of the base demand. Note in this model ϵ_is are i.i.d. random variables. Thus, the products are identical.

Stock quantity for the products are identical and denoted with I. Cost per unit produced is c. The variable I is determined before demand realization, whereas price p is determined after demand realization. For a given I, the profit function is

$$\begin{aligned} E[R(I)] &= E[\max_{p:A-p+\epsilon_i\geq 0} \sum_i \min\{A-p+\epsilon_i, I\}p] - 2cI \\ &= (1-\pi)^2 R_{ll}(I) + 2(1-\pi)\pi R_{hl}(I) + \pi^2 R_{hh}(I) - 2cI, \end{aligned}$$

where $R_{ll}(I)$ stands for the expected revenue when for both products, the base demand is “low", $R_{hl}(I)$ stand for the revenue when one product has “high” base demand whereas the other has “low" base demand, and $R_{hh}(I)$ is the revenue when both products have “high” base demand. The functions are expressed as:

$$\begin{aligned} R_{ll}(I) &= \max_{p:A-p\geq 0} \sum_i \min\{A-p, I\}p, \\ R_{hl}(I) &= \max_{p:A-p\geq 0} \min\{A-p, I\}p + \min\{A-p+b, I\}p, \\ R_{hh}(I) &= \max_{p:A-p+b\geq 0} \sum_i \min\{A-p+b, I\}p. \end{aligned}$$

Each of the $R_{ll}(I)$, $R_{ll}(I)$ and $R_{ll}(I)$ are obtained as follows.

Lemma E.0.1.

$$R_{ll}(I) = \begin{cases} 2I(A-I), & \textit{for } I \leq \frac{A}{2}, \\ \frac{A^2}{2}, & \textit{for } I \geq \frac{A}{2}. \end{cases}$$

Proof. The value $R_{ll}(I)$ is obtained as follows. Under low demand status for each product, let $D = A-p$. Then, expected profit for each product can alternatively be expressed as $D(A-D)$. Here D is the decision variable. This is a concave function of D. Note $D^* = \frac{A}{2}$, i.e., $p = A - D = \frac{A}{2}$ maximizes the profit. The variable D is constrained by I. For $I < \frac{A}{2}$, optimal demand quantity for each product is I and optimal price is $A-I$. For $I > \frac{A}{2}$, optimal price and demand is $\frac{A}{2}$. Total expected profit for the two products is as in the expression above. □

Obtaining $R_{hh}(I)$ follows similar lines with $R_{ll}(I)$.It is possible to show that,

$$R_{hh}(I) = \begin{cases} 2I(A-I+b), & \text{for } I \leq \frac{A+b}{2}, \\ \frac{(A+b)^2}{2}, & \text{for } I \geq \frac{A+b}{2}. \end{cases}$$

For $R_{hl}(I)$ (or R_{lh}), expected profit is expressed as in the following lemma.

Lemma E.0.2. *The function $R_{hl}(I)$ is expressed as follows:*

1. *For $A \leq \frac{b}{2}$*

$$R_{hl}(I) = \begin{cases} 2I(A-I), & \text{for } I \leq \frac{A}{3}, \\ \frac{(A+I)^2}{4}, & \text{for } A \geq I \geq \frac{A}{3}, \\ IA, & \text{for } b \geq I \geq A, \\ bA, & \text{for } I \geq b. \end{cases}$$

2. *For $\frac{b}{2} \leq A \leq b$*

$$R_{hl}(I) = \begin{cases} 2I(A-I), & \text{for } I \leq \frac{A}{3}, \\ \frac{(A+I)^2}{4}, & \text{for } A \geq I \geq \frac{A}{3}, \\ IA, & \text{for } b \geq I \geq A, \\ (2I-b)(A-I+b), & \text{for } \frac{A}{2} + \frac{3b}{4} \geq I \geq b, \\ \frac{(2A+b)^2}{8} & \text{for } I \geq \frac{A}{2} + \frac{3b}{4}. \end{cases}$$

3. *For $b \leq A$*

$$R_{hl}(I) = \begin{cases} 2I(A-I), & \text{for } I \leq \frac{A}{3}, \\ \frac{(A+I)^2}{4}, & \text{for } \frac{A+2b}{3} \geq I \geq \frac{A}{3}, \\ (2I-b)(A-I+b), & \text{for } \frac{A}{2} + \frac{3b}{4} \geq I \geq \frac{A+2b}{3}, \\ \frac{(2A+b)^2}{8}, & \text{for } I \geq \frac{A}{2} + \frac{3b}{4}. \end{cases}$$

Proof. The problem under the state hl can be expressed as follows:

$$\begin{aligned} R_{hl}(I) &= \max_p \min\{D_1, I\}p + \min\{D_2, I\}p \\ s.t. & \\ D_i &= A - p + \epsilon_i, \ i = 1, 2 \\ D_1 &\geq 0, \\ D_2 &\geq 0, \\ p &\geq 0. \end{aligned}$$

Note that the constraint $p \geq 0$ is redundant. D_i is as defined in Eq. E.0.1. The decision variable is p. The model can equivalently be expressed as follows.

$$
\begin{aligned}
R_{hl}(I) &= \max D_1 p + D_2 p \\
&s.t. \\
D_1 + p &\leq A + b, \\
D_2 + p &\leq A, \\
-D_1 &\leq 0, \\
D_1 &\leq I, \\
-D_2 &\leq 0, \\
D_2 &\leq I.
\end{aligned}
$$

Now, the decision variables are p, D_1 and D_2. Here D_1 and D_2 are not necessarily defined by Eq. E.0.1, but rather as $\min\{D_i, I\}$. Associate the dual variables $u_1, \cdots, u_6$ with each constraint. KKT conditions for optimality are:

$$
-\nabla f(x) + \sum_i u_i \nabla g(x) = \begin{bmatrix} -p + u_1 - u_3 + u_4 \\ -p + u_2 - u_5 + u_6 \\ -(D_1 + D_2) + u_1 + u_2 \end{bmatrix} = \mathbf{0},
$$

$$
\text{(Complementary slackness conditions)}
$$

$$
u_i \geq 0.
$$

Here we observe that in feasible optimal points:

1. Either $u_1 > 0$ or $u_2 > 0$.
2. $u_3 = 0$, and $u_4 \geq 0$.
3. $u_5 = 0$ or $u_6 = 0$.
4. Either $u_1 > 0$ or $u_4 > 0$.
5. Either $u_2 > 0$ or $u_6 > 0$.

Table E.2: Summary of E[R(I)] for $A \leq \frac{b}{2}$

Interval	E[R(I)]	Maximizer for interval
$b \leq I$	$\frac{(A+b)^2}{2}\pi^2 + 2Ab\pi(1-\pi)$ $+\frac{A^2}{2}(1-\pi)^2 - 2cI$	$I^* = b$
$\frac{A+b}{2} \leq I \leq b$	$\frac{(A+b)^2}{2}\pi^2 + 2AI\pi(1-\pi)$ $+\frac{A^2}{2}(1-\pi)^2 - 2cI$	$A\pi(1-\pi) - c \geq 0 \implies I^* = b$ $A\pi(1-\pi) - c \leq 0 \implies I^* = \frac{A+b}{2}$
$A \leq I \leq \frac{A+b}{2}$	$2I(A+b-I)\pi^2 + 2AI\pi(1-\pi)$ $+\frac{A^2}{2}(1-\pi)^2 - 2cI$	$\bar{I} = \frac{(A+b)\pi^2 + A\pi(1-\pi) - c}{2\pi^2}$ $I^* = \min(\frac{A+b}{2}, \max(A, \bar{I}))$
$\frac{A}{2} \leq I \leq A$	$2I(A+b-I)\pi^2 + \frac{(A+I)^2}{2}\pi(1-\pi)$ $+\frac{A^2}{2}(1-\pi)^2 - 2cI$	$\bar{I} = \frac{2(A+b)\pi^2 + A\pi(1-\pi) - 2c}{4\pi^2 - \pi(1-\pi)}$ $I^* = A \wedge \frac{A}{2} \wedge \bar{I}$
$\frac{A}{3} \leq I \leq \frac{A}{2}$	$2I(A+b-I)\pi^2 + \frac{(A+I)^2}{2}\pi(1-\pi)$ $+2I(A-I)(1-\pi)^2 - 2cI$	$\bar{I} = \frac{2(A+b)\pi^2 + A\pi(1-\pi) + 2A(1-\pi)^2 - 2c}{4\pi^2 - \pi(1-\pi) + 4(1-\pi)^2}$ $I^* = \frac{A}{3} \wedge \frac{A}{2} \wedge \bar{I}$
$0 \leq I \leq \frac{A}{3}$	$2Ib\pi^2$ $+2I(A-I) - 2cI$	$\bar{I} = \frac{b\pi^2 + A - c}{2}$ $I^* = \frac{A}{3} \wedge \bar{I}$

Table E.3: Summary of E[R(I)] for $\frac{b}{2} \leq A \leq b$

Interval	E[R(I)]	Maximizer for interval
$\frac{A}{2} + \frac{3b}{4} \leq I$	$\frac{(A+b)^2}{2}\pi^2 + \frac{(2A+b)^2}{4}\pi(1-\pi)$ $+\frac{A^2}{2}(1-\pi)^2 - 2cI$	$I^* = \frac{A}{2} + \frac{3b}{4}$
$b \leq I \leq \frac{A}{2} + \frac{3b}{4}$	$\frac{(A+b)^2}{2}\pi^2 + 2(2I-b)(A-I+b)\pi(1-\pi)$ $+\frac{A^2}{2}(1-\pi)^2 - 2cI$	$\bar{I} = \frac{A}{2} + \frac{3b}{4} - \frac{2c}{8\pi(1-\pi)}$ $I^* = \min(\frac{A}{2} + \frac{3b}{4}, \max(b, \bar{I}))$
$\frac{A+b}{2} \leq I \leq b$	$\frac{(A+b)^2}{2}\pi^2 + 2IA\pi(1-\pi)$ $+\frac{A^2}{2}(1-\pi)^2 - 2cI$	$A\pi(1-\pi) - c \geq 0 \Longrightarrow I^* = b$ $A\pi(1-\pi) - c \leq 0 \Longrightarrow I^* = \frac{A+b}{2}$
$A \leq I \leq \frac{A+b}{2}$	$2I(A+b-I)\pi^2 + 2AI\pi(1-\pi)$ $+\frac{A^2}{2}(1-\pi)^2 - 2cI$	$\bar{I} = \frac{(A+b)\pi^2 + A\pi(1-\pi) - c}{2\pi^2}$ $I^* = \min(\frac{A+b}{2}, \max(A, \bar{I}))$
$\frac{A}{2} \leq I \leq A$	$2I(A+b-I)\pi^2 + \frac{(A+I)^2}{2}\pi(1-\pi)$ $+\frac{A^2}{2}(1-\pi)^2 - 2cI$	$\bar{I} = \frac{2(A+b)\pi^2 + A\pi(1-\pi) - 2c}{4\pi^2 - \pi(1-\pi)}$ $I^* = A \wedge \frac{A}{2} \wedge \bar{I}$
$\frac{A}{3} \leq I \leq \frac{A}{2}$	$2I(A+b-I)\pi^2 + \frac{(A+I)^2}{2}\pi(1-\pi)$ $+2I(A-I)(1-\pi)^2 - 2cI$	$\bar{I} = \frac{2(A+b)\pi^2 + A\pi(1-\pi) + 2*A(1-\pi)^2 - 2c}{4\pi^2 - \pi(1-\pi) + 4*(1-\pi)^2}$ $I^* = \frac{A}{3} \wedge \frac{A}{2} \wedge \bar{I}$
$0 \leq I \leq \frac{A}{3}$	$2I(b)\pi^2 + 2I(A-I)$ $-2cI$	$\bar{I} = \frac{b\pi^2 + A - c}{2}$ $I^* = \frac{A}{3} \wedge \bar{I}$

Table E.4: Summary of E[R(I)] for $b \leq A$

Interval	E[R(I)]	Maximizer for interval
$\frac{A}{2} + \frac{3b}{4} \leq I$	$\frac{(A+b)^2}{2}\pi^2 + \frac{(2A+b)^2}{4}\pi(1-\pi)$ $+\frac{A^2}{2}(1-\pi)^2 - 2cI$	$I^* = \frac{A}{2} + \frac{3b}{4}$
$\frac{A+b}{2} \leq I \leq \frac{A}{2}+\frac{3b}{4}$	$\frac{(A+b)^2}{2}\pi^2$ $+2(2I-b)(A-I+b)\pi(1-\pi)$ $+\frac{A^2}{2}(1-\pi)^2 - 2cI$	$\bar{I} = \frac{A}{2} + \frac{3b}{4} - \frac{2c}{8\pi(1-\pi)}$ $I^* = \min(\frac{A}{2} + \frac{3b}{4}, \max(\frac{A+b}{2}, \bar{I}))$
$\max(\frac{A}{2}, \frac{A+2B}{3}) \leq$ $I \leq \frac{A+b}{2}$	$2I(A+b-I)\pi^2$ $+2(2I-b)(A-I+b)\pi(1-\pi)$ $+\frac{A^2}{2}(1-\pi)^2 - 2cI$	$\bar{I} = \frac{(A+b)\pi^2+(2A+3b)\pi(1-\pi)-2c}{2\pi^2+4\pi(1-\pi)}$ $I^* = \min(\frac{A+b}{2}, \max(\max(\frac{A}{2}, \frac{A+2B}{3}), \bar{I}))$
If $A \leq 4b$ $\frac{A}{2} \leq I \leq \frac{A+2B}{3}$	$2I(A+b-I)\pi^2 + \frac{(A+I)^2}{2}\pi(1-\pi)$ $+\frac{A^2}{2}(1-\pi)^2 - 2cI$	$\bar{I} = \frac{2(A+b)\pi^2+A\pi(1-\pi)-2c}{4\pi^2-\pi(1-\pi)}$ $I^* = \frac{A+2B}{3} \wedge \frac{A}{2} \wedge \bar{I}$
If $A \geq 4b$ $\frac{A+2B}{3} \leq I \leq \frac{A}{2}$	$2Ib\pi + 2b(2I-A-b)\pi(1-\pi)$ $+2I(A-I) - 2cI$	$\bar{I} = \frac{b\pi^2+3b\pi(1-\pi)+A-c}{2}$ $I^* = \frac{A+2B}{3} \wedge \frac{A}{2} \wedge \bar{I}$
$\frac{A}{3} \leq I \leq$ $\min(\frac{A}{2}, \frac{A+2B}{3})$	$2I(A+b-I)\pi^2 + \frac{(A+I)^2}{2}\pi(1-\pi)$ $+2I(A-I)(1-\pi)^2 - 2cI$	$\bar{I} = \frac{2b\pi^2-3A\pi(1-\pi)+2A-2c}{4\pi^2+4(1-\pi)^2-\pi(1-\pi)}$ $I^* = \frac{A}{3} \wedge \min(\frac{A}{2}, \frac{A+2B}{3}) \wedge \bar{I}$
$0 \leq I \leq \frac{A}{3}$	$2Ib\pi^2 + 2I(A-I) - 2cI$	$\bar{I} = \frac{b\pi^2+A-c}{2}$ $I^* = \frac{A}{3} \wedge \bar{I}$

All possible KKT points are enumerated in set B.

$$B = \{\{1,2\},\{1,2,5\},\{1,2,4\},\{2,5,4\},\{2,6,4\},\{2,4\}\}$$

other sets are either infeasible or degenerate. The active constraints are denoted as follows. If $u_i > 0$ then $i \in Ac$, and if $u_i = 0$ then $i \in Ac^c$, where $Ac \in B$.

Each set generates Table E.5. This completes the proof.

Table E.5: KKT points

Active Set	(D_1, D_2, p)	Lagrange values	$R_{hl}(I)$	condition
$\{1,2\}$	$(\frac{A}{2}+\frac{3b}{4}, \frac{A}{2}-\frac{b}{4}, \frac{A}{2}+\frac{b}{4})$	$u_1 = u_2 = \frac{A}{2}+\frac{b}{4}$	$\frac{(2A+b)^2}{8}$	$I \geq \frac{A}{2}+\frac{3b}{4}$, $2A \geq b$
$\{1,2,5\}$	$(b,0,A)$	$u_1 = A$, $u_2 = b-A$, $u_5 = b-2A$	bA	$I \geq b$, $2A \leq b$
$\{2,5,4\}$	$(I,0,A)$	$u_4 = A$, $u_2 = I$, $u_5 = I-A$	IA	$b \geq I \geq A$
$\{1,2,4\}$	$(I, I-b, A-I+b)$	$u_2 = A-I+b$, $u_1 = 3I-A-2b$, $u_4 = 2A+3B-4I$	$(2I-b)(A-I+b)$	$I \geq b$, $I \geq \frac{A+2b}{3}$, $I \leq \frac{2A+3b}{4}$
$\{2,4\}$	$(I, \frac{A-I}{2}, \frac{A+I}{2})$	$u_2 = u_4 = \frac{A+I}{2}$	$\frac{(A+I)^2}{4}$	$A \geq I$, $\frac{A+2b}{3} \geq I \geq \frac{A}{3}$
$\{2,4,6\}$	$(I,I,A-I)$	$u_2 = 2I$, $u_4 = A-I$, $u_6 = A-3I$	$2I(A-I)$	$I \leq \frac{A}{3}$

□

Combining $R_{ll}(I)$, $R_{hl}(I)$, $R_{lh}(I)$, and $R_{hh}(I)$ gives $E[R(I)]$:

$$E[R(I)] = [(1-\pi)^2 R_{ll}(I) + 2*(1-\pi)\pi R_{hl}(I) + \pi^2 R_{hh}(I)]$$

Table E.2 summarizes the equation details and maximizers for $E[R(I)]$.

APPENDIX F

AN ALTERNATIVE EXPLANATION OF THE TWO PERIOD DEMAND FUNCTION IN SECTION

Here, we explain the form of the second period demand without resorting to the Representative Consumer Theory.

Case(i) $D_i^1(\omega) \leq I_i,\ i = 1, 2.$

If at the end of Period 1, both products have remaining inventory, $I_i > D_i^1(\omega)$ (or $s_i > \epsilon_i(\omega)$), then transfer of overstock to understock is not possible. However, if $p_2 < p_1$ some of the customers that did not make any purchase in the first period, may make purchase in the second period. Thus in Period 2, there will be demand generated due to the price differential. For a given realization of ϵ_i, demand in Period 2 will be defined as

$$D_i^2 = (D_i^1 - I_i)^+ + (A_i - p_i^2 + \beta p_j^2 + \epsilon_i - D_i^1)^+,$$
$$i = 1, 2,\ i \neq j.$$

First term in the expression $(D_i^1 - I_i)^+$, denotes the unsatisfied demand in the first period, whereas the second term $(D_i^1 - (A_i - p_i^2 + \beta p_j^2 + \epsilon_i))$ denotes the additional demand generated due to price differential. Under uniform pricing we obtain,

$$D_i^2 = (1 - \beta)(p_1 - p_2),\ i = 1, 2.$$

So both products will be governed by the same demand function in Period 2, if there is inventory remaining for both products at the end of Period 1. Note here that without loss of optimality p_2 would be lower than p_1.

Case(ii) $D_i^1(\omega) \leq I_i, D_j^1(\omega) \geq I_j,\ i,j \in \{1,2\}, i \neq j.$

If product 1 is overstock while the other is understock, then demand in Period 2 can be expressed as follows:

$$\begin{aligned} D_1^2 &= (D_1^1 - I_1)^+ + (A_1 - p_1^2 + \beta p_2^2 + \epsilon_1 - D_1^1)^+, \\ D_2^2 &= (D_2^1 - I_2)^+ + (A_2 - p_2^2 + \beta p_1^2 + \epsilon_2 - D_2^1)^+ \end{aligned} \tag{F.0.1}$$

Note in this case, $(D_1^1 - I_1)^+ = (\epsilon_1 - s_1)^+ = 0$, whereas $(D_2^1 - I_2)^+ = (\epsilon_2 - s_2)^+ \geq 0$, with equality due to $\epsilon_2 - s_2 = 0$. In the expressions, in D_1^1 and D_2^1, $p_2^1 = p_1^1 = p_1$, whereas we set p_1^2 and p_2^2 not necessarily identical (we explain the reasoning below).

For Product 1, other than the price differential created in Period 2, there will be customers who opt to buy Product 1 due to unavailability of Product 2. The unmet demand for Product 2 will shift to Product 1 demand (after rationed by the substitution factor). Note that at the end of Period 1, unmet Product 2 demand is $\epsilon_2 - s_2$. This amount, multiplied by the substitution factor β will spill over to Product 1 demand. Note that, this is the maximum amount that can spill over, and the maximum spill over will result in a higher profit than any other spill over quantity. In fact, there is an implicit p_2^2, that determines the amount of spill over. Note that in Eq. F.0.1 above, Product 1 and Product 2 demand in Period 2 are expressed as follows:

$$\begin{aligned} D_1^2 &= (p_1 - p_1^2) - \beta(p_1 - p_2^2), \\ D_2^2 &= (\epsilon_2 - s_2) + (p_1 - p_2^2) - \beta(p_1 - p_1^2). \end{aligned}$$

A rational price setter will set the price for product 2, so that all the possible demand for product 2 will shift to Product 1 demand. So the p_2^2 that will set $D_2^2 = 0$ can be obtained as:

$$p_2^2 = p_1 + (\epsilon_2 - s_2) - \beta(p_1 - p_1^2).$$

Thus, D_1^2 and D_2^2 are:

$$\begin{aligned} D_1^2 &= \beta(\epsilon_2 - s_2) + (1 - \beta^2)(p_1 - p_2), \\ D_2^2 &= 0. \end{aligned}$$

Here since Product 2 does not have any inventory left, it is as if the manufacturer has the flexibility to charge Product 1 and Product 2 different prices in Period 2. Period 1 prices for both products must still be the same, at p_1. For this case, we simply denote p_1^2 with p_2. Here if we look at the coefficient of the price differential, which is $(1-\beta^2)$, we see that it takes into consideration the spillover due to both price and stockout occasion. In Period 1, the product 1 demand is spilled into product 2, and if there is stockout for Product 2, it is again spilled into Product 1.

Case(iii) $D_i > I_i^1, D_j > I_j^1, i, j \in \{1,2\}, i \neq j$

If both products are depleted by the beginning of period 2, then

$$D_i^2 = 0, i = 1, 2.$$

APPENDIX G

PARTITION ANALYSIS FOR TWO-PERIOD PROBLEM IN SELLER'S RESPONSE TO CUSTOMER-DRIVEN SUBSTITUTION

In the two-period problem, identifying the optimal values s_1, s_2, p_1 requires partitioning of the solution space $R \times R$. Consider the partition P in the following figure.

For each element in P, we identify which partition for second period is valid, for each possible net inventory level at the beginning of Period 2, (K_1, K_2).

G.1 Region: pp1-Area a2

Possible areas in partition B for each of the possible net inventory levels that satisfy **pp1-a2** are

$(s_1-\epsilon_1^h, s_2)$	$(s_1-\epsilon_1^h, s_2-\epsilon_2^h)$	(s_1, s_2)	$(s_1, s_2-\epsilon_2^h)$
b1	b1	b1	b1
b2	b2	b2	b2
b2	b2	b1	b2
b2	b2	b1	b1
b4	b4	b4	b4
b4	b4	b3	b4
b4	b4	b2	b2
b4	b4	b2	b3
b4	b4	b2	b4
b3	b3	b2	b2
b3	b3	b1	b2
b3	b3	b1	b2
b3	b4	b3	b3
b3	b4	b3	b4
b3	b4	b2	b2
b3	b4	b2	b3
b3	b4	b2	b4
b3	b4	b1	b2
b3	b4	b1	b3
b3	b4	b1	b4

G.1.0.1 Region: pp2

Region pp2 is defined as $\{(s_1, s_2) \in R \times R : 0 \leq s_1 \leq \epsilon_1^h, s_2 \geq \epsilon_2^h\}$. In this element of the partition P, expected profit function is defined as follows:

$$\begin{aligned}E[R] = (A_1 + A_2 - 2(1-\beta)p_1)(p_1 - c) + s_1(p_1\pi_1 - c) - c(s_2)\\
+ (1-\pi_1)(1-\pi_2)R_2^*(s_1, s_2, p_1)\\
+ \pi_1(1-\pi_2)R_2^*(s_1-\epsilon_1^h, s_2, p_1)(1-\pi_1)\pi_2R_2^*(s_1, s_2-\epsilon_2^h, p_1)\\
+ \pi_1\pi_2R_2^*(s_1-\epsilon_1^h, s_2-\epsilon_2^h), \quad (s_1, s_2) \in pp2\end{aligned}$$

In order to find which function governs R_2, we first look at possible net inventory levels at the beginning of period 2, given (s_1, s_2). Note that pp2 has intersection with a1 and a2.

Region: pp2: Area a1

In region pp2, area a1 is defined as:

$$a1 = \{(s_1, s_2) \in pp2 : s_1 - \epsilon_1^h \leq s_1 \leq s_2 - \epsilon_2^h \leq s_2\}$$

Hence resulting revenue function is:

$$\begin{aligned}
E[R] = &(A_1 + A_2 - 2(1-\beta)p_1)(p_1 - c) + s_1(p_1\pi_1 - c) - c(s_2) \\
&+ (1-\pi_1)(1-\pi_2)R^2(s_1, s_2, p_1) \\
&+ \pi_1(1-\pi_2)R^2(s_1 - \epsilon_1^h, s_2, p_1) + (1-\pi_1)\pi_2 R^2(s_1, s_2 - \epsilon_2^h, p_1) \\
&+ \pi_1\pi_2 R^2(s_1 - \epsilon_1^h, s_2 - \epsilon_2^h, p_1), \quad (s_1, s_2) \in pp2 \bigcap a1
\end{aligned}$$

For each possible net inventory level at the beginning of period 2, $(s_1, s_2 - \epsilon_2^h)$, (s_1, s_2), $(s_1 - \epsilon_1^h, s_2 - \epsilon_2^h)$, $(s_1 - \epsilon_1^h, s_2)$, we identify the elements of the partition B and D, which the tuple assumes. See Appendix.

Region: pp2: Area a2

In region pp2, area a2 is defined as:

$$a2 = \{(s_1, s_2) \in pp2 : s_1 - \epsilon_1^h \leq s_2 - \epsilon_2^h \leq s_1 \leq s_2\}$$

Hence resulting expected profit function is:

$$\begin{aligned}
E[R] = &(A_1 + A_2 - 2(1-\beta)p_1)(p_1 - c) + s_1(p_1\pi_1 - c) - c(s_2) \\
&+ (1-\pi_1)(1-\pi_2)R^2(s_1, s_2, p_1) \\
&+ \pi_1(1-\pi_2)R^2(s_1 - \epsilon_1^h, s_2, p_1) + (1-\pi_1)\pi_2 R^2(s_2 - \epsilon_2^h, s_1, p_1) \\
&+ \pi_1\pi_2 R^2(s_1 - \epsilon_1^h, s_2 - \epsilon_2^h, p_1), \quad (s_1, s_2) \in pp2 \bigcap a2
\end{aligned}$$

For each possible net inventory level at the beginning of period 2, $(s_2-\epsilon_2^h, s_1)$, (s_1, s_2), $(s_1-\epsilon_1^h, s_2-\epsilon_2^h)$, $(s_1-\epsilon_1^h, s_2)$, we identify the elements of the partition B and D, which the tuple assumes. See Appendix.

G.1.0.2 Region: pp3

Area pp3 is $\{(s_1, s_2) \in R \times R : s_1 \geq \epsilon_1^h, 0 \leq s_2 \leq \epsilon_2^h\}$

$$\begin{aligned} E[R] = &(A_1 + A_2 - 2(1-\beta)p_1)(p_1 - c) + s_2(p_1\pi_1 - c) - c(s_1) \\ &+ (1-\pi_1)(1-\pi_2)R_2^*(s_1, s_2, p_1) \\ &+ \pi_1(1-\pi_2)R_2^*(s_1 - \epsilon_1^h, s_2, p_1)(1-\pi_1)\pi_2 R_2^*(s_1, s_2 - \epsilon_2^h, p_1) \\ &+ \pi_1\pi_2 R_2^*(s_1 - \epsilon_1^h, s_2 - \epsilon_2^h, p_1), \qquad (s_1, s_2) \in pp2 \end{aligned}$$

In order to find which function governs R_2, we first look at the magnitudes comparison of (s_1, s_2). pp3 has intersection with a3, a4 and a5.

Area a3

Expected profit function is:

$$\begin{aligned} E[R] = &(A_1 + A_2 - 2(1-\beta)p_1)(p_1 - c) + s_2(p_1\pi_1 - c) - c(s_1) + (1-\pi_1)(1-\pi_2)R^2(s_1, s_2, p_1) \\ &+ \pi_1(1-\pi_2)R^2(s_1 - \epsilon_1^h, s_2, p_1) + (1-\pi_1)\pi_2 R^2(s_1, s_2 - \epsilon_2^h, p_1) \\ &+ \pi_1\pi_2 R^2(s_1 - \epsilon_1^h, s_2 - \epsilon_2^h, p_1), (s_1, s_2) \in pp3 \bigcap a3 \end{aligned}$$

The possible areas for Region pp3, Area a3 are:

$(s_1 - \epsilon_1^h, s_2)$	(s_1, s_2)
b1	b1
b2	b1
b2	b2
b3	b3
b3	b2
b3	b1
b4	b4
b4	b3
b4	b2

$(s_2-\epsilon_2^h, s_1-\epsilon_1^h)$	$(s_2-\epsilon_2^h, s_1)$
d1	d1
d2	d1
d2	d2

Area a4

Ditto for pp2-Area a2 replacing s_1 and s_2 with each other, and replacing ϵ_1^h and ϵ_2^h with each other.

Area a5

Ditto for pp2-Area a1 replacing s_1 and s_2 with each other, and replacing ϵ_1^h and ϵ_2^h with each other.

G.1.0.3 Region: pp4

Area pp4 is $\{(s_1, s_2) : 0 \leq s_1 \leq \epsilon_1^h, 0 \leq s_2 \leq \epsilon_2^h\}$

$$\begin{aligned}
E[R] &= (A_1 + A_2 - 2(1-\beta)p_1)(p_1 - c) - c(s_1 + s_2) \\
&\quad + (1-\pi_1)(1-\pi_2)R_2^*(s_1, s_2, p_1) \\
&\quad + \pi_1(1-\pi_2)R_2^*(s_1 - \epsilon_1^h, s_2, p_1)(1-\pi_1)\pi_2 R_2^*(s_1, s_2 - \epsilon_2^h, p_1), \quad (s_1, s_2) \in pp2
\end{aligned}$$

In order to find which function governs R_2, we first look at the magnitudes comparison of (s_1, s_2). pp4 has intersection with a2,a3 and a4.

Area a2

Expected profit function is:

$$\begin{aligned}
E[R] &= (A_1 + A_2 - 2(1-\beta)p_1)(p_1 - c) + s_1(p_1\pi_1 - c) - c(s_2) \\
&\quad + (1-\pi_1)(1-\pi_2)R^2(s_1, s_2, p_1) \\
&\quad + \pi_1(1-\pi_2)R^2(s_1 - \epsilon_1^h, s_2, p_1) \\
&\quad + (1-\pi_1)\pi_2 R^2(s_2 - \epsilon_2^h, s_1, p_1), \quad (s_1, s_2) \in pp4 \bigcap a2
\end{aligned}$$

$(s_2-\epsilon_2^h, s_1)$	$(s_1-\epsilon_1^h, s_2)$
d1	d1
d1	d2
d2	d1
d2	d2

(s_1, s_2)
b1
b2
b3
b4

Area a3

Expected profit function is:

$$\begin{aligned} E[R] = &(A_1 + A_2 - 2(1-\beta)p_1)(p_1 - c) + s_2(p_1\pi_1 - c) - c(s_1) \\ &+ (1-\pi_1)(1-\pi_2)R^2(s_1, s_2) \\ &+ \pi_1(1-\pi_2)R^2(s_1 - \epsilon_1^h, s_2, p_1) \\ &+ (1-\pi_1)\pi_2 R^2(s_1, s_2 - \epsilon_2^h, p_1), (s_1, s_2) \in pp3 \bigcap a3 \end{aligned}$$

$(s_2-\epsilon_2^h, s_1)$	$(s_1-\epsilon_1^h, s_2)$
d1	d1
d2	d1
d2	d2

(s_1, s_2)
b1
b2
b3
b4

Area a4

Ditto for pp4-Area a2 replacing s_1 and s_2 with each other, and replacing ϵ_1^h and ϵ_2^h with each other.

The following is the list of all possible second period net inventory levels, and the regions that generate those levels.

Table G.1: Feasible areas for second period demand

$(s_1, s2)$	$(s_1, s2 - \epsilon_2^h)$	$(s1 - \epsilon_1^h, s2)$	$(s1 - \epsilon_1^h, s2 - \epsilon_2^h)$	PP	A
d0	d0	d0	d0	nn1	
d2	d2	d2	d2	np1	
d1	d2	d1	d2	np1	
d1	d2	d2	d2	np1	
d1	d1	d1	d2	np1	
d1	d1	d1	d1	np1	
d2	d0	d2	d0	np2	
d1	d0	d1	d0	np2	
d1	d0	d2	d0	np2	
$\bar{d}2$	$\bar{d}2$	$\bar{d}2$	$\bar{d}2$	pn1	
$\bar{d}1$	$\bar{d}2$	$\bar{d}1$	$\bar{d}2$	pn1	
$\bar{d}1$	$\bar{d}2$	$\bar{d}2$	$\bar{d}2$	pn1	
$\bar{d}1$	$\bar{d}1$	$\bar{d}1$	$\bar{d}2$	pn1	
$\bar{d}1$	$\bar{d}1$	$\bar{d}2$	$\bar{d}2$	pn1	
$\bar{d}1$	$\bar{d}1$	$\bar{d}1$	$\bar{d}1$	pn1	
$\bar{d}2$	$\bar{d}2$	d0	d0	pn2	
$\bar{d}1$	$\bar{d}2$	d0	d0	pn2	
$\bar{d}1$	$\bar{d}1$	d0	d0	pn2	
b4	$\bar{b}4$	b4	b4	pp1	a2
b3	$\bar{b}4$	b3	b4	pp1	a2
b3	$\bar{b}4$	b4	b4	pp1	a2
b2	$\bar{b}4$	b3	b4	pp1	a2
b2	$\bar{b}4$	b4	b4	pp1	a2
b2	$\bar{b}3$	b3	b4	pp1	a2
b2	$\bar{b}3$	b4	b4	pp1	a2
b1	$\bar{b}3$	b3	b4	pp1	a2
b2	$\bar{b}2$	b3	b4	pp1	a2

Continued on next page

$(s_1, s2)$	$(s_1, s2 - \epsilon_2^h)$	$(s1 - \epsilon_1^h, s2)$	$(s1 - \epsilon_1^h, s2 - \epsilon_2^h)$	PP	A
b2	$\bar{b2}$	b4	b4	pp1	a2
b1	$\bar{b2}$	b3	b4	pp1	a2
b2	$\bar{b2}$	b3	b3	pp1	a2
b1	$\bar{b2}$	b3	b3	pp1	a2
b1	$\bar{b1}$	b3	b3	pp1	a2
b2	$\bar{b2}$	b2	b2	pp1	a2
b1	$\bar{b2}$	b2	b2	pp1	a2
b1	$\bar{b1}$	b2	b2	pp1	a2
b1	$\bar{b1}$	b1	b1	pp1	a2
b4	$\bar{b4}$	b4	$\bar{b4}$	pp1	a3
b3	$\bar{b4}$	b3	$\bar{b4}$	pp1	a3
b3	$\bar{b4}$	b4	$\bar{b4}$	pp1	a3
b2	$\bar{b4}$	b2	$\bar{b4}$	pp1	a3
b2	$\bar{b4}$	b3	$\bar{b4}$	pp1	a3
b2	$\bar{b4}$	b4	$\bar{b4}$	pp1	a3
b2	$\bar{b3}$	b2	$\bar{b4}$	pp1	a3
b2	$\bar{b3}$	b3	$\bar{b4}$	pp1	a3
b2	$\bar{b3}$	b4	$\bar{b4}$	pp1	a3
b1	$\bar{b3}$	b2	$\bar{b4}$	pp1	a3
b1	$\bar{b3}$	b3	$\bar{b4}$	pp1	a3
b2	$\bar{b3}$	b2	$\bar{b3}$	pp1	a3
b1	$\bar{b3}$	b1	$\bar{b3}$	pp1	a3
b1	$\bar{b3}$	b2	$\bar{b3}$	pp1	a3
b2	$\bar{b2}$	b2	$\bar{b2}$	pp1	a3
b1	$\bar{b2}$	b1	$\bar{b2}$	pp1	a3
b1	$\bar{b2}$	b2	$\bar{b2}$	pp1	a3
b1	$\bar{b1}$	b1	$\bar{b1}$	pp1	a3
$\bar{b4}$	$\bar{b4}$	b4	$\bar{b4}$	pp1	a4
$\bar{b3}$	$\bar{b4}$	b4	$\bar{b4}$	pp1	a4
$\bar{b2}$	$\bar{b4}$	b2	$\bar{b4}$	pp1	a4

Continued on next page

$(s_1, s2)$	$(s_1, s2-\epsilon_2^h)$	$(s1-\epsilon_1^h, s2)$	$(s1-\epsilon_1^h, s2-\epsilon_2^h)$	PP	A
$\bar{b2}$	$\bar{b4}$	b3	$\bar{b4}$	pp1	a4
$\bar{b2}$	$\bar{b4}$	b4	$\bar{b4}$	pp1	a4
$\bar{b3}$	$\bar{b3}$	b4	$\bar{b4}$	pp1	a4
$\bar{b2}$	$\bar{b3}$	b2	$\bar{b4}$	pp1	a4
$\bar{b2}$	$\bar{b3}$	b3	$\bar{b4}$	pp1	a4
$\bar{b2}$	$\bar{b3}$	b4	$\bar{b4}$	pp1	a4
$\bar{b1}$	$\bar{b3}$	b2	$\bar{b4}$	pp1	a4
$\bar{b1}$	$\bar{b3}$	b3	$\bar{b4}$	pp1	a4
$\bar{b2}$	$\bar{b3}$	b2	$\bar{b3}$	pp1	a4
$\bar{b1}$	$\bar{b3}$	b1	$\bar{b3}$	pp1	a4
$\bar{b1}$	$\bar{b3}$	b2	$\bar{b3}$	pp1	a4
$\bar{b2}$	$\bar{b2}$	b2	$\bar{b2}$	pp1	a4
$\bar{b1}$	$\bar{b2}$	b1	$\bar{b2}$	pp1	a4
$\bar{b1}$	$\bar{b2}$	b2	$\bar{b2}$	pp1	a4
$\bar{b1}$	$\bar{b1}$	b1	$\bar{b1}$	pp1	a4
b4	b4	d2	d2	pp2	a1
b3	b4	d1	d2	pp2	a1
b3	b4	d2	d2	pp2	a1
b3	b3	d1	d2	pp2	a1
b3	b3	d2	d2	pp2	a1
b2	b2	d1	d2	pp2	a1
b2	b2	d2	d2	pp2	a1
b1	b1	d1	d2	pp2	a1
b1	b1	d2	d2	pp2	a1
b3	b3	d1	d1	pp2	a1
b2	b2	d1	d1	pp2	a1
b1	b1	d1	d1	pp2	a1
b4	$\bar{b4}$	d2	d2	pp2	a2
b3	$\bar{b4}$	d1	d2	pp2	a2
b3	$\bar{b4}$	d2	d2	pp2	a2

Continued on next page

$(s_1, s2)$	$(s_1, s2-\epsilon^h_2)$	$(s1-\epsilon^h_1, s2)$	$(s1-\epsilon^h_1, s2-\epsilon^h_2)$	PP	A
b2	$\bar{b4}$	d1	d2	pp2	a2
b2	$\bar{b4}$	d2	d2	pp2	a2
b2	$\bar{b3}$	d1	d2	pp2	a2
b2	$\bar{b3}$	d2	d2	pp2	a2
b1	$\bar{b3}$	d1	d2	pp2	a2
b1	$\bar{b3}$	d2	d2	pp2	a2
b2	$\bar{b2}$	d1	d2	pp2	a2
b2	$\bar{b2}$	d2	d2	pp2	a2
b1	$\bar{b2}$	d1	d2	pp2	a2
b1	$\bar{b2}$	d2	d2	pp2	a2
b1	$\bar{b1}$	d1	d2	pp2	a2
b1	$\bar{b1}$	d2	d2	pp2	a2
b1	$\bar{b1}$	d1	d1	pp2	a2
b4	$\bar{d2}$	b4	$\bar{d2}$	pp3	a3
b3	$\bar{d2}$	b3	$\bar{d2}$	pp3	a3
b3	$\bar{d2}$	b4	$\bar{d2}$	pp3	a3
b2	$\bar{d2}$	b2	$\bar{d2}$	pp3	a3
b2	$\bar{d2}$	b3	$\bar{d2}$	pp3	a3
b2	$\bar{d2}$	b4	$\bar{d2}$	pp3	a3
b1	$\bar{d2}$	b1	$\bar{d2}$	pp3	a3
b1	$\bar{d2}$	b2	$\bar{d2}$	pp3	a3
b1	$\bar{d2}$	b3	$\bar{d2}$	pp3	a3
b1	$\bar{d1}$	b1	$\bar{d2}$	pp3	a3
b1	$\bar{d1}$	b2	$\bar{d2}$	pp3	a3
b1	$\bar{d1}$	b3	$\bar{d2}$	pp3	a3
b1	$\bar{d1}$	b1	$\bar{d1}$	pp3	a3
$\bar{b4}$	$\bar{d2}$	b4	$\bar{d2}$	pp3	a4
$\bar{b3}$	$\bar{d2}$	b4	$\bar{d2}$	pp3	a4
$\bar{b2}$	$\bar{d2}$	b2	$\bar{d2}$	pp3	a4
$\bar{b2}$	$\bar{d2}$	b3	$\bar{d2}$	pp3	a4

Continued on next page

$(s_1, s2)$	$(s_1, s2 - \epsilon_2^h)$	$(s1 - \epsilon_1^h, s2)$	$(s1 - \epsilon_1^h, s2 - \epsilon_2^h)$	PP	A
$\bar{b2}$	$\bar{d2}$	b4	$\bar{d2}$	pp3	a4
$\bar{b1}$	$\bar{d2}$	b1	$\bar{d2}$	pp3	a4
$\bar{b1}$	$\bar{d2}$	b2	$\bar{d2}$	pp3	a4
$\bar{b1}$	$\bar{d2}$	b3	$\bar{d2}$	pp3	a4
$\bar{b3}$	$\bar{d1}$	b4	$\bar{d2}$	pp3	a4
$\bar{b2}$	$\bar{d1}$	b2	$\bar{d2}$	pp3	a4
$\bar{b2}$	$\bar{d1}$	b3	$\bar{d2}$	pp3	a4
$\bar{b2}$	$\bar{d1}$	b4	$\bar{d2}$	pp3	a4
$\bar{b1}$	$\bar{d1}$	b1	$\bar{d2}$	pp3	a4
$\bar{b1}$	$\bar{d1}$	b2	$\bar{d2}$	pp3	a4
$\bar{b1}$	$\bar{d1}$	b3	$\bar{d2}$	pp3	a4
$\bar{b1}$	$\bar{d1}$	b1	$\bar{d1}$	pp3	a4
$\bar{b4}$	$\bar{d2}$	$\bar{b4}$	$\bar{d2}$	pp3	a5
$\bar{b3}$	$\bar{d2}$	$\bar{b3}$	$\bar{d2}$	pp3	a5
$\bar{b3}$	$\bar{d2}$	$\bar{b4}$	$\bar{d2}$	pp3	a5
$\bar{b2}$	$\bar{d2}$	$\bar{b2}$	$\bar{d2}$	pp3	a5
$\bar{b1}$	$\bar{d2}$	$\bar{b1}$	$\bar{d2}$	pp3	a5
$\bar{b3}$	$\bar{d1}$	$\bar{b3}$	$\bar{d2}$	pp3	a5
$\bar{b3}$	$\bar{d1}$	$\bar{b4}$	$\bar{d2}$	pp3	a5
$\bar{b2}$	$\bar{d1}$	$\bar{b2}$	$\bar{d2}$	pp3	a5
$\bar{b1}$	$\bar{d1}$	$\bar{b1}$	$\bar{d2}$	pp3	a5
$\bar{b3}$	$\bar{d1}$	$\bar{b3}$	$\bar{d1}$	pp3	a5
$\bar{b2}$	$\bar{d1}$	$\bar{b2}$	$\bar{d1}$	pp3	a5
$\bar{b1}$	$\bar{d1}$	$\bar{b1}$	$\bar{d1}$	pp3	a5
b4	$\bar{d2}$	d2	d0	pp4	a2
b3	$\bar{d2}$	d1	d0	pp4	a2
b3	$\bar{d2}$	d2	d0	pp4	a2
b2	$\bar{d2}$	d1	d0	pp4	a2
b2	$\bar{d2}$	d2	d0	pp4	a2
b1	$\bar{d2}$	d1	d0	pp4	a2

Continued on next page

$(s_1, s2)$	$(s_1, s2-\epsilon_2^h)$	$(s1-\epsilon_1^h, s2)$	$(s1-\epsilon_1^h, s2-\epsilon_2^h)$	PP	A
b1	$\bar{d2}$	d2	d0	pp4	a2
b1	$\bar{d1}$	d1	d0	pp4	a2
b1	$\bar{d1}$	d2	d0	pp4	a2
$\bar{b4}$	$\bar{d2}$	d2	d0	pp4	
$\bar{b3}$	$\bar{d2}$	d2	d0	pp4	
$\bar{b2}$	$\bar{d2}$	d2	d0	pp4	
$\bar{b1}$	$\bar{d2}$	d1	d0	pp4	
$\bar{b1}$	$\bar{d2}$	d2	d0	pp4	
$\bar{b3}$	$\bar{d1}$	d2	d0	pp4	
$\bar{b2}$	$\bar{d1}$	d2	d0	pp4	
$\bar{b1}$	$\bar{d1}$	d1	d0	pp4	
$\bar{b1}$	$\bar{d1}$	d2	d0	pp4	

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Töre, Nurşen
Nationality: Turkish (TC)
Date and Place of Birth: 25.04.1979, Eskişehir
Marital Status: Single
Phone: 0 538 4170691

EDUCATION

Degree	Institution	Year of Graduation
M.S.	Columbıa Unıversıty	2001
B.S.	Bılkent Unıversıty	2000
High School	Eskişehir Anadolu Lisesi	1997

PUBLICATIONS

Töre, N., Savaşaneril, S. and Serin, Y.(2013), Value Of Stock Level information And impact On Manufacturers' Substitution Decisions, *Communications*, 54-61.

Töre, N. and Savaşaneril, S.(in progress), Policies For The Single-Echelon N-Product System.

Töre, N. and Savaşaneril, S.(in progress), Seller's Response To Customer-Driven Substitution .

CONFERENCE PRESENTATIONS

Töre, N., S. Savaşaneril, Y. Serin, İki-Seviyeli Bir Üretim Sisteminde Envanter Yönetimi ve Ürün İkamesi, YA/EM 30. Congress, Sabancı University, İstanbul, 2010.

PROJECT WORK

Yedek Parça Yönetim Sistemlerinde Envanter Havuzlama ve Bilgi Paylaşımının Fayda Analizi. TUBİTAK Projects , 1 June 2008 - 1 June 2010. Project No: 108M004.

Üreticiler için Ürün İkame Politikalarının Belirlenmesi ve İkame Esnekliğinin Getiri Analizi, TUBİTAK Projects 01 November 2012-01 November 2013. Project No: 112M431.0