

ABSTRACT

Title of dissertation: ESSAYS IN RETAIL OPERATIONS AND
 HUMANITARIAN LOGISTICS

Kiran V. Panchamgam
Ph.D. Candidate, 2011

Dissertation directed by: Prof. Bruce L. Golden
 France-Merrick Chair in Management Science
 Prof. Itir Z. Karaesmen Aydin
 Assistant Professor of Operations Management

This dissertation introduces and analyzes research problems related to Retail Operations and Humanitarian Logistics. In Retail Operations, the inventory that ends up as unsaleable at primary markets can be significant (up to 20% of the retail product). Thus retailers look for strategies like selling in secondary markets at a discounted price. In such a setting, the decisions of how much to order for a product of limited shelf life and when (if at all) to start selling the product in the secondary market become critical because these decisions not only affect the retailer's cost of procurement and sales revenues obtained from the product but also affect utilization of shelf space, product rollover and assortment decisions of the retailer. Apart from using secondary markets, retailers that sell seasonal products or products with sales horizons shorter than the typical production/procurement lead time also enter into contractual agreements with suppliers. These contracts are in place to share risks associated with unknown or uncertain demand for the product. Presence of such

contracts does affect a retailer's order quantity as well as the time to start selling in the secondary market. In our two essays on retail operations, we analyze a retailer's optimal order quantity and when he/she starts selling in the secondary market. We refer to the former as the 'ordering decision' and the latter as the 'timing decision.' These two decisions are studied first without risk sharing contracts in Essay 1, and then in the presence of contracts in Essay 2.

In Essay 1, we build a two-stage model with demand uncertainty. The ordering decision is made in the first stage considering cost of procurement and expected sales revenue. The timing decision is made in the second stage and is conditional on the order quantity determined in the first stage. We introduce a new class of aggregate demand model for this model. We study the structural properties of the retailer's timing and ordering problem and identify optimality conditions for the timing decision. Finally, we complement our analytical results with computational experiments and show how retailer's optimal decisions change when problem parameters are varied.

In Essay 2, we extend the work in first essay to include the contracts between the retailer and a supplier. In this essay, we introduce a time-based Poisson demand model. We define three different types of contracts and investigate the effect of each of these contracts on the retailer's ordering and timing decisions. We investigate how the analytical structure of the retailer's decision changes in the presence of these contracts. For a given order quantity, we show that the timing decision depends on the type of contract. Our analytical results on the timing decision are complemented with computational experiments where we investigate the impact of contract type

on the optimal order quantity of the retailer.

In Humanitarian Logistics, non-profit organizations receive several-billion-dollars-worth of donations every year but lack a sophisticated system to handle their complex logistics operations; the absence of expertly-designed systems is one of the significant reasons why there has been a weak link in the distribution of relief aid. The distribution of relief aid is a complex problem as the goal is humanitarian yet at the same time, due to limited resources, the operations have to be efficient. In the two essays on humanitarian logistics, we study the distribution of aid using homogeneous fleet, with and without capacity restrictions.

In Essay 3, we discuss routing for relief operations using one vehicle without capacity restrictions. Contrary to the existing vehicle routing models, the key property of our routing models is that the nodes have priorities along with humanitarian needs. We formulate this model with d-Relaxed Priority rule that captures distance and response time. We formulate routing models with strict and relaxed forms of priority restrictions as Mixed Integer Programs (MIP). We derive bounds for this problem and show that this bound is attained in limiting condition for a worst-case example. Finally, we evaluate the optimal solutions on test problems for response time and distance and show that our vehicle routing model with priorities captures the trade-off between distance and response time unlike existing Vehicle Routing Problem (VRP) models without priorities.

In Essay 4, we extend the problem dealt in third essay to consider fleet consisting of multiple vehicles (homogeneous) with capacity and route length restrictions. First, we show that the humanitarian aspect imposes additional challenges

and develop routing models that capture performance metrics like fill rate, distance traversed, response time and number of victims satisfied. Proposed routing models are formulated as Mixed Integer Programs and are solved to optimality for small test problems. We conduct computational experiment and show that our models perform well on these performance metrics.

ESSAYS IN RETAIL OPERATIONS AND HUMANITARIAN LOGISTICS

by

Kiran Venkata Panchamgam

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2011

Advisory Committee:

Dr. Bruce L. Golden, Co-Chair

France-Merrick Chair in Management Science

Dr. Itir Z. Karaesmen Aydin, Co-Chair

Assistant Professor of Operations Management

Dr. Michael O. Ball

Orkand Corporation Professor of Management Science

Dr. Wedad J. Elmaghraby

Associate Professor of Management Science and Operations Management

Dr. Steven A. Gabriel, Dean's Representative

Associate Professor, Department of Civil and Environmental Engineering

© Copyright by
Kiran Venkata Panchamgam
2011

DEDICATION

To my generating function

Mom and Dad (Mani and Srinivasa Murthy)

& other moments

Brothers (Diwakar and Sashidhar)

Acknowledgments

This dissertation would not have reached this stage if not for efforts from many individuals. First and foremost, I would like to thank my advisors, Dr. Itir Karaesmen Aydin and Dr. Bruce Golden for their guidance and support in both academic and non-academic aspects during my doctoral studies. It has been a great pleasure working and learning from these exemplary individuals over the past few years. They have been steady source of inspiration and I can only hope to emulate their professional and personal achievements. I would also like to thank my committee members – Dr. Michael Ball, Dr. Wedad Elmaghraby, and Dr. Steven Gabriel for not only providing insightful feedback serving on the dissertation committee but also for laying the foundations through their classes.

I would like to thank the Department of Decision, Operations and Information Technologies, University of Maryland, College Park, for providing the financial support during my doctoral studies. I would like to thank Graduate School at University of Maryland for providing me with Ann G. Wylie Dissertation Fellowship. I wish to thank my PhD colleagues, faculty, and staff of Robert H. Smith School of Business who over the years have helped me in many ways.

I had many an interesting discussion with my classmates - William Mennell, Scott Nestler and Matthew Reindorp. My nocturnal working experience has been invigorated with presence of Dina, Animesh and ‘Coffee’! I would also like to thank all my colleagues for their support during my doctoral studies. Most of them have become good friends for life. Work would not have been enriching if I did not have

a chance to pursue my other passion - Tennis. I would like to acknowledge Mak, Panda, Pooja, Ravi, Rahul, Rodrigo and Manish for our ‘countably many’ tennis sessions.

I would like to express my gratitude to all my friends. Rapa for interesting math discussions, Mak for being my eternal sports partner, Gayatri for being ‘Tiffins’ mate, Reddy for the ‘Old Rag’, Raju for ‘political’ debates, Sasi for ‘traffic’ ideas, Harini for great desserts, Joe for ‘Delta’ training, and Prachi for being ‘very nice’ to me. Finally, I would like to acknowledge the support and love of my family for sharing and working towards my passion. I would like to express my deepest gratitude to my mom for her guidance, support, love and patience. Without her support and encouragement throughout my life, I would not have achieved this important milestone in my career. Regrettably, it is impossible to list all those who deserve credit and I apologize to those that I might have inadvertently left out.

Table of Contents

List of Tables	viii
List of Figures	xi
1 Introduction	1
2 Essay 1: When to Start Selling in a Secondary Market	7
2.1 Introduction	7
2.2 Problem Definition	10
2.3 Literature Review	12
2.4 Retailer's Expected Profit Function and Demand Models	20
2.5 Analysis of Second Stage: Retailer's Timing Decision	23
2.6 Analysis of First Stage: Retailer's Ordering Decision	33
2.7 Computational Experiments	34
2.7.1 Design of the experiment	35
2.7.2 The Optimal Ordering and Timing Decision	35
2.7.3 Effect of prices and state of the market	38
2.8 Conclusions	40
3 Essay 2: Secondary Markets and Supplier-Retailer Contracts	43
3.1 Introduction	43
3.2 Literature Review	45
3.3 Problem Definition	48
3.4 Retailer's Timing Decisions	54
3.4.1 Wholesale-Price (WP) Contract	54
3.4.2 Buyback (BB) Contract	57
3.4.3 Markdown-Money (MM) Contract	59
3.5 Comparison of Retailer's Timing Decisions	62
3.6 Retailer's Ordering and Supplier's Wholesale Pricing Decisions	64
3.7 Computational Experiments	67
3.7.1 Setup and Design	67
3.7.2 Summary of Results	69
3.7.3 Optimal Decisions under WP Contract	70
3.7.4 Optimal Decisions under BB Contract	75
3.7.5 Optimal Decisions under MM Contract	81
3.7.6 Discussion of Optimal Decisions across Contracts	87
3.7.7 Preferences of Supplier and Retailer for High Risk Markets	90
3.8 Conclusions	91
4 Essay 3: Humanitarian Vehicle Routing Problem: Uncapacitated Case	100
4.1 Introduction	100
4.2 Setup	104
4.2.1 Definition of Node Priorities	105

4.2.2	d-Relaxed Priority Rule	107
4.3	Literature Review	110
4.4	Single Uncapacitated VRP with Priorities (u-HVRP)	118
4.4.1	MIP Formulations	119
4.4.2	Bounds and Worst-case Examples for u-HVRP	122
4.4.2.1	Worst-case Example (Tree) when $d = 0$	125
4.4.2.2	Worst-case Example (Tree) where $d > 0$	125
4.4.2.3	Worst-case Example (Cluster) when $d = 0$	126
4.4.2.4	Worst-case Example (Cluster) where $d > 0$	129
4.4.3	Relation of u-HVRP to ATSP	130
4.5	Discussion of Computational Studies	131
4.6	Conclusions	138
4.7	Summary of Insights	139
5	Essay 4: Humanitarian Vehicle Routing Problem: Capacitated Case	142
5.1	Introduction	142
5.2	Setup	144
5.3	Literature Review	147
5.4	Routing Models	152
5.4.1	Single Capacitated VRP with Priorities (1-HVRP)	152
5.4.1.1	Two-stage Optimization	153
5.4.1.2	Weighted Objective	157
5.4.2	Multiple Capacitated VRP with Priorities (m-HVRP)	159
5.5	Discussion of Computational Studies	164
5.5.1	1-HVRP	164
5.5.2	m-HVRP	169
5.6	Conclusions	174
5.7	Summary of Insights	178
6	Conclusions and Future Work	183
A	Appendix for Essay 1	189
A.1	Proof for Componentwise Concavity of the Sample Path Revenue	189
A.2	Proof of Joint Concavity of the Sample Path Revenue Function	191
B	Appendix for Essay 2	197
B.1	Proofs for Retailer's Timing Problem with WP Contract	197
B.1.1	Proof for Proposition 1	197
B.1.2	Lemmas for Retailer's Timing Problem with Poisson Demand Model	199
B.2	Proofs for Retailer's Timing Problem with BB Contract	203
B.2.1	Proof for Proposition 2	203
B.3	Proofs for Retailer's Timing Problem with MM Contract	204
B.3.1	Proof for Proposition 3	204
B.4	Description of Algorithm for Computational Experiments	205

C	Appendix for Essay 3	207
C.1	Notation	207
C.2	Formulation MinDist(d) for u-HVRP	207
C.2.1	Proof for Proposition 1	208
C.2.2	Proof for Proposition 2	209
C.2.3	Valid Inequalities	210
C.3	Formulation MinMax and MinSum for u-HVRP	210
D	Appendix for Essay 4	212
D.1	Notation for all models	212
D.2	Appendix for 1-HVRP	212
D.2.1	Formulation 1-HVRP-MaxDemand	213
D.2.2	Formulation 1-HVRP-MinDist-i	215
D.2.3	Formulation 1-HVRP-MinDist-a	215
D.2.4	Formulation 1-HVRP-WeightedObj	216
D.3	Appendix for m-HVRP	217
D.3.1	Formulation for Stage I with Local Timing Rule	217
D.3.2	Formulations for Stage II with Local Timing Rule	219
D.3.3	Formulation for Stage I with Global Timing Rule	220
D.3.4	Formulations for Stage II with Global Timing Rule	222
E	Networks used in Essays 3 and 4	224

List of Tables

2.1	Probability distribution of the demand model	35
2.2	Optimal Decisions: Q^* and β_s^*	36
3.1	Price Conditions for Rebate Situations under MM contract	59
3.2	Evaluation of Revenue Function for MM contract	60
3.3	WP Contract - Results for different Price Paths with $(\theta_l, \theta_h) = (0.1, 0.9)$	72
3.4	WP Contract - Results for different Price Paths with $(\theta_l, \theta_h) = (0.7, 0.3)$	72
3.5	WP Contract - Results for different WTP distribution parameters . .	72
3.6	WP Contract - Results under different Market State Probabilities for (100, 60)	74
3.7	WP Contract - Results under different market state probabilities for (120, 80)	75
3.8	BB Contract - Results for different Buyback Credits with $(p_1, p_2) =$ (100, 60)	76
3.9	BB Contract - Results for different Price Paths with $(\theta_l, \theta_h) = (0.1, 0.9)$	77
3.10	BB Contract - Results for different Price Paths with $(\theta_l, \theta_h) = (0.7, 0.3)$	78
3.11	BB Contract - Results for different WTP distribution parameters . .	79
3.12	BB Contract - Results under different Market State Probabilities for (100, 60)	80
3.13	BB Contract - Results under different Market State Probabilities for (120, 80)	81
3.14	MM Contract - Results for different Profit Margins with $(p_1, p_2) =$ (100, 60)	81
3.15	MM Contract - Results for different Price Paths with $(\theta_l, \theta_h) = (0.1, 0.9)$	84
3.16	MM Contract - Results for different Price Paths with $(\theta_l, \theta_h) = (0.7, 0.3)$	85
3.17	MM Contract - Results for different WTP distribution parameters . .	86

3.18	MM Contract - Results under different Market State Probabilities for (100, 60)	86
3.19	MM Contract - Results under different Market State Probabilities for (120, 80)	87
3.20	Comparison of Supply Chain Profits under different Contracts with (θ_l, θ_h) = (0.1, 0.9)	89
4.1	Optimal Results for u-HVRP (with inequalities) 134	
4.2	Comparison of distribution of service times for u-HRP 137	
5.1	Network Characteristics for 1-HVRP and m-HVRP	165
5.2	Optimal Results for 1-HVRP using Two-stage Optimization on a 21 node network 166	
5.3	Optimal Results for 1-HVRP using Two-stage Optimization on a 30 node network 168	
5.4	Optimal Results for 1-HVRP using Weighted Objective Approach on a 21 node network 169	
5.5	Optimal Results for m-HVRP using Two-stage Optimization and Lo- cal Timing Rule on 21 nodes network 171	
5.6	Optimal Results for m-HVRP using Two-stage Optimization and Lo- cal Timing Rule on 30 nodes network 172	
5.7	Optimal Results for m-HVRP using Weighted Objective Approach and Local Timing Rule on 21 nodes network 172	
5.8	Optimal Results for m-HVRP using Global Timing Rule on 21 nodes network	174

5.9	Service Times for m-HVRP using Local & Global Timing Rules (after Stage II)	175
A.1	Feasible region for first order derivatives (in sample-path) for the Linear Demand Model	190
A.2	First order derivative wrt β (in sample-path) for the Linear Demand Model	191
A.3	Proof of joint concavity in sample-path for the Linear Demand Model	192
B.1	Evaluation of the difference functions $V(j+1, k) - V(j, k)$ and $V(j, k+1) - V(j, k)$	198
B.2	Difference Function for BB contract	203

List of Figures

2.1	Timeline of Retail Operations	13
2.2	Retailer's Timing Decisions	14
2.3	Illustration of the joint concavity of the sample-path revenue function	27
2.4	Lack of Super or Submodularity of Expected Revenue Function . . .	31
2.5	Effect of inventory on Optimal Timing	32
2.6	Effect of higher discounts on Optimal Expected Profits	39
2.7	Effect of higher discounts on β_{low}^* and β_{high}^*	40
2.8	Effect of higher discounts on Q^*	41
3.1	Timeline of Retail Operations with Retailer-Supplier Contracts	50
3.2	Optimal Order Quantity vs. Wholesale Price for WP contract	92
3.3	Supplier and Optimal Retailer Expected Profit vs. Wholesale Price for WP contract	92
3.4	Optimal Order Quantity vs. Wholesale Price for BB contract	93
3.5	Supplier and Optimal Retailer Expected Profit vs. Wholesale Price for BB contract	93
3.6	Optimal Order Quantity vs. Wholesale Price for MM contract	94
3.7	Supplier and Optimal Retailer Expected Profit vs. Wholesale Price for MM contract	94
3.8	Optimal Low State Transfer Time for Different Contracts	95
3.9	Optimal High State Transfer Time for Different Contracts	95
3.10	Optimal Order Quantity for $(p_1, p_2) = (100, 80)$ for Different Contracts	96
3.11	Optimal Order Quantity for $(p_1, p_2) = (100, 60)$ for Different Contracts	96
3.12	Optimal Order Quantity for $(p_1, p_2) = (100, 40)$ for Different Contracts	97
3.13	Optimal Order Quantity for $(p_1, p_2) = (100, 20)$ for Different Contracts	97

3.14	Optimal Retailer Expected Profits for Different Contracts	98
3.15	Optimal Supplier Expected Profits for Different Contracts	98
3.16	Supplier and Retailer Expected Profits for Different Low State Prob- abilities	99
4.1	Illustration of Urgencies and Priorities for Disasters	106
4.2	Illustration of the \mathbf{d} -Relaxed Priority Rule	109
4.3	Proof for bound on optimal u-HVRP tour and Worst-case Example .	124
4.4	Worst-case Example for HTSP with $\mathbf{d} = 0$ (any P) and $\mathbf{d} = 1$ ($P = 3$)	128
4.5	Transformation of u-HVRP with $\mathbf{d}=0$ into ATSP	131
4.6	Transformation of u-HVRP into ATSP	132
4.7	Optimal results for different u-HVRP formulations on 11 nodes network	136
5.1	Illustration of Order of Demand Fulfillment Rule	146
5.2	Illustration of Stage II Methods	158
5.3	Illustration of Global vs. Local Timing Rules for m-HVRP	163
5.4	Optimal results of m-HVRP on 21 nodes using the Local Timing Rule	176
5.5	Optimal results of m-HVRP on 21 nodes using the Global Timing Rule	177
A.1	Sample path revenue as a function of β	189
C.1	Proof for \mathbf{d} -Relaxed Priority Rule Constraint	209
E.1	30-node network used in Chapters 4 and 5	224
E.2	40-node network used in Chapter 4	225

Chapter 1

Introduction

The field of operations is claimed to be as old as civilization (Singhal et al. (2007)) but has found its niche during the Industrial Revolution early in the 20th century. The problems and ideas came from manufacturing and military in those days. When Elwood Buffa, a pioneer, coined the term “operations management” in 1950s, he referred to both manufacturing and non-manufacturing activities. In his 1961 textbook, Buffa discussed the increasing role of services in the economy and provided many examples where production-related concepts equally apply to services (Buffa (1961)). Despite Buffa’s vision, majority of the research papers that appear in academic outlets remained focused on manufacturing-related problems for a long time: A survey by Pannirselvam et al. (1999) shows that only a handful of articles on services were published in operations management journals during 1990s. The interest on services have been growing, with several journals publishing special issues on topics related to service industries in the recent years (e.g. Management Science journal published a Special Issue on Call Center Management in 2008, Production and Operations Management journal announced two special issues, one on Financial Services and the other on Retail Operations in the last two years). Two of the emerging subfields within services, retail operations and humanitarian logistics, are studied in this dissertation.

Retail industry is comprised of individuals and companies that sell finished products to end users. Many different products are sold at retail, for instance, coffee at a coffee shop such as Starbucks, fashion apparel at a department store such as Nordstrom, electronics at an online retailer such as Amazon.com. The total sales for the U.S. Retail Industry in 2009 was \$4,131 billion. This total includes \$690.5 billion in automotive sales, \$575.8 billion in food and beverage, and \$208.6 billion in clothing and accessories.¹ Companies use various strategies to sell the goods to end users at retail. For instance, some products are sold by infomercials, some are sold in stores, some through catalogs and some on the internet. Despite the differences in these sales strategies, retailers face a common set of questions in managing their operations: Which products should the retailer sell? What should be the stock levels of these products? How frequently should the inventory be replenished? How should the goods be procured from the vendors? What should be the prices of the products at any point in time?

Finding the ‘right’ answers to these questions is difficult given the level of uncertainty regarding consumer demand. A retailer has to make many logistical decisions including procurement, transportation, and inventory-related ones, long before he/she knows how many customers will purchase the products, and when and where they will make a purchase. For retailers that sell goods that are seasonal, perishable or face obsolescence, the answers to the above-mentioned questions are even more critical as the time window to sell the goods at a profitable level is typically shorter and the opportunity cost of not selling them before their effective

¹http://www.nrf.com/modules.php?name=Pages&sp_id=1237

sales season ends is high. Furthermore, a retailer that sells an assortment of seasonal goods has to manage product rollovers effectively from one season to the other.

Another complication in the retail industry arises from the heterogeneity of the consumers who are willing to pay different prices for a given product and who differ in their urgency in making a purchase (e.g., customers who want to purchase an apparel early in the season vs. the ones that purchase the product at a discount at the end of the season). To manage demand of such heterogeneous customers effectively, some retailers take advantage of multiple distribution channels. One common arrangement is selling products in primary and secondary markets. A retailer can offer an assortment of products in its own stores with a particular sales and pricing strategy and can offer some (if not all) of these products to a different customer segment using another channel/strategy. An example of an alternative or secondary channel is the internet store of the retailer where the goods can be marked down and offered for a longer period. Another example is the outlet store of a retailer. Clearly, selling the products at a discount in a secondary market, such as the internet store or the outlet, can cannibalize the sales in the primary market. To prevent this, retailers typically do not offer the same product at their primary and secondary markets simultaneously and offer the products in the secondary markets with a delay.

The first essay in this dissertation addresses questions for a retailer that has access to both primary and secondary markets. Should the retailer selling a seasonal product offer it only in the primary market? If not, then what is the best time to start offering the product in the secondary market? The answers to these questions

are obtained by developing an analytical model that focuses on the logistical decisions made by the retailer, specifically, the amount of goods to have in stock at the beginning of the sales season and the time to switch the sales from the primary to the secondary market. The second essay investigates these decisions further by considering the retailer-vendor relations. Retailers enter into agreements with vendors to share various supply chain costs. The agreements are in the form of contracts. In this essay, we study the effects of three types contracts on the decisions of a retailer who makes use of primary and/or secondary markets.

The second research topic, humanitarian logistics, has received wide-spread recognition in academia and industry especially after natural disasters like Hurricane Katrina in 2005 and tsunami in South-East Asia in 2004. The complex operational challenges in humanitarian logistics came to the forefront due to these natural disasters. The entire world mobilized to donate \$13 billion in response to the tsunami (e.g., Thomas & Fritz (2006)). Yet, according to Doctors without Borders, "... we have already received as much money as we can spend... What is needed are supply-managers without borders: people to sort goods, identify priorities, track deliveries and direct the traffic of a relief effort in full gear" (e.g., The Economist [Jan 2005]). These challenges are not limited to natural disasters like the South-East Asia Tsunami in 2004 or Hurricane Katrina in 2005 – relief agencies encounter similar and additional challenges in military environments like disease outbreaks, civil conflicts, war-zones, and terrorist attacks.

Compared to well-studied business logistics, humanitarian logistics presents unanswered challenges in terms of movement of relief goods like food, bedding and

shelter, medical care, and clothes. The business supply chain and humanitarian relief chain differ in aspects like revenue sources, goals, stakeholders, performance measurements, demand characteristics, and customer characteristics. The goal of a business supply chain is to maximize profit. For humanitarian relief chain, the goal is to maximize timely delivery in order to minimize the loss of life. A typical humanitarian relief chain consists of a primary supply hub such as a port of entry, a central warehouse, and local distribution centers. Finally, the relief goods are distributed to people in need (or demand points) from the distribution center. This last stage is commonly known as the “Last Mile Distribution” in humanitarian logistics, and this is the main focus of our third and fourth essays.

Our purpose is to consider the urgency of locations along with other humanitarian needs and design delivery routes that cater to performance metrics relevant for humanitarian relief operations. Traditional or non-humanitarian routing models do not capture the humanitarian issues and, thus, perform at a poor level on one or all of metrics such as distance, latest response time, fill rate (or percentage demand satisfied), and number of customers serviced. The combination of urgencies for locations in a vehicle routing problem poses interesting research questions in different relief situations: How do we design optimization models to capture relevant performance metrics discussed above? How much is the route altered when the urgency restrictions are strict? Is there a structured way to model relaxed forms of the urgency restrictions? In the third essay, we design delivery routes for the distribution of relief good with a single vehicle that are efficient *as well as* that meet the ultimate humanitarian goal of mitigating human loss in a timely manner.

In the fourth essay, we extend our model discussed in third essay to consider the distribution system with single and multiple vehicles constrained by capacity and route-length restrictions.

This dissertation is organized as follows. In Chapter 2, we consider a retailer facing primary and secondary markets and analyze the retailer's logistical decisions. In Chapter 3, we discuss the implications of supplier-retailer contracts. In Chapter 4, we consider the vehicle routing problem for humanitarian relief operations with a single vehicle without capacity restrictions. In Chapter 5, we extend the vehicle routing problem for relief operations with single and multiple vehicles having capacity and route-length restrictions. In Chapter 6, we summarize the contributions of our research and discuss future research topics in Retail Operations and Humanitarian Logistics.

Chapter 2

Essay 1: When to Start Selling in a Secondary Market

2.1. Introduction

Mismatch of demand and supply is commonplace at retailers due to the coupled effect of demand uncertainty and lead times in the supply chain. According to Tibben-Lembke (2004), at the primary market of a retailer, anywhere from 5% to 20% of the products may end up as unsaleable. Overstocking, inadequate pricing, damage, and customer returns all contribute to these percentages. Retailers look to offer products that are unsaleable in primary markets to appropriate customers at apt time and price by marking them down, refurbishing the goods, relocating the inventory, or changing the distribution channel. In all these cases, the good enters the “secondary market”. Secondary markets thus provide an opportunity for the retailers to increase sales, decrease inventory levels and costs, and to increase profits.

Retailers selling seasonal products such as fashion apparel make product assortment (merchandising) and ordering decisions well ahead of the selling season with limited information and inaccurate forecasts of demand. They have better demand forecasts at the start of the sales season, as opposed to at the time goods are ordered from a vendor, as they gather market indicators until the sales season starts. Change in demand forecasts, unfortunately, does not give the retailers an

opportunity to change their order quantities when they have long supply lead times. Therefore they rely on other tactics to manage demand and increase profits: They can match demand with the quantity on hand by changing the prices of the goods or reallocating them across different channels during the sales season. For instance, a product that is not selling as fast at “primary market” (e.g., the department store), can be offered at a discounted price in a “secondary market” (e.g., outlet store) to increase the sales and profits. Primary and secondary markets typically appeal to different customer segments. The former can have customers that are willing to pay premium prices to purchase a product early in the season, while the customers for the latter are bargain hunters that do not have “urgent” needs for a product.

As a concrete example, consider Saks Fifth Avenue and Nordstrom, each of which has its retail stores but also sells products through its outlet stores, OFF 5th and Nordstrom Rack, respectively. Some of the products that do not sell well in the primary market (e.g., Nordstrom) are moved to the secondary market (e.g., Nordstrom Rack) and offered at a 30-70% discount (Sanders (2010)). Department stores are not the only ones that make use of outlets as secondary markets: Apparel brands that manufacture and sell products in the company’s own stores, such as Polo Ralph Lauren and Calvin Klein have primary stores and outlet stores. In their detailed study of outlets, Coughlan & Soberman (2004) mention that Polo Ralph Lauren is one brand that offers its products at its own stores as the primary market and offers the same products with a time delay at its outlets and at a discount of 29%. Similarly, Nautica offers products at its own retail stores first and later in its outlets at a discount of 47%.

Yet, another example of secondary markets in fashion apparel is the existence of third-party discount retailers such as TJ Maxx. TJ Maxx buys merchandise that have been returned from department stores (e.g., Nordstrom returning excess inventory to Ralph Lauren, which in return sells it to TJ Maxx) and sells it at a discount at its own stores (Wahba (2010)). Ongoing economic recession has boosted the sales in the secondary markets, including discount retailers and outlet stores, while the department store sales have been declining (Birchall (2010)). Retail analysts point out (Birchall (2010)) that “vendor who does not sell to [the secondary markets] runs the risk of not being able to move excess inventory.”

In this chapter, we focus on a decision maker that has the option of selling products through primary and/or secondary markets that is common for products with limited shelf-life such as high-end fashion apparel. The decision maker retains the ownership of goods in both the primary and the secondary markets; this is the case for the department stores and their outlets (i.e., Nordstrom and Nordstrom Rack). We refer to the decision maker as the retailer henceforth. Specifically, we study how much the retailer should order prior to the sales season and when (if at all) he/she should offer the product in the secondary market. If the product is offered too early in the secondary market, then the retailer loses out on the revenue from customers paying higher prices at the primary market. If the product is offered very late in the sales horizon, when the product is not selling well in primary market, then the retailer loses out on two fronts: First, the valuable shelf-space in primary market is held up by goods with low profit potential. Second, the retailer loses out on potential sales in the secondary market eventually forcing retailer to, possibly,

salvage excess inventory at a loss.

The remainder of this chapter is organized as follows. We briefly define the problem in Section 2.2 and discuss the literature pertinent to secondary markets in Section 2.3. We formulate the retailer's decision problem in Section 2.4. We analyze the retailer's timing and ordering decisions in Sections 2.5 and 2.6. We discuss computational experiments in Section 2.7. Finally, we summarize our findings and discuss the contributions of our research in Section 2.8.

2.2. Problem Definition

Consider a single retailer selling a single product over a fixed sales horizon $[0, T]$. The retailer has to place an order from a vendor, long before the selling season starts. After giving the order but before the season, the retailer gathers market indicators. This allows him/her to have better information about the state of the market, hence the demand, at the start of the season. Once the goods are received, the retailer decides how best to use two different channels, primary and secondary, with his/her strategy being that the product will not be offered in these two markets simultaneously.

At the time of ordering, the retailer knows the demand/market is uncertain and predicts it will be in one of a finite number of states, denoted s . For simplicity, we assume only two states in this chapter. The knowledge of the state does not provide perfect information on the demand, but provides perfect information about the probability distribution of demand in the primary and secondary markets. For

each state and market, the demand is a random variable, characterized by a known probability distribution. Without loss of generality, we refer to the states as *low* and *high*. At the time of ordering, the retailer knows that the market will be in a low state, denoted l , with probability θ_l and in a high state, denoted h , with probability θ_h where $\theta_l + \theta_h = 1$. The state of the system is known to the retailer only once the season begins, i.e., the parameters of the underlying random demand process/distribution are known only at the start of the selling season.

One can think of this model as follows: The retailer orders merchandise to be sold during the Holiday Season in the US at the end of the summer. At that time, the retailer thinks the economic recession will continue through out the end of the year or a recovery will start prior to holiday season, with probabilities θ_l and θ_h , respectively. For each scenario, the retailer predicts a different demand distribution. At the start of the holiday season, the retailer realizes which state of the recession (continuing vs. recovery) best characterizes the demand.

We formulate the retailer's problem as a two-stage optimization problem where the first-stage decision is the order quantity Q , determined prior to revelation of the market state and before the selling season. The second-stage decision is the length of time to offer the product in the primary market, after which the goods will be offered only in the secondary market. Once the retailer is at the beginning of the selling season, he/she¹ finds himself in a particular state of demand market $s \in \{l, h\}$, with Q units of inventory on hand. At that point, the retailer chooses β , $0 \leq \beta \leq 1$

¹Hereafter, the decision makers (retailer in this and next chapter and the supplier in the next chapter) are referred as 'he'.

which is the fraction of the entire sales season during which the goods are offered at the primary market. By design, $\beta = 1$ and $\beta = 0$ indicate the goods are offered only in the primary and only in the secondary market, respectively. The time intervals during which the goods are sold at the primary and secondary markets is $[0, \beta T)$ and $[\beta T, T]$. Once the selling season ends at time T , any unsold goods are salvaged. During the sales season, the retailer is committed to predetermined unit prices of p_1 (e.g., manufacturer's suggested retail price, MSRP) and p_2 (e.g., markdown) in the primary and secondary market, respectively. We have $p_1 \geq p_2 \geq 0$.

The sequence of decision making for the retailer is illustrated in Figure 2.1. Notice that, depending on the state of the market, the retailer's optimal timing decision may be different. We use the notation β_s^* to denote the optimal length of time the goods are sold at the primary market when the state of the market is $s = l, h$. The second-stage decision follows a temporal model as illustrated in Figure 2.2. Before we formulate the retailer's expected profit function and analyze his ordering and timing decisions, we provide a review of the relevant literature in the next section.

2.3. Literature Review

Considering the nature of the retailer's problem introduced above, there are several research fields including Retail Operations, Secondary Markets, Dynamic Pricing, and Timing/Switching that are relevant.

Retail Operations. Retail operations consists of complex decisions at dif-

Retailer's Ordering Decision

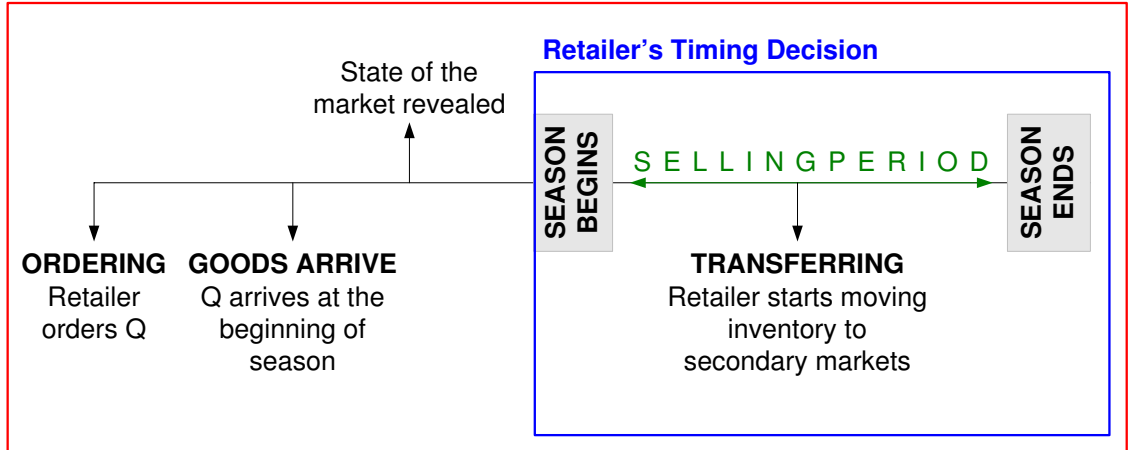


Figure 2.1: Timeline of Retail Operations

ferent stages (for example, ordering, assortment, forecasting, and pricing) in a retail environment. First we look at the relevance of our model for assortment planning problem (we look at the ordering and pricing problems later in this section). Retail assortment planning is a problem that has been well studied from an optimization perspective; Kök et al. (2009) provide a survey of research on optimal retail assortment decisions. Retailer needs to plan on the assortment to procure based on past and forecasted sales such that the gross margin is maximized. But, the retailer is constrained by the limited budget, limited shelf space for displaying products and desire to offer brand choice for a product to the customer. Research surveyed by Kök et al. (2009) focus only on a single market, whereas our focus is on primary and secondary markets. Grewal & Levy (2007) and Grewal & Levy (2009) conduct a survey of research pertaining to retail published in Journal of Retailing for years 2002-2007. They categorize research on retail into ten broad categories as price, promotion, brand/product, service, loyalty, consumer behavior, channel, organiza-

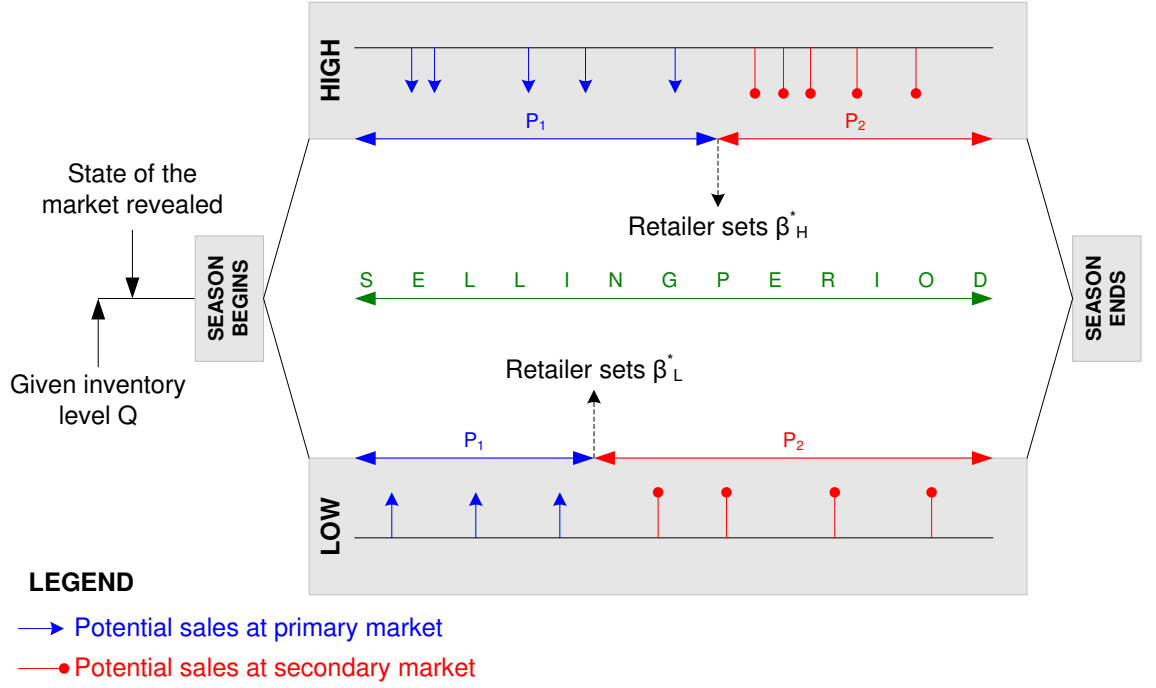


Figure 2.2: Retailer's Timing Decisions

tional, Internet, and others. Notice that the secondary market in our context can be interpreted as an alternative channel for selling the product. These two surveys discuss the research under broad category of 'channels' while focusing on topics like conflict, trust and risk, the need for standardizing merchandise, pricing and promotion across the multiple channels. A comprehensive survey of literature on channel design and coordination has been discussed in Sa Vinhas et al. (2010). In this paper, the authors point out that the complexity of the channels has increased tremendously and there is a need to analyze the channels together so as not to *compete* with one another. Thus, timing decision for moving to the product to the alternative channels becomes critical in this context.

A recent industry report on retail operations, SAS (2007), discusses the current practical challenges for retailers, including fashion merchandise planning life cycle:

During the preseason, the retailer develops a merchandise plan, assortment strategy, assortment plan and space plan, among other things. During the season, allocation, market-level price changes, market-level promotional markdown, adjustments and exit strategies are of interest. Our problem, including ordering and timing decisions is inline with the timeline of activities outlined in this report. The adjustment and exit strategies during season are analogous to our timing decision that ends the sales in the primary market. The report, on the other hand, is a white paper and does not go beyond discussing the practical problems.

Secondary Markets. While the literature that mentions secondary markets is rich, there is no standard definition, hence no uniform model, for this phenomenon. This is not surprising: The definition of secondary markets in Tibben-Lembke (2004) includes markdowns, use of outlets, third parties, or auctions to sell or dispose of excess inventory, managing customer returns and selling original vs. used products (as in closed-loop supply chains). For instance, Lee & Whang (2002) discuss the dynamics of n re-sellers in a secondary market for high-tech products. The motivation for their research is based on an internet-based secondary market TradingHubs.com, launched by Hewlett-Packard Company in 1999. The authors consider a simple newsvendor model with combined demand over two time periods. In this paper, the re-sellers trade in excessive supply in a “secondary market” at the end of first period. Ghose et al. (2005) refer to an exchange for used merchandise like CDs, DVDs, videos, and books as a secondary market. They consider a two-period problem with new goods being sold in both first and second periods but used goods, with quality degradation being sold only in second period. Nocke & Peitz (2003)

investigate whether possibility of future trade in a secondary market for durable goods (e.g., collector's items, coins, stamps) affects consumer behavior in a primary market. In contrast to our work, these papers do not address the timing aspect of transferring to secondary markets which is important for many brands, especially in high-end fashion apparel (e.g., Coughlan & Soberman (2004)).

Dynamic Pricing. In our problem, the retailer's decision of transferring the goods from the primary to the secondary market corresponds to the time of markdown for the product. In that respect, the second-stage decision can also be viewed as a pricing decision that is made at the start of the sales season. Consequently, research involving pricing decisions is relevant to our problem. In the pricing literature, there are numerous papers that determine the optimal markdown pricing policy for a retailer; we refer the reader to Elmaghraby & Keskinocak (2003), Bitran & Caldentey (2003), and Talluri & van Ryzin (2004) for a review of that literature. A number of researchers considered and studied dynamic models to determine the optimal prices and the optimal price path for a single product in a finite time horizon. In a way, our problem is a simplified dynamic pricing problem, where the price path is fixed and our decision is dynamic to the market state *before* to the sales season starts but not to the amount of sales, as is typically the case in other research papers. Others that consider fixed price paths are Feng & Gallego (1995), Bitran & Mondschein (1997), Feng & Gallego (2000) and Petruzzi & Monahan (2003). Feng & Gallego (1995) propose a dynamic programming formulation for determining the optimal starting time for end-of-season sales or beginning-of-promotional fares for a single product with finite horizon. Feng & Gallego (2000) extend this work to

consider a demand that is modeled by a Markovian, time-dependent rate. These two papers assume on-hand inventory at the beginning of the season is predetermined, i.e., no discussion of ordering decisions is presented. Yang et al. (2007) focus on the effect of competition on multiple retailers when each retailer is allowed to choose the timing of a markdown. Bitran & Mondschein (1997) study both dynamic pricing and fixed discounting policies (where an initial price is determined and then predetermined markdowns are applied) for a retailer in a multi-period, dynamic setting. They also discuss the choice of optimal order quantity for each of the pricing policies. Bitran et al. (1998), on the other hand, consider a retailer that commits to predetermined times to revise prices. They develop a stochastic dynamic programming formulation where the retailer coordinates prices multiple stores in their problem and inventory can be transferred from one store to another at the end of each period. In our work, apart from the price changes, we consider the timing of price changes and the choice of order quantity as well.

The article that is most relevant for our work is by Petruzzi & Monahan (2003). They discuss the retailer's recourse strategy of selling in secondary markets by analyzing a dynamic model wherein the decision after each period is whether or not to terminate sales in the primary market. They consider demand to be a function of price in linear form with additive and multiplicative uncertainty (as in Petruzzi & Dada (1999)). Contrasting our model to this work, we determine the optimal time to stop selling in the primary market after having acquired information on the state of the demand but before the sales at the primary market start.

Timing/Switching. Timing is also a key decision when a seller has multiple

products to offer: In sports and entertainment industry (e.g., Drake et al. (2008)) determining when to switch from selling bundles to individual tickets *a priori*, is significant as it can increase early cash flow and donation rates. Khouja & Robbins (2005) discuss book publishing and the switch from hard-bound to paper-back versions. In this paper, the product offered in the secondary market is geared towards price-sensitive customers and thus of different quality (i.e., the paperback version).

Hugos & Thomas (2005) discuss about the different types of distribution channel mechanisms like, direct channel, retailer channel, wholesaler channel, agent or broker channel, and dual or multiple channel for the retail industry. Our problem can be considered as a dual channel with the product offered sequentially in the channels, corresponding to primary and secondary markets. Lehmann & Weinberg (2000) discuss the application of sequential channel management for theater release and video release of the movie. In this paper, the authors discuss when should the movie distributor release the video, after the movie is released in the theaters. Contrasting to our model, we consider timing decision from the perspective of retailer (e.g., Blockbuster). The exponential decaying demand model considered in their paper is specific to movies. Apart from this article, most of channel management research has focused on the design and understanding the conflict/competition among the channels. From this perspective, our problem helps the retailer/seller in dynamic channel management, where timing decision corresponds to when to offer the product sequentially in the alternate channel. Timing decision has not received much attention from a dynamic channel management perspective and this is the focus of our research.

Timing is a key decision when we consider product development and product rollovers in a market. New product development is a complex process; Krishnan & Ulrich (2001) provide an exhaustive literature survey on product development. One of the important decisions of product rollover is determining the timing of withdrawal for the current generation product. Erhun et al. (2007) discuss the art of managing new product transitions. Their research mentions that due to supply constraints, the transition is not smooth but timing is critical so as not to cannibalize the current or new generation product. Our research helps in product rollover as it determines the timing of when to offer the product in secondary markets and effectively end the life of the product in the primary market and thus having room for next generation of products in this market.

Our Contribution. The research literature falls short on analyzing the operational decisions of the retailer with secondary markets in a temporal setting. As is evident in the survey by Coughlan & Soberman (2004), many brands in fashion apparel offer their products sequentially in primary and secondary markets. However, it is not clear whether this is an optimal action given a short sales horizon and limited inventories. Our novel model enables us to study when a brand/retailer benefits from sequential access to primary and secondary markets. Taking an optimization perspective, we are able to identify the optimal time to start selling goods in the secondary market. Unlike the extant literature, where timing or ordering is the sole decision for a retailer, we study both of these decisions. We show that the retailer may not benefit from the sequential channel approach all the time; this depends on the prevailing prices and the potential demand in each market. We show that this is

true even if the retailer chooses the inventory level to maximize the total expected profits in the primary and secondary markets.

2.4. Retailer's Expected Profit Function and Demand Models

We formulate the retailers problem in two stages. The first stage determines the optimal order quantity, Q^* , based on expected profit maximization:

$$\pi^* = \max_{Q \geq 0} \theta_l R_l^*(Q) + \theta_h R_h^*(Q) - wQ \quad (2.1)$$

where w is the unit cost purchasing the product from a vendor and $R_s^*(Q)$ is the expected revenue from primary and secondary market sales. π^* is the optimal expected total profit for the retailer. We assume that there are no capacity constraints, so the retailer can order and receive as much as ordered.² $R_s^*(Q)$ is determined by solving the second stage optimization problem:

$$R_s^*(Q) = \max_{0 \leq \beta \leq 1} E[R_s(\beta|Q)] \quad (2.2)$$

$$E[R_s(\beta|Q)] = p_1 E[\min(Q, N_1^s(\beta))] + p_2 E[\min((Q - N_1^s(\beta))^+, N_2^s(\beta))] \quad (2.3)$$

for $s = l, h$. $N_1^s(\beta)$ and $N_2^s(\beta)$ are two random variables that denote the demand observed at the primary and secondary markets during $[0, \beta T)$ and $[\beta T, T]$, respectively, given state s and the timing decision β . Without loss of generality, any unsold

²If there were capacity constraints, it can be imposed as upper bound on the order quantity Q .

goods have zero salvage value at the end of the season.³

To capture the essence of the problem, the demand function has to be defined carefully. The demand in the primary and secondary markets cannot be arbitrary; the demand has to be dependent on the length of time goods are sold in each market. Naturally, the longer the time, the higher (or “not lower” to be precise) should be the demand. That is, demand is a random variable which has a monotonic relation with parameter β . The larger β is, the longer (shorter) the goods are sold in the primary (secondary) market, hence the higher $N_1(\beta)$ is and the lower $N_2(\beta)$ is. We use monotonicity not in a strict sense: We say “increasing” (or higher) for non-decreasing (or not lower).

In this chapter, we introduce a demand model where the demand in any market is “proportional” to the time the product is offered in the market given the state of the system, s . Let $\tilde{X}_1^s(t)$ and $\tilde{X}_2^s(t)$ be two non-negative stochastic processes (not necessarily independent), defined over a known probability space. If the product is offered for sale during $[t_1, t_2]$ in market i ($i = 1, 2$) at state s ($s = l, h$), then the demand during this time period is $\int_{t_1}^{t_2} X_2^s(t)dt$; which is a random variable. In our case, the total (random) demand in the primary and the secondary market is then $N_1^s(\beta) = \int_0^\beta X_1^s(t)dt$ and $N_2^s(\beta) = \int_\beta^1 X_2^s(t)dt$, respectively.

In this chapter, we analyze a special case where the aggregate demand is

³Suppose every unit of inventory unsold at time T is salvaged at a value ρ such that $p_1 \geq p_2 > \rho$. The structure of the problem with positive salvage value is the same as the one with zero salvage value with appropriate change in parameters: The optimal ordering and timing decisions are obtained by setting $w \leftarrow w - \rho$ for the first stage and $p_i \leftarrow p_i - \rho$, $i = 1, 2$ in the second stage.

$X_1^s \triangleq \int_0^1 X_1^s(t)dt$ and $X_2^s \triangleq \int_0^1 X_2^s(t)dt$, and

$$N_1^s(\beta) = \beta X_1^s, \quad (2.4)$$

$$N_2^s(\beta) = (1 - \beta)X_2^s. \quad (2.5)$$

That is, random demand during $[0, \beta]$ in the primary market is assumed to be linear in the total market potential, X_1^s , and the length of the time interval, β . Similarly, random demand during $(\beta, 1]$ in the secondary market is linear in the total market potential, X_2^s , and the length of the time interval, $(1 - \beta)$. Starting with the general model we introduced above, this proportional model is valid when (i) $X_1^s(t)$ and $X_2^s(t)$ are time-invariant, or (ii) changes (trends) in $X_1^s(t)$ and $X_2^s(t)$ over time are small so that $N_1(\beta)$ and $N_2(\beta)$ are good approximations for the demand in the primary and the secondary markets, respectively, for any $\beta \in [0, 1]$.

Notice that the expected value, $E[N_1^s(\beta)] = \beta E[X_1^s]$, is increasing and linear in β . Similarly, the expected value, $E[N_2^s(\beta)] = (1 - \beta)E[X_2^s]$, is decreasing and linear in β . Note that the primary and secondary market demand as represented by random variables X_1^s and X_2^s can be stochastic functions of p_1 and p_2 . For instance, the aggregate demand in any market in any state can be modeled by the classical price-demand curve with additive uncertainty: $X_i^s = a_i^s - b_i^s p_i + \varepsilon_i^s$ for $i = 1, 2$, $s = l, h$ where a_i^s, b_i^s are the parameters of the demand curve and ε_i^s is a random variable. Such a model poses no challenges in our analysis because p_1 and p_2 are fixed and are *not* decision variables.

The above demand model builds on the overall market potential and uses

static information about aggregate market demand. We develop a different stochastic model in the next chapter. In this chapter, we use this linear demand model introduced in Equations (2.4) and (2.5), and assume the probability distribution of $X_i^s (i = 1, 2 \text{ and } s = l, h)$ is available.

2.5. Analysis of Second Stage: Retailer's Timing Decision

At the beginning of the sales season, the retailer finds himself in a particular state of the market and decides on the optimal length of time to offer the product in the primary market. At this stage the retailer has $0 \leq Q < \infty$ units on hand. The formulation of the problem is presented in Equations (2.2) and (2.3). In this section, we drop the superscript s from consideration as the problem for low and high states has the same structure. Basically, the optimization problem is

$$\begin{aligned} & \max_{0 \leq \beta \leq 1} E[R(\beta|Q)] \\ &= \max_{0 \leq \beta \leq 1} p_1 E[\min(Q, N_1(\beta))] + p_2 E[\min((Q - N_1(\beta))^+, N_2(\beta))] \quad (2.6) \end{aligned}$$

The linear demand model has several advantages: We can establish structural properties of the expected revenue function with minimal restrictions on the probability distribution of X_1 and X_2 . We only require the demand to be non-negative, bounded, and proportional to the length of the time interval. Critical to our analysis in the remainder of this section are the properties of the *sample path* revenues. Suppose we take a random sample from the given probability distributions; $k_1 \geq 0$ is a realization of X_1 and $k_2 \geq 0$ is a realization of X_2 . Then the corresponding primary

and secondary market demand values are βk_1 and $(1 - \beta)k_2$. The revenue is then

$$R(\beta|Q) = p_1 \min(Q, \beta k_1) + p_2 \min((Q - \beta k_1)^+, (1 - \beta)k_2). \quad (2.7)$$

which is a deterministic function of β and Q . The sample path revenue function can be written as

$$R(\beta|Q) = \begin{cases} h_1(\beta, Q) = p_1 Q & Q < \beta k_1 \\ h_2(\beta, Q) = (p_1 - p_2)\beta k_1 + p_2 Q & \beta k_1 \leq Q < \beta k_1 + (1 - \beta)k_2 \\ h_3(\beta, Q) = p_1 \beta k_1 + p_2(1 - \beta)k_2 & \beta k_1 \leq \beta k_1 + (1 - \beta)k_2 \leq Q \end{cases} \quad (2.8)$$

The function consists of three pieces (planes), defined by functions $h_i(\beta, Q)$ for $i = 1, 2, 3$, where each $h_i(\beta, Q)$ is linear (affine) and separable with respect to β and Q . Note that the revenue function itself is not separable with respect to β and Q : The conditions under which the functions $h_i(\beta, Q)$, $i = 1, 2, 3$ are defined depend on values of β and Q . We summarize the analytical properties of the sample-path revenue function below. Note that this is not an exhaustive list; we state the properties that are later used in the analysis of the expected revenue function.

Proposition 1 (*Sample Path Properties*) *Sample path revenue is defined as a function $R(\beta|Q) : [0, 1] \times [0, \infty) \rightarrow \mathbb{R}$ which is (i) continuous over $[0, 1] \times [0, \infty)$, (ii) Lipschitz continuous with respect to $\beta \in [0, 1]$ for a given Q , (iii) piecewise linear in $\beta \in [0, 1]$, (iv) piecewise linear in $Q \in [0, \infty)$, (v) componentwise concave and non-decreasing in Q for a given value of β , (vi) componentwise concave in β for a given value of Q , and (vii) jointly concave in $\beta \in [0, 1]$ and $Q \in [0, \infty)$.*

Proof (i) Continuity is proved by checking the end points in (2.8): We first set $Q = \beta k_1$ and see that $h_1(\beta, Q) = h_2(\beta, Q)$ when $Q = \beta k_1$. We then set $Q = \beta k_1 + (1 - \beta)k_2$ and see that $h_2(\beta, Q) = h_3(\beta, Q)$. (ii) Lipschitz continuity with respect to β follows from demand samples $k_1, k_2 < \infty$ because the random variables X_1 and X_2 have finite support. Piecewise linearity in (iii) and (iv) is immediately observed because $h_i(\beta, Q)$ is linear and separable in β and Q for $i = 1, 2, 3$. (v) The function is nondecreasing and componentwise concave in Q for a given β because

$$\frac{\partial h_1(\beta, Q)}{\partial Q} = p_1 > \frac{\partial h_2(\beta, Q)}{\partial Q} = p_2 > \frac{\partial h_3(\beta, Q)}{\partial Q} = 0 \quad (2.9)$$

and $h_i(\beta, Q)$, $i = 1, 2, 3$ are defined for $0 \leq Q < \beta k_1$, $\beta k_1 \leq Q < \beta k_1 + (1 - \beta)k_2$, and $\beta k_1 \leq \beta k_1 + (1 - \beta)k_2 \leq Q$, respectively. (vi) Componentwise concavity with respect to β for a given Q requires inspecting

$$\frac{\partial h_1(\beta, Q)}{\partial \beta} = 0, \quad (2.10)$$

$$\frac{\partial h_2(\beta, Q)}{\partial \beta} = (p_1 - p_2)k_1, \quad (2.11)$$

$$\frac{\partial h_3(\beta, Q)}{\partial \beta} = p_1 k_1 - p_2 k_2 \quad (2.12)$$

and performing a case-by-case analysis on the values of k_1 , k_2 and Q to determine the sign and the magnitude of the partial derivatives $\frac{\partial h_i(\beta, Q)}{\partial \beta}$. We refer the reader to Appendix A.1 for a complete description and treatment of each case. (vii) Joint concavity is also proved by case-by-case analysis; the details are provided in A.2. •

The monotonicity of the sample path revenue with respect to Q is intuitive: If inventory is so low that you can never sellout in the primary market, then procuring an additional unit of the product gathers additional revenue of p_1 . Similarly, if the inventory is enough to meet primary market demand but not enough to meet all of secondary market demand, then marginal revenue generated is p_2 . Finally, if the inventory is too high and exceeds the total demand in the primary and secondary markets, then the marginal value of an additional unit is 0.

Notice that joint concavity of the revenue function with respect to β and Q implies componentwise concavity with respect to each parameter. Therefore concavity claims in (v) and (iv) do not require a separate proof when property (vi) is proved. We discuss componentwise concavity above because componentwise properties are easier to observe from the piecewise structure of the sample-path revenue. An illustration of the joint concavity is provided in Figure 2.3 for $p_1 = 10$, $p_2 = 8$ and sample $k_1 = 13$, $k_2 = 10$.

The real advantage of the sample-path analysis is that many of the properties established for the sample path revenue do generalize to the expected revenue function $E[R(\beta|Q)]$ without making further assumptions about the random variables. We present the next set of results without a proof because continuity, monotonicity, and concavity are all preserved under expectation. The results below are immediate from the sample-path properties.

Proposition 2 (*Properties of the Expected Revenue Function*) *Expected revenue function $E[R(\beta|Q)]$ defined as a function $E[R(\beta|Q)] : [0, 1] \times [0, \infty) \rightarrow \mathbb{R}$ is (i)*

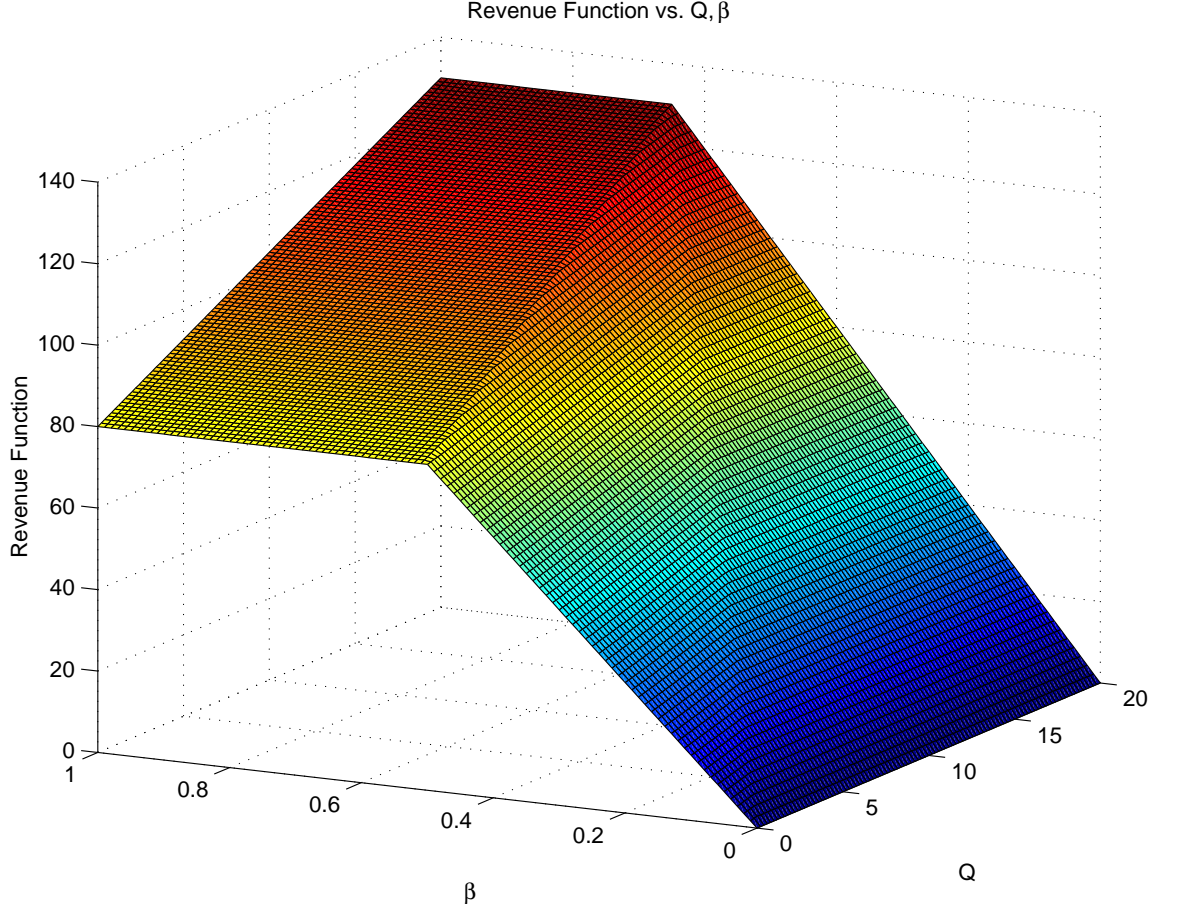


Figure 2.3: Illustration of the joint concavity of the sample-path revenue function *continuous over $[0, 1] \times [0, \infty)$, (ii) componentwise concave and non-decreasing in Q for a given value of β , (iii) componentwise concave in β for a given value of Q , and (iv) jointly concave in $\beta \in [0, 1]$ and $Q \in [0, \infty)$.*

Note that the sample-path revenue function is not differentiable everywhere for $\beta \in [0, 1]$ given Q . Consequently, $E[R(\beta|Q)]$ need not be differentiable for all $\beta \in [0, 1]$, for instance, when the random variables X_1 and X_2 have discrete probability distributions. However, $R(\beta|Q)$ is differentiable almost everywhere. When the random variables X_1 and X_2 have continuous probability distributions, then $E[R(\beta|Q)]$ is differentiable (see for e.g. Kleywegt & Shapiro (2001)). In that case,

the first order condition with respect to β is sufficient to identify the optimal length of time for the goods to be sold at the primary market.

Proposition 3 *Suppose $E[R(\beta|Q)]$ is differentiable with respect to $\beta \in [0, 1]$. Then*

(a) For a given inventory level $Q < \infty$, the first order derivative with respect to β is

$$\begin{aligned} \frac{\partial E[R(\beta|Q)]}{\partial \beta} &= E[p_1 X_1 - p_2 X_2 | \beta X_1 + (1 - \beta) X_2 \leq Q] \\ &\quad + (p_1 - p_2) E[X_1 | \beta X_1 \leq Q < \beta X_1 + (1 - \beta) X_2]. \end{aligned} \quad (2.13)$$

(b) Let the optimal length of time for the goods to be sold at the primary market be denoted $\beta^(Q)$. If $\frac{\partial E[R(\beta|Q)]}{\partial \beta}|_{\beta=0} < 0$, then $\beta^*(Q) = 0$ and if $\frac{\partial E[R(\beta|Q)]}{\partial \beta}|_{\beta=1} > 1$, then $\beta^*(Q) = 1$. Otherwise, $0 < \beta^*(Q) < 1$ and is determined by solving $\frac{\partial E[R(\beta|Q)]}{\partial \beta} = 0$.*

Proof (a) Because $R(\beta|Q)$ is differentiable almost everywhere and Lipschitz continuous with respect to β , we can interchange differentiation and expectation to derive the first order condition (Kleywegt & Shapiro (2001), Glasserman & Tayur (1995)):

$$\begin{aligned} \frac{\partial E[R(\beta|Q)]}{\partial \beta} &= E\left[\frac{\partial R(\beta|Q)}{\partial \beta}\right] \\ &= E\left[\frac{\partial h_1(\beta, Q)}{\partial \beta} | Q < \beta X_1\right] + E\left[\frac{\partial h_2(\beta, Q)}{\partial \beta} | \beta X_1 \leq Q < \beta X_1 + (1 - \beta) X_2\right] \\ &\quad + E\left[\frac{\partial h_3(\beta, Q)}{\partial \beta} | \beta X_1 \leq \beta X_1 + (1 - \beta) X_2 \leq Q\right] \\ &= 0 + E[(p_1 - p_2) X_1 | \beta X_1 \leq Q < \beta X_1 + (1 - \beta) X_2] \\ &\quad + E[p_1 X_1 - p_2 X_2 | \beta X_1 + (1 - \beta) X_2 \leq Q]. \end{aligned} \quad (2.14)$$

(b) follows from the concavity of the function and the first order condition. •

In summary, we can easily determine the optimal timing using the first order conditions when the expected revenue function is differentiable, or using a simple search procedure (such as Fibonacci search, see Bazaraa et al. (2006)) when the function is not differentiable. Depending on the demand distributions, prices, and the order quantity chosen, we can have an interior solution to the timing problem which means that the retailer needs to utilize both the primary and secondary markets in order to maximize expected revenues. However, there are cases where the retailer's optimal decision is to operate in only one of the markets. Analytically speaking, when the function $E[R(\beta|Q)]$ is nondecreasing (or nonincreasing) for $\beta \in [0, 1]$, then it is optimal to set $\beta^*(Q) = 1$ ($\beta^*(Q) = 0$) and sell the goods exclusively in the primary (secondary) market.

Note that we have not established strict concavity and cannot guarantee a unique optimal solution to the timing problem. In fact, we can see that multiple optima exists when we consider the following example: Suppose $p_1 E[X_1] = p_2 E[X_2]$ with X_1 and X_2 being non-negative random variables bounded above by $Q/2$, i.e. $X_1 + X_2 < Q$, almost surely. Then the first order condition reduces to $\frac{\partial E[R(\beta|Q)]}{\partial \beta} = E[p_1 X_1 - p_2 X_2] = 0$ and is satisfied by any $\beta \in [0, 1]$.

We next turn our attention to the sensitivity of the optimal timing decision to other problem parameters, especially, to the order quantity Q . In parametric optimization, it is customary to investigate comparative statics for this purpose. In our case, we are interested in determining whether the optimal length of time the

goods are sold in the market is monotonic in the order quantity. Intuitively, $\beta^*(Q)$ is not monotonic in Q . This is because the retailer is not interested in ‘clearing’ out the inventory by the end of selling season but increasing his profits. When the retailer starts the sales season with few units on hand, then he will extend the primary market exposure and limit secondary market sales (which may not even be needed because the retailer can sell out the inventory at the primary market at price p_1). Suppose the starting inventory level, Q is high, p_2 is low, and the demand is very high in the secondary market. If the retailer were to use ‘clearance’ policy, then he would switch to the secondary market as early as possible. But in our case, the retailer would like to extract as much revenue as possible and hence may not switch to the secondary market since the revenue potential is negligible in this market. Thus, when the retailer has high inventories, he may not *always* benefit from secondary market sales (depends on potential revenue generated). This means that the optimal switching time may or may not decrease in Q .

To analytically prove that $\beta^*(Q)$ is non-increasing or non-decreasing in Q , a number of ‘general tools’ are available. For instance, if the expected revenue function $E[R(\beta|Q)]$ has increasing (decreasing) differences in β and Q or when $\frac{\partial^2 E[R(\beta|Q)]}{\partial \beta \partial Q} \geq 0$ (≤ 0) in case of twice continuous differentiability, then $\beta^*(Q)$ would be non-decreasing (non-increasing) in Q ; see Van Zandt (2002) and Vives (2000). Increasing or decreasing differences are structural properties that, when established at the sample-path level, would generalize to the expected revenue function. Unfortunately, our problem does not satisfy neither the decreasing nor the increasing differences properties. We show this using a numerical example. First, we provide a formal

definition of increasing differences property:

Definition 1 A function $f : A \times \Theta \rightarrow \mathbb{R}$ satisfies the increasing differences property in (a, θ) if for all $a', a'' \in A$, the function $f(a', \theta) - f(a'', \theta)$ is nondecreasing in $\theta \in \Theta$ (see Chapter 10 of Sundaram (1996)).

The example we chose has $p_1 = 10$, $p_2 = 8$, $X_1 = 5$ with probability 1 and $X_2 = 8$ with probability 1. The expected revenue function $E[R(\beta|Q)]$ is plotted for $\beta = 0$ and $\beta = 1$ for varying values of Q in Figure 2.4. Notice that the difference between the two curves are, increasing, decreasing, and then non-decreasing as Q increases. This observation rules out applicability of the well-known results on monotonic comparative statics to prove the monotonicity of the timing decision.

Next, we provide a numerical example to show non-monotonicity of $\beta^*(Q)$ in Q . Due to lack of strict concavity, there may be multiple optima for optimal timing decisions. Figure 2.5 shows the two curves: $\beta^{*\min}(Q)$, and $\beta^{*\max}(Q)$. The

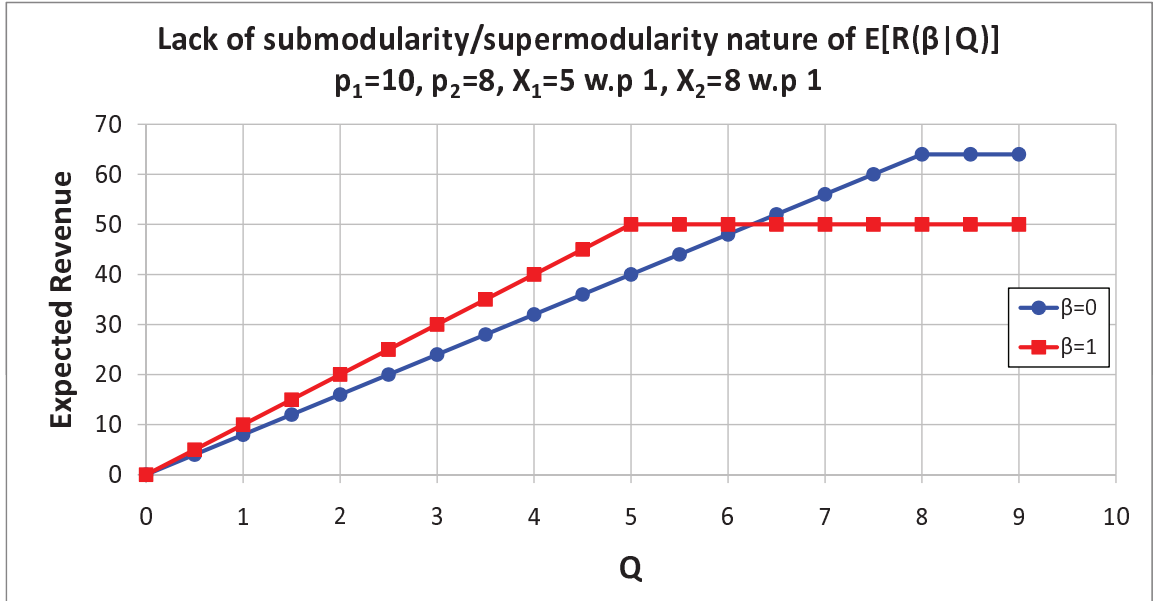


Figure 2.4: Lack of Super or Submodularity of Expected Revenue Function

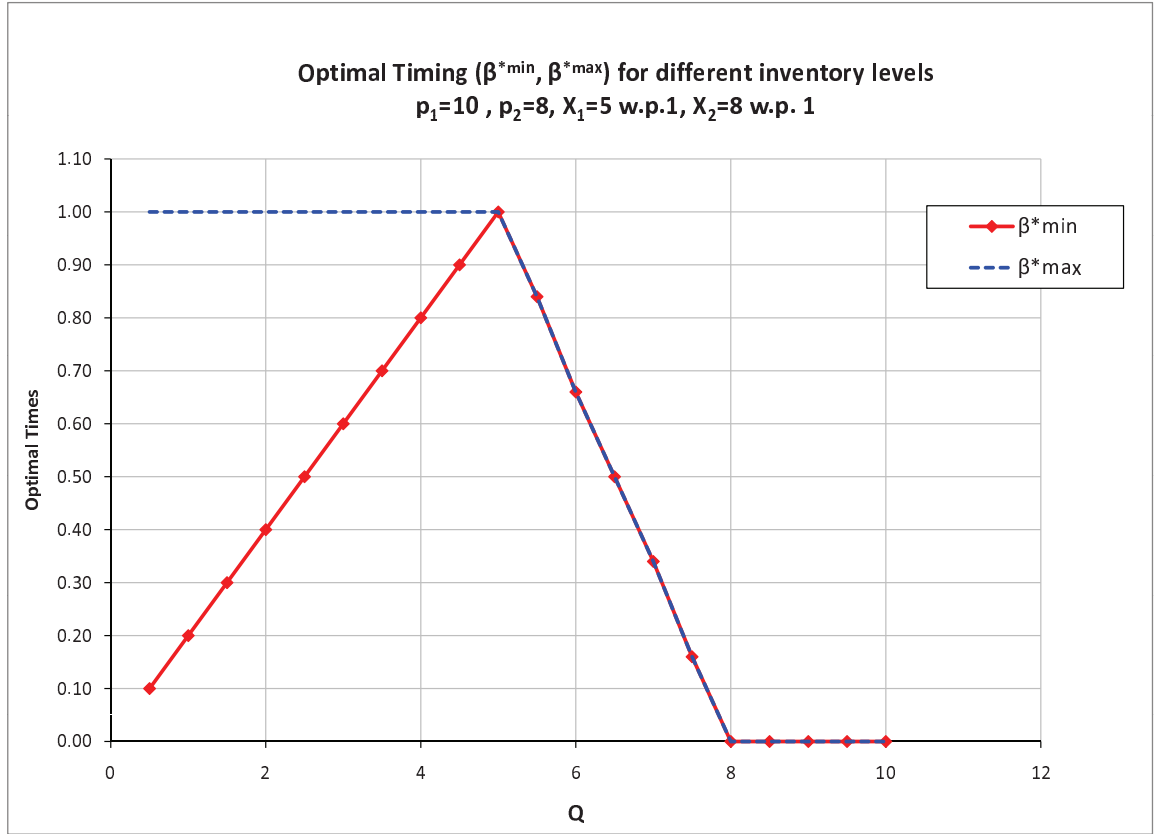


Figure 2.5: Effect of inventory on Optimal Timing

range $[\beta^{*\min}(Q), \beta^{*\max}(Q)]$ represents the set of optimal timing decisions (i.e., $\beta \in [\beta^{*\min}(Q), \beta^{*\max}(Q)]$ is optimal). From the figure, although $\beta^{*\max}$ is non-increasing in Q , we see that $\beta^*(Q) \in [\beta^{*\min}(Q), \beta^{*\max}(Q)]$, in general is not monotonic in Q . This is because the potential revenue and not just demand determine the optimal timing. We next focus on the retailer's optimal choice of Q in order to maximize the expected profits.

2.6. Analysis of First Stage: Retailer's Ordering Decision

The retailer's first-stage problem was introduced in Equation (2.1) as:

$$\pi^* = \max_{Q \geq 0} \theta_l R_l^*(Q) + \theta_h R_h^*(Q) - wQ$$

where w is the variable cost of product per unit and π^* is the optimal expected total profit for the retailer. Let

$$TP(Q) = \theta_l R_l^*(Q) + \theta_h R_h^*(Q) - wQ$$

be the retailer's expected total profit as a function of Q . We investigate the properties of $TP(Q)$ in this section. We have not been able to analytically prove the concavity of the function relying on the properties of the second-stage problem. However, through many computational experiments we have not come across any example where $TP(Q)$ is non-concave in Q . So we conjecture that for many practical choices of problem parameters and demand distributions, the expected total profit is concave. We rely on computational methods to solve for Q^* .

Notice that the optimal order quantity Q^* may not be the optimal for state l or h , which makes the timing decision important (and non-trivial) in the second stage. If we consider a special case where $w \rightarrow 0$, then the optimal first-stage and second-stage solutions can be obtained analytically:

Proposition 4 *If $w = 0$, then $Q^* \rightarrow \infty$ and $\beta_s^*(Q^*) \in \{0, 1\}$.*

Proof For this special case, the retailer solves:

$$\pi^* = \max_{Q \geq 0} \theta_l \max_{0 \leq \beta \leq 1} E[R_l(\beta|Q)] + \theta_h \max_{0 \leq \beta \leq 1} E[R_h(\beta|Q)]. \quad (2.15)$$

The $Q^* \rightarrow \infty$ because the expected revenue function $E[R_s(\beta|Q)]$ is non-decreasing in Q for $s = l, h$. When $Q^* \rightarrow \infty$ and the random variables $X_1^s + X_2^s < Q$ almost surely for $s = l, h$, then the expected revenue function reduces to

$$E[R_l(\beta|Q)] = p_1 E[X_1^s \beta] + p_2 E[X_2^s (1 - \beta)] = p_2 E[X_2^s] + (p_1 E[X_1^s] - p_2 E[X_2^s]) \beta$$

which is linear in β . Then, $\beta^* = 0$ if $p_1 E[X_1^s] \leq p_2 E[X_2^s]$ and $\beta^* = 1$ otherwise. •

The last result leads to a trivial case but is relevant for an industry where the retailer has extremely high profit margin (i.e., $p_1, p_2 \gg w \approx 0$). In this case, the retailer does not benefit from having two separate markets. He is better off with only one market, the one that has the highest potential expected sales revenue.

2.7. Computational Experiments

In this section we further investigate the retailer's timing and ordering decisions. Specifically, we (i) provide examples on the optimal ordering and timing decisions, and perform sensitivity analysis to test (ii) the effect of price at the secondary market (p_2) and (iii) the effect of probabilities of the state of demand (θ_l, θ_h) on the optimal ordering and timing decisions of the retailer.

2.7.1 Design of the experiment

In our experiments, the aggregate demand in the primary and secondary markets is a linear function of price with additive uncertainty. That is, $X_1^s = a_P^s - b_P^s p + \epsilon_P^s$ and $X_2^s = a_S^s - b_S^s p + \epsilon_S^s$, where, $a_P^s, b_P^s, a_S^s, b_S^s > 0$ and $s = l, h$ is the state of the system. There are more price-sensitive customers in the secondary market, and we choose the parameter values so that $a_P/b_P < a_S/b_S$. The uncertainty $\epsilon_P^s, \epsilon_S^s$ can follow any distribution but here we use independent discrete probability distributions as given in Table 2.1. We use the following price-paths: $p_1 = 2$ and $p_2 = \{1.75, 1.5, 1.25\}$. We selected three values for p_2 to reflect different levels of discounts. And, for both market states, we use $a_P = 4, b_P = \{3, 2, 1\}$ and $a_S = 20, b_S = 4$. We set wholesale price to $w = 1$. In the experiments, primary market demand is always dominated by secondary market. Primary market demand is robust to the state of the market (i.e., aggregate demand in the primary market has the same probability distribution for both high and low states). This enables us to isolate the effect of fluctuations in the secondary market.

Table 2.1: Probability distribution of the demand model

State 0 (low)			
$P(\epsilon_P = 5) = 0.2$	$P(\epsilon_P = 7) = 0.8$	$P(\epsilon_S = 2) = 0.3$	$P(\epsilon_S = 5) = 0.7$
State 1 (high)			
$P(\epsilon_P = 5) = 0.2$	$P(\epsilon_P = 7) = 0.8$	$P(\epsilon_S = 6) = 0.4$	$P(\epsilon_S = 9) = 0.6$

2.7.2 The Optimal Ordering and Timing Decision

In this section, we have $a_P = 4, b_P = 1, a_S = 20, b_S = 4$ and $w = 1$. Table 2.2 shows the optimal order quantity and the corresponding optimal timing decisions

when only one state of the market is possible (i.e. $\theta_l \in \{0, 1\}$). We compute the optimal values for $p_2 \in \{1.25, 1.5, 1.75\}$.

Table 2.2: Optimal Decisions: Q^* and β_s^*

p_1	p_2	θ_l	π^*	Q^*	β_l^*	β_h^*	$R(\beta_l^* Q)$	$R(\beta_h^* Q)$
2	1.75	0	14.4	22	-	0	-	36.4
2	1.75	1	11.925	18	0	-	29.925	-
2	1.5	0	10	20	-	0	-	30
2	1.5	1	8.3645	11.75	0.61	-	20.1145	-
2	1.25	0	8.2	9	-	1	-	17.2
2	1.25	1	8.2	9	1	-	17.2	-

What is interesting in these results is that depending on the level of mark-down and the market state, the retailer can choose an order quantity to sell only in the primary market, to sell in both markets or to sell only at the secondary market. The level of markdown and the corresponding revenue potential are the main determinants for utilizing only one of the markets in this experiment.

Note that the aggregate primary market demand is $X_1 = \{7, 9\}$. The unit revenue is $p_1 = 2$. If the retailer sells exclusively in the primary market, then total expected revenue, by ordering high enough, is $2 * 8.6 = \$17.2$. This serves as a benchmark for the other cases.

For $p_2 = 1.75$, we have $X_2^l = \{15, 18\}$ and $X_2^h = \{19, 22\}$. If retailer sells exclusively in the secondary market, then the expected revenue is $1.75 * 17.1 = \$29.925$ for low state and $1.75 * 20.8 = \$36.4$ for high state, assuming order quantity is high enough. Both of these quantities yield a higher profit compared to the profit that would be obtained by ordering and selling exclusively at the primary market. Therefore, when $p_2 = 1.75$, it is optimal to order as high as the secondary market demand and start selling the goods at the secondary market immediately at the

start of the sales horizon.

For $p_2 = 1.25$, we have $X_2^l = \{17, 20\}$ and $X_2^h = \{21, 24\}$. If the retailer sells exclusively in the secondary market, then total expected revenue is $1.25 * 19.1 = \$23.875$ for low state and $1.25 * 22.8 = \$28.5$ for high state, assuming the retailer's order quantity is high enough. For either state, the expected profit of satisfying the primary market demand only is $\$8.6$. In comparison, the expected profit of satisfying the secondary market demand only is $0.25*19.1=\$4.775$ in the low state and $0.25*22.8=\$5.7$ in the high state. Thus, even though secondary market has higher demand compared to the primary market, the level of markdown is too high and the loss in revenue due to markdowns cannot be compensated with the resulting sales in the secondary market. Therefore the optimal decision in this case is to order only enough to satisfy the demand in the primary market and only operate at the primary market.

For $p_2 = 1.5$, $X_2^l = \{16, 19\}$ and $X_2^s = \{20, 23\}$. If retailer sells exclusively in the secondary market, then total expected revenue is $1.5*18.1 = \$27.15$ for low state and $1.5 * 21.8 = \$32.70$ for high state, assuming his order quantity is high enough. The primary market profit potential at any state is $\$8.6$ and this is not significantly different than the profit potential of selling exclusively at the secondary market (specifically at a profit of $0.5*18.1=\$9.05$ in the low state). In this case, operating in a single market, i.e., exclusively, is not the dominant strategy for the retailer: He chooses an order quantity that is slightly higher than the primary market demand but is not high enough to satisfy all the demand in the secondary market, and accordingly, the optimal choice is $\beta_i^* = 0.61$.

2.7.3 Effect of prices and state of the market

To study the effect of higher discounts, we fix $p_1 = 2$ and drop p_2 from 1.75 to 1.50 to 1.25. This represents discounts of 12.5% off, 25% off, and 37.5% off. For higher discounts, let us observe the Q^* , β^* and total profit. We also vary the market state probability θ_l from 0 to 1 in 0.10 increments.

In Figure 2.6, we see that as discounts are steeper, the total expected profits drop steadily. As the low state is more likely, retailer's total expected profits drop. This is because, in general, the demand for the product is low when market is in low state. Notice that this observation is significant as the retailer's order quantity does not necessarily decrease for steeper discounts. Figure 2.8 shows the optimal order quantity for different discount levels and state probabilities. Steeper discounts do not always mean higher order quantities as the revenue potential drives the order quantity. For example, if $p_2 = 0$, then theoretically, demand is very high, but the revenue potential is 0, since there is no gain for the retailer. For $p_2 = 1.75$, the secondary market demand is comparable to the primary market. When p_2 drops to 1.5, the secondary market demand increases, and the revenue potential determines the order quantity. The interplay of primary and secondary markets is evident in such a scenario. Though the optimal order quantities increase or decrease with steeper discounts depending on revenue potential as discussed, but with the low state more likely, the optimal order quantities decrease. The effect of higher discounts on optimal timings is shown in Figure 2.7. With steeper discounts, for example, when $p_2 = 1.25$, the decision to switch to the secondary market is delayed or in fact, the

retailer never switches. On the other hand, if $p_2 = 1.50$, depending on the state of the market, retailer might switch to secondary market earlier than he would when, say $p_2 = 1.75$. This is again, because, of the revenue potential in each market for each state. There is a trade-off between generating higher demand at a discount vs. a lower demand at the highest price.

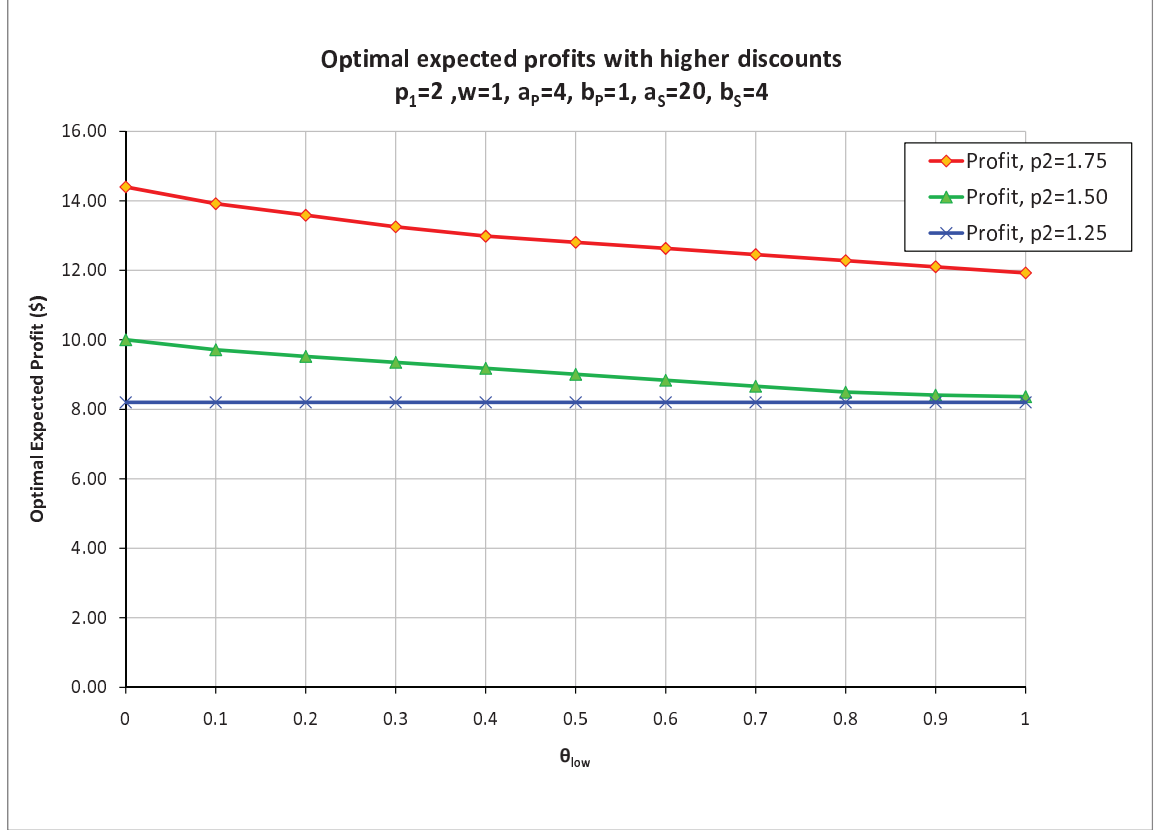


Figure 2.6: Effect of higher discounts on Optimal Expected Profits

From Figure 2.8, we can study how the optimal order quantity is influenced by the state probabilities. As the low state is more likely, the retailer would generally put in a request for lower order quantities. This is intuitive as the market is more likely to be in low state, then, the likelihood of generating high demand is decreased, which in turn effects the sales of the retailer. Since sales are decreased, the retailer

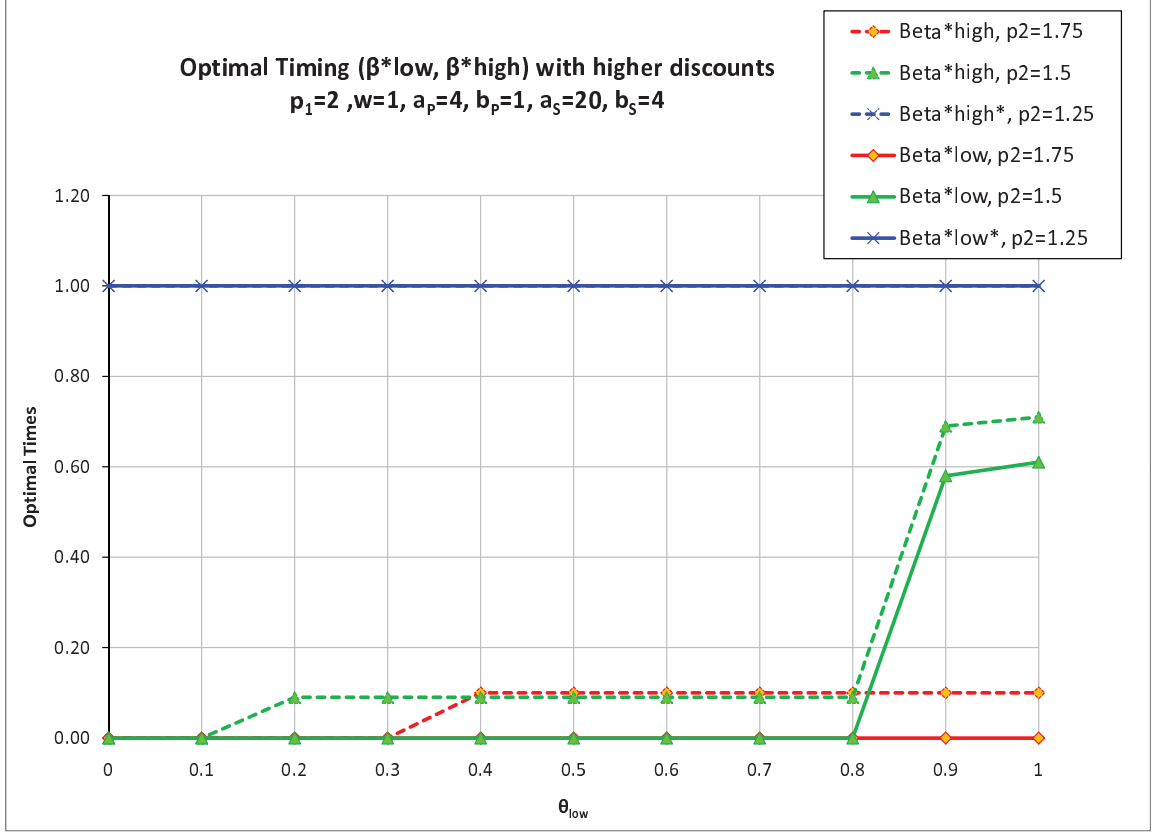


Figure 2.7: Effect of higher discounts on β^*_{low} and β^*_{high}

has no incentive to place higher order quantities. From the Figure 2.8, we can study how the optimal timing decision is influenced by the state probabilities. Optimal timing decision inversely follows the optimal order quantity curve.

2.8. Conclusions

In this chapter, we focus on decisions that are prevalent in high-end fashion retail: Many retailers make use of their own stores (brands) as well as outlets. It is typical in practice to sell the same goods in the stores and in the outlets, but the retailers choose to sell the goods first in their primary stores and then at the outlets with a delay. We build a model to answer the following questions: Should the

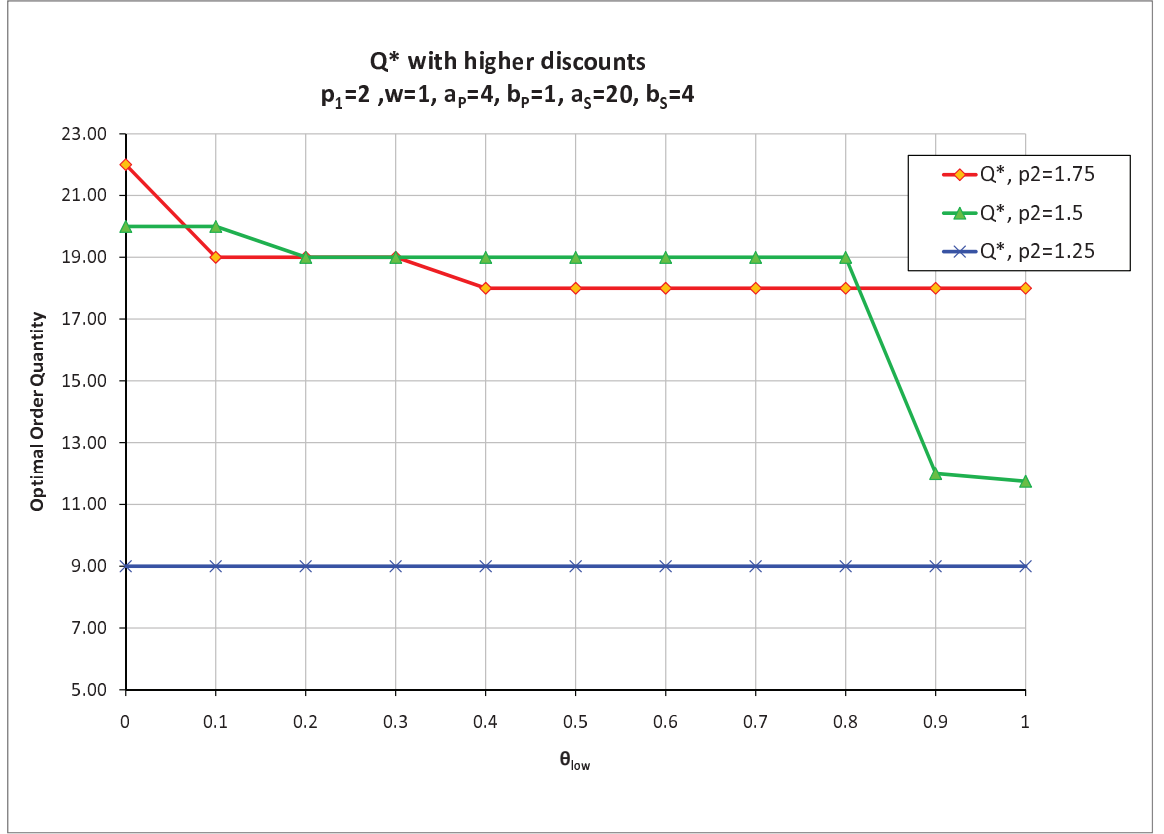


Figure 2.8: Effect of higher discounts on Q^*

retailer selling a seasonal product offer it only in the primary market? If not, then what is the best time to start offering the product in the secondary market? Our model is novel that it captures the temporal (sequential) aspect of the sales channels and also analyzes the optimal order quantity for the retailer that faces both market and demand uncertainty. Critical to our model and its structural properties is a demand model where the demand in each market is proportional to the length of time the product is offered in that market.

We model the retailer's decision in two stages: first stage involves the ordering decision which is made with uncertain market information while the second stage involves the timing decision after observing the market state and having received

the order. We establish structural properties of the retailer's profit function in the first stage and the revenue function in the second stage. Second-stage analysis makes use of sample path results, hence minimal assumptions are made on the random demand variables. We also show the retailer's expected revenue function does not exhibit additional properties (for e.g., sub or super modularity) to derive results on monotone comparative statics. Computational studies show that higher optimal order quantities in general result in decreasing the optimal timing decision to switch to the secondary market. Steeper discounts do not necessarily result in higher order quantities as the revenue potential plays an important part. If the market is more likely to be in the low state, then the retailer would place lower order quantities and thus consequently earn lower total profits. Most importantly, it may or may not be optimal for the retailer to delay the sales of the good in the secondary market depending on the sales and revenue potential in both primary and secondary markets.

Most often, a retailer that faces market and/or demand uncertainty enters into contractual agreements with his/her vendors to share the risk on unsold inventory at the end of the sales season. Intuitively, contractual agreements affect the retailer's "selling aggressively" vs. safety-first approaches. Thus, the next essay discusses the effect of contracts on the retailer's timing decision and ordering decision.

Chapter 3

Essay 2: Secondary Markets and Supplier-Retailer Contracts

3.1. Introduction

Based on the latest published Annual Retail Trade Survey (up to the year 2008) apparel businesses, in general, have a gross margin in the low 40%'s, department stores make around 30% (U. S. Census Bureau (2010)). For large retailers, this translates into roughly 3% profit margin (National Retail Federation (2010)). The retailers like Dillards, Federated Department Stores request the vendors (or suppliers) to share the cost of unsold inventory or lost revenue due to heavy discounts or markdowns on their selling floors (e.g., Mantrala et al. (2005), Lee & Rhee (2008)). The argument for such requests is that the selling season of a perishable product is short with uncertain demand, and there is negligible salvage at the end of selling horizon with no risk borne by the suppliers.

It is evident that retailers are faced with huge sales risks; if the demand for product is low, the retailer ends up with huge, profitless unsold inventory. This is valid even if retailers use different sales tactics and engage in business models that allow them to sell their goods in primary and secondary markets. On the other hand, if the demand for the product is high, the retailer should have sufficient inventory to meet the demand otherwise the retailer faces potential lost sales and revenues. Thus retailers argue that the request for rebates or return credits is justifiable as some of

the sales and inventory risk can, now, be shared with the vendor. The vendors buy into this argument because the vendor-retailer relationship goes beyond one selling season. Small vendors, who want to improve their brand image and market share of their products, accept because of the retailer's superior negotiation power. It is evident that different risk-sharing mechanisms between the retailer and the supplier affect the retailer's operational decisions like how much to order and when (if at all) to offer the products in a secondary market.

Our goal in this essay is to understand how different contracts between retailers and their vendors affect retailer's decisions and the supply chain's overall profitability. Specifically, we answer the following questions: What is the optimal length of time for the retailer to offer the good in a primary market when the retailer has a contract with a vendor to share the risk of excess inventory? Does the retailer benefit from having access to both primary and secondary markets when the risk of excess inventory is shared with a vendor? What is the effect of risk-sharing contracts on the retailer's ordering decision? How does the supply chain perform under different contracts? How does the supplier choose his contract parameters knowing the retailer's decision process? For vendors of high-risk, high-return products (e.g., innovative products that are sold at high margins) which contract type is most beneficial? If a vendor is interested in maximizing his/her products' primary market exposure, what contract should he/she offer to the retailer? And, at the firm level, we try to answer: Who, the supplier or the retailer is better off, in terms of profits and in what type of contract?

We answer these questions extending/adjusting the temporal model we in-

troduced in Chapter 2. Traditionally, research models in operations management, study contracts and investigate supply chain coordination for a given contract. Our research makes a contribution to the literature by (i) developing a model that represents primary and secondary markets in the retail industry, (ii) introducing a demand model that captures the trade off involved in selling products in primary vs. secondary markets during a short sales horizon, (iii) investigating the effect of contracts on a retailer’s timing and ordering decisions, and (iv) studying a contract type that is less well known in the research literature but is widely used in industry.

The remainder of the chapter is organized as follows. The next section discusses pertinent literature to contracts. Section 3.3 formally defines the problem and we perform analysis for the retailer decisions in Section 3.4. We compare the optimal decisions for the three contracts in Section 3.5. The supplier’s problem is discussed in Section 3.6. Results of computational experiments are presented in Section 3.7. Finally, we conclude in Section 3.8.

3.2. Literature Review

We focus on the literature pertaining to contracts in this section; the literature on secondary markets is reviewed in Chapter 2. There is a vast amount of literature concerning varied types of supplier-retailer contracts (e.g., Lariviere (1998), Cachon (2003)). However, most of the existing literature pertaining to contracts focuses on analyzing structural properties of contract parameters with respect to achieving supply chain coordination. A contract is said to coordinate the supply chain, if

thereby the partners' optimal local decisions lead to optimal system-wide (supply-chain) performance (see Cachon (2003)). We study three types of contracts in this paper: Wholesale-Price (WP), Buyback (BB) and Markdown-Money (MM). We refer the reader to Cachon (2003) for a thorough investigation of the first two contracts. Based on the literature and practice we describe the forms of these three contracts for this research work in the following paragraphs.

Wholesale-Price (WP) contract is the most basic form of contract without any cost (or risk) sharing mechanism between the supplier and the retailer. In this, the supplier charges a price of w per unit purchased by the retailer. This is in fact the model we introduced and analyzed for the retailer in Chapter 2. Compared to WP contract, in Buyback (BB) contract, the supplier shares some of the sales and inventory risk: the retailer receives a fixed return credit per unit, on all or portion of the inventory that is leftover at the end of the selling-season. Pasternack (1985) is the first to study this type of contract. The author analyzes this contract in the most general form, that is with two contract parameters: percentage of leftover that qualify for buyback and the buyback credit. One of the key results in this paper is that channel coordination is achieved if the manufacturer offers partial credit on all items unsold at the end of the season.

Markdown-Money (MM) contract received very little attention in literature and has been the subject of many controversies in the industry. Saks Fifth Avenue ran into legal trouble with its vendor, International Design Concepts LLC for allegedly collecting more than \$31 million in penalties for so-called markdown money for discounted merchandise (Byron (2005)). Internal Revenue Service (IRS) defines

“markdown money” as the payment by a vendor to a retailer to compensate the retailer for losses incurred because of the need to reduce the selling price of the vendor’s merchandise. The practice of “markdown money” also known as “guaranteed profit margin” has been severely criticized in industry articles (Gottlieb (2005), Insider (2006)) partly due to the vagueness surrounding what qualifies for the rebate and also due to the implementation (Agins (2006)). On the contrary, the research literature has few articles (Tsay (2001)) that particularly look at how contracts affect the operational decisions of the firms involved. In a MM contract, the supplier charges w per unit purchased by the retailer, but pays the retailer rebates for the qualified merchandise. From practice and the literature, all the merchandise that the retailer prices at less than a pre-specified, agreed upon, profit margin of g , qualifies for the rebate. From this, definitely all goods that are *sold* at price less than g margin qualify for rebate, and, depending on the agreement between the retailer and the supplier, the goods left unsold at the end of the horizon may qualify for rebate or other forms of chargeback. Krishnan & Soni (1997) look at the power of retailer and the vendor in a guaranteed profit margin setting. They consider two competing retailers and manufacturers and discuss the credibility of the threat posed by the different substitutable brands of the two manufacturers. Tsay (2001) develops a model for MM contract, considering a newsvendor model for the retailer where units that are salvaged qualify for MM rebates. Contrasting to this article, in our work, we develop a temporal model for the retailer, where the model includes timing and ordering decisions. Similar to our work, Mantrala et al. (2005) discuss the impact of MM contract on the profits of the retailer and the vendor when both

agree to a guaranteed profit margin mechanism. However, their model does not capture the uncertainty in secondary market demand. Lee & Rhee (2008) show that the MM contract cannot coordinate the supply chain; this is based on single-period, newsvendor-type setting.

The majority of the research models considered for the retail setting are single-period models, which are not adequate for representing the retailer's markdowns/pricing tactics or use of a secondary market. These are, however, crucial modeling elements especially if one is interested in the MM contracts. To the best of our knowledge, our model with primary and secondary markets with different revenue potential is the most comprehensive model built to study MM contracts. In addition to studying MM contracts in detail, the model also allows us to compare retailer's decisions and the performance of the supply chain with other types of contracts.

3.3. Problem Definition

We consider the single product, single retailer setting discussed in Chapter 2. The product is sold over the horizon $[0, T]$. The retailer can sell the good at a primary market or a secondary market (but not both) during this sales horizon. The retailer has to place an order from a supplier, long before the selling season. At the time of ordering, there is uncertainty about the state of the markets: there are two possible states. The primary and secondary markets can both be in a low state with probability θ_l and high state with probability θ_h ($\theta_l + \theta_h = 1$). The demand in

each of the markets and for each state is a random variable. The retailer knows the probability distribution of demand for each market and state. The retailer receives the goods from his vendor at the start of the sales season, after which the state of the system is revealed. At that point in time, the retailer decides the optimal length of time for the goods to be sold at the primary market (i.e., the optimal time to transfer the product from the primary to the secondary market). Prior to giving an order, the retailer and his supplier engage in a contract. While the retailer's decisions involve ordering and timing, the supplier's decision involves the choice of the wholesale price (corresponding to the the retailer's unit cost) for the contract.

The order of events is as follows: First, the retailer and the supplier agree to a contract and contract parameters, specifically, the buyback credit b in BB and the guaranteed profit margin g in MM. This stage is given and exogenous to our problem. Then, the supplier chooses w , which is the unit wholesale price of the good. Given a unit cost of w and the contract type, the retailer chooses his order quantity Q to maximize his expected profits for the entire sales season. After the order is given and after the state of the market is revealed, the retailer chooses β such that the goods are sold at the primary and secondary markets during $[0, \beta T)$ and $[\beta T, T]$, respectively. The sales revenues are obtained throughout the sales season in the corresponding markets: each unit sold for p_1 at the primary market and for p_2 in the secondary market, with $p_1 \geq p_2 \geq 0$. Once the sales season ends, the supplier pays rebates/credits (if any) to the retailer in accordance with the contract type and the contract parameters. This entire setup is illustrated in Figure 3.1. This sequence of events is very common for fashion apparel as discussed in “Lifecycle of

variable/fashion products” in SAS (2007).

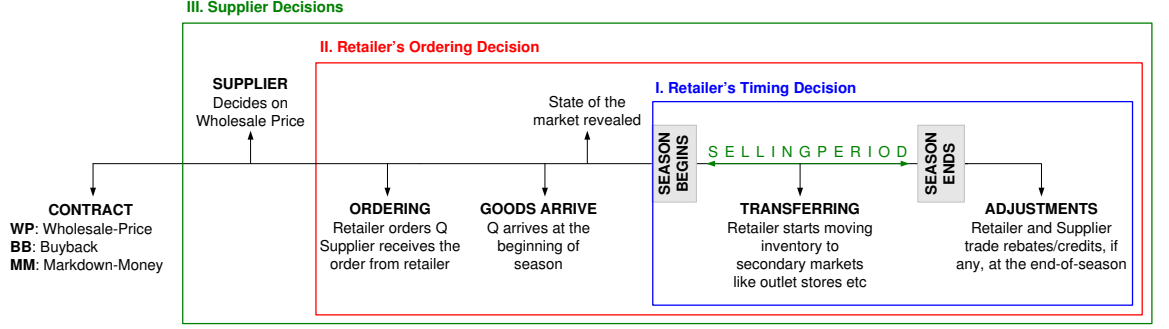


Figure 3.1: Timeline of Retail Operations with Retailer-Supplier Contracts

We provide a three-stage model below to solve for the supplier's and the retailer's optimal decisions.

Stage 1. Supplier Decisions. The supplier determines the optimal wholesale price w based on expected profit maximization where the supplier's revenues are based on the quantity sold and the wholesale price, and his costs are based on the contract terms (e.g., rebates to retailer) and cost of production. The supplier carries no inventory and produces to order. Let π_{sup}^{*c} be the optimal expected profit for the supplier for a given contract type c and unit cost of production ν . The supplier's problem is

$$\pi_{sup}^{*c} = \max_{w \geq \nu} E[(w - \nu)Q^{c*}(w) - ADJ^c] \quad (3.1)$$

The exact definition of cost of adjustments, ADJ^c , depends on the type of contract in place between the retailer and the supplier. It is equal to the rebates offered to the retailer in the case of MM contract or in the case of BB contract, it is the return credits offered to the retailer. $Q^{c*}(w)$ is the retailer's optimal order quantity for a

given wholesale price w and contract c .

Stage 2. Retailer's Ordering Decision. The retailer decides on the optimal order quantity Q^* based on expected profit maximization:

$$\pi_{ret}^{c*}(w) = \max_{Q \geq 0} \theta_l R_l^{c*}(Q) + \theta_h R_h^{c*}(Q) - wQ \quad (3.2)$$

where $\pi_{ret}^{c*}(w)$ is defined as the optimal expected profit for the retailer for a given contract c and wholesale price w . $R_s^{c*}(Q)$ represents the optimal expected revenue for the retailer for a given contract type c and order quantity Q when the state of the system is $s = l, h$.

Stage 3. Retailer's Timing Decision. The order is received and the state of the system is realized at the beginning of the season. The retailer next decides on the optimal length of time to sell the goods at the primary market, after which the sales will continue in the secondary market until time T . Thus, we have the following optimization problem for the retailer:

$$R_s^{c*}(Q) = \max_{0 \leq \beta \leq 1} E[R_s^c(\beta|Q)] \quad (3.3)$$

where $E[R_s^c(\beta|Q)]$ is the retailer's expected revenue for a given state of system s , contract c , and order quantity Q . To solve the retailer's timing problem, we have to discuss the details of the demand model and the end-of-season adjustments for each contract so that we can appropriately formulate the expected revenue function.

Demand Model. Similar to our discussion in Chapter 2, we note that the demand

model is critical in capturing the trade off between primary and secondary market sales. The model we analyzed in Chapter 2 took a static view of the aggregate market demand and assumed demand in a market was proportional to the length of time the product was offered in that market. In this chapter, we assume demand in each market during $[0, T]$ is governed by a *stochastic process*. Consequently, the demand in the primary and secondary markets are obtained by considering the evolution of the corresponding demand processes during $[0, \beta T)$ and $[\beta T, T]$, respectively. In the remainder of this chapter, we use a demand model that is based on a Poisson process.

Without loss of generality, let us normalize the length of the sales horizon to $T = 1$. Let $\lambda_1^s(t)$ and $\lambda_2^s(t)$ be the rate of demand for the primary market and the secondary market for state s , respectively. The retailer starts offering the product in the secondary market at time $\beta \in [0, 1]$. Let $N_1^s(\beta)$ and $N_2^s(\beta)$ be the demand at primary market in interval $[0, \beta)$ and demand at secondary market in interval $[\beta, 1]$. Then, according to Poisson model, we have, $N_1^s(\beta) \sim \text{Poisson}(\int_0^\beta \lambda_1^s(t)dt)$ and $N_2^s(\beta) \sim \text{Poisson}(\int_\beta^1 \lambda_2^s(t)dt)$. As a special case, we assume a stationary Poisson process. That is, assume $\lambda_1^s(t) = \lambda_1^s$ and $\lambda_2^s(t) = \lambda_2^s$ for all $t \in [0, 1]$. Then, we know, $N_1^s(\beta) \sim \text{Poisson}(\lambda_1^s\beta)$ and $N_2^s(\beta) \sim \text{Poisson}(\lambda_2^s(1 - \beta))$. Let us define $\mathbf{P}_j^1(\beta) = P(N_1^s(\beta) = j)$ and $\mathbf{P}_k^2(\beta) = P(N_2^s(\beta) = k)$. Then the probability mass

function of demand is

$$\mathbf{P}_j^1(\beta) = P(N_1^s(\beta) = j) = (\lambda_1^s \beta)^j \frac{e^{-\lambda_1^s \beta}}{j!}, \quad (3.4)$$

$$\mathbf{P}_k^2(\beta) = P(N_2^s(\beta) = k) = (\lambda_2(1 - \beta))^k \frac{e^{-\lambda_2(1 - \beta)}}{k!}. \quad (3.5)$$

In this model, we assume the demand in each market and state is an independent Poisson distributed random variable. Note that the expected demand in each market is linear in β : $E[N_1^s(\beta)] = \lambda_1^s \beta$ and $E[N_2^s(\beta)] = \lambda_2(1 - \beta)$. In contrast to the linear demand model introduced in Chapter 2, the demand takes only discrete values. Therefore, it is natural to restrict the order quantity to non-negative integers here.

Expected Revenue Function. Given the demand model, we are ready to formulate the expected revenue function for each contract type. For a WP contract the formulation is:

$$\begin{aligned} R_s^{WP*}(Q) &= \text{Max}_{0 \leq \beta \leq 1} E[R_s^{WP}(\beta|Q)] \\ &= \text{Max}_{0 \leq \beta \leq 1} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \left\{ p_1 \min(j, Q) + p_2 \min(k, (Q - j)^+) \right\} \mathbf{P}_j^1(\beta) \mathbf{P}_k^2(\beta) \end{aligned} \quad (3.6)$$

For a BB contract, the formulation is:

$$\begin{aligned} R_s^{BB*}(Q) &= \text{Max}_{0 \leq \beta \leq 1} E[R_s^{BB}(\beta|Q)] \\ &= \text{Max}_{0 \leq \beta \leq 1} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \left\{ p_1 \min(j, Q) + p_2 \min(k, (Q - j)^+) \right. \\ &\quad \left. + b(Q - j - k)^+ \right\} \mathbf{P}_j^1(\beta) \mathbf{P}_k^2(\beta) \end{aligned} \quad (3.7)$$

For a MM contract, the retailer and vendor agree on profit margin level of g . If the retailer has to drop prices below the agreed profit margin level, that is, if $p \leq w(1+g)$ where the unit price is p , then the retailer receives rebates at rebate per unit of $w(1+g) - p$. The formulation is:

$$\begin{aligned}
R_s^{MM*}(Q) &= \text{Max}_{0 \leq \beta \leq 1} E[R_s^{MM}(\beta|Q)] \\
&= \text{Max}_{0 \leq \beta \leq 1} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \left\{ (p_1 + (w(1+g) - p_1)^+) \min(j, Q) \right. \\
&\quad \left. + (p_2 + (w(1+g) - p_2)^+) \min(k, (Q - j)^+) \right\} \mathbf{P}_j^1(\beta) \mathbf{P}_k^2(\beta).
\end{aligned} \tag{3.8}$$

In the next section, we analyze the retailer's optimal decisions for different contract types.

3.4. Retailer's Timing Decisions

We analyze and discuss the retailer's ordering and timing decisions for WP, BB, and MM. We start with WP, analysis of which constitutes a stepping stone for the other two contracts. We drop the superscript s from the notation in this section because the problems have the same structure regardless of the market state s .

3.4.1 Wholesale-Price (WP) Contract

The retailer's model analyzed in Chapter 2 is in fact the model for the WP contract. We repeat the analysis under the new demand model, i.e., where the

demand is based on a Poisson process. We introduce new notation and state the structural properties of the retailer's timing problem below. Defining

$$V^{WP}(j, k) = p_1 \min(Q, j) + p_2 \min(k, (Q - j)^+), \quad (3.9)$$

we can rewrite the optimization problem in Eq. (3.6) as

$$R^{WP*}(Q) = \max_{0 \leq \beta \leq 1} E[R_s^{WP}(\beta|Q)] = \max_{0 \leq \beta \leq 1} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} V^{WP}(j, k) \mathbf{P}_j^1(\beta) \mathbf{P}_k^2(\beta) \quad (3.10)$$

Proposition 1 (a) *The expected revenue function $E[R_s^{WP}(\beta|Q)]$ is differentiable for $\beta \in [0, 1]$ and the partial derivative of the function with respect to β is*

$$\begin{aligned} \frac{\partial E[R(\beta|Q)]}{\partial \beta} &= (p_1 \lambda_1 - p_2 \lambda_2) P(N_1(\beta) + N_2(\beta) \leq Q) \\ &\quad + \lambda_1 (p_1 - p_2) P(N_1(\beta) \leq Q < N_1(\beta) + N_2(\beta)), \end{aligned} \quad (3.11)$$

(b) *For a given inventory level $Q < \infty$, $E[R(\beta|Q)]$ is quasiconcave in β .*

Proof (a) Differentiability of the function with respect to β is observed from the fact that β only appears in the Poisson probability components $\mathbf{P}_j^1(\beta) \mathbf{P}_k^2(\beta)$. For first order partial derivative and proof of quasiconcavity in part (b), see Appendix B.1.1. •

Corollary 1 *The optimal timing decision, denoted by β^* satisfies (i) $0 \leq \beta^* \leq 1$ if $p_1 \lambda_1 - p_2 \lambda_2 \leq 0$ and (ii) $\beta^* = 1$ if $p_1 \lambda_1 - p_2 \lambda_2 \geq 0$.*

This corollary follows from the first order condition above: When $p_1\lambda_1 - p_2\lambda_2 \geq 0$, then the expected revenue function is a nondecreasing function of β , leading to $\beta^* = 1$, otherwise, we have $0 \leq \beta^* \leq 1$.

The Poisson demand model and its properties are closely related to ‘marginal revenues’ and inventory on hand. Observe that the term $(p_1\lambda_1 - p_2\lambda_2)$ represents the marginal revenue received if we switch from the primary market to the secondary market provided we have enough inventory on hand to meet any increase in demand. Suppose by delaying the decision to transfer from the primary to the secondary market, the marginal revenue is negative. If the retailer had enough inventory to meet demand in both the markets then in such a case, the retailer would immediately transfer the product to the secondary market since keeping it in the primary market is not beneficial. But if the retailer does not have enough inventory, then, the trade-off between demand and the inventory level becomes important. Offering the product very early in secondary market can result in higher demand with insufficient inventory and thus lost sales. Offering too late in the season can result in lower demand with excess inventory unsold. Thus the retailer would like to time the transfer in such a way that the maximum revenues are obtained from both the markets with this inventory level and hence, the optimal timing of offering product in secondary market is given by an optimal value between 0 and 1. If the marginal revenue received is positive or it means you gain revenue when you delay the decision to transfer the product and hence you never switch from the primary market to the secondary market, that is, $\beta^* = 1$. To illustrate this, think of p_1 as a very high positive price and $p_2 = 0$, here, the timing decision is trivial; since it is never

optimal to transfer the goods to the secondary market.

3.4.2 Buyback (BB) Contract

The key difference between this contract and WP contract is that the retailer gets credit of b per unit, on all the items that are leftover at the end of the season. Since the retailer cannot profit from just leftover inventory hence it is reasonable to assume that $b \leq p_2$, as mentioned before in Section 3.2.

Let us define the following:

$$V^{BB}(j, k) = V^{WP}(j, k) + b(Q - j - k)^+ \quad (3.12)$$

where V^{WP} is defined as in (3.9). Then we can rewrite the optimization problem for BB:

$$R^{BB*}(Q) = \max_{0 \leq \beta \leq 1} E[R^{BB}(\beta|Q)] = \max_{0 \leq \beta \leq 1} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} V^{BB}(j, k) \mathbf{P}_j^1(\beta) \mathbf{P}_k^2(\beta) \quad (3.13)$$

Similar to WP contract we determine the optimal timing policy (or policies) by deriving the first and second order derivative of the expected revenue function.

Proposition 2 (a) *The first order derivative of the expected revenue function $E[R^{BB}(\beta|Q)]$ with respect to β is*

$$\begin{aligned} \frac{\partial E[R^{BB}(\beta|Q)]}{\partial \beta} = & \left[(p_1 - b)\lambda_1 - (p_2 - b)\lambda_2 \right] P(N_1(\beta) + N_2(\beta) \leq Q) \\ & + \lambda_1 [p_1 - p_2] P(N_1(\beta) \leq Q < N_1(\beta) + N_2(\beta)). \end{aligned} \quad (3.14)$$

(b) For a given inventory level Q , the expected revenue function $E[R^{BB}(\beta|Q)]$ is quasiconcave in β .

Proof See Appendix B.2.1. •

Corollary 2 If $p_1\lambda_1 - p_2\lambda_2 + b(\lambda_2 - \lambda_1) \geq 0$ then we have $\beta^* = 1$.

The proof of this corollary, for the optimality conditions follows from re-writing the expression for $\frac{\partial E[R^{BB}(\beta)]}{\partial \beta}$ given by Equation (3.14).

$$\begin{aligned} \frac{\partial E[R^{BB}(\beta|Q)]}{\partial \beta} &= \left[p_1\lambda_1 - p_2\lambda_2 + b(\lambda_2 - \lambda_1) \right] P(N_1(\beta) + N_2(\beta) \leq Q) \\ &\quad + \lambda_1 \left[p_1 - p_2 \right] P(N_1(\beta) \leq Q < N_1(\beta) + N_2(\beta)). \end{aligned}$$

This is immediate from the sign of the first term in $\frac{\partial E[R^{BB}(\beta|Q)]}{\partial \beta}$: When $p_1\lambda_1 - p_2\lambda_2 + b(\lambda_2 - \lambda_1) \geq 0$, then the function is non-decreasing in β and $\beta^* = 1$.

The above results highlight the importance of buyback credit in relation to marginal revenues when we compare BB to WP contract. That is, the marginal revenue (with respect to the timing decision), $p_1\lambda_1 - p_2\lambda_2$, in the WP contract is increased by a positive term $b(\lambda_2 - \lambda_1)$ in the BB contract. That is, when the demand in both the markets is less than the inventory level, then by delaying the decision to transfer to the secondary market the retailer gets an additional revenue of $b(\lambda_2 - \lambda_1)$ per unit. This shows that the retailer can bear the demand risk in the primary market for a longer period in the BB contract as the retailer is assured of a higher marginal revenue (compared to the WP contract) if the decision to transfer

to secondary market is delayed.

3.4.3 Markdown-Money (MM) Contract

In our model of the MM contract, the retailer gets rebates on all goods that are *sold* at less than the pre-specified margin, g regardless of the market they are sold in (we discussed there could be different arrangements between the retailer and the supplier under a MM contract in practice). Based on the prices the retailer charges at the primary and secondary markets, there are three possible cases: (i) The retailer gets rebate on all goods sold in both markets if $p_2 \leq p_1 \leq w(1 + g)$. The goods sold in the primary and secondary markets qualify for unit rebates of $p_1 - w(1 + g)$ and $p_2 - w(1 + g)$, respectively. (ii) The retailer gets a rebate for the products sold in only the secondary market if $p_2 \leq w(1 + g) < p_1$. This situation is typical to industry, where in a retailer prices aggressively in the primary market but marks down the prices towards the end of the selling season to avoid excess, unsold inventory. Each unit sold in the secondary market qualifies for a rebate of $p_2 - w(1 + g)$ in this case. (iii) If $w(1 + g) < p_2$, then the retailer gets no rebate. This is a trivial, and impractical scenario from a MM contract perspective. We summarize the conditions on the parameters for these three cases in Table 3.1:

Table 3.1: Price Conditions for Rebate Situations under MM contract

Rebate Situation	Condition on p_1	Condition on p_2
Rebate in both markets	$p_1 \leq w(1 + g)$	$p_2 \leq w(1 + g)$
Rebate in one market	$p_1 > w(1 + g)$	$p_2 \leq w(1 + g)$
No Rebate	$p_1 > w(1 + g)$	$p_2 > w(1 + g)$

Let us now define

$$V^{MM}(j, k) = V^{WP}(j, k) + (w(1 + g) - p_1)^+ \min(j, Q) \\ + (w(1 + g) - p_2)^+ \min(k, (Q - j)^+). \quad (3.15)$$

The function $V^{MM}(j, k)$ takes different forms in the case of different rebate situations; we provide the expressions for each case in Table 3.2.

Table 3.2: Evaluation of Revenue Function for MM contract

Rebate Situation	$V^{MM}(j, k)$
Rebate in both markets	$w(1 + g)\min(j, Q) + w(1 + g)\min(k, (Q - j)^+)$
Rebate in one market	$p_1\min(j, Q) + w(1 + g)\min(k, (Q - j)^+)$
No Rebate	$p_1\min(j, Q) + p_2\min(k, (Q - j)^+) = V^{WP}(j, k)$

Given the function $V^{MM}(\cdot)$, we can rewrite the retailer's timing problem as:

$$R^{MM*}(Q) = \max_{0 \leq \beta \leq 1} E[R^{MM}(\beta|Q)] = \max_{0 \leq \beta \leq 1} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} V^{MM}(j, k) \mathbf{P}_j^1(\beta) \mathbf{P}_k^2(\beta) \quad (3.16)$$

Based on Table 3.2, we note the following in the case of different rebate situations.

- (i) When the retailer qualifies for rebate in both markets, the expected revenue function of the retailer is similar to one in the WP contract with prices in both markets replaced by $w(1 + g)$.
- (ii) When the retailer qualifies for rebate for the goods sold in the secondary market only, then we have to replace p_2 of WP with $w(1 + g)$ as the unit revenue obtained in the secondary market under MM contract.
- (iii) When no sales qualifies for the rebate in the MM contract, then the situation

is exactly the same as the WP contract. Therefore, we drop this third case from further consideration.

Proposition 3 (a) *The first order derivative of the expected revenue function with respect to β for MM contract is given as*

(i) *When $p_2 \leq p_1 \leq w(1 + g)$,*

$$\frac{\partial E[R^{MM}(\beta|Q)]}{\partial \beta} = w(1 + g)[\lambda_1 - \lambda_2] P(N_1(\beta) + N_2(\beta) \leq Q). \quad (3.17)$$

(ii) *When $p_2 \leq w(1 + g) < p_1$,*

$$\begin{aligned} \frac{\partial E[R^{MM}(\beta|Q)]}{\partial \beta} = & [p_1 \lambda_1 - w(1 + g) \lambda_2] P(N_1(\beta) + N_2(\beta) \leq Q) \\ & + \lambda_1 [p_1 - w(1 + g)] P(N_1(\beta) \leq Q < N_1(\beta) + N_2(\beta)). \end{aligned} \quad (3.18)$$

(b) *For a given inventory level Q and wholesale price w , the retailer's expected revenue function for a MM contract is quasiconcave in β .*

Proof See Appendix B.3.1. •

In contrast to the WP contract, we notice from the marginal revenue terms in Equations (3.17) and (3.18) that the retailer would like to take rebates if it is possible at all to avail the rebates.

Corollary 3 *The optimal timing for MM contract satisfies:*

(i) *If $p_2 < p_1 \leq w(1 + g)$ then $\beta^* = 0$*

(ii) If $p_2 \leq w(1 + g) < p_1$ and $p_1\lambda_1 - w(1 + g)\lambda_2 > 0$ then $\beta^* = 1$.

This follows from the first order derivatives: In part (i), the first order derivative is negative which implies $\beta^* = 0$. In part (ii), it is positive, hence $\beta^* = 1$.

3.5. Comparison of Retailer's Timing Decisions

In this section we analytically compare the retailer's optimal timing of offering the product in the secondary market for different contracts for a given order quantity Q . As the results hold true for any state s , we will drop the subscript s in this section.

Proposition 4 *For a given wholesale price w and order quantity Q , let the optimal timing in WP and BB contracts be denoted $\beta_{WP}^*(Q)$ and $\beta_{BB}^*(Q)$, respectively. Then, we have $\beta_{WP}^*(Q) \leq \beta_{BB}^*(Q)$.*

Proof The proof for this follows from comparing the first order derivatives for WP and BB contract in Equations (3.11) and (3.14). •

This result is intuitive since when we have the BB contract, the retailer is willing to be more aggressive (in terms of price) by keeping the product in the primary market for a longer time, as the vendor is sharing some of his/her risk of unsold items. Two corollaries that follow from the above proposition are:

Corollary 4 *If we have $\beta_{WP}^*(Q) = 1$ then $\beta_{BB}^*(Q) = 1$.*

Corollary 5 *If we have $\beta_{BB}^*(Q) = 0$ then $\beta_{WP}^*(Q) = 0$.*

Proposition 5 *For a given wholesale price w and order quantity Q , let the optimal timing in WP and MM contracts be denoted $\beta_{WP}^*(Q)$ and $\beta_{MM}^*(Q)$, respectively. Then, we have $\beta_{MM}^*(Q) \leq \beta_{WP}^*(Q)$.*

Proof In the case of the rebate situation corresponding to “rebate in both markets,” we know that $\beta_{MM}^*(Q) = 0$ (see Corollary 3) where as $\beta_{WP}^*(Q)$ can take any value between $[0, 1]$. Consider the case of rebate situation corresponding to “rebate in one market”. We can derive the following relations in this rebate situation: $-w(1 + g)\lambda_2 \leq -p_2\lambda_2$ and $[p_1 - w(1 + g)]\lambda_1 \leq [p_1 - p_2]\lambda_1$. From Equations (3.18) and (3.11) and using the relations just derived, we see that $\frac{\partial E[R^{MM}(\beta)]}{\partial \beta} \leq \frac{\partial E[R^{WP}(\beta)]}{\partial \beta}$ for “rebate in one market” situation. This completes the proof. •

From this result we see that unlike a BB contract, the retailer does not delay offering the product in the secondary market under a MM contract compared to a WP contract. This is because there are three possible situations. When both the prices are higher than the agreed margin of $w(1 + g)$, MM contract behaves just like WP contract. When both the prices are lower than the agreed margin of $w(1 + g)$, then the retailer would like to take rebate as early as possible since the rebate is provided on all items that are sold; and the retailer receives highest demand for lower price, which happens in the secondary market. In the case of the primary market price being higher than the cutoff price and the secondary market price being lower than cutoff, there is a trade-off between delaying offering the product in secondary market to receive rebates on higher demand to receiving revenue for higher price but possibly lower demand in primary market. Why would the retailer like to offer the

product in the secondary market earlier under a MM contract in this latter rebate situation, compared to WP contract? In the case of a WP contract, the retailer receives marginal revenue of $p_1\lambda_1 - p_2\lambda_2$ per unit time delay in offering the product in the secondary market, where as in a MM contract in this rebate situation, the retailer receives a marginal revenue of $p_1\lambda_1 - w(1+g)\lambda_2$ and since $p_2 \leq w(1+g)$, the retailer's marginal revenue is lower if the product is offered in the secondary market at a later time when compared to a WP contract.

Proposition 6 *For a given Q and w , $\beta_{MM}^*(Q) \leq \beta_{BB}^*(Q)$.*

Proof Follows from Propositions 4 and 5 . •

While these results convey practical information on the use of secondary markets, they are valid when the same order quantity, Q , is used in all contracts. However, the retailer's optimal order quantity can be different for each contract. Comparison of optimal timing decisions for optimal order quantities (corresponding to each contract) are done later in this chapter using a computational experiment. Furthermore, the supplier's optimal choice of w depends on the optimal order quantity of the retailer and vice versa. Hence, the relation among the optimal timing decisions for different contract types is not straightforward when we consider the actions of both the retailer and the supplier.

3.6. Retailer's Ordering and Supplier's Wholesale Pricing Decisions

We solve the three-stage model starting from the third stage and going backwards. Once the expected revenue from sales and rebates is formulated for the

retailer, the optimal time to transfer the goods can be determined. Once the optimal timing decision for a given order quantity is known, then retailer's Stage-2 problem is solved for each contract type:

$$\pi_{ret}^{c*}(w) = \text{Max}_{Q \geq 0} \quad \theta_l R_l^{c*}(Q) + \theta_h R_h^{c*}(Q) - wQ \quad (3.19)$$

where $\pi_{ret}^{c*}(w)$ is the optimal expected profit for the retailer for a given contract c and wholesale price w . The complexity of the model does not allow us to obtain closed-form solutions for the retailer's ordering problem. We determine Q^* or optimal order quantity using computational methods. Note that we restrict ourselves to discrete order quantities because the demand is discrete.

Finally, we can analyze the actions of the supplier, knowing the retailer's optimal actions for a given contract type and a wholesale price. The supplier's wholesale pricing problem has a different formulation for each contract. In case of a WP contract, we can write the supplier's optimization problem as:

$$\pi_{sup}^{WP*} = \text{Max}_{w \geq \nu} (w - \nu)Q^{WP*}(w) \quad (3.20)$$

$Q^{WP*}(w)$ is the optimal order quantity for a retailer for a given wholesale price w .

In case of a BB contract, the supplier's expected profit maximization problem

is as follows:

$$\begin{aligned} \pi_{sup}^{BB*} = \text{Max}_{w \geq \nu} & (w - \nu)Q^{BB*}(w) \\ & - b \sum_{s=l,h} \theta_s \left(\sum_{j=0}^{\infty} \sum_{k=0}^{\infty} (Q^{BB*}(w) - j - k)^+ \mathbf{P}_j^1(\beta_s^*) \mathbf{P}_k^2(\beta_s^*) \right) \end{aligned} \quad (3.21)$$

The first term is the profit from delivering $Q^{BB*}(w)$, the retailer's optimal order quantity for a given wholesale price w and β_s^* is the corresponding optimal time of transfer of goods from the primary to the secondary market, for state s . The second term denotes the expected return credits offered to the retailer over all possible demand states.

In the case of a MM contract, we can write the supplier's optimization problem as:

$$\begin{aligned} \pi_{sup}^{MM*} = \text{Max}_{w \geq \nu} & (w - \nu)Q^{MM*}(w) \\ & - \sum_{s=l,h} \theta_s \left\{ \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \left[(w(1+g) - p_1)^+ \min(Q^{MM*}(w), j) \right. \right. \\ & \left. \left. + (w(1+g) - p_2)^+ \min((Q^{MM*}(w) - j)^+, k) \right] \mathbf{P}_j^1(\beta_s^*) \mathbf{P}_k^2(\beta_s^*) \right\} \end{aligned} \quad (3.22)$$

The first term is the profit from delivering $Q^{MM*}(w)$, the optimal order quantity for a retailer for a given wholesale price w and β_s^* is the corresponding optimal time of transfer of goods from the primary to the secondary market, for state s . The second term denotes the expected rebates offered to the retailer over all possible demand

states.

We conduct extensive numerical experiments to understand the properties of the expected profit functions of the retailer and the supplier in the next section.

3.7. Computational Experiments

We conduct computational experiments to understand the effects of WP, BB and MM contracts on the optimal order quantity, expected profit and supplier's choice of wholesale price.

3.7.1 Setup and Design

We design the computational experiment to investigate the effect of various problem parameters on the retailer and supplier optimal actions. The values considered for different parameters are as follows:

- State: Market state probabilities (θ) vary from 0.1 to 0.9. The state probabilities are varied in steps of 0.1 in the numerical exercises.
- Demand Model: In our experiments, we use the Poisson demand model where the demand in the primary and secondary markets can also be a function of prices in the respective markets. Consider the Poisson arrival process and assume that each arriving customer has a willingness-to-pay (WTP). A customer in the primary or secondary market purchases the product if his/her WTP exceeds the price p_1 or p_2 in the respective markets. This effectively reduces the sales rate in corresponding markets to $\lambda_1^s \bar{F}_1(p_1)$ or $\lambda_2^s \bar{F}_2(p_2)$, where

$\bar{F}_1(p_1) = 1 - F_1(p_1)$ and $\bar{F}_2(p_2) = 1 - F_2(p_2)$. Here, $F_1(\cdot)$ and $F_2(\cdot)$ represent the WTP distribution of the customers in the respective markets. This joint WTP and Poisson arrival type of demand model is commonly used in the pricing literature (e.g., Bitran & Mondschein (1997)). In our model, this translates to primary and secondary market demand distributions of $N_1^s(\beta) \sim \text{Poisson}(\lambda_1^s \bar{F}_1(p_1)\beta)$ and $N_2^s(\beta) \sim \text{Poisson}(\lambda_2^s \bar{F}_2(p_2)(1 - \beta))$, respectively.

- Demand Parameters: We use a common demand rate (λ) for Poisson process. λ takes one of the following values: 10, 30 and 50 in our experiments. Consumer's WTP (F) follows from this: $U[0, 150]$, $U[0, 200]$ and $U[0, 300]$. This affects the overall sales rates as mentioned above. The parameter λ , prices p_1 and p_2 , and the WTP determine the probability distribution of the demand in any market.
- Prices: We consider two sequences of price paths (p_1, p_2) for the two markets. In first sequence we have price paths of (100, 80), (100, 60), (100, 40) and (100, 20). In second sequence we have the following price paths: (120, 100), (120, 80), (120, 60) and (120, 40).
- Contract Parameters: Guaranteed profit margin (g) varies from 0 to 1.2 in steps of 0.1. Buyback credit (b) varies from 0 to 30 in steps of 5.
- Cost of Supplier: Finally, the marginal cost of production for supplier (ν) is assumed to be 15 .

- Optimal decisions: In each experiment, we determine the optimal wholesale price, w^* , for the supplier, and the optimal order quantity, Q^* and timing, β^* for the retailer for each contract. The algorithm used in optimization is described in Appendix B.4.

For each of the contracts we conduct the following experiments: In the *first experiment*, we study the effect of contract parameters (if any) on the optimal decisions of the retailer and the supplier. In the *second experiment* we discuss the properties and behavior of optimal decisions for a set of parameter values. In the *third experiment*, we study the effect of the % discount in secondary market's price on optimal decisions of the supplier and retailer, that is, p_1 vs. p_2 . In the *fourth*, we study the effect of WTP distribution parameters. Finally, in the *fifth*, we study the effect of market state probabilities, that is, θ_l, θ_h .

We conduct a comprehensive computational experiment. Before we provide the numerical results, we discuss the major findings in the next Section 3.7.2. In Sections 3.7.3 - 3.7.5, we discuss the optimal decisions for the WP, BB and MM contracts. In Section 3.7.6, we compare the optimal decisions across the three contracts. In Section 3.7.7, we discuss the preference of the retailer and the supplier for each of the contracts when products are of a high-risk and high-return nature.

3.7.2 Summary of Results

We conduct a comprehensive series of experiments, with different goals in mind. Here is a summary of our main observations: (1) The time to transfer the

products from the primary to secondary markets is affected by contractual agreements - BB transfers later and MM transfers earlier compared to WP contract. (2) The retailer places higher order quantities under a MM contract compared to a WP contract. However, the order quantities are marginally higher in a BB contract when compared to a WP contract. (3) The retailer's profits under BB contract can be lower compared to WP contract. In contrast, the supplier is always better off with a BB contract compared to a WP contract. A MM contract is more beneficial for the retailer compared to a WP contract but can also be a win-win situation for both parties in the supply chain. (4) Considering the uncertainty in the state of the market: When the potential demand market is likely to be in a low state, then the retailer prefers MM contract and on the other hand the supplier prefers the BB contract. This is a significant finding for the suppliers and the retailers as it helps them to quantify the effectiveness of the individual contracts under potential demand market scenarios.

3.7.3 Optimal Decisions under WP Contract

For the WP contract, the *first* and *second* experiments are coupled: we investigate the properties of optimal decisions for the retailer and the supplier. We use the following values for the parameters: $\nu = 15$, and $F^l, F^h \sim U[0, 150]$. The demand arrival rates are given as follows: In the primary market we have $\lambda_1^l = 10(1 - \frac{p_1}{150})$, $\lambda_1^h = 50(1 - \frac{p_1}{150})$ and in secondary market we have $\lambda_2^l = 10(1 - \frac{p_2}{150})$, $\lambda_2^h = 50(1 - \frac{p_2}{150})$. Tables 3.3 and 3.4 show the results of the optimization with state probabilities as

(0.1, 0.9) and (0.7, 0.3) respectively. In general we note the following observations from this numerical exercise:

- The optimal order quantity, Q^* decreases with increasing w . We illustrate this for the different price paths in Figure 3.2. Notice that Q^* stays constant for a certain range of wholesale prices and drops to lower Q when w is increased. This happens because we considered Q as discrete for this research.
- The optimal retailer's expected profit, π_{ret}^* decreases with increasing w . This is shown in Figure 3.3.
- The supplier's expected profit, π_{sup} is not concave as shown in Figure 3.3. We believe this happens because (i) Q is discrete, and (ii) for multi-stage optimization problems, some of the essential properties like continuity and differentiability are lost (in general) from one stage to the next stage. However, based on computational experiments, we observe that the supplier's expected profit function exhibits a uni-modal structure, allowing us to find the optimal wholesale price. Note that the w^* does not necessarily lie between the two prices, p_2 and p_1 .

In the *third experiment*, we varied the discounted price in the secondary market, keeping other parameters the same as the *second experiment*. The results are shown in Tables 3.3 and 3.4. We observe that:

- The retailer's optimal time to transfer to the secondary market increases with p_2 in this experiment. However, this need not be true in other cases and the

Table 3.3: WP Contract - Results for different Price Paths with $(\theta_l, \theta_h) = (0.1, 0.9)$

(p_1, p_2)	w^*	π_{sup}^*	π_{ret}^*	Q^*	β_l^*	β_h^*
(100, 80)	76	793.000	187.380	13	0.0003	0.9433
(100, 60)	76	793.000	185.373	13	0.0003	0.9834
(100, 40)	76	793.000	182.578	13	0.9997	0.9997
(100, 20)	76	793.000	182.548	13	0.9997	0.9997
(120, 100)	58	645.000	432.708	15	0.0003	0.0003
(120, 80)	88	584.000	153.353	8	0.0003	0.8989
(120, 60)	87	576.000	155.532	8	0.0003	0.9452
(120, 40)	85	560.000	161.516	8	0.2583	0.9809

Table 3.4: WP Contract - Results for different Price Paths with $(\theta_l, \theta_h) = (0.7, 0.3)$

(p_1, p_2)	w^*	π_{sup}^*	π_{ret}^*	Q^*	β_l^*	β_h^*
(100, 80)	46	186.000	144.297	6	0.0003	0.9721
(100, 60)	60	180.000	82.303	4	0.9202	0.9928
(100, 40)	75	180.000	36.984	3	0.9997	0.9997
(100, 20)	75	180.000	36.975	3	0.9997	0.9997
(120, 100)	65	200.000	85.451	4	0.0003	0.9008
(120, 80)	62	235.000	90.720	5	0.0003	0.9352
(120, 60)	47	224.000	142.933	7	0.0003	0.9534
(120, 40)	36	147.000	173.298	7	0.4217	0.9840

Table 3.5: WP Contract - Results for different WTP distribution parameters

(F^l, F^h)	w^*	π_{sup}^*	π_{ret}^*	Q^*	β_l^*	β_h^*
$U[0, 150], U[0, 150]$	78	504.000	130.184	8	0.9753	0.9753
$U[0, 150], U[0, 200]$	83	748.000	145.784	11	0.9521	0.9997
$U[0, 150], U[0, 300]$	81	990.000	212.085	15	0.8150	0.9997
$U[0, 200], U[0, 150]$	80	520.000	118.470	8	0.9997	0.9753
$U[0, 300], U[0, 150]$	80	520.000	118.754	8	0.9997	0.9753

optimal order quantity (in particular a jump in Q as p_2 changes) can positively or negatively impact β^* .¹

- The retailer's optimal expected profit does not show any well-defined trend with a drop in the discounted price (or higher discounts). But the supplier's optimal expected profit exhibits quasi-concavity in the discounted price.

In the *fourth experiment* the goal is to study the effect of parameters of the WTP distribution on the optimal solutions. we consider price path of $(p_1, p_2) = (100, 60)$ with the following parameters: $\nu = 15$, $\theta_l = 0.1$, $\theta_h = 0.9$, $\lambda^l = \lambda^h = 30$ and $F^l, F^h \sim U[0, 150], U[0, 200], U[0, 300]$. The demand rates for state s are given by $\lambda_1^s = 30\bar{F}^s(p_1)$ and $\lambda_2^s = 30\bar{F}^s(p_2)$. The results as shown in Table 3.5. We observe the following:

- The optimal timing of transfer to secondary market is affected by the variance (range) of demand in the primary market. As the variance increases, there is a higher probability of high-price paying customers in the primary market. Consequently, switch to the secondary market is delayed.
- The supplier's optimal expected profits are non-decreasing in the variance of the demand in the primary market. This is also true for the retailer's optimal expected profits. The supplier's expected profit increases with the variance of the WTP distribution in the secondary market, although retailer observes decline in expected profits in this case.

¹ β^* is not monotonic in general. Consider $p_1 = 47, p_2 = \{11, 12, 13\}$; $F^l \sim U[0, 50], F^h \sim U[0, 50]$; $\lambda_1^l = 3/5, \lambda_1^h = 3$; $\lambda_2^l = 10(1 - p_2/50), \lambda_2^h = 50(1 - p_2/50)$; $\theta_l = 0.1$. Then, β_l^* is 0.7123, 0.5244, and 0.6634 when p_2 is 11, 12, and 13, respectively. This is due to the fact that Q^* is 2, 3, and 2 when p_2 is 11, 12, and 13, respectively.

- This analysis tells us that the suppliers prefer market states which have higher variance.

In the *fifth experiment* the goal is to understand the effect of different θ_l and θ_h for two price paths in primary and secondary markets: $(p_1, p_2) = (100, 60)$ and $(120, 80)$. For both price paths we assume that $\nu = 15$, $F^l \sim U[0, 150]$ and $F^h \sim U[0, 150]$. The demand rates are given as follows: In primary market we have $\lambda_1^l = 10(1 - \frac{p_1}{150})$, $\lambda_1^h = 50(1 - \frac{p_1}{150})$ and in secondary market we have $\lambda_2^l = 10(1 - \frac{p_2}{150})$, $\lambda_2^h = 50(1 - \frac{p_2}{150})$. We vary the market state probabilities in steps of 0.1 from 0.1 to 0.9. The results

Table 3.6: WP Contract - Results under different Market State Probabilities for (100, 60)

(θ_l, θ_h)	w^*	π_{sup}^*	π_{ret}^*	Q^*	β_l^*	β_h^*
(0.1, 0.9)	76	793.000	185.373	13	0.0003	0.9854
(0.2, 0.8)	72	684.000	151.005	12	0.0003	0.9853
(0.3, 0.7)	63	576.000	177.019	12	0.0003	0.9853
(0.4, 0.6)	54	468.000	202.032	12	0.0003	0.9853
(0.5, 0.5)	48	363.000	195.283	11	0.0003	0.9865
(0.6, 0.4)	39	264.000	221.208	11	0.0003	0.9865
(0.7, 0.3)	60	180.000	82.303	4	0.9202	0.9928
(0.8, 0.2)	71	168.000	43.761	3	0.9458	0.9928
(0.9, 0.1)	68	159.000	47.357	3	0.9458	0.9928

are shown in Tables 3.6 and 3.7. Here is a summary of our observations:

- The optimal order quantity, Q^* , decreases and β^* increases as θ_l increases.

This is because the overall market demand decreases with θ_l , reducing the number of units that can be sold. Notice that when $\theta_l = 0.1$, high state is very likely, and retailer places an order of 13 units. It is interesting to note that if retailer finds himself in the low state, then he would switch immediately to

Table 3.7: WP Contract - Results under different market state probabilities for (120, 80)

(θ_l, θ_h)	w^*	π_{sup}^*	π_{ret}^*	Q^*	β_l^*	β_h^*
(0.1, 0.9)	88	584.000	153.353	8	0.0003	0.8989
(0.2, 0.8)	79	512.000	170.829	8	0.0003	0.8989
(0.3, 0.7)	79	448.000	124.900	7	0.0003	0.9139
(0.4, 0.6)	69	378.000	149.282	7	0.0003	0.9139
(0.5, 0.5)	69	324.000	111.850	6	0.0003	0.9258
(0.6, 0.4)	70	275.000	78.540	5	0.0003	0.9352
(0.7, 0.3)	62	235.000	90.720	5	0.0003	0.9352
(0.8, 0.2)	54	195.000	102.901	5	0.0003	0.9352
(0.9, 0.1)	46	155.000	115.081	5	0.0003	0.9352

secondary market. This looks counter-intuitive since $p_2 = 60$, is lower than $w^* = 76$. This can be explained when one calculates the expected profits, taking into account the order quantity.

- The supplier's optimal expected profits are decreasing as the lower market state is more probable. The retailer's optimal expected profits do not exhibit monotonic behavior.
- This analysis tells us that the supplier prefers higher market states where as the retailer's preference is influenced not only by the state of the market but also the supplier's choice of the optimal wholesale price.

3.7.4 Optimal Decisions under BB Contract

In the *first experiment* for studying the effect of buyback credit on the optimal solutions we consider the following parameters: $\nu = 15$, $F^l \sim U[0, 150]$, $F^h \sim$

$U[0, 150]$ and price path as $(p_1, p_2) = (100, 60)$. The demand rates are given as follows: In primary market we have $\lambda_1^l = \frac{10}{3}$, $\lambda_1^h = \frac{50}{3}$ and in secondary market we have $\lambda_2^l = 6$, $\lambda_2^h = 30$. The buyback credit is dropped from 40 to 0. From the results

Table 3.8: BB Contract - Results for different Buyback Credits with $(p_1, p_2) = (100, 60)$

b	w^*	π_{sup}^*	π_{ret}^*	Q^*	β_l^*	β_h^*	$RetRev_l^*$	$RetRev_h^*$
40	81.5	867.219	163.246	14	0.9997	0.9997	426.633	23.464
30	83.0	845.196	130.398	13	0.9997	0.9997	289.975	10.896
25	82.0	838.663	136.931	13	0.9997	0.9997	241.646	9.080
20	81.0	832.131	143.464	13	0.9997	0.9997	193.317	7.264
18	80.5	828.218	147.377	13	0.9997	0.9997	173.985	6.538
15	79.5	819.098	156.496	13	0.9997	0.9997	144.988	5.448
12	79.0	816.478	159.116	13	0.9997	0.9997	115.990	4.358
10	78.5	812.565	163.029	13	0.9997	0.9997	96.659	3.632
5	77.5	807.441	170.900	13	0.0003	0.9909	35.033	1.729
0	76.0	793.000	185.373	13	0.0003	0.9834	0.000	0.000

of optimization as shown in Table 3.8 we can observe that:

- In general, the optimal timing does not vary with change in buyback credit.

If we look at the expression $p_1\lambda_1 - p_2\lambda_2 + b(\lambda_2 - \lambda_1)$ for low demand state it equals $\frac{1}{3}(8b - 80)$ and for high demand state it is $\frac{1}{3}(100b - 400)$. From Corollary 2, we note that for low state $\beta^* = 1$ when $b \geq 10$ and for high state when $b \geq 4$. When the buyback credit approaches 0 then BB contract behaves like WP contract and we see optimal timings close to what we observed in WP contract.

- One of the important observations is that the supplier's optimal expected profit increases monotonically in the buyback credit. The retailer obtains

higher profits at lower buyback credit. The highest profit attained by the retailer corresponds to that of a WP contract. Notice that the total supply chain profit remains constant for buyback credit between 30 and 0 as the order quantity does not change. The supply chain profit just gets re-distributed between the retailer and the supplier.

Table 3.9: BB Contract - Results for different Price Paths with $(\theta_l, \theta_h) = (0.1, 0.9)$

(p_1, p_2)	w^*	π_{sup}^*	π_{ret}^*	Q^*	β_l^*	β_h^*	$RetRev_l^*$	$RetRev_h^*$
(100, 80)	78.0	807.743	172.567	13	0.0003	0.9596	83.343	3.247
(100, 60)	78.5	812.565	163.029	13	0.9997	0.9834	96.659	3.632
(100, 40)	78	806.069	169.509	13	0.9997	0.9997	96.654	3.629
(100, 20)	78	806.072	169.476	13	0.9997	0.9997	96.650	3.626
(120, 100)	58	663.643	452.846	16	0.0003	0.0003	126.671	12.989
(120, 80)	43	609.868	612.189	23	0.0003	0.0003	183.342	17.554
(120, 60)	89	586.917	144.426	8	0.0003	0.9590	23.150	3.076
(120, 40)	87	565.875	152.422	8	0.9997	0.9997	59.989	4.585

In the *second experiment* we study the behavior and properties of optimal decisions using the following values for the parameters: $\nu = 15$, $F^l \sim U[0, 150]$ and $F^h \sim U[0, 150]$. We assume buyback credit of 10 for this experiment. The demand rates are given as follows: In primary market we have $\lambda_1^l = 10(1 - \frac{p_1}{150})$, $\lambda_1^h = 50(1 - \frac{p_1}{150})$ and in secondary market we have $\lambda_2^l = 10(1 - \frac{p_2}{150})$, $\lambda_2^h = 50(1 - \frac{p_2}{150})$. Tables 3.9 and 3.10 show the results of the optimization with state probabilities as $(0.1, 0.9)$ and $(0.7, 0.3)$ respectively. In general, the following is observed:

- The optimal order quantity decreases as w increases. One key difference in this behavior compared to WP contract is that the optimal order quantity

approaches infinity for very low wholesale prices. This is because the retailer is assured of revenue on the leftover inventory at the end of the horizon due to the BB contract. This is illustrated in Figure 3.4.

- Similar to the WP contract, the retailer's optimal expected profit drops with increasing wholesale price and the supplier's optimal expected profit is not concave as shown in Figure 3.5.

Table 3.10: BB Contract - Results for different Price Paths with $(\theta_l, \theta_h) = (0.7, 0.3)$

(p_1, p_2)	w^*	π_{sup}^*	π_{ret}^*	Q^*	β_0^*	β_1^*	$RetRev_0^*$	$RetRev_1^*$
(100, 80)	46	199.365	156.542	7	0.0003	0.9596	25.178	0.034
(100, 60)	63	184.191	77.689	4	0.9997	0.9834	11.156	0.001
(100, 40)	63	184.193	77.679	4	0.9997	0.9997	11.153	0.001
(100, 20)	63	184.194	77.662	4	0.9997	0.9997	11.151	0.001
(120, 100)	69	208.160	77.290	4	0.0003	0.9095	11.163	0.087
(120, 80)	64	237.681	88.036	5	0.0003	0.9440	10.345	0.256
(120, 60)	50	233.514	133.387	7	0.0003	0.9652	15.708	1.634
(120, 40)	67	147.472	65.103	3	0.9997	0.9997	12.169	0.033

In the *third experiment* we varied the discounted price, keeping other parameters the same as in the *second experiment* and results are shown in Tables 3.9 and 3.10.

We observe that:

- Similar to the WP contract, optimal timing decision is not monotonic in higher discounts. Based on Corollary 2 we see that for price paths of (100, 40) and (100, 20) the optimal timing is 1 as $p_1\lambda_1 - p_2\lambda_2 + b(\lambda_2 - \lambda_1) > 0$.
- The retailer's and the supplier's optimal expected profit show similar behavior

as discussed in the WP contract. π_{ret}^* does not show any trend and π_{sup}^* exhibits quasi-concavity.

The goal of the *fourth experiment* is to study the effect of parameters of the WTP distribution on the optimal solutions. We consider price path of $(p_1, p_2) = (100, 60)$ and assume the following parameters: $\nu = 15$, $b = 10$, $\theta_l = 0.1$, $\theta_h = 0.9$, $\lambda^l = \lambda^h = 30$ and $F^l, F^h \sim U[0, 150], U[0, 200], U[0, 300]$. The demand rates for state s is given by $\lambda_1^s = 30\bar{F}^s(p_1)$ and $\lambda_2^s = 30\bar{F}^s(p_2)$. We obtain the results as shown in Table 3.11.

- From Tables 3.11 and 3.12, it is optimal to keep the product in the primary market as long as possible. This is an effect of the buyback credit.
- The supplier's optimal expected profits are increasing in the variance of the WTP distribution in either market. This is not true for the retailer's optimal expected profits.

Table 3.11: BB Contract - Results for different WTP distribution parameters

(F_l, F_h)	w^*	π_{sup}^*	π_{ret}^*	Q^*	β_l^*	β_h^*	$RetRev_l^*$	$RetRev_h^*$
$U[0, 150], U[0, 150]$	80	515.402	118.568	8	0.9997	0.9997	4.598	4.598
$U[0, 150], U[0, 200]$	85	765.872	127.803	11	0.9997	0.9997	18.327	2.551
$U[0, 150], U[0, 300]$	83	1012.646	188.768	15	0.9997	0.9997	51.019	2.503
$U[0, 200], U[0, 150]$	81	523.832	114.446	8	0.9997	0.9997	0.295	4.598
$U[0, 300], U[0, 150]$	82	531.861	106.701	8	0.9997	0.9997	0.011	4.598

In the *fifth experiment* the goal is to understand the effect of different θ_l and θ_h for two price paths in primary and secondary markets: $(p_1, p_2) = (100, 60)$ and

Table 3.12: BB Contract - Results under different Market State Probabilities for (100, 60)

(θ_l, θ_h)	w^*	π_{sup}^*	π_{ret}^*	Q^*	β_l^*	β_h^*	$RetRev_l^*$	$RetRev_h^*$
(0.1, 0.9)	78.5	812.565	163.029	13	0.9997	0.9997	96.659	3.632
(0.2, 0.8)	71.0	705.763	176.803	13	0.9997	0.9997	96.659	3.632
(0.3, 0.7)	63.0	592.460	197.078	13	0.9997	0.9997	96.659	3.632
(0.4, 0.6)	58.0	480.072	180.593	12	0.9997	0.9997	86.659	2.108
(0.5, 0.5)	48.0	378.855	224.627	13	0.9997	0.9997	96.659	3.632
(0.6, 0.4)	42.0	271.162	220.394	12	0.9997	0.9997	86.659	2.108
(0.7, 0.3)	63.0	184.191	77.689	4	0.9997	0.9997	11.156	0.001
(0.8, 0.2)	74.0	172.658	38.901	3	0.9997	0.9997	5.427	0.000
(0.9, 0.1)	71.0	163.115	43.014	3	0.9997	0.9997	5.427	0.000

$(p_1, p_2) = (120, 80)$. We assume the following values for the parameters: $\nu = 15$, $b = 10$, $F^l \sim U[0, 150]$ and $F^h \sim U[0, 150]$. The demand rates are given as follows: In primary market we have $\lambda_1^l = 10(1 - \frac{p_1}{150})$, $\lambda_1^h = 50(1 - \frac{p_1}{150})$ and in secondary market we have $\lambda_2^l = 10(1 - \frac{p_2}{150})$, $\lambda_2^h = 50(1 - \frac{p_2}{150})$. We vary the market state probabilities in steps of 0.1 from 0.1 to 0.9. Performing the optimization we obtain the results as shown in Table 3.12 and 3.13. The optimization results provide us with useful observations:

- The optimal order quantity decreases as θ_l increases. Depending on the realized state of the market, the retailer may benefit from secondary market sales entirely or prefer to sell in both the markets.
- The supplier's optimal expected profits are decreasing in θ_l . The retailer's optimal expected profits do not exhibit monotonic behavior.

Table 3.13: BB Contract - Results under different Market State Probabilities for (120, 80)

(θ_l, θ_h)	w^*	π_{sup}^*	π_{ret}^*	Q^*	β_l^*	β_h^*	$RetRev_l^*$	$RetRev_h^*$
(0.1, 0.9)	43	609.868	612.189	23	0.0003	0.0003	183.342	17.554
(0.2, 0.8)	82	527.101	155.629	8	0.0003	0.9126	34.173	2.581
(0.3, 0.7)	81	453.477	119.380	7	0.0003	0.9258	25.182	1.384
(0.4, 0.6)	73	395.097	132.148	7	0.0003	0.9258	25.182	1.384
(0.5, 0.5)	73	339.132	96.704	6	0.0003	0.9358	17.089	0.646
(0.6, 0.4)	64	283.488	115.844	6	0.0003	0.9358	17.089	0.646
(0.7, 0.3)	64	237.681	88.036	5	0.0003	0.9440	10.345	0.256
(0.8, 0.2)	57	201.672	96.226	5	0.0003	0.9440	10.345	0.256
(0.9, 0.1)	49	160.664	109.417	5	0.0003	0.9440	10.345	0.256

3.7.5 Optimal Decisions under MM Contract

In the *first experiment* to study the effect of profit margin on optimal solutions, consider the following values for the parameters: $\nu = 15$, $\theta_l = 0.1$, $\theta_h = 0.9$, $F_l \sim U[0, 150]$, $F_h \sim U[0, 150]$ and price path as $(p_1, p_2) = (100, 60)$. The demand rates are: In primary market we have $\lambda_1^l = \frac{10}{3}$, $\lambda_1^h = \frac{50}{3}$ and in secondary market we have $\lambda_2^l = 6$, $\lambda_2^h = 30$. Performing the optimization for different profit margins, g , varying from 0 till 1.2, we obtain the results as shown in Table 3.14.

Table 3.14: MM Contract - Results for different Profit Margins with $(p_1, p_2) = (100, 60)$

g	w^*	π_{sup}^*	π_{ret}^*	Q^*	β_l^*	β_h^*	$RebRev_l^*$	$RebRev_h^*$
0.0	78.0	800.291	174.441	13	0.0003	0.9026	107.862	8.803
0.2	77.0	809.904	209.683	15	0.0003	0.7016	194.314	111.850
0.4	61.5	660.209	429.708	17	0.0003	0.7041	156.549	127.373
0.6	50.0	531.169	595.364	18	0.0003	0.7292	124.760	105.950
0.8	42.5	430.367	719.663	19	0.0003	0.7367	98.969	91.374
1.2	33.0	284.706	881.677	20	0.0003	0.7405	75.576	75.263

From the results of the optimization we observe that:

- For $g > 0$, note that the optimal timing is decreasing as g increases. This is intuitive as the retailer is assured of a higher profit margin level and hence can be more aggressive in pricing thus delaying the discount. The situation of $g = 0$ is peculiar in that the retailer asks for guarantee of selling up to w , which is the purchase price for the retailer. It can represent any one of the three possibilities: (a) When $p_2 \leq p_1 \leq w$, it can represent an unrealistic situation as the retailer is acting like an extension of the supplier and does not earn any profit; (b) When $w < p_2 \leq p_1$, it represents a situation exactly like a WP contract; (c) When we have $p_2 \leq w < p_1$, it represents a situation corresponding to part (ii) of corollary 3. For the given parameter values, we see that here with $g = 0$, case (c) is applicable. Thus calculating $p_1\lambda_1 - w\lambda_2$ we see that this expression is negative (for low state it is $-\frac{80}{3}$ and for high state it is $-\frac{400}{3}$) and hence $\beta^* \in [0, 1]$.
- The supplier's optimal expected profits decrease monotonically in profit margin. On the contrary, the retailer's profits increase with a higher profit margin. This is because the MM contract naturally favors the retailer compared to other contracts. Also note that the total supply chain profits increase with higher profit margin.

In the *second experiment* we assume the following values for the parameters: $\nu = 15$, $g = 0.4$, $F^l \sim U[0, 150]$ and $F^h \sim U[0, 150]$. Based on the surveys of retailing industry for previous years we felt 0.4 represents a reasonable value. The demand

arrival rates are given as follows: In primary market we have $\lambda_1^l = 10(1 - \frac{p_1}{150})$, $\lambda_1^h = 50(1 - \frac{p_1}{150})$ and in secondary market we have $\lambda_2^l = 10(1 - \frac{p_2}{150})$, $\lambda_2^h = 50(1 - \frac{p_2}{150})$. Tables 3.15 and 3.16 show the results of optimization with state probabilities as (0.1, 0.9) and (0.7, 0.3) respectively. In general, we note the following results:

- The optimal order quantity decreases with increasing wholesale price until a cutoff wholesale price as illustrated in Figure 3.6. Beyond this cutoff wholesale price, the optimal order quantity increases with increasing wholesale price. This is vastly different behavior compared to other two contracts. At sufficiently high wholesale prices, the retailer is assured of revenue up to guaranteed profit margin level and hence would like to order enough to meet the maximum possible demand that can occur at the given prices.
- Unlike the other two contracts, the retailer's optimal expected profits is quasi-convex (only a conjecture) in the wholesale price as shown in Figure 3.7. Retailer's profits drop until a cutoff wholesale price but beyond which since the retailer orders higher quantity, the profit increases.
- The supplier's expected profit, π_{sup} is not concave in w as shown in Figure 3.7. Notice how the supplier's profit drops rapidly beyond a cutoff wholesale price. This is because the retailer can meet (almost surely) entire demand in the secondary market and take rebates straightaway as this falls under "rebate in two markets" situation.

In the *third experiment*, we varied the discounted price keeping other parameters the same as *second experiment* and the results are shown in Tables 3.15 and 3.16.

Table 3.15: MM Contract - Results for different Price Paths with $(\theta_l, \theta_h) = (0.1, 0.9)$

(p_1, p_2)	w^*	π_{sup}^*	π_{ret}^*	Q^*	β_l^*	β_h^*	$RebRev_l^*$	$RebRev_h^*$
(100, 80)	67.0	755.826	337.528	19	0.0003	0.0216	64.380	250.818
(100, 60)	61.5	660.209	429.708	17	0.0003	0.7041	156.549	127.373
(100, 40)	57	547.723	512.580	18	0.0003	0.7549	291.761	199.001
(100, 20)	52	429.145	597.779	18	0.0003	0.8200	457.313	212.360
(120, 100)	78	711.335	264.277	13	0.0003	0.0003	30.657	116.221
(120, 80)	70	738.980	352.687	19	0.0003	0.0153	83.974	330.692
(120, 60)	64	559.323	430.245	25	0.0003	0.0141	177.544	719.914
(120, 40)	57	340.250	495.162	13	0.0003	0.7674	290.463	196.357

We observe that:

- The effect of higher discounting on the optimal timing is complex; there is no monotonic behavior.
- The retailer's optimal expected profit shows quasi-convex behavior with a drop in the discounted price. The supplier's optimal expected profit shows quasi-concavity in the discounted price.

The goal of the *fourth experiment* is to study the effect of parameters of the WTP distribution on the optimal solutions. We consider price path of $(p_1, p_2) = (100, 60)$ with the following parameters: $\nu = 15$, $g = 0.4$, $\theta_l = 0.1$, $\theta_h = 0.9$, $\lambda^l = \lambda^h = 30$ and $F^l, F^h \sim U[0, 150], U[0, 200], U[0, 300]$. The demand rates for state s are given by $\lambda_1^s = 30\overline{F}^s(p_1)$ and $\lambda_2^s = 30\overline{F}^s(p_2)$. Performing the optimization we obtain the results as shown in Table 3.17.

- We observe that the optimal timing is monotonically non-decreasing in the

Table 3.16: MM Contract - Results for different Price Paths with $(\theta_l, \theta_h) = (0.7, 0.3)$

(p_1, p_2)	w^*	π_{sup}^*	π_{ret}^*	Q^*	β_l^*	β_h^*	$RebRev_l^*$	$RebRev_h^*$
(100, 80)	46	186.000	144.297	6	0.0003	0.9721	0.000	0.000
(100, 60)	65	190.713	110.770	6	0.0003	0.9026	156.080	0.104
(100, 40)	42	152.603	173.812	9	0.0003	0.9740	129.024	0.269
(100, 20)	27	143.959	262.992	12	0.9997	0.9997	0.048	0.023
(120, 100)	65	200.000	85.451	4	0.0003	0.9008	0.000	0.000
(120, 80)	69	223.551	102.080	5	0.0003	0.8876	65.816	1.258
(120, 60)	52	209.378	157.383	7	0.0003	0.9245	69.486	3.273
(120, 40)	54	142.597	190.246	8	0.0003	0.9064	233.110	20.752

upper limit of the WTP distribution. This occurs because the probability of high-price paying customers increases.

- The supplier's and retailer's optimal expected profits are non-decreasing in the upper limit of the WTP distribution.
- This analysis tells us that both the supplier and the retailer prefer markets which have higher variance as it leads to more profits.

In the *fifth experiment* the goal is to understand the effect of different θ_l and θ_h for two price paths in primary and secondary markets: $(p_1, p_2) = (100, 60)$ and $(p_1, p_2) = (120, 80)$. For both price paths we assume the following values for the parameters: $\nu = 15$, $g = 0.4$, $F_l \sim U[0, 150]$ and $F_h \sim U[0, 150]$. The demand rates are given as follows: In primary market we have $\lambda_1^l = 10(1 - \frac{p_1}{150})$, $\lambda_1^h = 50(1 - \frac{p_1}{150})$ and in secondary market we have $\lambda_2^l = 10(1 - \frac{p_2}{150})$, $\lambda_2^h = 50(1 - \frac{p_2}{150})$. We vary the market state probabilities in steps of 0.1 from 0.1 to 0.9. Performing the optimization

Table 3.17: MM Contract - Results for different WTP distribution parameters

(F_l, F_h)	w^*	π_{sup}^*	π_{ret}^*	Q^*	β_l^*	β_h^*	$RebRev_l^*$	$RebRev_h^*$
$U[0, 150],$ $U[0, 150]$	58	390.610	322.239	11	0.6509	0.6509	82.390	82.390
$U[0, 150],$ $U[0, 200]$	62	578.181	424.771	14	0.1043	0.8006	334.278	51.546
$U[0, 150],$ $U[0, 300]$	65	816.113	518.470	18	0.0003	0.8707	505.639	37.023
$U[0, 200],$ $U[0, 150]$	59	398.710	328.995	12	0.9114	0.5025	9.049	142.650
$U[0, 300],$ $U[0, 150]$	58	403.405	337.183	13	0.9997	0.4067	0.006	172.883

we obtain the results as shown in Table 3.18 and 3.19.

Table 3.18: MM Contract - Results under different Market State Probabilities for (100, 60)

(θ_l, θ_h)	w^*	π_{sup}^*	π_{ret}^*	Q^*	β_l^*	β_h^*	$RetRev_l^*$	$RetRev_h^*$
(0.1, 0.9)	61.5	660.209	429.708	17	0.0003	0.7041	156.549	127.373
(0.2, 0.8)	66.0	643.203	285.116	16	0.0003	0.6421	194.331	167.414
(0.3, 0.7)	63.0	554.680	232.476	13	0.0003	0.8306	168.984	26.607
(0.4, 0.6)	51.0	438.553	267.745	13	0.0003	0.9358	68.313	3.536
(0.5, 0.5)	42.0	351.000	265.716	13	0.0003	0.9834	0.000	0.000
(0.6, 0.4)	39.0	264.000	221.208	11	0.0003	0.9865	0.000	0.000
(0.7, 0.3)	65.0	190.713	110.770	6	0.0003	0.9026	156.080	0.104
(0.8, 0.2)	56.0	148.578	101.890	5	0.1556	0.9540	70.526	0.007
(0.9, 0.1)	54.0	134.549	96.053	4	0.5802	0.9634	23.835	0.001

The optimization results provide us with useful observations:

- The optimal timing is not monotonic in the probability of the lower market state. This is evident when one observes the β_h^* for different θ_l .
- The supplier's optimal expected profits are monotonically non-increasing as

Table 3.19: MM Contract - Results under different Market State Probabilities for (120, 80)

(θ_l, θ_h)	w^*	π_{sup}^*	π_{ret}^*	Q^*	β_l^*	β_h^*	$RetRev_l^*$	$RetRev_h^*$
(0.1, 0.9)	70.0	738.980	352.687	19	0.0003	0.0153	83.974	330.692
(0.2, 0.8)	77.0	661.611	242.937	17	0.0003	0.0316	129.693	458.063
(0.3, 0.7)	82.0	442.101	163.795	8	0.0003	0.6547	159.457	65.803
(0.4, 0.6)	63.0	365.168	207.664	8	0.0003	0.8657	37.573	5.432
(0.5, 0.5)	61.0	309.076	172.513	7	0.0003	0.8957	24.199	1.648
(0.6, 0.4)	62.0	264.106	135.193	6	0.0003	0.9058	29.175	0.971
(0.7, 0.3)	69.0	223.551	102.080	5	0.0003	0.8876	65.816	1.258
(0.8, 0.2)	54.0	195.000	102.901	5	0.0003	0.9352	0.000	0.000
(0.9, 0.1)	65.0	165.667	71.788	4	0.0003	0.9164	38.124	0.217

the lower market state is more probable. The retailer's optimal expected profits do not exhibit monotonic behavior.

- This analysis tells us that the supplier prefers higher market states where as the retailer's preference is not trivially obvious.

3.7.6 Discussion of Optimal Decisions across Contracts

In this section we discuss the decisions across WP, BB and MM contracts when both the retailer and supplier make optimal choices. We will consider the following numerical experiment for this discussion. Consider the first sequence of prices in the corresponding markets: (100, 80), (100, 60), (100, 40) and (100, 20). We assume the following values for other parameters: $\theta_l = 0.7$, $\theta_h = 0.3$, $b = 10$, $g = 0.4$ and demand rates for effective Poisson process are as follows: $\lambda_1^l = \frac{10}{3}$, $\lambda_1^h = \frac{50}{3}$, $\lambda_2^l = \frac{1}{15}(150 - p_2)$ and $\lambda_2^h = \frac{1}{3}(150 - p_2)$.

Timing of Transfer to Secondary Market: This is illustrated in Figures 3.8 and 3.9. When we compare the optimal timing of transfer to secondary market across different contracts for same price path, we see that the relation between the optimal timing across the three contracts, as discussed in Propositions 4 through 6, holds true *even* when all the parties involved make optimal ordering decisions. This observation tells us BB contract gives the product a longer exposure in the primary market.

Order Quantity: Referring to Figures 3.10 through 3.13, when we compare the optimal order quantities for BB and WP contract under the same price path we note that the retailer orders slightly higher or same amount for BB contract. When we compare MM and WP contract, we notice that the order quantities are equal or higher under a MM contract. There is no apparent relation between the order quantities under BB and MM contract.

Retailer's Expected Profit: As shown in Figure 3.14, comparing the retailer's optimal expected profits for BB and WP contracts, we see that a BB contract need not be beneficial to the retailer. However, the retailer's profits under MM contract are at least as high as WP contract.

Supplier's Expected Profit: Contrary to expectation, we notice from Figure 3.15 that the supplier is better off in terms of profits for a BB contract when compared to a WP contract. This happens because the supplier has the first-mover advantage and thus can negate the buyback credit requested by the retailer by choosing an optimal wholesale price. When we compare the MM and the WP contracts, we see that the

relation is not obvious as the supplier can earn lower or higher profits under the MM contract depending on the profit margin requested. Again, the relation between the optimal profits of the supplier under BB and MM contract is not straightforward.

Total Supply Chain Profit: We compare the total supply chain profits across different contracts as shown in Table 3.20. In general, we observe that the total supply chain profits between the BB and the WP contracts remain more or less similar. Another important observation is that for a certain range of buyback credit values, the total supply chain profit remains constant but it gets re-distributed differently between the retailer and the supplier. On the other hand, in general, MM contract performs better in terms of supply chain profits. Notice that the profits are higher in a MM contract compared to the BB or the WP contract.

Table 3.20: Comparison of Supply Chain Profits under different Contracts with $(\theta_l, \theta_h) = (0.1, 0.9)$

Price Path	Contract	Parameter	Supply Chain Profits	π_{sup}^*	π_{ret}^*
(100, 60)	WP	—	978.373	793.000	185.373
(100, 60)	BB	$b = 25.0$	975.594	838.663	136.931
(100, 60)	BB	$b = 20.0$	975.595	832.131	143.464
(100, 60)	BB	$b = 10.0$	975.594	812.565	163.029
(100, 60)	BB	$b = 5.0$	978.341	807.441	170.900
(100, 60)	BB	$b = 0.0$	978.373	793.000	185.373
(100, 60)	MM	$g = 0.4$	1089.917	660.209	429.708
(100, 60)	MM	$g = 0.6$	1126.533	531.169	595.364

3.7.7 Preferences of Supplier and Retailer for High Risk Markets

To analyze the preferences of the supplier and the retailer under different contracts towards products with potential higher demand or lower demand, we compare the preference of the supplier and the retailer under different low state probabilities. Recall that the general state of the market is unknown at the time of the ordering decision. This reflects the notion that both the suppliers and the retailers are uncertain whether the product (e.g., new designer shirt, new music player) that will be launched during the season turns out to be very popular or be a flop. The popular product resembles the high demand state and on the other hand the flop item is equivalent to the low demand state. Thus, when we vary the occurrence of the low state probabilities under different contracts we can see which state is preferred by the supplier and the retailer. With different contracts, consider the following prices in the corresponding markets: $(100, 60)$. We assume the following values for other parameters: $\nu = 15$, $b = 10$, $g = 0.4$ and demand rates for an effective Poisson process are as follows: $\lambda_1^l = \frac{10}{3}$, $\lambda_1^h = \frac{50}{3}$, $\lambda_2^l = 6$ and $\lambda_2^h = 30$. This is illustrated in Figure 3.16. Comparing the profits we observe that:

- In general, the supplier's and the retailer's profits decreases when the low state is more likely to occur. This means that both the retailer and the supplier would like to see the product be a popular item under any contract.
- When we observe the retailer's behavior we notice that they prefer MM contract when the low state is more probable. That is, when the product is likely to be a flop, the retailer would like to utilize the better risk-sharing mechanism,

which is the MM contract.

- Observing the supplier's behavior, we notice that they prefer BB contract under all low state probabilities. This looks counter-intuitive since WP contract does not offer any risk sharing mechanism but is still not preferred. This happens because we observed that BB contract is not an effective risk-sharing mechanism for the retailer, since retailer's profits are adversely affected as shown in Table 3.20.

3.8. Conclusions

In this essay, we develop a model to study the effect of contract types on a retailer's use of primary and secondary markets to sell his product. Specifically, we discuss the implications of WP, BB and MM on the optimal timing of transfer of goods from the primary to the secondary market and the optimal order quantity. We also discuss the effect of these contracts on the supplier and on the supply chain performance as a whole. We show that for a given order quantity, the time to transfer the goods from the primary market to the secondary market varies depending on the contract type - BB contract transfers later and MM contract transfers earlier compared to a WP contract. Considering the optimal order quantities for the retailer, the retailer places the highest amount of orders under the MM contract. While this is advantageous for the supplier, increasing his own sales, it also poses challenges because the potential for paybacks from the suppliers to the retailer are also higher when the order quantity is high. In general, the retailer benefits more

from the MM contract but there can also be a win-win situation for both parties in the supply chain under MM. On the other hand, the retailer's order quantities are marginally higher in BB contract when compared to WP contract. The retailer's profits under BB contract can be lower compared to WP contract while the supplier is always better-off compared to WP contract.

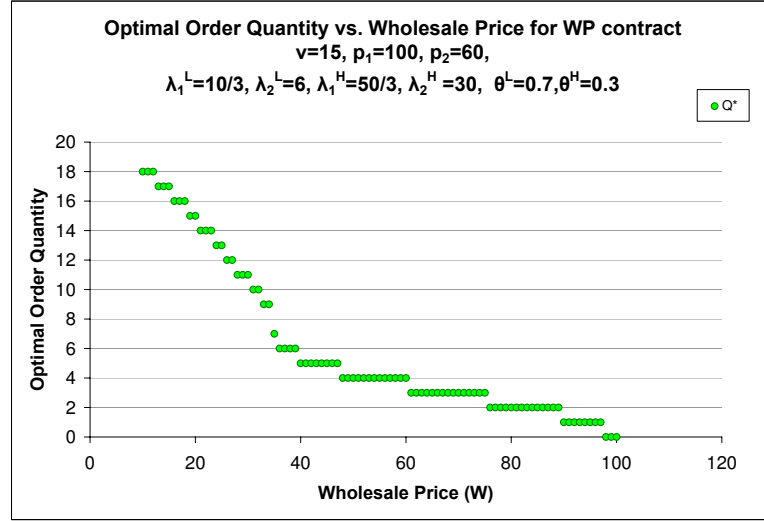


Figure 3.2: Optimal Order Quantity vs. Wholesale Price for WP contract

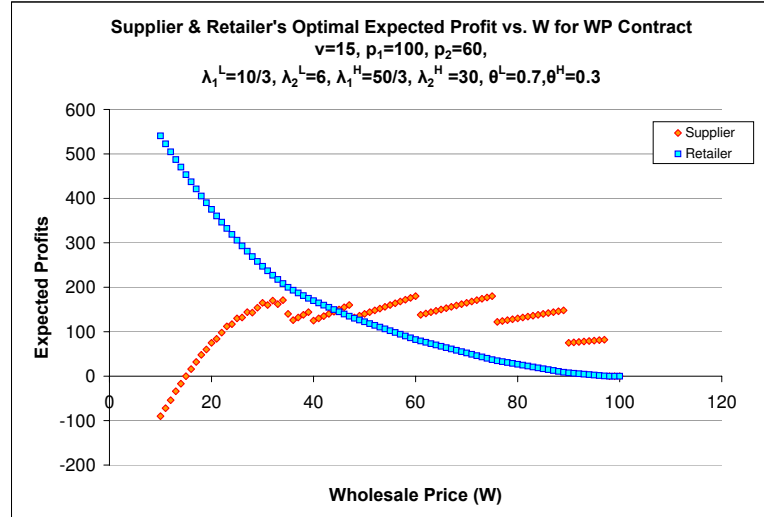


Figure 3.3: Supplier and Optimal Retailer Expected Profit vs. Wholesale Price for WP contract

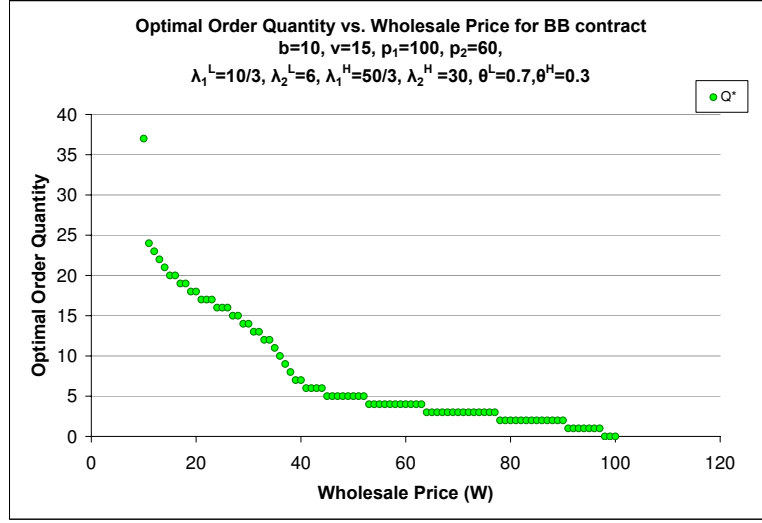


Figure 3.4: Optimal Order Quantity vs. Wholesale Price for BB contract

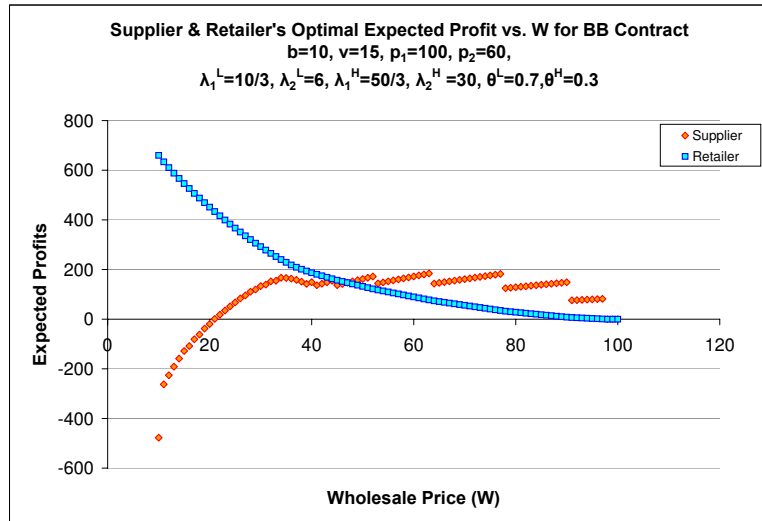


Figure 3.5: Supplier and Optimal Retailer Expected Profit vs. Wholesale Price for BB contract

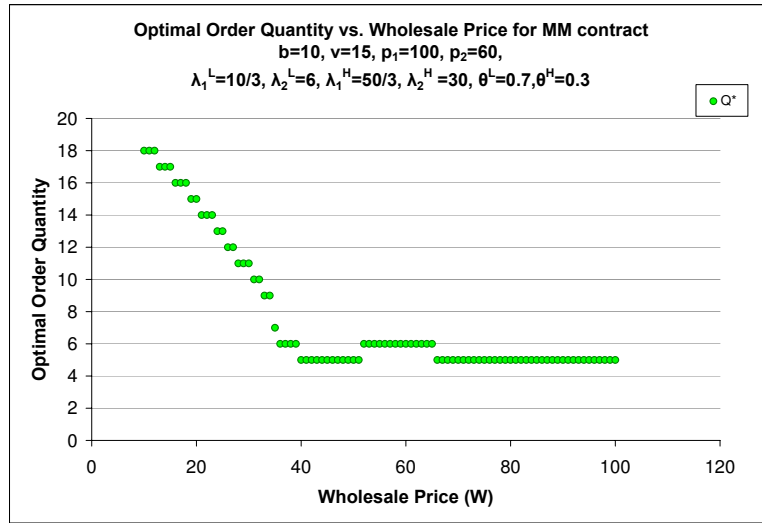


Figure 3.6: Optimal Order Quantity vs. Wholesale Price for MM contract

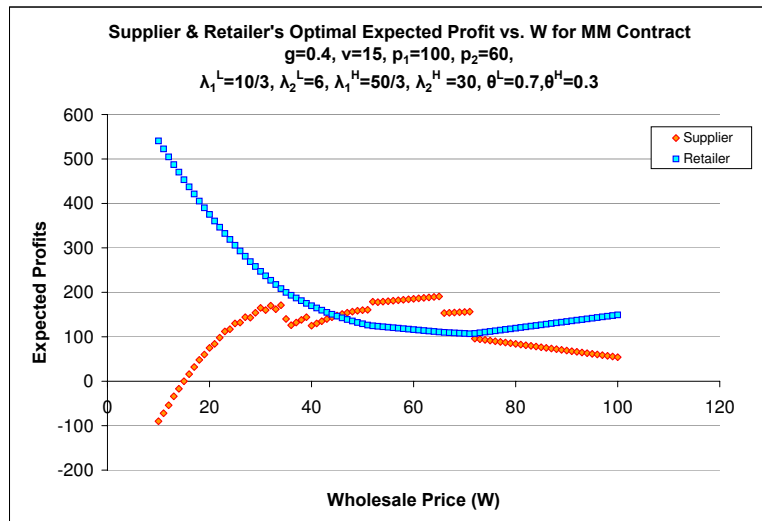


Figure 3.7: Supplier and Optimal Retailer Expected Profit vs. Wholesale Price for MM contract

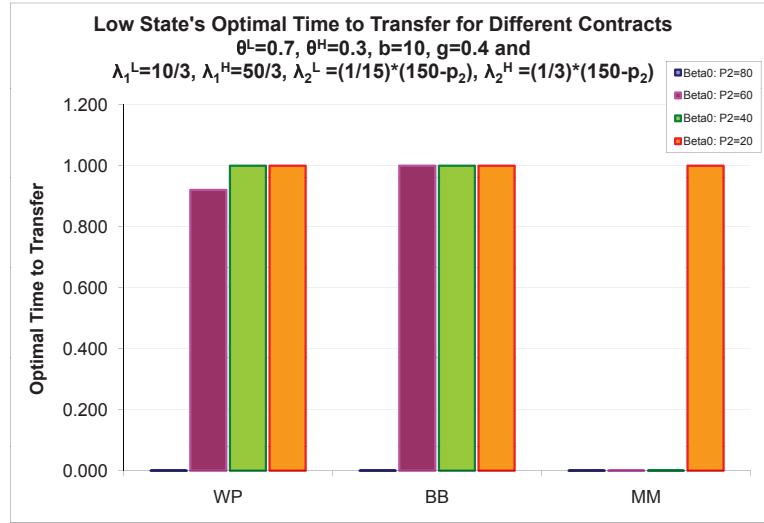


Figure 3.8: Optimal Low State Transfer Time for Different Contracts

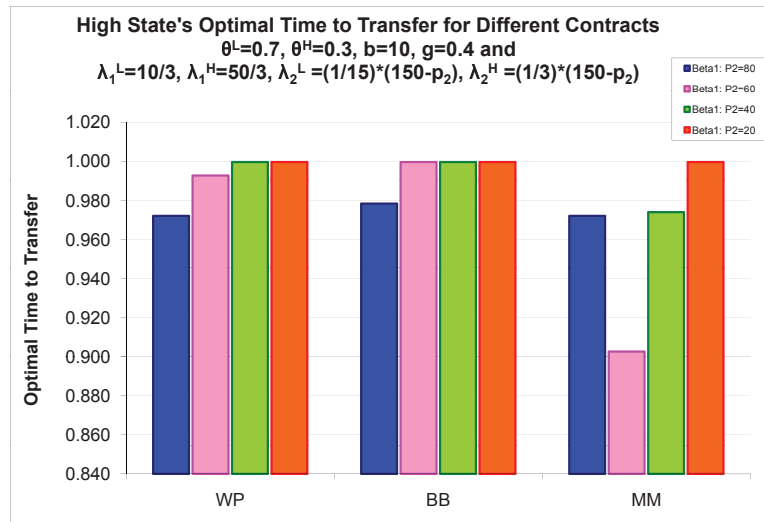


Figure 3.9: Optimal High State Transfer Time for Different Contracts

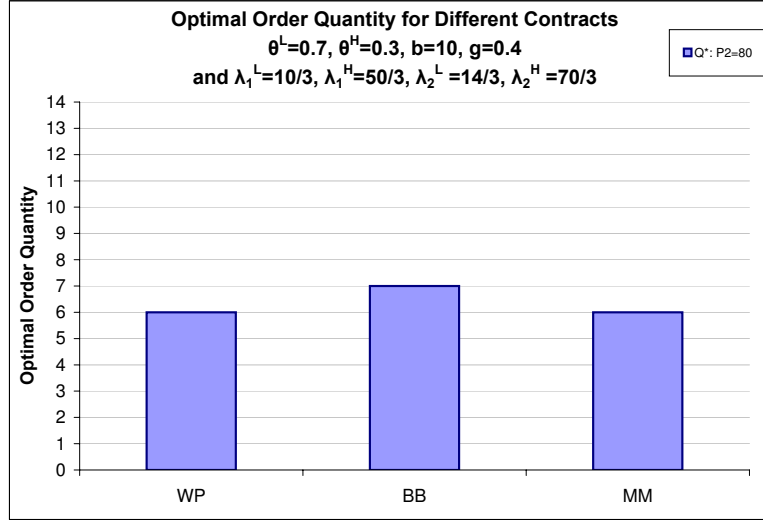


Figure 3.10: Optimal Order Quantity for $(p_1, p_2) = (100, 80)$ for Different Contracts

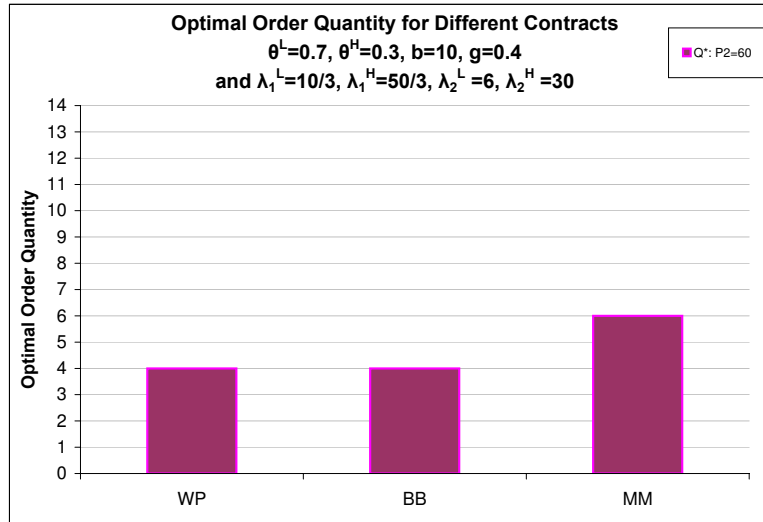


Figure 3.11: Optimal Order Quantity for $(p_1, p_2) = (100, 60)$ for Different Contracts

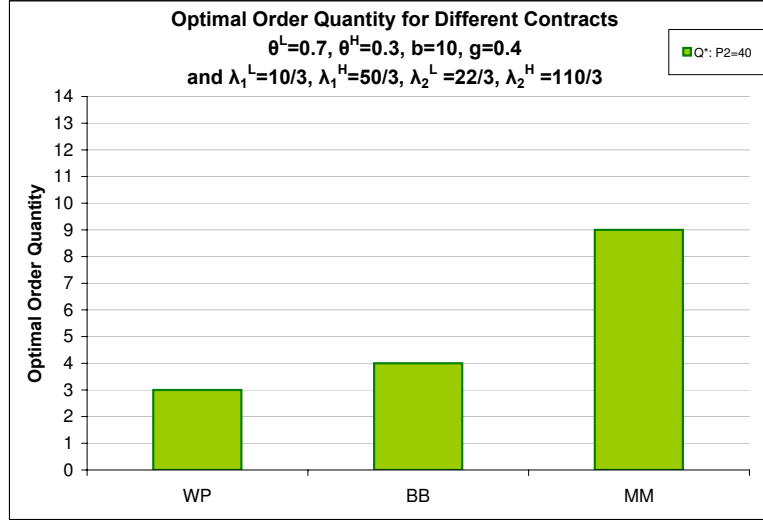


Figure 3.12: Optimal Order Quantity for $(p_1, p_2) = (100, 40)$ for Different Contracts

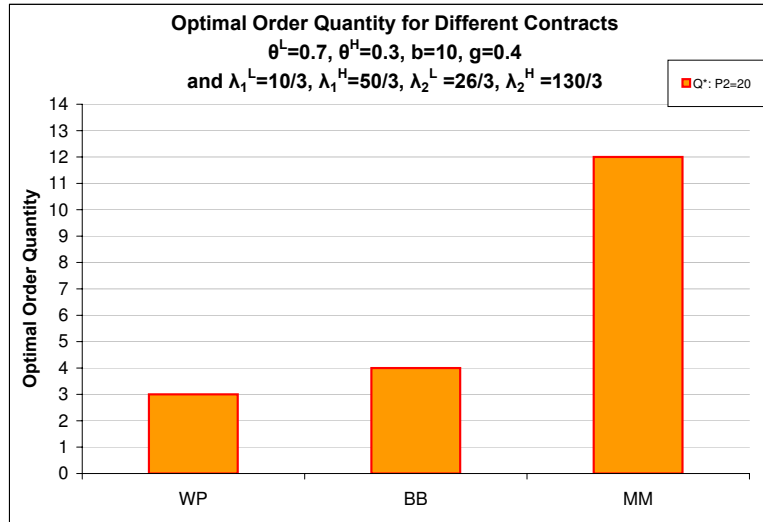


Figure 3.13: Optimal Order Quantity for $(p_1, p_2) = (100, 20)$ for Different Contracts

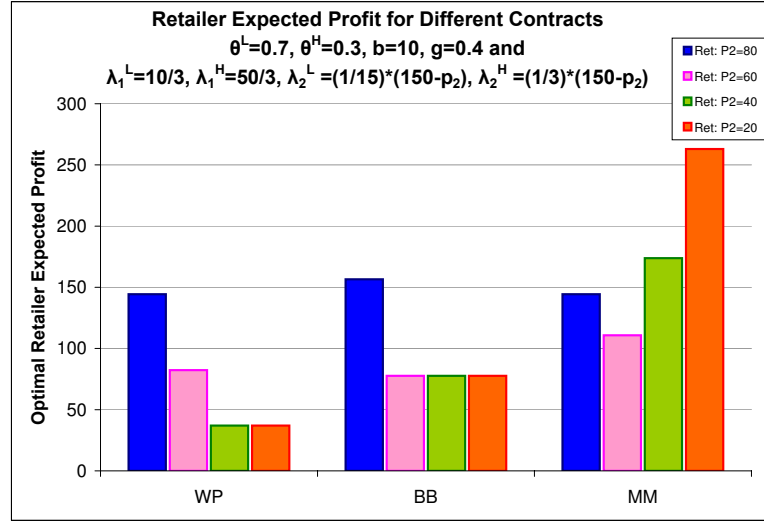


Figure 3.14: Optimal Retailer Expected Profits for Different Contracts

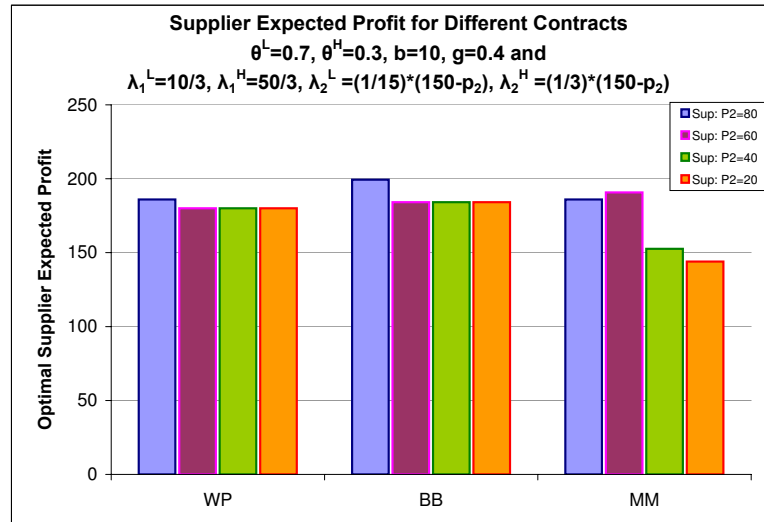


Figure 3.15: Optimal Supplier Expected Profits for Different Contracts

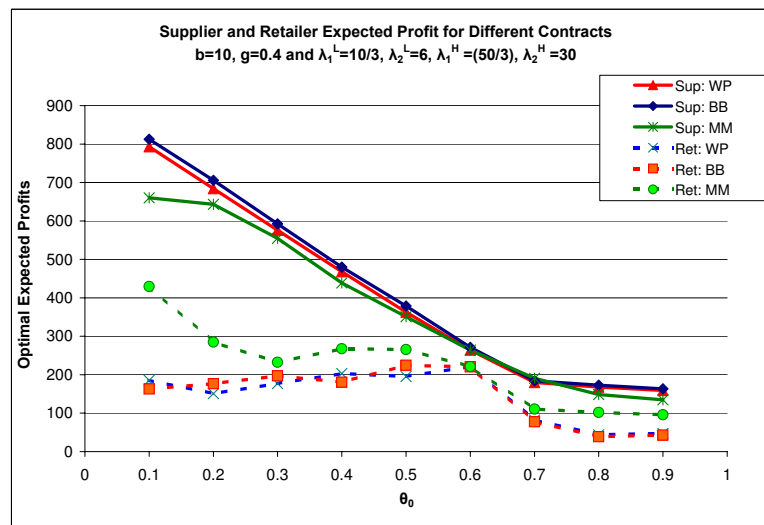


Figure 3.16: Supplier and Retailer Expected Profits for Different Low State Probabilities

Chapter 4

Essay 3: Humanitarian Vehicle Routing Problem: Uncapacitated Case

4.1. Introduction

Rapid urbanization coupled with environmental degradation and the spread of HIV/AIDS is unfortunately, expected to increase the scope of man-made and natural disaster relief five-fold in the next 50 years (Thomas & Kopczak (2005)). Complex operational challenges in humanitarian logistics came to the forefront when an earthquake and resulting tsunami crippled South Asia on December 26, 2004 (Russell (2005), Fritz Institute (2006)). The entire world mobilized to donate \$13 billion (Thomas & Fritz (2006)), but according to Doctors without Borders, "... we have already received as much money as we can spend... What is needed are supply managers without borders: people to sort goods, identify priorities, track deliveries and direct the traffic of a relief effort in full gear" (Economist (Jan 2005)). In the aftermath of the tsunami, Sri Lanka struggled to manage the large number of cargo-laden humanitarian flights and distribution of relief goods to warehouses; in India, the transportation pipelines were clogged; in Indonesia, the damaged infrastructure with excessive assistance created unforeseen logistical challenges in the context of humanitarian relief operations. These challenges are not limited to natural disasters

like the tsunami in 2004 or hurricane Katrina in 2005. Relief agencies encounter similar and additional challenges in military environments like civil conflicts, war-zones, and terrorist attacks.

Compared to well-studied business logistics, humanitarian logistics presents unanswered challenges in terms of movement of relief goods like food, bedding and shelter, medical care, and clothes (Sheu (2007*a*), Kovács & Spens (2007)). The challenges are magnified when the humanitarian relief operations need to maintain neutrality and impartiality (Tomasini & van Wassenhove (2009)). The business supply chain and humanitarian relief chain differ in aspects like revenue sources (e.g., investments vs. philanthropic donations), goals (e.g., profit vs. loss of life), stakeholders (e.g., businesses vs. Non-Government Organizations), performance measurements (e.g., financial metrics vs. social welfare), demand characteristics (e.g., erratic in disasters), and customer characteristics (e.g., no competing brands in disasters). The goal of a business supply chain is to maximize profit; for humanitarian relief chains, the goal is to maximize timely delivery in order to minimize the loss of life. A typical humanitarian relief chain consists of a primary supply hub such as a port of entry, a central warehouse (or a secondary hub), and local distribution centers (Balcik et al. (2008)). Finally, the relief goods are distributed to people in need (or demand points) from the distribution centers. This last stage is commonly known as the “Last Mile Distribution” in humanitarian logistics, and this is the main focus of our research.

This research aims to design delivery routes for the distribution of relief goods that meet the ultimate humanitarian goal of mitigating human loss in a timely man-

ner and that are efficient on performance metrics relevant with respect to humanitarian relief operations. Davidson (2006) identifies multiple performance metrics for humanitarian relief operations but for this work, we focus on two applicable measures - distance and response time. Traditional or non-humanitarian vehicle routing models do not capture these metrics very well and, thus, perform at a poor level on these metrics. One of the key principles identified in the manual, prepared by Emergency Preparedness and Disaster Relief Program of PAHO¹, for the distribution phase is that not all victims are the same and the delivery of assistance must be proportionate and equitable (Cuadra (2001)). Towards this end, we consider the node priorities, a proxy for urgency of locations along with other humanitarian needs and develop mathematical models for designing efficient delivery routes. The scope of this work is not only limited to addressing the operational challenges of relief operations due to natural disasters, but also has other applications in situations that gives primary importance to social welfare with customers at varying urgency levels. For example, medical agencies encounter similar challenges when they have to deliver vaccines to people in distress during disease outbreaks. If a vaccine is available for Swine influenza or other types of Avian influenza, then the agencies would need to dispatch the vaccine with minimal loss of life but in an efficient manner.

As mentioned above, one of the key components of the humanitarian relief chain is the delivery of aid to the recipients. In this research, we are concerned with the delivery of relief goods from a distribution center to the people in need as efficiently as possible with the ultimate objective of minimizing the loss of life.

¹Pan American Health Organization

In this work, we consider distribution of a single product using a single vehicle with unlimited capacity and no route length restrictions. One can imagine such a situation when the entire network's demand can be serviced by a single vehicle in one particular day. Campbell et al. (2008) consider such a problem but without taking into account the urgencies for the nodes. This research is (we think) the first of its kind pertaining to humanitarian relief routing, as we address the issue of priorities for locations along with the humanitarian demands. The combination of priorities for locations in a vehicle routing problem poses interesting research questions: (1) How do we simultaneously capture performance metrics like distance and response time in our models? (2) How much is the route altered when the priority restrictions are strict? (3) Is there a way to model relaxed forms of the priority restrictions? If we can incorporate the relaxations, then our models can provide a lot of flexibility to the decision maker in terms of handling political and social issues. We plan to answer these research questions by developing mathematical models that produce optimal results for small to mid-size problems. It should be noted that to evaluate the performance of a humanitarian relief operation, one needs not a single, but multiple metrics (Beamon & Balcik (2008)). As the vehicle does not have capacity restrictions, and based on humanitarian literature, in our work, we plan to compare routes using the following metrics: (i) Distance, (ii) Response Time (Earliest, Latest, Average). Also, we compare our models to VRP models without priorities used in Campbell et al. (2008).

One important contribution of this research work to the literature is that we consider the effect of priorities on vehicle routing problems in the context of

humanitarian relief operations. To the best of our knowledge, previous work in humanitarian relief operations does not address the issue of priorities for the efficient distribution of relief goods using a mixed integer program (MIP). Not only is humanitarian logistics an emerging area of research, but it is of high practical importance as evidenced by natural disasters such as hurricane Katrina and the tsunami in South Asia in 2004, discussed earlier. Non-Governmental Organizations and non-profit organizations receive billions of dollars in donations every year, but lack a sophisticated system to handle their complex logistics operations; the absence of expertly-designed systems is one of the significant reasons why there has been such a weak link in the distribution of relief aid. Thus, this work is of significant value to both researchers and practitioners as it strives to improve the delivery of relief aid with the ultimate goal of saving lives.

4.2. Setup

To answer our research questions, we adopt the following framework for the vehicle routing problem. Customers are represented as nodes and a depot is denoted by node 1. Apart from the humanitarian needs, we assume that the nodes have priorities. It should be mentioned that priorities or variants have been used in the literature (see Section 4.3) in a variety of applications. Priorities are proxies for urgencies at each location. This is illustrated in Figure 4.1 for different relief situations with three priority classes. The relief situation can be to deal with a natural disaster such as a tsunami, earthquake, hurricane, or flood. Also, it can

arise when agencies have to deliver vaccines in case of a disease outbreak. Priorities are a natural representation of how urgent the situation is, due to a disaster or how badly people are affected. For example, certain locations urgently need life-saving drugs, while other locations, that need precautionary drugs, can be considered less urgent. The needs at higher priority nodes have to be satisfied before the vehicle can visit the lower priority nodes. Without loss of generality, we assume that lower numbers represent locations with higher priorities. That is, a node with priority 1 needs to be served before we proceed to serve nodes with priority 2. The key rule in our models that is not present in traditional routing models is the **d**–Relaxed Priority Rule. We plan to develop MIPs for the **d**–Relaxed Priority Rule for single vehicle and solve the networks to optimality. Before we discuss this rule further, let us take a closer look at the definition for node priorities.

4.2.1 Definition of Node Priorities

The definition of priority for node i , p_i can vary based on different social, political criteria set by decision makers in conjunction with on-the-ground realities. We briefly discuss this aspect to highlight the underlying issues. In a simple assignment, it can be just an ordinal number based on the urgency for the node. Without loss of generality, we can assume lower urgency values (for e.g., u_i for node i can be probability of node survival) represent nodes that require immediate response (thus, we can assign the depot an urgency of 0). Priority can also be based on a combination of distance and urgency; this second definition can result in substantial

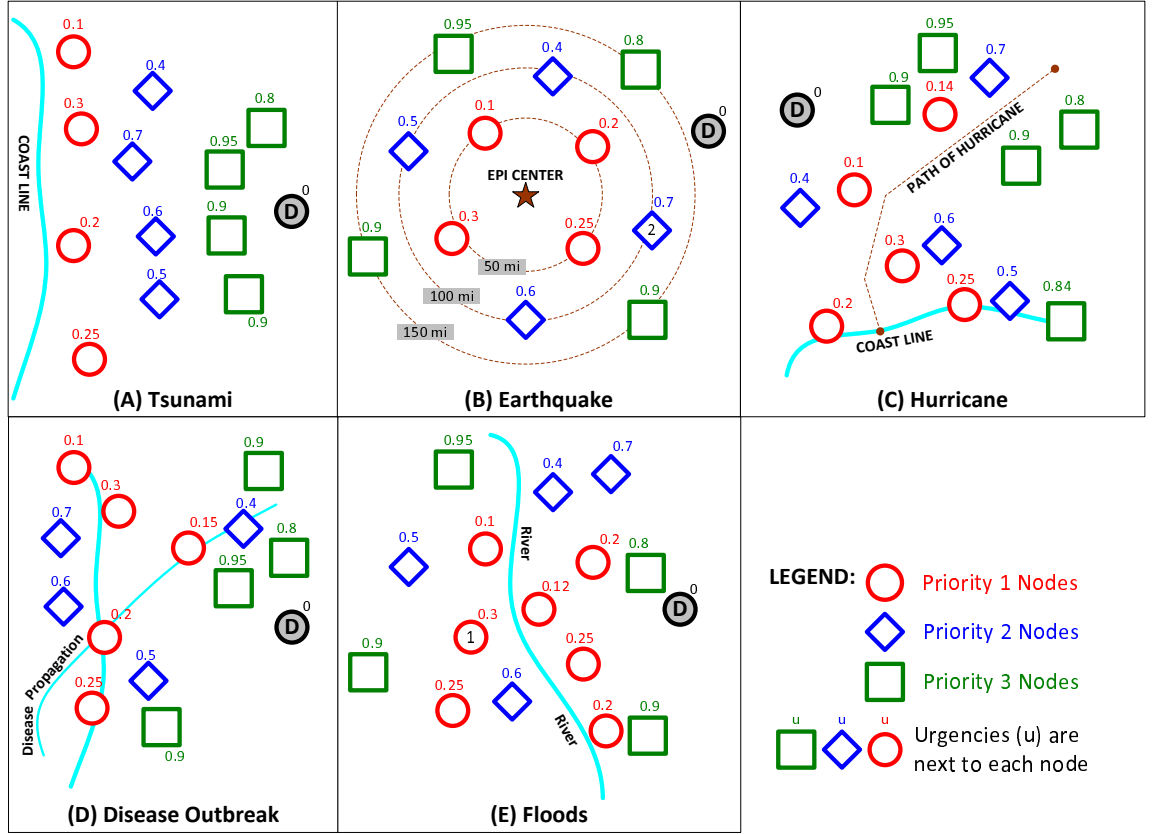


Figure 4.1: Illustration of Urgencies and Priorities for Disasters

distance savings, as it can affect how the vehicle trades off distance for priority. In most situations, a certain set of nodes with urgencies lying between a lower limit of urgency, u_l^k , and an upper limit of urgency, u_h^k , can be considered to be of the same priority, k , as illustrated in Figure 4.1. In such a case we can define the following: $p_i = k$ if $u_l^k \leq u_i \leq u_h^k$ where $k = 1, 2, 3, \dots, p^{max}$. One can define as many clusters (p^{max}) as needed by the decision maker. Note that, in this definition, we assume that a node cannot be assigned two priorities. In this research, we do not focus on how to devise priorities but rather on the routing problem with priorities. We believe that the assignment of priorities is better handled at a planning level, rather than at an operational level.

4.2.2 d –Relaxed Priority Rule

This rule depending on the relaxation, balances the trade-off between operational efficiency and meeting humanitarian goals. In general, the d –Relaxed Priority Rule means that the vehicle can satisfy demand at a node with priority, d lower than a higher priority node before it satisfies all higher priority nodes. One can think of d as the degree of relaxation. To illustrate this rule, let us consider a special variant of this rule called the 0–Relaxed Priority Rule where, $d = 0$ represents the hardest version of priority restrictions. Here, all the higher priority nodes must be satisfied before the vehicle can proceed to satisfy any lower priority node. To explain further, let us say, a vehicle starts servicing nodes with priority p and on its way it encounters a node with priority $p + 1$. According to this 0–Relaxed Priority Rule, the vehicle will have to come back later to satisfy the node with priority $p + 1$. The vehicle first finishes servicing all nodes with priority p before proceeding to nodes with priority $p + 1$. If we had the 1–Relaxed Priority Rule, then a vehicle servicing nodes with priority p can satisfy nodes with priority $p + 1$, even if it has not finished servicing all nodes with priority p . However, the vehicle cannot service nodes with priority greater than $p + 1$ until it finishes servicing all nodes of priority p .

Figure 4.2 illustrates the optimal routes under different relaxations for a 21-node network. It represents a relief operation with a single vehicle of unlimited capacity. As expected, we can see from the figure that the optimal distance decreases with higher d values. In the case of $d = 2$, one can notice that the vehicle does not observe any priorities. Thus, we can consider the traditional vehicle routing problem

or the problem without any priorities as a problem with the d –Relaxed Priority Rule with highest d value (which is one less than the total number of priority classes, $p^{max} - 1$). With higher d values, though distance decreases, the service times for different priority nodes are not equitably distributed. For example, the latest service time for priority 1 nodes, which are the nodes requiring urgent response, goes up from 10.85 to 12.42 to 14.22 for $d = 0, 1$, and 2 , respectively. On the other hand, the latest service time for priority 3 nodes, which are not so urgent, goes down from 28.57 to 21.56 to 16.27 for $d = 0, 1$, and 2 , respectively. Delaying rescue response to priority 1 nodes, for example, could be fatal and thus, the decision maker needs to effectively devise efficient and timely routes. Hence, we believe that this form of modeling is unique to humanitarian relief operations and has not been addressed in the existing vehicle routing literature.

Apart from this key restriction, we make assumptions that are typical to uncapacitated vehicle routing problems. For the sake of completeness, we write down all the assumptions: (i) the depot is designated as node 1 with a service time of 0. Vehicle(s) can leave and return to the depot at most once after servicing the nodes. (ii) Customers are designated as nodes $\{2, \dots, n\}$. Each node can be serviced at most by one vehicle. (iii) Vehicle does not have capacity restrictions. (iv) Vehicles does not have any route-length restriction (in terms of distance traversed) due to fuel, working hours, etc. (v) All vehicle routes will have to obey the d –Relaxed Priority Rule. There can be two types of underlying networks that we consider: (a) Basic (in the form of trees). This can occur in rural areas where there is no complex road network or when most road links are temporarily destroyed or deemed dangerous for

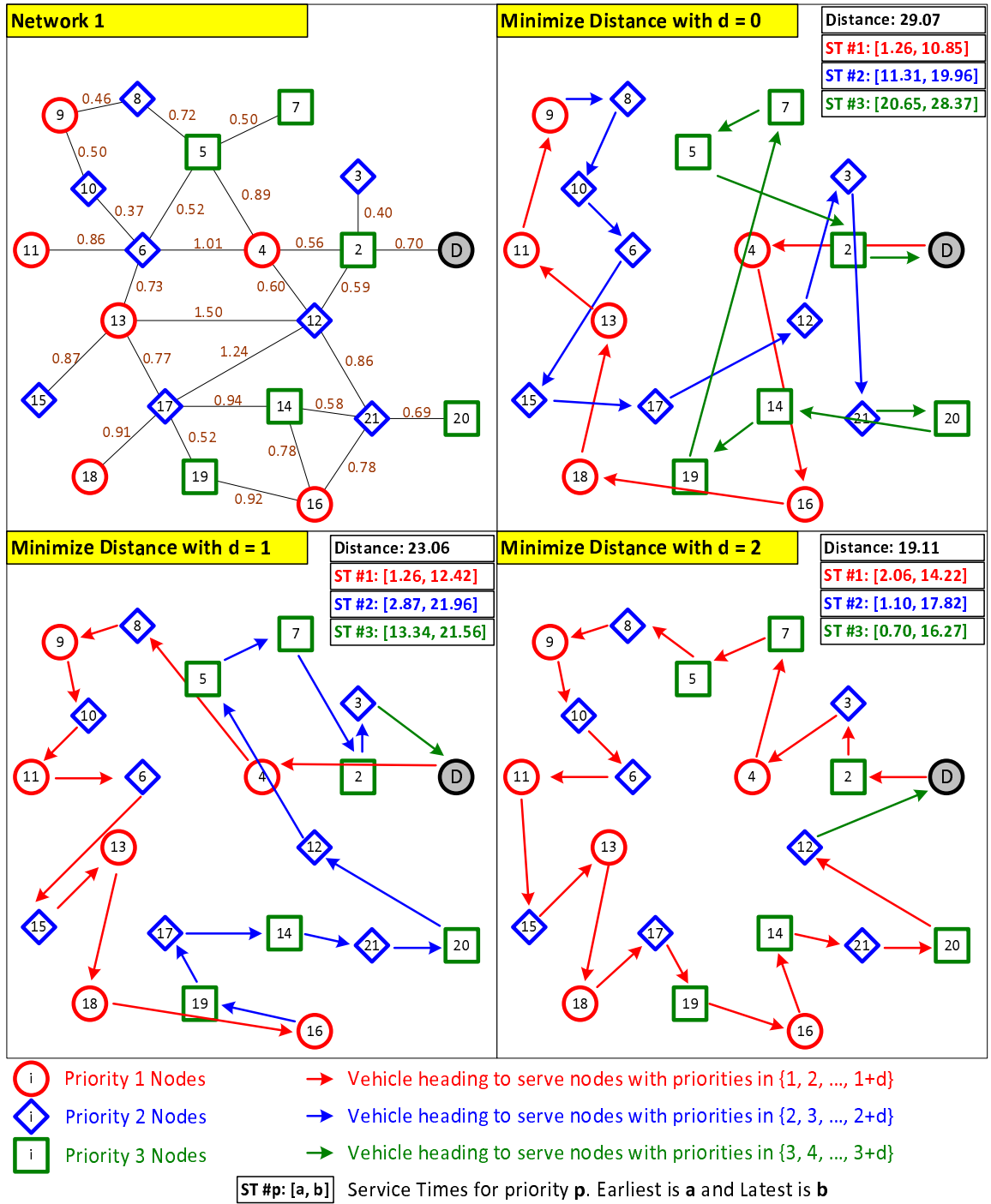


Figure 4.2: Illustration of the d -Relaxed Priority Rule

relief operations; (b) Sophisticated Networks (generic graphs). This occurs in most urban areas or when the road network is largely intact after disaster. In general, for this research, we assume type (b) – general undirected graphs. However, note

that the vehicle will always use the shortest path of distance when servicing nodes of priorities satisfying the d -Relaxed Priority Rule. As we are dealing with undirected graphs with non-negative costs, we use the Floyd-Warshall algorithm to calculate the shortest path and corresponding distances and travel times between any pair of nodes i and j . As a result, we transform the original graph into a complete graph with edge costs as shortest path distances between the nodes in the original graph. Henceforth, we will use the complete graph for our MIP formulations.

The outline for the paper is as follows. In Section 4.3 we discuss the relevant literature. In Section 4.4, we develop mathematical models with different objectives using Mixed Integer Programming (MIP) techniques, derive bounds for VRP with d -Relaxed Priority Rule and conduct worst-case analysis. In this research, we use a small set of test problems and compare our models with methods from the literature with respect to the performance metrics discussed. Computational Study is discussed in Section 4.5 and finally we present conclusions in Section 4.6.

4.3. Literature Review

Though there is a large literature associated with the standard non-humanitarian vehicle routing problem, the field of humanitarian relief operations has received widespread recognition after the Southeast tsunami in 2004. In this section, we discuss the literature associated with humanitarian relief logistics and variants of vehicle routing problems relevant for humanitarian relief situations. Finally we end the section with a summary of our contribution to the VRP literature.

Several applications pertaining to service and military operations have discussed concept of separating the customers or ranking the targets, in a similar vein to the concept of priorities that was introduced in Section 4.2. Shetty et al. (2008) introduce ‘value’ for a target in a military setting. The objective is to design a route for unmanned combat aerial vehicles such that it delivers maximum total weighted payload. Payload delivered at each node is weighted by a ‘value’ for that node. This approach does not work well for humanitarian relief operations as, a higher-valued node can be substituted with a lower-valued node having an appropriately designed payload. To illustrate, consider two nodes having values and payloads as (5, 10) and (10, 5). This route is reasonable for military operations since both contribute weighted payload of 50 but for humanitarian operations, one cannot ignore the fact that second node is of higher importance.

In the context of service delivery, Schmitz & Niemann (2009) discuss a Traveling Salesman Problem with Sequence Priorities (TSPwSP). The objective is to minimize the cost of traveling but at the same time, the cities are to be serviced strictly in an order. This order is dictated by the arrival time of the delivery requests (that is, nodes $1, \dots, m$ correspond to $1, \dots, m^{th}$ service delivery request). When node i is visited at a_i , a penalty of $\max(i - a_i, 0)$ is calculated for node i . One issue with this penalty function is that it can ignore the order if the total penalty value is same. For example, if we had 3 nodes, then this penalty function cannot distinguish between routes 2, 3, 1 and 3, 1, 2 as the total penalty is 2 but in the first route, a priority 1 node is serviced last! If the order in service context is violated, the firm incurs an additional cost but for humanitarian relief operations it can be fatal.

Smith et al. (2010) discuss a dynamic vehicle routing problem in which there are n vehicles and m priority classes of service demands. Customer requests for service arrive randomly over time at different locations in Euclidean space and require a random amount of on-site service. Each customer request is assigned to one of the m priority classes. The goal is to minimize the convex combination of expected delay for each priority class. Under heavy-load conditions, they show that ignoring priorities makes the expected delay unbounded. Since the vehicle provides only service (no pick up or delivery), their work ignores the vehicle capacity restrictions. Compared to service industry, in most disasters, customers or victims cannot communicate in real-time and put in a request for service. But notice that our idea of priorities for demand is similar to the definition the authors used in their work.

Recently, concepts similar to priorities have been discussed in the context of humanitarian relief operations. Chiu & Zheng (2007) develop a model that minimizes total travel time for dynamically assigning response resources and evacuation groups in the presence of multi-priority emergency response resources and evacuation groups. Example of multi-priority evacuation groups are elderly, patients or nursing home residents, etc. Their work provides an another interpretation for the concept of priorities. Sheu (2010) extended prior work in Sheu (2007*b*) to develop a dynamic relief-demand management model for large-scale disasters. The methodology consists of three recursive mechanisms (1) dynamic relief-demand forecasting (2)affected-area grouping, and (3) identification of area-based relief-demand urgency. The first mechanism predicts relief-demand over time, the second mechanism groups the affected areas through multi-criteria fuzzy clustering technique and

the third mechanism associates urgency with each clustered affected area utilizing multi-criteria decision-making methodology called TOPSIS (Technique For Order Preference by Similarity to Ideal Situation). The author tests this methodology at the massive Chichi earthquake, which occurred in central Taiwan on September 21, 1999 and concludes that the forecasted model performs satisfactorily. This paper complements our work as it identifies distribution priorities for affected areas and our work comes up different routing models for distribution that take into account the priorities.

Next, let us look at the literature that specifically discuss vehicle routing models for humanitarian relief operations. Yi & Kumar (2007) solve a variant of simultaneous pickup-delivery problem, where the problem is to transport prioritized commodities to survivors and deliver health care services to injured people by transporting them to emergency units. The goal is to design routes that minimize the weighted sum of unsatisfied demand over all commodities and that of unserved wounded people at demand nodes and emergency units. The authors propose an ant colony optimization based heuristic that works in two phases: the vehicle routes construction and the multi-commodity dispatch in disaster relief distribution. Balci et al. (2008) discuss the last mile distribution in humanitarian relief chain. The objective is to minimize the total costs which is the unweighted sum of routing costs and penalty costs over all days in the planning horizon for two types of products. Type 1 has very large demand with backlogs allowed (e.g., tents, blankets). Type 2 represents demand for regularly consumed items and demand is lost if it is not satisfied (e.g., food, hygiene kits). Penalty costs are linear in unmet demand and

accumulate for lost or back-ordered demand. The problem is solved in two phases. In Phase I, the authors develop all feasible demand combinations for each vehicle. Then, the authors develop routes with minimum travel time (by solving a TSP) for each of these demand combinations. In Phase-II, they determine which demand locations to visit on any day and how much to allocate on each day at a demand location. Though the authors consider two types of products over a planning horizon, they do not take into account the priorities of nodes. Campbell et al. (2008) work is most relevant to our work. They discuss routing for relief efforts using two types of objective functions. In the first, they minimize the maximum arrival time at a node. This means that nodes that are far away from the depot can be reached as early as possible for service. However, this is not an appropriate objective function to cater to networks with nodes of varying urgency levels, as nodes that are not that far out do not affect the objective function. The focus is on a particular node that is the farthest from the depot and this may not be the node that requires urgent attention. In the second formulation, they minimize the average arrival time. It is an aggregate measure, so when all nodes have more or less the same level of urgency this will work well but there is no guarantee that all high-priority nodes will be visited before low-priority nodes are visited. Jotshi et al. (2009) take an integrated view for dispatching emergency vehicles in a disaster situation using data fusion. Information on casualties and road and traffic conditions is combined to dispatch emergency vehicles and deliver patients to appropriate hospitals. The authors conduct a simulation study to minimize the service time for the aftermath of an earthquake with a large number of casualties needing medical attention. The au-

thors consider a pick-up and delivery problem with real-time information. However, in most humanitarian relief operations, the decision maker has limited or uncertain information from multiple sources. In Shen, Ordóñez & Dessouky (2009), the authors discuss the stochastic vehicle routing problem. They develop routing model for large-scale emergencies that minimizes unmet demand with uncertain demand and travel time, with predefined service deadline and limited supply at the depot. They formulate three models Deterministic Model, Chance-constrained Program and Robust optimization to develop the preplanned vehicle routes and compare the results. The authors extended the work to include planning stage and operational stage in Shen, Dessouky & Ordóñez (2009). In this two-stage approach, the routes are developed in the planning stage and in the operational stage, using a recourse strategy, the actual routes are decided with adjustments based on the information revealed. Ngueveu et al. (2010) discuss cumulative capacitated vehicle routing problem (CCVRP). In this problem, the objective is to minimize the sum of arrival times at customers, instead of the traditional total length, subject to vehicle capacity constraints. This type of objective assumes significance when importance is attached to the satisfaction of customer need, e.g., supply of necessary goods or rescue after a natural disaster. In this paper, the authors discuss upper and lower bounding procedures for this new problem using memetic algorithm and properties of the CCVRP. In Nolz et al. (2010) the authors discuss a covering tour problem (CTP) where in the routes have to satisfy three criteria: (1) minimizes the sum of distances between all members of a population and their nearest facility (2) a tour length criterion and (3) minmax routing criterion that minimizes the latest arrival

time at a population center point. The authors develop Pareto-optimal solutions for this bi-objective CTP. Hentenryck et al. (2010) discuss the single commodity allocation problem (SCAP) for disaster recovery. SCAPs are stochastic optimization problems that combine resource allocation, warehouse routing, and parallel fleet routing. The objective function aims at minimizing three factors (1) the amount of unsatisfied demands (2) the time it takes to meet those demands (3) the cost of storing the commodity. For the fleet routing, the objective is to minimize the latest delivery time. They investigate the performance of a novel algorithm and validate it on hurricane disaster scenarios generated by Los Alamos National Laboratory. To summarize one can note that there is no unifying framework in the development of routing models. However, one commonality among the articles is that typical humanitarian routing model is multi-objective in terms of unmet demand, response time and other performance measures. Thus, in this work, we focus on development of routing models with different objectives and look at the performance of the models in the humanitarian relief operations.

As can be observed, there are numerous challenges present in distribution of humanitarian aid. There are many inherently conflicting objectives for humanitarian relief operations, right from the procurement phase to deciding on the inventory or supply levels at the distribution center. Sheu (2007*a*) mentions the various challenges in humanitarian logistics but specifically, discusses how crucial is the 3-day period right after the occurrence of a disaster. This highlights the importance of effectively utilizing the limited response period. Balcik & Beamon (2008) discuss the importance of positioning of distribution centers in a relief network (facility

decisions) and the amount of supplies to be stocked at each center (inventory decisions) in case of a quick-onset disaster using a variant of maximal covering location model. For this research, we assume that the distribution center location and inventory level are given and thus we focus on the distribution or routing model for relief operations.

Another stream of literature that is related to vehicle routing problem with priorities is vehicle routing with time windows or precedence relations. Variants of this problem have been extensively studied (e.g., Bräysy & Gendreau (2005*a*), Bräysy & Gendreau (2005*b*)). There is a similarity among the vehicle routing problem with priorities, the vehicle routing problem with time windows, and the vehicle routing problem with precedence constraints. In the case of humanitarian relief operations, lack of information on road network conditions and the actual state of the people in need do not make it viable for the customer to be assigned a time window. The assignment of priorities is realistic as it is based on the effects of the disaster that struck the region. Note that one can theoretically convert the network with priorities into a network with suitable time windows, but this is a complex task as we need to ensure that feasibility and optimality are maintained in the transformed network. There is some level of similarity between our problem and TSP with Precedence constraints. But, later in the paper, we show that our model is much more powerful than a TSP with Precedence constraints.

To summarize, concept of priorities has been used in the context of service operations, military operations and more recently in the humanitarian operations. Compared to service or military operations, priorities assume primary importance

in humanitarian relief operations since ignoring the priorities can be fatal to the survival of a customer. Some have studied the vehicle routing models in the context of humanitarian relief operations in conjunction with inventory allocation, and location of distribution center problems. Few articles have studied the vehicle routing problem but ignored the effect of node priorities. In this work, not only we study the effect of priorities on the vehicle routing problem for humanitarian relief operations but also propose a novel way (d-Relaxed Priority Rule) to model the relaxed form for enforcement of priorities. This provides the decision maker with optimal solutions that highlight the tradeoff between metrics like distance and response time. Thus our work is of tremendous value to practitioners and we make a significant contribution to the vehicle routing literature as it provides new variants for the VRP community to study.

4.4. Single Uncapacitated VRP with Priorities (u-HVRP)

In this simplified version of the relief operations, we assume that the vehicle has unlimited capacity, and no route-length restriction. Hence, all the nodes will be visited. One can imagine such a situation when the entire network's size and total demand needs permit it to be serviced by a single vehicle in one particular day. This model also serves as a base model for developing complex models with capacity and other restrictions. Campbell et al. (2008) consider such a model in their research work. Based on the literature we identify three possible objectives: (1) Minimize Total Distance (2) Minimize the Latest Arrival Time (3) Minimize

the Average Arrival Time. Corresponding to the three objectives, we develop three different formulations: MinDist(\mathbf{d}), MinMax, and MinSum respectively. In the first formulation MinDist(\mathbf{d}), we consider priorities for nodes and enforce with different \mathbf{d} -Relaxed Priority Rules. Campbell et al. (2008) consider MinMax and MinSum objectives in their paper and to evaluate these objectives, we ignore priorities for nodes and write down formulations MinMax and MinSum. This helps us compare the optimal solution(s) of our formulation MinDist with the \mathbf{d} -Relaxed Priority Rule with the optimal solutions using MinMax and MinSum objectives.

We use u-HVRP (Uncapacitated Humanitarian Vehicle Routing Problem) to denote this routing problem. In Section 4.4.1, we discuss the MIP formulations. In Section 4.4.2 we derive the worst-case bounds for this routing problem, and in Section 4.4.3, we illustrate the powerfulness and compactness of u-HVRP compared to a variant of Asymmetric Traveling Salesman Problem (ATSP).

4.4.1 MIP Formulations

All the notation used in this work is defined in Appendix C.1. MIP formulations for MinDist(\mathbf{d}) and MinMax and MinSum are provided in Appendix C.2 and C.3 respectively. The complete mathematical formulations are provided in appendices but we discuss the key observations for the MIP formulations².

There are a few observations that are worth-noting in these formulations. Firstly, notice that for all of the formulations, we do not need to explicitly write

²In this paper, please note that we use different typeface to differentiate parameters or given data and *decision variables* in the MIP formulations

down demand constraints as the vehicle has no capacity restrictions. Secondly, notice that all formulations do not have sub-tour elimination constraints. This is because service time constraints, as described for each of the three formulations, rule out the possibility of sub-tours in the optimal solution as shown in Proposition 1.

Proposition 1 *Optimal solution(s) to formulation MinDist, MinMax, and MinSum contain no sub-tours.*

Proof See Appendix C.2.1. •

The proof for the proposition works on the idea that any sub-tour will always contain the depot. Though these set of constraints are of $O(n^2)$, but it should be noted that they are not facet-inducing. The main difference between MinMax, MinSum and our MinDist(\mathbf{d}) is the constraints corresponding to \mathbf{d} –Relaxed Priority Rule that take into account the node priorities. Intuitively, one can write these constraints as $s_i \leq s_j \forall i \in \mathbf{N}_p, j \in \mathbf{N}_{p+\mathbf{d}+1}$. However, when the triangle inequality holds true for travel times between two nodes, we can strengthen these constraints, as shown in Proposition 2.

Proposition 2 *Constraints defined by equation (C.7), enforces the \mathbf{d} –Relaxed Priority Rule without affecting the optimality of the route as long as the triangle inequality holds true for travel time (T_{ij}) between node i and j for all $i, j \in \mathbf{N}$.*

Proof See Appendix C.2.2. •

This proposition gives us tighter version of \mathbf{d} –Relaxed Priority Rule constraints.

Apart from the strengthening of these constraints, depending on the \mathbf{d} level, one can write few additional inequalities. These are illustrated in Appendix C.2.3. For example, if $\mathbf{d} = 0$, then vehicle can never travel from a lower priority node to the higher priority node. Polyhedron defined by our MIP formulation is not full-dimensional and some of the inequalities used to described are not facets. There is a lot of literature (e.g., Miller et al. (1960), Desrochers & Laporte (1991), Gouveia & Pires (1999), Sherali & Driscoll (2002)) that try to tighten the formulations by developing facets or better bounds in similar problems, but, in this paper, we do not delve in depth into improving the solution time for reaching optimality. We stop our discussion of MIP formulations here, since the focus of our work is to understand the formulation(s) that can handle the different performance metrics for humanitarian relief operations. Further work is needed to strengthen and improve the MIP formulation for this new variant of routing problem.

Intuitively, we can see that the MinMax and MinSum objectives fail to give us a reasonable distribution of service times for nodes. These do not consider the relative differences in urgencies of nodes; they only focus on a particular node (MinMax) or assign the same level of urgency to all nodes (MinSum). On the contrary, the \mathbf{d} -Relaxed Priority Rule should capture the trade-off between the two performance metrics: distance and response times in a reasonable manner as it considers the urgencies of the nodes. In the next section, we study the distance traveled in the worst-case using MinDist(\mathbf{d}) formulation.

4.4.2 Bounds and Worst-case Examples for u-HVRP

In this section, we discuss the bounds for u-HVRP with $\text{MinDist}(\mathbf{d})$ formulation and show that the bound is tight as it is attainable in a the worst-case situation. Let $Z_{\mathbf{d},P}^*$ be the total distance traversed in the optimal tour by the vehicle with \mathbf{d} -Relaxed Priority Rule in a network with P priority classes. If the network has P priorities, then $\mathbf{d} = P - 1$ corresponds to u-HVRP without any consideration for the priorities. In this situation, it just becomes a standard TSP, so let us denote the optimal tour length in this situation as Z_{TSP}^* .

Theorem 1 *Let $Z_{\mathbf{d},P}^*$ and Z_{TSP}^* be the optimal tour length (distance) for u-HVRP with \mathbf{d} -Relaxed Priority Rule and for u-HVRP without any priorities, respectively. If the triangle inequality holds for the distances, then (a) $Z_{0,P}^* \leq PZ_{TSP}^*$, and (b) $Z_{\mathbf{d},P}^* \leq (P - \mathbf{d}) Z_{TSP}^*$.*

Proof (a) We will prove this by contradiction. Assume that $Z_{0,P}^* > PZ_{TSP}^*$. This implies that for any feasible tour τ for u-HVRP with 0-Relaxed Priority Rule, we have: $Z_{0,P}^\tau \geq Z_{0,P}^* > PZ_{TSP}^* \Rightarrow Z_{0,P}^\tau > PZ_{TSP}^*$. Let $\tau = (D, R_1, \dots, R_r, B_1, \dots, B_b, G_1, \dots, G_g, \dots, D)$. Figure 4.3(a) illustrates such a route for $P = 3$ classes. Let us construct the tour τ as follows. Consider a subgraph consisting of only priority 1 nodes and depot. Then we know that optimal length of such a tour, Z^{1*} , cannot exceed the optimal length of TSP tour for the entire graph. Let's denote such a tour as D, R_1, \dots, R_r, D and we have: $Z^{1*} \leq Z_{TSP}^*$. We do the same for priority 2 nodes and we get the optimal tour as D, B_1, \dots, B_b, D and optimal tour lengths are related as: $Z^{2*} \leq Z_{TSP}^*$. Similarly, for all other priority classes, we can construct TSP tours

as mentioned and for each priority class, the optimal length of the tour cannot exceed optimal length of TSP tour for the entire graph. From this set of P TSP tours, let us construct one single TSP tour for u-HVRP that satisfies 0–Relaxed Priority Rule. First, let us combine priority 1 TSP tour and priority 2 TSP tour as follows. Remove the two edges R_r, D and D, B_1 , and instead add edge R_r, B_1 . Since triangle inequality holds, replacing the two edges by a single edge will result in a tour for priority 1 and priority 2 that travels at most $Z^{1*} + Z^{2*}$. This is illustrated in the figure for $P = 3$ classes. Continue in this manner until all P TSP tours are combined into a single TSP tour. Thus, we obtain tour τ for u-HVRP with 0–Relaxed Priority Rule such that $Z_{0,P}^\tau \leq Z^{1*} + \dots + Z^{P*}$. And we also know that $Z^{1*} + \dots + Z^{P*} \leq PZ_{TSP}^*$. This implies that $Z_{0,P}^\tau \leq PZ_{TSP}^*$, which is a contradiction and hence proved. (b) Consider a graph with P classes and with degree of relaxation as \mathbf{d} . Recall that \mathbf{N}_p is the set of nodes with priority p . In the optimal tour for this problem, notice that any such set with priority p gets partitioned except \mathbf{N}_1 and \mathbf{N}_P . \mathbf{N}_1 and \mathbf{N}_P do not get partitioned because \mathbf{N}_1 is the highest priority class and \mathbf{N}_P is the lowest priority class, so no nodes need to be visited before and after these priority classes, respectively. We call this optimal partition C_p^* for priority p . This is illustrated for $P = 4$ classes for $\mathbf{d} = 1$ in Figure 4.3(c). This happens because each node needs to be visited only once with \mathbf{d} –Relaxed Priority Rule. This rule says that nodes of a priority p can be visited before nodes of priority $p - \mathbf{d}$. Figure 4.3(b) shows how the \mathbf{d} –Relaxed Priority Rule splits each of the node sets. Effectively, with \mathbf{d} –Relaxed Priority Rule, we are partitioning all nodes into new priority classes $\eta_1, \eta_2, \dots, \eta_{P-\mathbf{d}}$. Notice that, now the nodes in set η_p are visited before nodes in any

other set η_q where $q > p$. This means that original problem is equivalent to this new problem with $P - d$ priority classes and 0-Relaxed Priority Rule. From part(a), we know that the bound for this problem is $(P - d)Z_{TSP}^*$. Thus, we have proved that $Z_{d,P}^* \leq (P - d)Z_{TSP}^*$. •

Theorem 1 gives us an upper bound on the optimal distance for u-HVRP tour. In the next section, we show that this bound is tight for any number of priority classes when $d = 0$. For $d > 0$, we were able to provide worst-case examples to show that this bound is tight, for up to $P = 3$ priority classes.

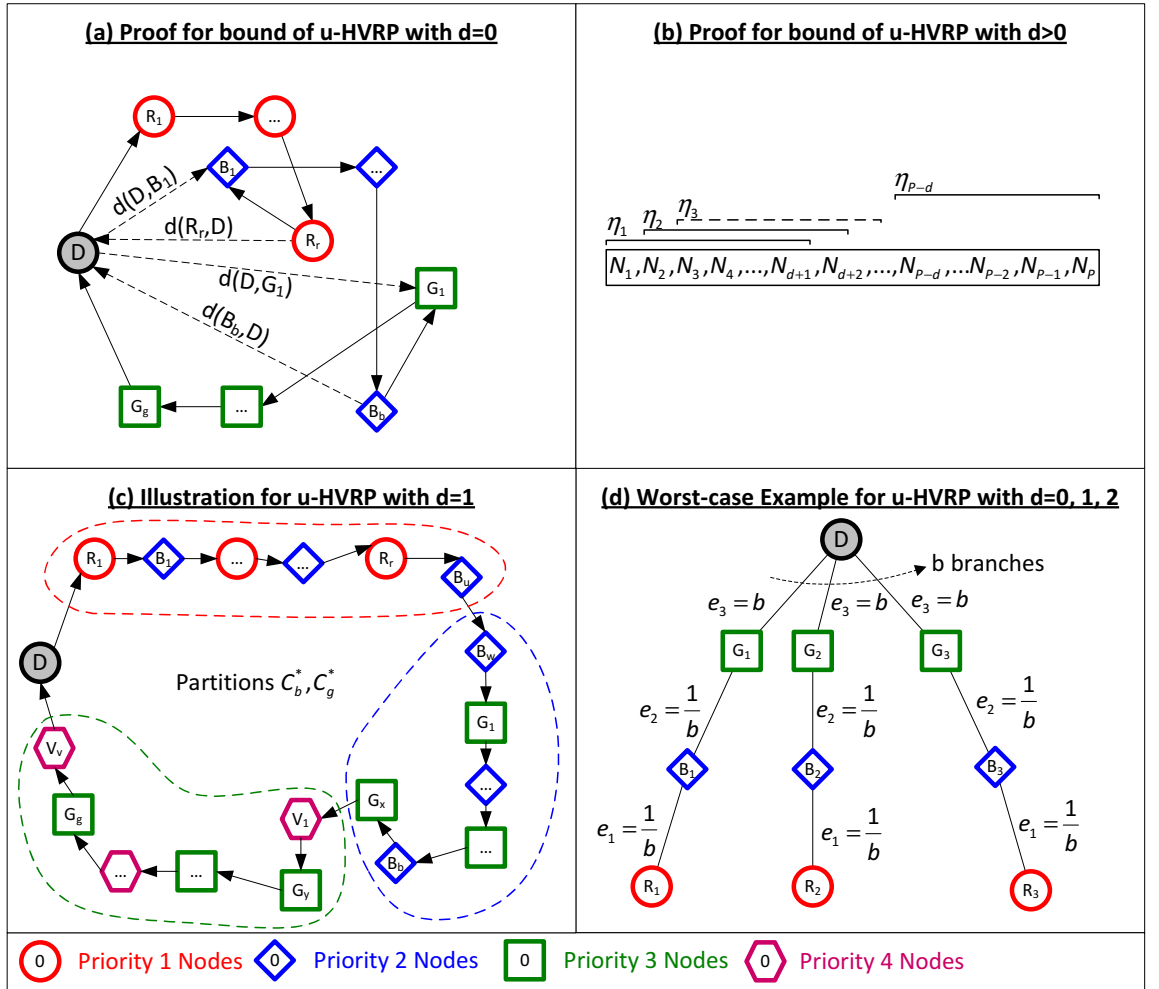


Figure 4.3: Proof for bound on optimal u-HVRP tour and Worst-case Example

4.4.2.1 Worst-case Example (Tree) when $d = 0$

For this situation, we devise a worst-case example, where in, the bound of PZ_{TSP}^* is attained in the limiting condition. Consider the network as shown in Figure 4.3(d). The network is a tree with distances and priorities for nodes as shown. The figure is an illustration just for $P = 3$ but one can trivially extend this to P priorities. Then, for such a network with P priority classes and $d = 0$, we can calculate Z_{TSP}^* and $Z_{0,P}^*$ by inspection. Thus we get, $Z_{TSP}^* = b(2e_1 + \dots + 2e_P)$ and $Z_{0,P}^* = (b)2e_1 + (2b - 1)2e_2 + (3b - 2)2e_3 + \dots + (Pb - (P - 1))2e_P$. Let the edge distance with P priority classes be $e_1 = \frac{1}{b}, \dots, e_{P-1} = \frac{1}{b}, e_P = b$ as illustrated in the figure. Substituting these values, we get: $Z_{TSP}^* = 2(b - 1) + 2b^2$ and $Z_{0,P}^* = (P - 1)P - \frac{(P-2)(P-1)}{b} + (Pb - (P - 1))2b$. Now consider the ratio of distances for optimal tour in u-HVRP and TSP, in the limiting condition when $b \rightarrow \infty$.

$$\lim_{b \rightarrow \infty} \frac{Z_{0,P}^*}{Z_{TSP}^*} = \lim_{b \rightarrow \infty} \frac{(P - 1)P - \frac{(P-2)(P-1)}{b} + (Pb - (P - 1))2b}{2(b - 1) + 2b^2} = P$$

Thus, for $d = 0$, the ratio is P and in fact we have found a worst-case example where in this bound is attainable for u-HVRP.

4.4.2.2 Worst-case Example (Tree) where $d > 0$

Theorem 1 says that in such a case, the optimal tour distance is at most $(P - d)Z_{TSP}^*$. The authors have been able to prove the tightness of the bound for up to $P = 3$ priority classes. For $P > 3$, the authors were unable to find worst-case examples.

Consider $P = 2$. The bound is attained trivially for $P = 2$ with $\mathbf{d} = 1$ as u-HVRP is equivalent to standard TSP in this case. Consider $P = 3$. Again here, if $\mathbf{d} = 2$ then the u-HVRP is equivalent to standard TSP and bound is attained trivially. Let us consider the interesting case of $\mathbf{d} = 1$. Let us consider the example shown in Figure 4.3(d). In this case, the optimal tour distances for u-HVRP and TSP can be calculated by inspection. Thus we get, $Z_{TSP}^* = b(2e_1 + 2e_2 + 2e_3)$ and $Z_{1,3}^* = (2b)e_1 + (2b)e_2 + (4b + 2)e_3$. Let the edge distance with $P = 3$ priority classes be $e_1 = e_2 = \frac{1}{b}, e_3 = b$ as illustrated in the figure. Substituting these values, we get: $Z_{TSP}^* = 4 + 2b^2$ and $Z_{1,3}^* = 4 + 4b^2 + 2b$. Consider the ratio of distances for optimal tour in u-HVRP and TSP, in the limiting condition when $b \rightarrow \infty$.

$$\lim_{b \rightarrow \infty} \frac{Z_{1,3}^*}{Z_{TSP}^*} = \lim_{b \rightarrow \infty} \frac{4 + 4b^2 + 2b}{4 + 2b^2} = 2$$

Thus, for $\mathbf{d} = 1$, the ratio is $P - 1 = 2$ and in fact we have found a worst-case example where in, this bound is attainable for u-HVRP.

4.4.2.3 Worst-case Example (Cluster) when $\mathbf{d} = 0$

Consider the network in Figure 4.4. The network displays a cycle with $n + 1$ (where $n \geq 2$) clusters. All the clusters are identical and contain P nodes. Each of the P priority classes is represented in each cluster. For such a network with $\mathbf{d} = 0$, we can calculate Z_{TSP}^* and $Z_{0,P}^*$ by inspection. For the former, we obtain,

$$Z_{TSP}^* = e_1 + \dots + e_{n+2} + 2\delta(n + 1)P = 2 + n - 2 + 2\delta(n + 1)P = n + 2\delta(n + 1)P.$$

The optimal HTSP tour is illustrated for priority 1 and 2 classes in the figure. The vehicle starts service for priority 1 in cluster 1 and ends service for priority 1 in cluster $n + 1$. Since all priority 1 nodes are serviced, the vehicle begins service for priority 2 nodes in cluster $n + 1$ itself and ends service in cluster n . This continues for the remaining priorities and the vehicle finishes service for all priority P nodes in cluster $n + 2 - P$. From this cluster, there are two possible routes for the vehicle to return to the depot, as shown in the figure. We denote the distance traversed as $dist^{top}$ and $dist^{bot}$. Notice that the distance traversed in each cluster is the same as that of the TSP, that is, $2\delta(n + 1)P$. Thus, we can write down the expression for the optimal tour distance as

$$\begin{aligned}
Z_{0,P}^* &= \overbrace{e_1 + e_2 + \dots + e_{n+1}}^{\text{Priority 1}} + \overbrace{e_{n+2} + e_1 + \dots + e_n}^{\text{Priority 2}} + \overbrace{e_{n+1} + e_{n+2} + e_1 + \dots + e_{n-1}}^{\text{Priority 3}} + \dots \\
&\quad + \overbrace{e_{n+4-P} + \dots + e_1 + e_2 + \dots + e_{n+2-P}}^{\text{Priority } P} + \min(dist^{top}, dist^{bot}) + 2\delta(n + 1)P \\
&= (n - \epsilon) + (n - 1 + \epsilon) + \overbrace{(n - 1) + \dots + (n - 1)}^{(P-2) \text{ times}} + \min(dist^{top}, dist^{bot}) + 2\delta(n + 1)P.
\end{aligned}$$

Notice that $dist^{top} = n + 1 - P$ and $dist^{bot} = P - 1$. When we choose $n \geq 2(P - 1)$ then $dist^{bot} \leq dist^{top}$. Thus, we can write:

$$Z_{0,P}^* = 2n - 1 + (n - 1)(P - 2) + P - 1 + 2\delta(n + 1)P = nP + 2\delta(n + 1)P.$$

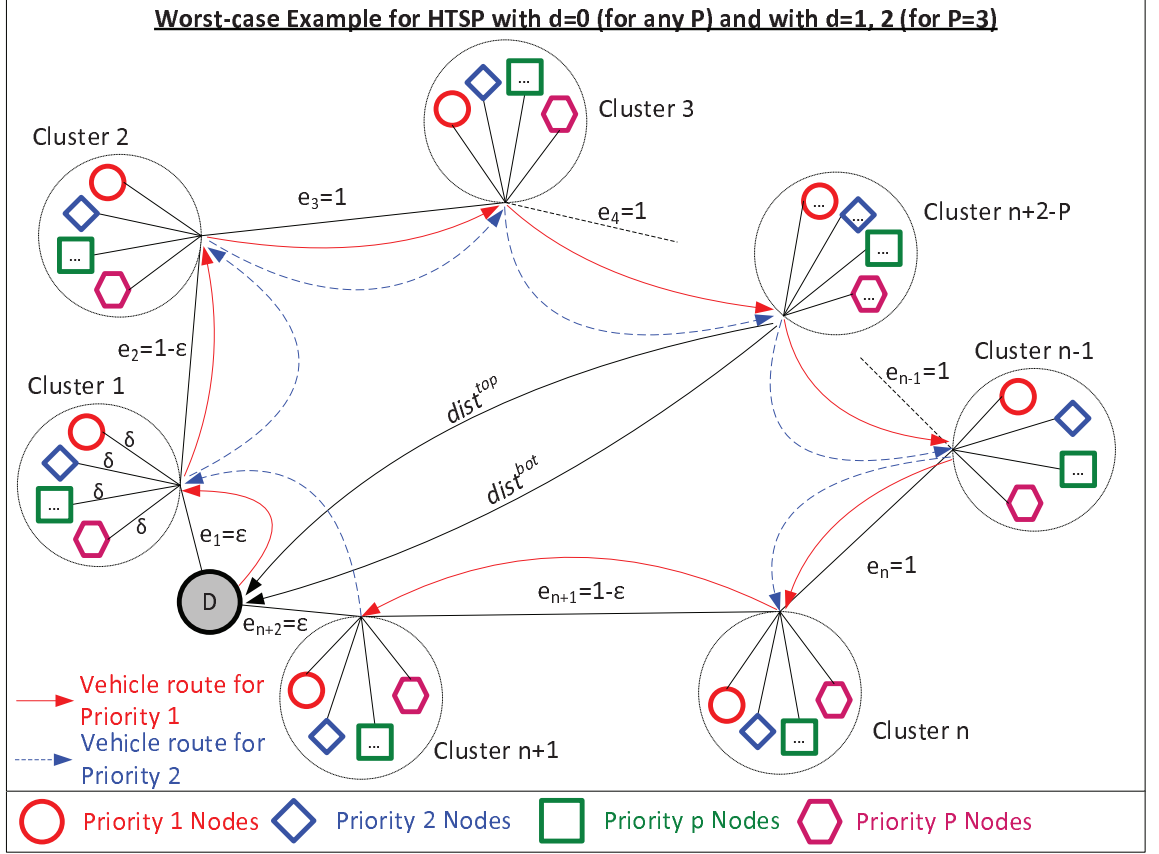


Figure 4.4: Worst-case Example for HTSP with $d = 0$ (any P) and $d = 1$ ($P = 3$)

Now consider the ratio of optimal lengths for the HTSP vs. the TSP, where $n \geq 2(P - 1)$, and in the limiting condition when $\delta \rightarrow 0$, we obtain

$$\lim_{\delta \rightarrow 0} \frac{Z_{0,P}^*}{Z_{TSP}^*} = \lim_{\delta \rightarrow 0} \frac{nP + 2\delta(n+1)P}{n + 2\delta(n+1)P} = \frac{nP}{n} = P.$$

Thus, for $d = 0$, the ratio is P and we have found a worst-case example where the bound is attainable.

4.4.2.4 Worst-case Example (Cluster) where $d > 0$

Theorem 1 says that in such a case, the optimal HTSP distance is at most $(P - d)Z_{TSP}^*$. The authors have been able to prove the tightness of the bound for up to $P = 3$ priority classes. For $P > 3$, the tightness of the bound is an open research question.

Consider $P = 2$ classes. When $d = 1$, the bound is attained as the HTSP is equivalent to the TSP. Consider $P = 3$. When $d = 2$, the HTSP is again equivalent to the TSP and the bound is attained. The remaining case is $d = 1$. To prove that the bound is attained in this case, refer to the example shown in Figure 4.4. The optimal tour distances for HTSP and TSP can be calculated by inspection. From Section 4.4.2.3, we have $Z_{TSP}^* = n + 6(n + 1)\delta$. In the optimal HTSP tour, the vehicle will travel to cluster 1 to service nodes of priority 1 and 2, then to cluster 2, and so on, until cluster $n + 1$. At this cluster, since all priority 1 nodes are serviced, it starts servicing nodes of priority 3 (only priority 3 nodes because all priority 2 and priority 1 nodes have already been serviced together) and keeps doing it until it reaches cluster n . Then we can write: $Z_{1,3}^* = n - \epsilon + n - 1 + \epsilon + 1 + 6(n + 1)\delta = 2n + 6(n + 1)\delta$. Consider the limiting ratio of distances as $\delta \rightarrow 0$, and this yields

$$\lim_{\delta \rightarrow 0} \frac{Z_{1,3}^*}{Z_{TSP}^*} = \lim_{\delta \rightarrow 0} \frac{2n + 6(n + 1)\delta}{n + 6(n + 1)\delta} = \frac{2n}{n} = 2.$$

Thus, for $P = 3$ and $d = 1$, the ratio is $P - d = 2$ and the bound is attained.

4.4.3 Relation of u-HVRP to ATSP

In this section, we study the relation between u-HVRP with $\text{MinDist}(\mathbf{d})$ formulation and show that our model is much more powerful than an ATSP or Precedence ATSP.

For u-HVRP with 0-Relaxed Priority Rule: Consider small example illustrated in Figure 4.5. Figure 4.5(a) represents the original graph on which we would like to solve for u-HVRP with 0-Relaxed Priority Rule. Figure 4.5(b) shows the corresponding transformation needed and we can easily see that the optimal tours in both are of length 23. In a general network, one can easily transform u-HVRP into equivalent ATSP as shown in Figure 4.6(a). Trivially, finding an optimal ATSP tour on this transformed network is equivalent to finding an optimal u-HVRP tour in the original problem. To our knowledge, there are no known tight bounds for ATSP exhibiting such a structure described. Our worst-case analysis for $\mathbf{d} = 0$ provides an attainable bound for this variant of ATSP.

For u-HVRP with \mathbf{d} -Relaxed Priority Rule where $\mathbf{d} > 0$: It is not easy to find an obvious transformation of u-HVRP into an equivalent ATSP when $\mathbf{d} > 0$. To illustrate, consider Figure 4.6(b) when $\mathbf{d} = 1$. Here, one needs to find *that optimal* partition C_b^* and solve for optimal ATSP tour. This is not a trivial exercise as there are $2^{|B|}$ partitions! One can possibly imagine that this complexity grows rapidly as \mathbf{d} gets higher. The powerfulness of our model is illustrated here as it combines the optimization on these two aspects: it partitions the nodes *optimally* as discussed

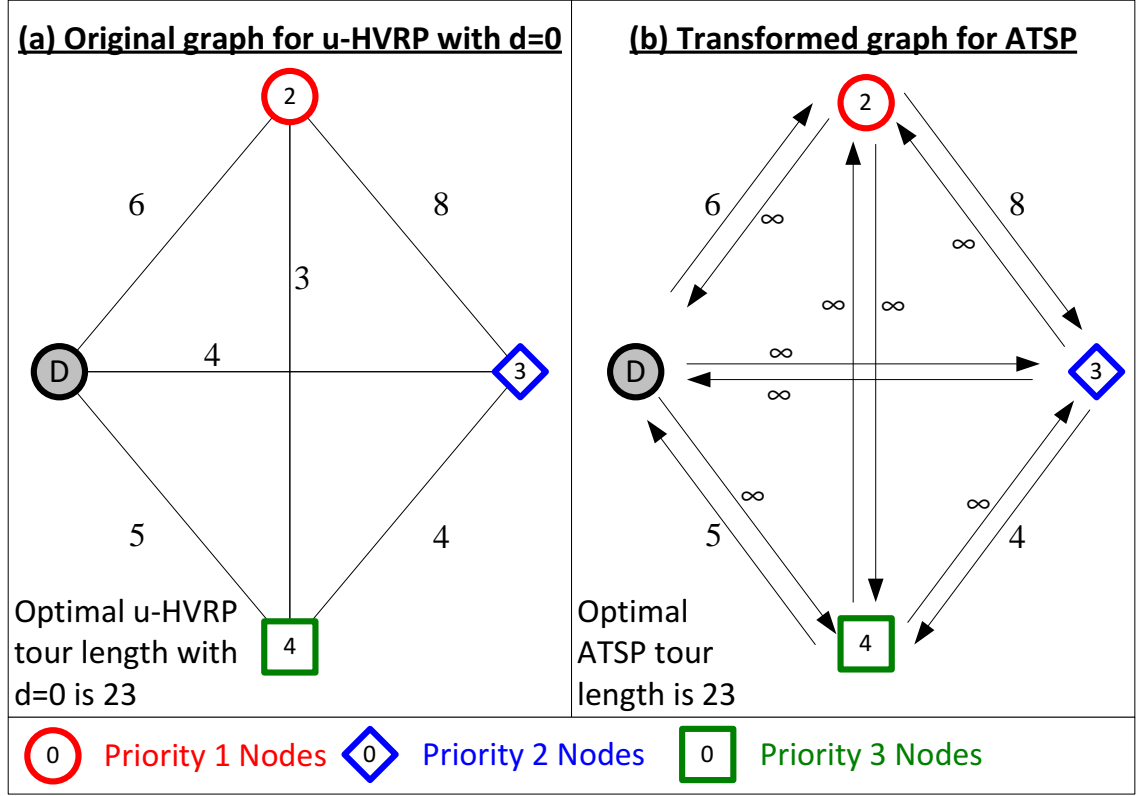


Figure 4.5: Transformation of u-HVRP with $d=0$ into ATSP

earlier and then finds an optimal ATSP tour. If one had to find such an optimal tour theoretically using just ATSP, first, one needs to find all partition of the nodes and then solve for least cost ATSP tour over all partitions. Thus, our model is much more powerful and compact as it combines the partitioning and finding ATSP tour. In the next section, we conduct computational experiment to study how our model captures the performance metrics like distance and response time.

4.5. Discussion of Computational Studies

The goal of our computational studies is to understand the appropriateness of our model with priorities to different objectives for existing routing models with

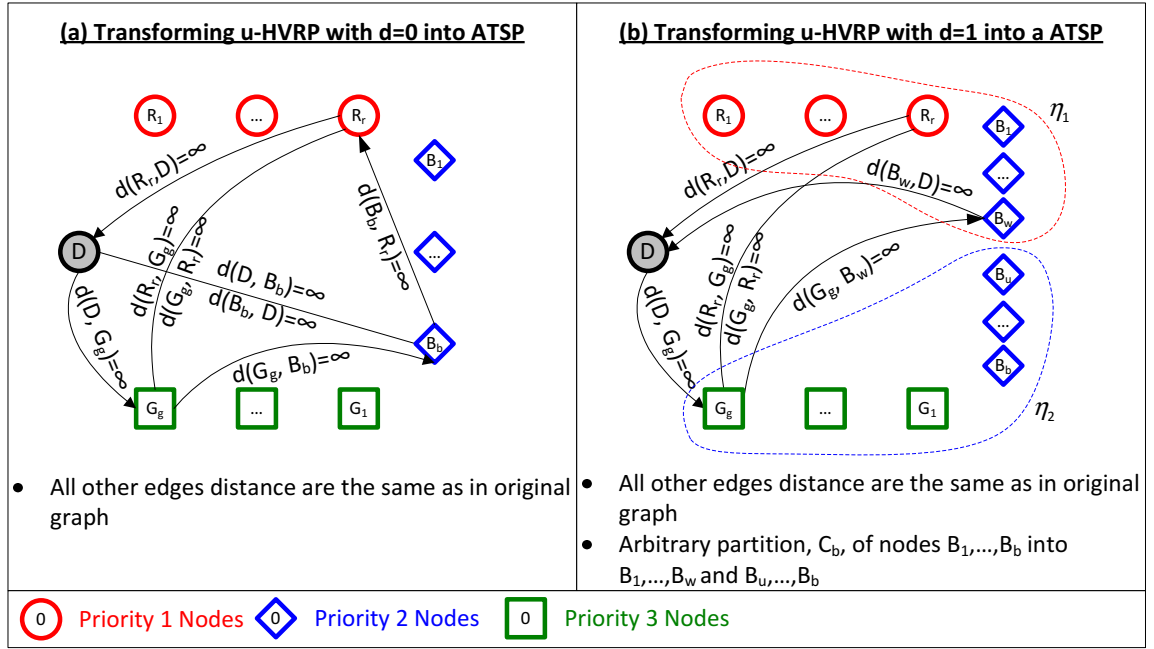


Figure 4.6: Transformation of u-HVRP into ATSP

respect to performance measures like Total Distance, and Latest Response Time. Towards this goal, we focus on optimality rather than on computational speed, so that we can analyze the applicability of our routing models for relief operations, in terms of these performance measures.

We run the MIP formulations proposed in the previous section, with different objectives on 11, 21, 30 and 40 nodes networks. The networks are assumed to have 3 priority classes and the unloading times are assumed to be negligible for this study. The travel times between nodes are assumed to be same as the distance between the nodes. The MIP formulations are coded using OOPs concept in C++. We use optimization software CPLEX 11.0 (©IBM-ILOG) on a 2.61GHz machine with 3.25GB RAM, and a AMD Athlon 64 X2 Dual Core Processor 5000+. The optimal solutions, computational times, and other performance measures from these

networks are discussed in the remainder of this section.

Table 4.1 shows the optimal results of computational runs on 21, 30, and 40 nodes networks using formulations MinDist with d -Relaxed Priority Rule in Appendix C.2 along with valid inequalities in C.2.3 and formulations MinMax, and MinSum given in Appendix C.3. The computational complexity increases as the d value increases and this is evident from the computation times to reach optimality. From Table 4.1, we see that as d increases, the total distance traveled decreases at the expense of distribution of service times. The optimal tours for network 1 with 21 nodes is illustrated in Figure 4.2. The total distance traversed in the strictest form of d -Relaxed Priority Rule is, 52.12%, 56.15%, 68.39% higher than the ones where the tour does not have to follow priorities for 21 nodes, 30 nodes and 40 nodes networks. This is reasonable for humanitarian relief operations, but as discussed in previous, the ratio go upto 300%. Apart from distance, applying d -Relaxed Priority Rule also gives us different distribution for service times for priority classes. If we enforce the strictest or the 0-Relaxed Priority Rule, we see that though overall response time is a bit longer, but the nodes are served in the order of their urgency. If severely damaged locations are not served immediately, it may prove fatal for these locations. As one can observe, it is not always required to enforce the order strictly as noticed in the 1-Relaxed Priority Rule results for these three networks. The advantage of using this rule is that it provides optimal solutions to the decision maker with a tradeoff between distance and service times. Thus, this helps in pooling resources for better serving the nodes or customers in a disaster.

Figure 4.7 illustrates the optimal solutions for different objectives discussed

Table 4.1: Optimal Results for u-HVRP (with inequalities)

ID(n)	d	Distance	Time ³	Priority 1 Service Times [min, max]	Priority 2 Service Times [min, max]	Priority 3 Service Times [min, max]
1 (21)	0	29.07	0.14s	[1.26, 10.85]	[11.31, 19.96]	[20.65, 28.37]
	1	23.06	5.57m	[1.26, 12.42]	[2.87, 21.96]	[13.34, 21.56]
	2	19.11	0.11m	[2.06, 14.22]	[1.10, 17.82]	[0.70, 16.27]
3 (30)	0	59.32	0.09m	[1.26, 19.90]	[20.53, 36.99]	[37.62, 58.62]
	1	45.08	3.76h	[2.06, 20.68]	[1.10, 40.44]	[22.12, 44.38]
	2	37.99	2.96h	[5.69, 36.73]	[1.10, 34.12]	[1.50, 35.34]
4 (40)	0	82.61	0.41m	[1.26, 29.79]	[37.79, 54.25]	[54.88, 81.91]
	1	62.14	77.52h	[1.26, 29.90]	[1.10, 57.50]	[33.90, 61.44]
	2	49.06 ^a	91.07h	[2.06, 44.17]	[1.10, 47.77]	[0.70, 46.22]

h:hours, m:min, s:sec, ^a1.12% gap

for u-HVRP. It represents a relief situation for an 11-node network that is served by a single vehicle without capacity or route-length restriction. With increasing d values, the distance decreases from 13.62 to 9.99. In the case of $d=0$, the latest service time is 12.92, compared to 8.89 for $d=1$ and 7.72 for $d=2$. Comparing these three models to MinMax model, we see that the total distance traversed remains the same as for $d=1$ but total service time is 6.86. Similarly MinSum is comparable to $d=2$. This means that MinSum behaves just like a standard VRP model without any consideration for priorities of the nodes.

The distribution of service times for priority classes can be compared by looking at measures like earliest service time, latest service time, and average service time. First, we compare by looking at the start times for each priority class. Consider the five models for an 11-node network show in Figure 4.7 corresponding to MinDist with $d=0, 1$, and 2, MinMax and MinSum formulations. When we compare the start times for priority 1 nodes and priority 3 nodes for the five models,

we see that it is (1.26, 1.26, 2.06, 2.06, 2.06) vs. (10.47, 6.04, 0.70, 0.70, 0.70) for MinDist with $d=0$, 1, and 2, MinMax, and MinSum respectively. The first two models perform reasonably well, but the last three models – MinDist ($d=2$), MinMax and MinSum can prove fatal to the people badly affected by a disaster as urgent nodes are kept waiting compared to nodes which do not require immediate assistance. When we observe the latest service times for each of the priority classes, we again observe that the service times for priority 1 nodes are higher than priority 3 nodes for the last three models – MinDist ($d=2$), MinMax and MinSum. Thus, when one considers the earliest and latest service times for the three priority classes in MinDist($d=2$) model, MinMax, and MinSum, the results show us that all classes are treated as of equal urgency. This behavior is undesirable as delaying service to urgent nodes can further deteriorate the crisis situation and result in more nodes needing urgent assistance. For networks with 21, 30 and 40 nodes, we could not solve MinMax and MinSum formulations to optimality and hence do not report these results.

Next, we compare average service times by priority class for the four models. Table 4.2 shows the average service times for different models on 11, 21, 30, and 40 node networks. For an 11 nodes network, we see that MinMax gives us an unreasonable distribution of service times as urgent nodes are kept waiting compared to MinDist ($d=1$). Intuitively, MinMax focuses on *that* one particular node; whereas our model can also be visualized as an multi-node extension of the MinMax model. For this specific 11 nodes network, MinDist ($d=1$) is better and a reasonable optimal solution. Now, let us look at how the d -Relaxed Priority Rule affects the

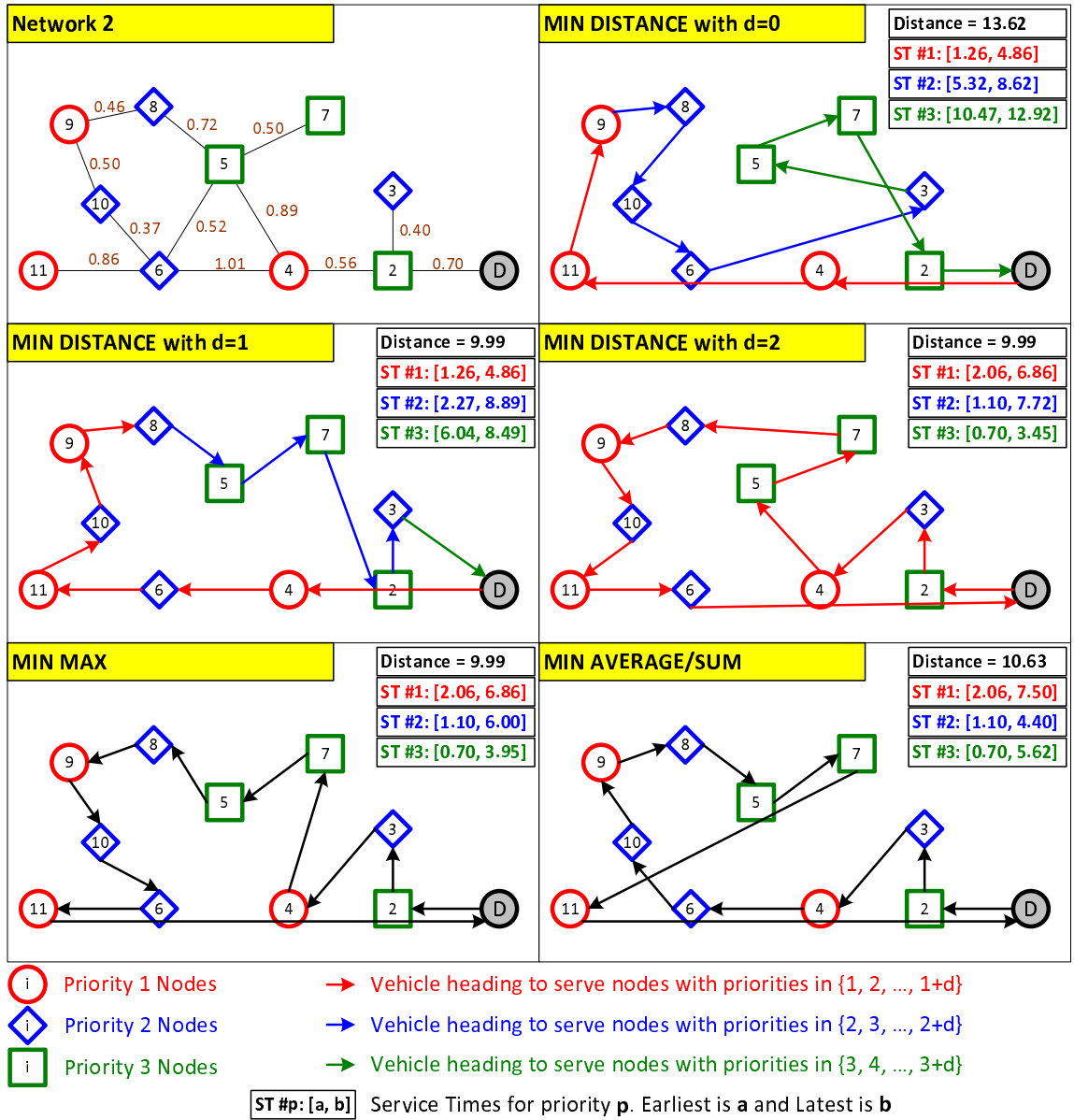


Figure 4.7: Optimal results for different u-HVRP formulations on 11 nodes network average service times for 21, 30 and 40 node networks. For the 21 node network, with MinDist ($d=2$), all the priority classes are serviced at more or less at the same time. Contrarily, the average service times for the MinDist ($d=0$) model show the sequential serving of customers based on their priority class. With the MinDist ($d=1$) model, Priority 1 nodes (which require urgent service) are visited first com-

pared to Priority 3 nodes at the cost of 20.7% increase in distance traversed. For 30 nodes, MinDist ($d=2$) gives us average service time as 24.23, 18.34, 16.29 compared to MinDist($d=1$) which gives us 10.85, 20.21 and 35.64 for Priority 1, 2, and 3 classes, respectively. The strictest model, MinDist($d=0$), gives us average service times increasing in priority class as 9.64, 28.12, and 50.45, respectively. One can observe similar behavior in the distribution of service times for 40 nodes. Thus, this rule provides the decision maker with a set of Pareto-optimal solutions in terms of distance and average response times. The unique aspect of this rule is that it systematically captures the urgencies in the optimal solution.

Table 4.2: Comparison of distribution of service times for u-HRP

ID(n)	Model	Distance	Priority 1 Avg. Service Time	Priority 2 Avg. Service Time	Priority 3 Avg. Service Time
2(11)	MinDist ($d=0$)	13.62	3.08	6.72	11.45
2(11)	MinDist ($d=1$)	9.99	3.96	5.58	7.27
2(11)	MinDist ($d=2$)	9.99	4.68	4.78	2.37
2(11)	MinMax	9.99	4.96	3.55	2.33
2(11)	MinSum	10.63	4.50	3.00	3.81
1(21)	MinDist ($d=0$)	29.07	6.35	15.44	24.61
1(21)	MinDist ($d=1$)	23.06	6.76	10.72	17.42
1(21)	MinDist ($d=2$)	19.11	8.10	9.33	8.91
3(30)	MinDist ($d=0$)	59.32	9.64	28.12	50.45
3(30)	MinDist ($d=1$)	45.08	10.85	20.21	35.64
3(30)	MinDist ($d=2$)	37.99	24.23	18.34	16.29
4(40)	MinDist ($d=0$)	82.61	13.89	43.87	70.10
4(40)	MinDist ($d=1$)	62.14	16.17	29.00	61.44
4(40)	MinDist ($d=2$) ^a	49.06	22.28	23.53	23.07

h:hours, m:min, s:sec, ^a1.12% gap

To conclude, the objectives MinMax and MinSum do not consider the relative differences in urgencies of nodes; they only focus on a particular node (MinMax) or assign the same level of urgencies for all nodes (MinSum). On the other hand,

as we can see, the d -Relaxed Priority Rule captures the trade-off between the two performance metrics – distance and response time.

4.6. Conclusions

In this research, we discussed the challenges that are faced in the last mile distribution of the aid to disaster victims. We developed routing models that consider the priorities or a proxy for urgency of a location, to cater to humanitarian relief operations. We formulated a d -Relaxed Priority Rule which provides flexibility to the decision maker. This rule captures the trade-off between operational efficiency and the humanitarian nature of the operations. We develop MIP formulations for a single vehicle without capacity restrictions. We derive the bounds for the u-HVRP and provide with a worst-case example that attains this bound in the limiting condition. Also, we show that our model is much richer as it encompasses ATSP with precedence constraints. As a result, we were able to provide a tight bound for this ATSP variant. Computationally, the routing models are successfully run on test problems with 21, 30 and 40 nodes network to optimality (or nearly optimal) using CPLEX optimization software. Finally, we discuss the performance of the MIP formulations on the performance metrics – distance traversed and response time (earliest, average, and latest). We show that VRP models which ignore priorities like standard MinMax or MinSum gives us unreasonable distribution of service times. The objectives MinMax and MinSum do not consider the relative differences in urgencies of nodes; they only focus on a particular node (MinMax) or

assign the same level of urgencies for all nodes (MinSum). On the other hand, as we can see from this work, the **d**–Relaxed Priority Rule captures the trade-off between the two performance metrics – distance and response time. Our work is of practical value as it provides a richer model for the distribution of aid in humanitarian relief operations. Also, our work is of tremendous research value to the vehicle routing problem (VRP) literature as we have defined a new class of VRPs.

There are many possible extensions to this work. In this work, we have ignored the capacity limitations for the vehicle. Imposing capacity limitations can potentially affect the other humanitarian relief operations’ performance measures like demand satisfied, and number of customers served. In the presence of priorities, relaxing the single to multiple vehicles can pose new challenges to the distribution system. In this work, we ignored the computational speed of solving these routing models; hence future research can be geared towards to solving these problems in real-time, efficiently to optimality or near-optimality.

4.7. Summary of Insights

In this chapter, we discussed the last mile distribution in the humanitarian relief chain. The key insights from this chapter are:

- Developed a new routing model that considers the node priorities. Priorities are a proxy for urgency of a location. Introduced a new constraint, **d**–Relaxed Priority Rule that provides flexibility to the decision maker. This rule captures the trade-off between operational efficiency and the humanitarian nature of the

operations.

- Developed MIP formulations for a single vehicle without capacity restrictions. To reduce the computation time, valid inequalities were developed.
- Derived the bounds for the u-HVRP. Also, provided worst-case examples for $d = 0$, with any number of priority classes and for $d = 1, 2$, for up to 3 priority classes. Thus, the tightness of the bound is proved in these situations.
- For $d = 0$, we show that the u-HVRP can be transformed into ATSP. And for $d > 0$, there is no *a priori* transformation (as far as we know). Our model performs the partition of priority classes and then devises the optimal tour for these partitions with precedence relations.
- Computationally, the routing models are successfully run on test problems with 11, 21, 30, and 40 nodes network to optimality (or nearly optimal) using CPLEX optimization software.
- Finally, discussed the performance of the MIP formulations on the performance metrics – distance traversed and response time (earliest, average, and latest). Showed that VRP models which ignore priorities like standard MinMax or MinSum give us unreasonable distribution of service times. The objectives MinMax and MinSum do not consider the relative differences in urgencies of nodes; they only focus on a particular node (MinMax) or assign the same level of urgencies for all nodes (MinSum). On the other hand, from this research, the d -Relaxed Priority Rule captures the trade-off between the two performance

metrics – distance and response time.

- Our work is of practical value as it provides a richer model for the distribution of aid in humanitarian relief operations. Also, our work is of value to the vehicle routing problem (VRP) literature as we have defined a new class of VRPs.

Chapter 5

Essay 4: Humanitarian Vehicle Routing Problem: Capacitated Case

5.1. Introduction

As discussed in previous Chapter, the goal of a humanitarian relief operation is to deliver as much aid as possible but at the same time, due to limited resources, the operations have to be efficient on other metrics as well. In the case of u-HVRP, we highlighted the fact that the routing models need to incorporate node priorities so that routing operations are efficient on the response time as well the distance. With additional restrictions on the vehicle(s) like capacity and route-length, humanitarian relief routing poses new challenges in the presence of node priorities. Due to capacity or route-length restrictions, the vehicle may be able to satisfy only a subset of the victims. For any humanitarian relief operations, satisfying a subset of customers is never ideal, but relief organizations face such a problem if one were to consider a snapshot of the humanitarian relief operations spanning many days. Thus, the need for operational efficiency assumes utmost importance given such limited resources.

Towards this goal, measuring operational efficiency of humanitarian relief operations is not trivial. For example, consider a single vehicle with only capacity restriction. Route that was designed to satisfy as much demand as possible does not necessarily imply that the route is efficient in terms of other measures like distance, response time and number of customers satisfied. Such issues are not just

limited to routing operations with a single vehicle. To illustrate, consider routing model with multiple vehicles having only capacity restriction. With multiple vehicles at its disposal, the relief organization face a new challenge. Should they use all vehicles to service all the nodes simultaneously? Or should the vehicles be split up to service the nodes separately? Thus, multiple vehicles gives rise to the problem of plenty, where in, the resources can be used simultaneously or sequentially. For example, one can imagine two routes which satisfy as much demand as possible but these routes may differ on the response time for the nodes. Thus, due to the capacitated nature of the problem, along with humanitarian restrictions, imposes new challenges for the routing models. It should be noted that to evaluate the performance of a humanitarian relief operation, one needs not a single, but multiple metrics (e.g., Beamon & Balcik (2008)). Based on humanitarian literature, in our work, we plan to compare routes using the following four metrics: (i) Distance, (ii) Response Time (Earliest, Latest, Average), (iii) Fill Rate (or % Demand Satisfied), (iv) Number of Customers Serviced. Traditional or non-humanitarian routing models do not capture the performance metrics for humanitarian operations and, thus, perform at a poor level on these metrics.

In this research, we are concerned with the delivery of a relief good (single product) from a distribution center to the people in need as efficiently as possible with the ultimate objective of minimizing the loss of life. With additional restrictions in the presence of node priorities poses interesting research questions for the vehicle routing problem: (1) How do we simultaneously capture performance metrics like distance, demand satisfied, number of customers served and response time in our

models? (2) How does the route perform on these measures with the **d**–Relaxed Priority Rule? In this research, we systematically analyze the trade-offs between the four metrics for different routing models. Thus, we believe that our models give a lot of flexibility to the decision maker in terms of handling political and social issues.

5.2. Setup

We plan to answer the research questions posed in Section 5.1 by extending the setup for u-HVRP that was discussed in the previous Chapter. Specifically, we consider a more realistic situation where the routing operations can be performed with a single or multiple vehicles consisting of homogeneous fleet with capacity and route-length restrictions. Not only all the routes have to satisfy the **d**–Relaxed Priority Rule but also other restrictions. Apart from the **d**–Relaxed Priority Rule and other restrictions, we impose a key additional rule, called “Order of Demand Fulfillment”, that caters to the humanitarian aspect of the routing problem.

The Order of Demand Fulfillment rule states that at the end of the day (or time horizon) all higher priority nodes need to be satisfied if a vehicle has enough available capacity. In this research, as we are dealing with humanitarian relief operations, the primary goal is to deliver as much supply as possible to the people who are most needy, and thus, to this effect, we enforce this rule. It means that at the end of the time period (can be a day, week, etc.), if there is enough capacity of the vehicle available, then all the higher priority nodes have to be satisfied before the vehicle

can satisfy *any* node of lower priority. This ensures that the higher priority nodes are *not* left unsatisfied at the expense of providing supplies to the lower priority nodes. For example, consider a vehicle of capacity 120. Let the network have two priority classes with total demand of 100 units for priority 1 and 50 units for priority 2. Then, when we impose this rule, it will eliminate solutions in which the vehicle supplies 80 units to priority 1 nodes and 20 units to priority 2 nodes. Such solutions are undesirable as priority 1 represents people with urgent needs which have to be met when there is enough vehicle capacity. If $d \geq 0$, the two rules together imply that lower priority nodes may be visited before higher priority nodes on a given day, provided that all higher priority nodes are visited by the end of the day. This rule is illustrated in Figure 5.1 for a 13-node network with two priority classes. We have one vehicle with capacity of 50 units and total demand for priorities 1, 2 is 40 and 43 respectively. With 1–Relaxed Priority Rule, though both the routes utilized the entire capacity of the vehicle, the route on the left satisfies the Order of Demand Fulfillment rule whereas, the route on the right violates this rule as priority 1 nodes are left unsatisfied at the end of the day.

Apart from these two key restrictions – d –Relaxed Priority Rule and Order of Demand Fulfillment, we make assumptions that are typical to vehicle routing problems with capacity and route-length restrictions. For the sake of completeness, we write down all the assumptions: (i) the depot is designated as node 1 with a service time of 0. Vehicle(s) can leave and return to the depot at most once after servicing the nodes. (ii) Customers are designated as nodes $\{2, \dots, n\}$. Each node can be serviced at most by one vehicle. (iii) Capacity of vehicle cannot be exceeded.

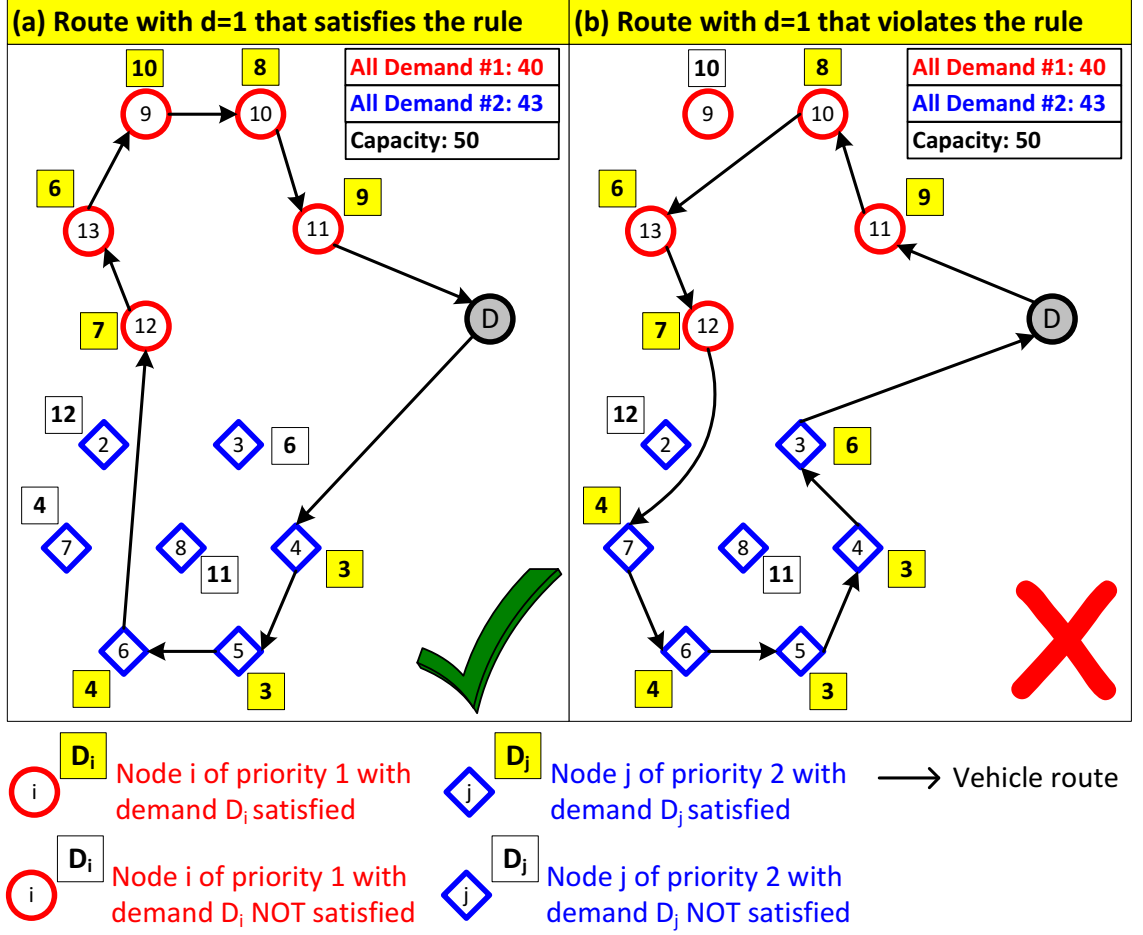


Figure 5.1: Illustration of Order of Demand Fulfillment Rule

(iv) Vehicles have route-length restriction (in terms of distance traversed) due to fuel, working hours, etc. (v) All vehicle routes have to obey the d -Relaxed Priority Rule, and Order of Demand Fulfillment. We assume the routing is done on complete graphs for reasons mentioned in the previous Chapter. Depending on the availability of the number of vehicles, the total supply and d -Relaxed Priority Rule for the relief operations, one can consider different models. Primarily, we consider two different models - Single Capacitated VRP with Priorities and Multiple Capacitated VRP with Priorities. We use the following shorthand notation, 1-HVRP and m-HVRP

to denote these models respectively.

The outline of this work is as follows. In Section 5.3 we discuss the literature relevant for these two models. In Section 5.4 we highlight the new challenges and develop mathematical models using Mixed Integer Programming (MIP) techniques for 1-HVRP and m-HVRP. Next, we use small-size test problems (up to 30 nodes) and compare optimal solutions from our models on the four performance metrics discussed. Computational Study is discussed in Section 5.5 and finally we present Conclusions in Section 5.6.

5.3. Literature Review

In the previous Chapter, we discussed literature associated with vehicle routing problems that consider the concepts similar to priorities in the service industry, military and a few humanitarian applications. In this section, we discuss relevant literature to our two problems 1-HVRP and m-HVRP.

Our two problems 1-HVRP and m-HVRP, have similarity to Orienteering or Team-Orienteering Problem in VRP literature. In an orienteering problem, the goal is to maximize the total score, which is obtained by collecting the reward at each node visited on the tour subject to a limit on the total distance that can be traveled. In case of the Team-Orienteering Problem, we have multiple vehicles with route-length restrictions that can collect the prizes. This problem and its variants have been well-studied in the literature. For most of the models in the literature, the focus has been on improving the MIP formulations, focus on improving the branch-

and-bound algorithm, and developing fast heuristics that generate optimal or near-optimal solutions. For example, Boussier et al. (2007) discuss exact algorithms for the team orienteering problem. In their work, the authors propose a branch-and-price algorithm for solving a team orienteering problem. However, as in this paper, and others, they do not consider priorities in their mathematical models. Our work on 1-HVRP and m-HVRP provides a new variant of Orienteering Problem that considers the node priorities.

Next, let us look at the literature that uses concepts similar to priorities in the context of orienteering problems. Few such variants are Orienteering Problems with Time Windows or Orienteering Problems with Precedence Relations. As shown in the previous Chapter, our model is much more powerful as d -Relaxed Priority Rule designs the partitions optimally such that the sequence for visiting nodes is optimally determined.

Let us look at the routing problems in the context of humanitarian relief operations. As discussed in the earlier Section, designing routing model is non-trivial as one needs to consider the multiple performance measures. Shetty et al. (2008), discussed earlier, consider a team orienteering problem in their paper. However, their model does not reflect the humanitarian relief situation as the Order of Demand Fulfillment Rule is violated and the model does not capture the other performance metrics (for example, response time). Schmitz & Niemann (2009), Smith et al. (2010), discussed earlier, do not consider the capacity or route-length restriction for the vehicles. In Liu et al. (2007), the authors develop routing model for distribution of medical supplies in large-scale emergencies. They develop an objective function

that is a combination of unmet demand and total delay but it is unclear how the weights are designed for each of the objectives. Tzeng et al. (2007) discuss development of relief-distribution model at the planning stage that optimizes multiple objectives like minimizing total cost, minimizing the total travel time and maximizing the minimal satisfaction. The last objective is reflective of fairness compared to first two objectives which reflect the efficiency of the model. Balcik et al. (2008) discuss the last mile distribution in humanitarian relief chain. The objective is to minimize the total costs which is the unweighted sum of routing costs and penalty costs over all days in the planning horizon. Campbell et al. (2008) discuss routing for relief efforts using two types of objective functions. In the first, they minimize the maximum arrival time at a node. This means that nodes that are far away from the depot can be reached as early as possible for service. However, this is not an appropriate objective function to cater to networks with nodes of varying urgency levels, as nodes that are not that far out do not affect the objective function. In the second formulation, they minimize the average arrival time. It is an aggregate measure, so when all nodes have more or less the same level of urgency this will work well but there is no guarantee that all high-priority nodes will be visited before low-priority nodes are visited. In Shen, Ordóñez & Dessouky (2009), the authors discuss the stochastic vehicle routing problem. They develop routing model for large-scale emergencies that minimizes unmet demand with uncertain demand and travel time, with predefined service deadline and limited supply at the depot. They formulate three models Deterministic Model, Chance-constrained Program and Robust optimization to develop the preplanned vehicle routes and compare the results. The

authors extended the work to include planning stage and operational stage in Shen, Dessouky & Ordóñez (2009). In this two-stage approach, the routes are developed in the planning stage and in the operational stage, using a recourse strategy, the actual routes are decided with adjustments based on the information revealed. Ngueveu et al. (2010) discuss cumulative capacitated vehicle routing problem (CCVRP). In this problem, the objective is to minimize the sum of arrival times at customers, instead of the traditional total length, subject to vehicle capacity constraints. This type of objective assumes significance when importance is attached the satisfaction of customer need, e.g., supply of necessary goods or rescue after a natural disaster. In this paper, the authors discuss upper and lower bounding procedures for this new problem using memetic algorithm and properties of the CCVRP. In Nolz et al. (2010) the authors discuss a covering tour problem (CTP) where in the routes have to satisfy three criteria: (1) minimizes the sum of distances between all members of a population and their nearest facility (2) a tour length criterion and (3) minmax routing criterion that minimizes the latest arrival time at a population center point. The authors develop Pareto-optimal solutions for this bi-objective CTP. Hentenryck et al. (2010) discuss the single commodity allocation problem (SCAP) for disaster recovery. SCAPs are stochastic optimization problems that combine resource allocation, warehouse routing, and parallel fleet routing. The objective function aims at minimizing three factors (1) the amount of unsatisfied demands (2) the time it takes to meet those demands (3) the cost of storing the commodity. For the fleet routing, the objective is to minimize the latest delivery time. They investigate the performance of a novel algorithm and validate it on hurricane disaster scenarios

generated by Los Alamos National Laboratory.

To summarize one can note that there is no unifying framework in the development of routing models for humanitarian relief operations. However, one commonality among the articles is that typical humanitarian routing model is multi-objective in terms of unmet demand, response time and other performance measures. Thus, in this work, we focus on development of routing models, 1-HVRP and m-HVRP, that consider the different objectives and look at the performance of the models in the humanitarian relief operations.

Based on the literature, we notice that the issue of priorities has not been addressed in orienteering problems. In this work, we study the effect of priorities on the vehicle routing problem for humanitarian relief operations. Also, traditional orienteering problems look at maximizing only the total reward. In our problem, if demand is considered as the reward, then maximizing demand does not necessarily mean the routes are optimal on other performance metrics like response time, distance and number of customers satisfied. Thus, our work nicely bridges the gap in the vehicle routing problems literature for humanitarian relief operations, where in we develop routing models that cater to multiple objectives. Also, our work is of value to vehicle routing literature as it provides new variants for the VRP community.

5.4. Routing Models

In this section, we discuss the two different models: Single Capacitated VRP with Priorities (1-HVRP) and Multiple Capacitated VRP with Priorities (m-HVRP). We highlight the issues with optimization methods for these routing models and discuss the key aspects of MIP formulations¹ for these two models. The complete mathematical formulations are provided in Appendices D.2 and D.3.

5.4.1 Single Capacitated VRP with Priorities (1-HVRP)

To state this model, consider a single capacitated vehicle with a restriction on the total route-length it can travel in one tour. As the vehicle has limited capacity and route-length restriction, it can service only a subset of nodes. Each node of the network has a fixed demand for the relief item. There can be different categories of goods like housing and shelter, water and sanitation etc with varying level of need but in our problem we assume that there is only good or one package of such items to be delivered. Assume that the relief agency has single vehicle with finite capacity for each category of good. Apart from the capacity for goods; assume that the vehicle also has a route-length restriction due to fuel and working hours, etc. In our model, apart from the distance traveled, we also consider the distance equivalent of unloading time at the customer location in the design of the route. As one can see, this model represents a more realistic version of relief operations compared to the problem discussed in the previous Chapter. In this case, it is possible that the

¹In this work, please note that we use different typeface to differentiate parameters or given data and *decision variables* in the MIP formulations

vehicle may not be able to meet the entire demand. If the vehicle is able to meet all of the demand, then this becomes the uncapacitated problem or u-HVRP.

The goal of relief operation in such a setting is to deliver maximum total demand and perform this task efficiently. Note that designing an objective function for this goal is a non-trivial task. We discuss two different approaches, along with their limitations. The first approach is to solve this problem using a two-stage procedure. The second approach is to use an objective function that is a weighted combination of two objectives - Total Distance and Total Demand Delivered. We write down the formulations for these two approaches and show that the second approach does not always find the *pareto optimal* solutions that maximize the total demand delivered and also minimize the total distance.

5.4.1.1 Two-stage Optimization

The general idea behind this two-stage optimization is that we can sequentially tackle two of the metrics used to measure performance of a humanitarian relief operation. If one wants the least total distance traversed, then it is theoretically zero, which corresponds to least total demand met, which is zero. If one wants to achieve maximum total demand met, then total distance will increase correspondingly. That is, both these metrics Total Demand Delivered and Total Distance Traversed increase or decrease in each other. However, we are interested in maximizing the Total Demand Delivered but also minimizing the Total Distance Traversed. To guarantee such pareto optimal solutions, we perform optimization in

two stages, described as follows:

Stage I: Maximize Total Demand Delivered. The objective in this stage reflects the humanitarian nature of the routing problem. The optimal solution gives us the maximum that can be delivered to the nodes subject to capacity of the vehicle or the route-length restriction. If route-length is not an issue but capacity of the vehicle is binding, then total demand met is just the capacity of the vehicle. On the other hand, if the vehicle has capacity that exceeds the total demand but with a limit on the route-length it can travel, then the route-length restriction determines how much demand can be served. The MIP formulation for this stage is denoted as 1-HVRP-MaxDemand.

Stage II: Minimize Total Distance Traversed. In this stage, we develop vehicle route that traverse minimal distance, given that we know the maximum demand that can be delivered in Stage I. If the route-length restriction is not binding, then the route that is generated in Stage I may not be efficient since maximal demand assignment corresponding to optimal solution in Stage I need not traverse the network efficiently to serve these nodes. This is illustrated in Figure 5.2 for a network of 10 nodes. The route-length restriction of 30 units is not binding as the vehicle traverses 29 units in Stage I optimization. Thus, one can pose the following question: Given the demand assignment corresponding to maximum demand delivered, what is the most efficient route with respect to the total distance? To obtain efficiency with respect to the total distance, we minimize total distance traversed subject to the maximum demand delivered in Stage I. It is not clearly evident how the maximum demand can be enforced as a constraint in this stage. In fact, the maximum demand

delivered constraint can be added to this model in two different ways:

Stage II(i): Individual Assignment. In this model, we enforce the constraint that the nodes that had been served in Stage I *need to be* serviced in this stage as well. This is equivalent to creating a subgraph with the subset of nodes that were serviced in Stage I and solving u-HVRP on these subset of nodes. The MIP formulation in this stage is denoted as 1-HVRP-MinDist-i.

Stage II(a): Aggregate Assignment. In this model, we impose the constraint that the total demand met in this formulation must be at least the total maximum demand obtained in Stage I. Here, we solve the problem on the entire network, since we are aggregating all demand for a priority class. The MIP formulation in this stage is denoted as 1-HVRP-MinDist-a.

One can expect that Individual Assignment might perform poorly on total distance metric relative to Aggregate Assignment, since it is more restrictive. However, one should also note that Individual Assignment can also result in a higher number of customers serviced relative to the Aggregate Assignment. Distance and Number of Customers serviced are two performance metrics, and the policy maker can decide on the appropriateness of the model to be applied. Irrespective of the Stage II method employed, this two-stage procedure gives us an efficient and humanitarian routing solution.

The Two-Stage optimization approach is illustrated for two networks - Network 1 and Network 2 in Figure 5.2. First consider Network 1, where, we have a 10 nodes with a total demand of 120 units and one priority class. The nodes are served by a single vehicle with capacity of 100 units and route-length limit of 30 units. After

Stage I, notice that 7 out of 8 nodes are served since capacity available is slightly lower than the total demand and the total distance traversed is 29 units. After Stage II using Individual Assignment, the number of nodes served is 7 and the distance traversed is 27. One can see that this helps in making the route efficient and yet serve the maximum demand. After Stage II using Aggregate Assignment, one can see that total distance traversed is 25 units, better than using Individual Assignment, but the number of nodes served is 6, lower than using Individual Assignment.

Individual Assignment is more restrictive on the distance traversed but it *can* serve a higher number of customers compared to Aggregate Assignment. Consider another example, as illustrated for Network 2 in Figure 5.2. This is a 9-nodes tree network with one priority class, served by one vehicle with a capacity of 50 units and a route-length limit of 18 units. Though distance is an important criterion, it cannot be the sole measure to be used in deciding the route as this example illustrates. Individual Assignment gives us an optimal route that traverses 8 units but serves 4 customers whereas Aggregate Assignment gives us a lower distance of 6 units, but serves only 3 customers. And from an operations perspective one cannot choose between serving 3 customers or 4 customers. As shown in this example, it is not obvious that Aggregate Assignment should be used in all situations as Individual Assignment can potentially perform well on another performance metric – number of customers served. Thus, for humanitarian operations, one can not trivially rule out one approach in favor of another just based on the perspective of operations. However, in this research, our goal is to put forth operationally *efficient* routing models in front of the decision maker to help him/her arrive at a decision

appropriate to the crisis situation.

The notation and formulations for the two stages are defined in the Appendices D.2.1, D.2.2, and D.2.3 respectively. We develop a MIP formulation for Stage I based on the orienteering problem. The key aspect of orienteering problem is the profit or prize collected when vehicle services a node. Here, we use the demand at each node as the prize or profit collected when vehicle services the node. One important constraint that is enforced here is the Order of Demand Fulfillment since all higher priority nodes have to be serviced if the vehicle has enough capacity. Stage II(i) formulation is similar to MinDist(d) formulation for u-HVRP. Stage II(a) formulation is similar to u-HVRP, except for aggregate demand constraint. Since the focus of this research is designing relevant models for humanitarian relief operations, we do not delve on improving the computational speed by improving the MIP formulations or designing clever heuristics. The computational results for this approach are illustrated in Section 5.5.

5.4.1.2 Weighted Objective

Another approach common in the literature is to use a weighted combination of the two objectives: Total Demand Delivered and Total Distance Traversed. In this situation, the feasible region for the MIP formulation is same as in formulation 1-HVRP-MaxDemand, except for the objective function, which is as follows:

$$\text{Max } \alpha \text{ Total Demand Delivered } - \beta \text{ Total Distance Traversed}$$

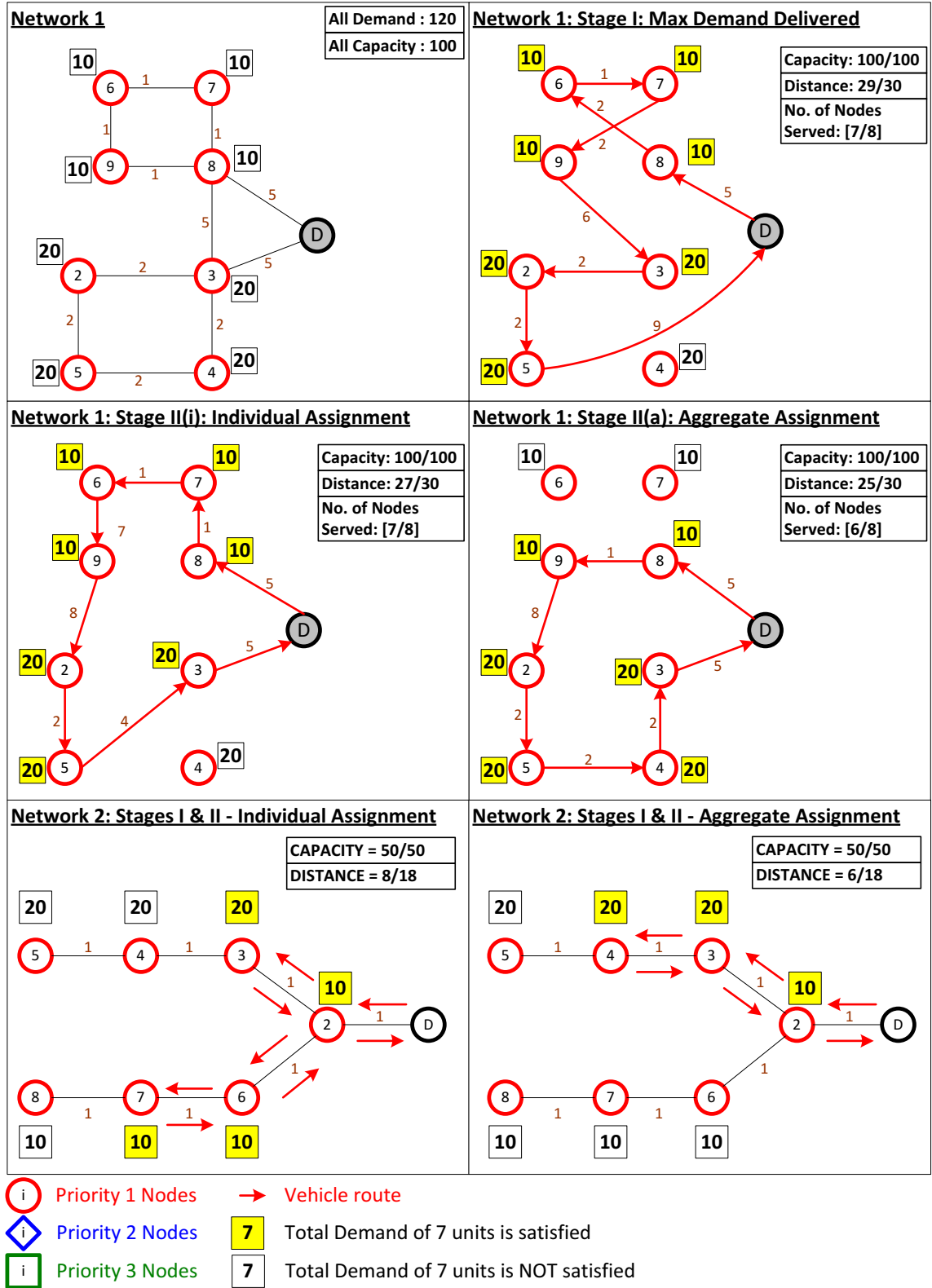


Figure 5.2: Illustration of Stage II Methods

where, $0 \leq \{\alpha, \beta\} \leq 1$. The complete formulation 1-HVRP-WeightedObj is provided in Appendix D.2.4. In this research, we examine this approach and find that this approach may not find that optimal solution which delivers maximum demand but is efficient in terms of distance. The intuition behind this can be explained as follows. When Total Demand Delivered has higher weight, the route is inefficient in terms of distance as long as the maximum demand is delivered. On the other hand, when weight for Total Distance Traversed is higher, the route places too much emphasis on the route-length and thus will start delivering at lower demand levels. This transition defeats the purpose of a humanitarian relief operation as nodes are left unsatisfied even though the vehicle has enough supply. For further analysis, we drop this weighted objective approach. The computational results for this approach are illustrated in Section 5.5.

5.4.2 Multiple Capacitated VRP with Priorities (m-HVRP)

This problem is similar to 1-HVRP except that now we have multiple vehicles that can handle the distribution of relief good. One can imagine such a problem in relief situations where the distribution center needs to dispatch a homogeneous batch of vehicles carrying one type of item. Each vehicle is limited by its capacity and the tour route-length. Note that the homogeneous assumption can be very easily relaxed in the MIP formulations.

With multiple vehicles, one additional issue that needs resolving is the enforcement of the d-Relaxed Priority Rule. Since we have multiple vehicles, one needs

to clarify whether this rule should be imposed globally or locally. The two types of enforcement are important because they impact two of the performance metrics – Latest Response Time and Number of Customers Satisfied, as explained below:

- (1) **Local Timing Rule:** In this, the route of each vehicle is consistent with the d –Relaxed Priority Rule. So, in this it is possible that the service time of a node with priority p is higher than a node with priority $p + d + 1$, since they can be serviced by two different vehicles. When we have the Local Timing Rule, although an individual vehicle is consistent with the d –Relaxed Priority Rule, a lower priority node can be serviced earlier than a higher priority node. This may be undesirable, but on the positive side, since the nodes are distributed between multiple vehicles, it is possible that all vehicles together cover a much bigger area of the entire network.
- (2) **Global Timing Rule:** In this, we enforce the d –Relaxed Priority Rule across *all* the vehicles. So, the service time for a node with priority p is *always* lower than for a node with priority $p + d + 1$, irrespective of which vehicle services these nodes. When we enforce d –Relaxed Priority Rule across *all* routes, then essentially all vehicles are dispatched as a batch to nodes of the same priority. This can result in vehicles reaching route-length restrictions at the same time, and, thus, all vehicles together may not be able to service the entire network. However, one big advantage to this approach is that the a priority class can be serviced in much less time.

These two approaches are relevant in different situations. When vehicle route-length is not a binding constraint, then the Global Timing Rule makes sense since it can service the entire priority class in much shorter duration as any point in the route, the focus of the relief fleet is servicing nodes of a priority class or for a more

general \mathbf{d} –Relaxed Priority Rule, nodes in $p, p+1, \dots, p+\mathbf{d}+1$ priority classes. On the other hand, if the route-length restriction is binding resulting in fewer number of customers served, then the Local Timing Rule helps in serving higher number of customers. For example, if the nodes of different priority classes are very far apart then dispatching all the vehicles to serve nodes of a priority class will result in all vehicles hitting the route-length restriction before they can hit the capacity restriction.

The two rules are illustrated in Figure 5.3. In this example, we have 12 nodes with two priority classes that are far apart. Total demand for priority 1 class is 40 units and for priority 2 is 47 units. The network is served by two identical vehicles of capacity 50 units and each with a route-length limit of 25 units. Let us impose the 0–Relaxed Priority Rule. With Local Timing Rule for enforcing priorities, after performing Stage I and Stage II optimization, the first vehicle serves *only* Priority 1 customers and second vehicle serves *only* Priority 2 customers. With Global Timing Rule for enforcing priorities, *both* vehicles first proceed to serve the priority 1 class and then proceed to serve Priority 2 customers, provided there is still available capacity and the route-length limit has not been reached. Comparing total distance traversed, with Local Timing Rule, first vehicle traverses 16 units and the second vehicle traverses 17 units whereas with Global Timing Rule, both the vehicles traverse 25 units. Capacities of the two vehicles are under-utilized in Global Timing Rule model as vehicles hit route-length limits even before the capacity is exhausted.

The main advantage of using the Global Timing Rule is that the range of

service times for each priority class is decreased as seen for each class in this example. Comparing the two rules, we can see that with Global Timing Rule, the route-length limit is more restrictive, whereas with the Local Timing Rule, the capacity of vehicle becomes more restrictive. Unlike this intentionally worst-case example where the two priority classes are very far apart, if applying the two rules do not result in leaving out customers, then one should appropriately choose the model based on the whether all vehicles should be directed towards servicing nodes in $p, p + 1, \dots, p + d + 1$ priority classes or the resources can be split up to service all priority classes simultaneously.

Similar to 1-HVRP, we employ the two-stage optimization procedure for this situation as well. We write MIP formulations for the two-stage optimization as discussed in Section 5.4.1.1, keeping in mind the two different ways one can enforce the d -Relaxed Priority Rule. Our goal in this study is to understand the appropriateness of our formulations in capturing the different performance metrics rather than on reducing the computational time by improving the MIPs or devising heuristics. The formulations are provided in Appendices D.3.1 and D.3.2 for the Local Timing Rule, and Appendices D.3.3 and D.3.4 for the Global Timing Rule. The MIP formulations corresponding to Local Timing Rule are denoted as: m-HVRP-local-MaxDemand for Stage I and m-HVRP-local-MinDist-i or m-HVRP-local-MinDist-a for Stage II Individual or Aggregate Assignment, respectively. Similarly, for the Global Timing Rule, the MIP formulations are denoted as m-HVRP-global-MaxDemand for Stage I and m-HVRP-global-MinDist-i or m-HVRP-global-MinDist-a for Stage II Individual or Aggregate Assignment, respectively.

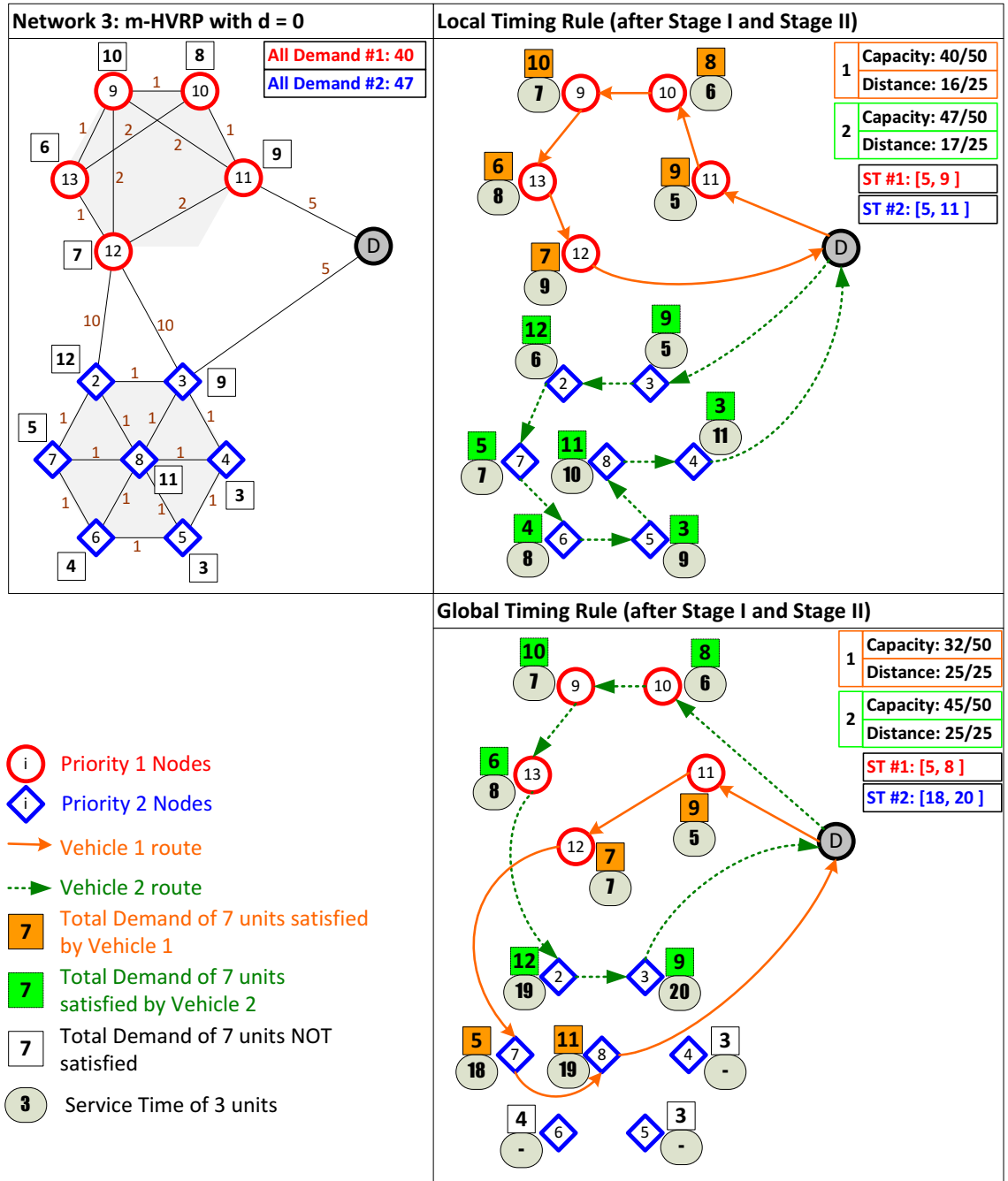


Figure 5.3: Illustration of Global vs. Local Timing Rules for m-HVRP

5.5. Discussion of Computational Studies

The goal of our computational studies is to understand the relevance of different routing models with respect to performance measures like Total Distance, Total Demand Met, Latest Response Time, and Number of Customers Served. Towards this goal, we focus on optimality rather than on computational time, so that we can analyze the applicability of our routing models for relief operations, in terms of these performance measures.

We run the MIP formulations proposed in the previous sections, with different objectives on 21, and 30 nodes networks. The networks are assumed to have 3 priority classes and the unloading times are assumed to be negligible for this study. The travel times between nodes are assumed to be same as the distance between the nodes as shown in the networks. The optimal solutions, computational times, and other measures from these networks are discussed in the next few sections. The MIP formulations are developed using OOPs concept in C++. We use optimization software CPLEX 11.0 (©IBM-ILOG) on a 2.61GHz machine with 3.25GB RAM, and a AMD Athlon 64 X2 Dual Core Processor 5000+.

5.5.1 1-HVRP

Here, we run the computational studies on 21 and 30 node networks. The network characteristics are given in the Table 5.1 (see also Appendix E). In the case of 21 nodes, the total demand is 132 units and in the case of 30 nodes, the total demand is 198 units. The results of computational runs on 21 and 30 nodes networks

using two-stage optimization procedure with formulations given in Appendix D.2.1, D.2.2, and D.2.3 are shown in Tables 5.2 and 5.3 for different route length limits and capacity. Computational time for reaching optimality is reasonable for 21 and 30 nodes networks. However, for a 40-node network, CPLEX did not reach optimality and hence are not reporting those results.

Table 5.1: Network Characteristics for 1-HVRP and m-HVRP

ID(n)	Priority Class	No. of Nodes	Total Demand	Demand at Individual Nodes (<i>Node #:Demand</i>)
4 (21)	1	6	44	4:12, 9:6, 11:3, 13:2, 16:12, 18:9
	2	8	49	3:4, 6:11, 8:7, 10:2, 12:8, 15:6, 17:3, 21:8
	3	6	39	2:10, 5:7, 7:4, 14:1, 19:1, 20:16
5 (30)	1	9	63	4:12, 9:6, 11:3, 13:2, 16:12, 18:9, 22:5, 23:10, 24:4
	2	11	76	3:4, 6:11, 8:7, 10:2, 12:8, 15:6, 17:3, 21:8, 28:10, 29:8, 30:9
	3	9	59	2:10, 5:7, 7:4, 14:1, 19:1, 20:16, 25:7, 26:9, 27:4

From Table 5.2, we see that as d increases, the total distance traveled decreases but at the expense of an inequitable distribution of service times. When the vehicle capacity is 100 units for 21 nodes network, it is never enough to satisfy total demand of 132 units; hence at best it can satisfy all demand of Priority 1 and 2 but only 7 units of Priority 3 nodes. Applying Stage II optimization (either Individual or Aggregate Assignment method) results in lowering the total distance traversed compared to just Stage I optimization. For this table, we have combined both the Stage II methods since they serve the same number of customers for these examples. But, in reality, with different problem data, it is possible to see why these Stage II methods assume importance as shown in Figure 5.2. Let us compare the average service times for different priority classes applying the d -Relaxed Priority

Table 5.2: Optimal Results for 1-HVRP using Two-stage Optimization on a 21 node network

d	L^{max}	CAP	Distance after Stage I (Stage II)	Demand met by Priority (#1, #2, #3)	Average Service Time by Priority (#1, #2, #3)	Time
0	20	100	19.55 (19.55)	44, 45, 0	6.35, 14.11, -	0.44s(0.03s)
1	20	100	19.82 (16.72)	44, 49, 7	8.02, 7.47, 14.57	4.08s(3.38s)
2	20	100	19.87 (16.15)	44, 49, 7	7.27, 7.70, 2.95	1.20s(1.28s)
0	36	100	29.95 (22.79)	44, 49, 7	5.72, 15.79, 20.64	0.06s(0.05s)
1	36	100	30.94 (16.72)	44, 49, 7	7.22, 9.03, 13.77	0.27s(3.61s)
2	36	100	31.15 (16.15)	44, 49, 7	7.27, 7.69, 2.95	0.03s(1.02s)
0	20	150	19.55 (19.55)	44, 45, 0	6.35, 14.11, -	0.53s(0.05s)
1	20	150	19.88 (19.88)	44, 49, 33	7.27, 7.87, 16.21	0.45h(5.30m)
2	20	150	19.75 (19.11)	44, 49, 39	7.59, 10.69, 8.11	0.76m(4.20s)
0	36	150	35.31 (29.07)	44, 49, 39	6.35, 15.44, 24.61	2.26s(0.36s)
1	36	150	34.70 (23.06)	44, 49, 39	6.76, 11.08, 17.55	5.69s(7.33m)
2	36	150	34.76 (19.11)	44, 49, 39	7.59, 10.69, 8.11	0.13s(4.78s)

h:hours, m:min, s:sec

Rule when the vehicle capacity and route-length limit are constrained. For a 21 node network with a vehicle capacity of 100 units and a route limit of 20 units, the optimal routes satisfy 89, 100, and 100 units of demand for different d values. In this situation, where the problem is severely constrained by both capacity and route-length, applying 0–Relaxed Priority Rule resulted in not *fully* utilizing the vehicle capacity and hence it may not be a preferred solution in humanitarian relief operations. On the contrary, the 1–Relaxed Priority Rule or the 2–Relaxed Priority Rule fully utilize the capacity. Thus, let us look at how these routes perform on distribution of service times. The average service times for Priority 1 are 8.02, 7.27, for Priority 2 are 7.47, 7.70 and for Priority 3 are 14.57, 2.95 across the two models. To ensure equitable distribution of service times, one might choose the

1–Relaxed Priority Rule. When the vehicle has the same capacity level but the route-length limit is increased to 36 units, the route limit is never binding in for $d=1,2,3$ as the capacity is fully utilized. Comparing the average service times, we see that 1–Relaxed Priority Rule performs well as the average service times are 7.22, 9.03, and 13.77 for Priority 1, 2 and 3 respectively. Let us consider the next situation when vehicle capacity is 150 units and thus never a constraint. When vehicle route-length is limited to 20 units, we can see that the 0–Relaxed Priority Rule and the 1–Relaxed Priority Rule *effectively* result in vehicle not fully utilizing its capacity. Though these models give optimal solutions that are result in ordered distribution of service times, these solutions may not be preferred in humanitarian relief operations since vehicle capacity is not fully utilized. On the other hand, when the vehicle route-length is at most 36 units, we see that the vehicle is able to satisfy all of the demand. Among the three models, 1–Relaxed Priority Rule results in an equitable distribution of service times for compared to other models. Table 5.3 shows the optimal results for two-stage optimization for a 30-node network. One can make similar observations from optimal results for 30 nodes network as reported in this table.

To summarize, for both networks, based on demand met for each priority class, notice that the optimal solutions satisfy the Order of Demand Fulfillment Rule. When vehicle capacity and route-length are restrictive, one should apply the d –Relaxed Priority Rule and not only look at the optimal solutions in terms of service times but also with respect to the other metrics like demand satisfied. These models thus help in developing optimal solutions that trade-off the pros and cons

of satisfying demand of urgent nodes at a faster pace vs. satisfying all nodes at a slower pace. These models can help the decision maker in the objective analysis of humanitarian relief operations.

Using Two-Stage optimization, we obtain Pareto-optimal solutions that maximize demand and are also efficient in terms of distance traversed. Such solutions are not trivial to obtain when we use the Weighted Objective approach. We run Weighted Objective approach using the formulation given in Appendix D.2.4 for 1-HVRP on a v21 node network and show the results in Table 5.4 for a single vehicle with capacity of 100 units and route-length limit of 36 units. As one can see from Table 5.4, arriving at the desired optimal solution depends on the weights associated with the two objectives. If the weights are not properly designed or if the step size is too high, it is possible that the Pareto-optimal solution might be missed in the transition of domination of one objective to the other. Compared to this ambiguous approach, the two-stage optimization approach *guarantees* us finding the desired Pareto-optimal solutions.

Table 5.3: Optimal Results for 1-HVRP using Two-stage Optimization on a 30 node network

d	L^{max}	CAP	Distance after Stage I (Stage II)	Demand met by Priority (#1, #2, #3)	Time
0	35	175	34.70 (34.65)	63/63, 68/76, 0/59	5.56s(1.84s)
1	35	175	34.82 (31.15)	63/63, 76/76, 36/59	0.20h(0.19h)
2	35	175	34.95 (30.57)	63/63, 76/76, 36/59	0.73m(0.39h)
0	35	200	34.65 (34.65)	63/63, 68/76, 0/59	6.20s(2.47s)
1	35	200	34.50 (34.50)	63/63, 76/76, 47/59	1.78h(5.23h)
2	35	200	34.67 (34.67)	63/63, 76/76, 55/59	0.02h(7.16h)

h:hours, m:min, s:sec

Table 5.4: Optimal Results for 1-HVRP using Weighted Objective Approach on a 21 node network

$\alpha(\beta)$	Distance	Demand Met	Priority 1 % Dem Met	Priority 2 % Dem Met	Priority 3 % Dem Met
1.0 (0.0)	34.59	100	44/44	49/49	7/39
0.9 (0.1)	22.79	100	44/44	49/49	7/39
...
0.3 (0.7)	22.79	100	44/44	49/49	7/39
0.2 (0.8)	18.61	87	44/44	43/49	0/39
0.1 (0.9)	2.52	12	12/44	0/49	0/39

5.5.2 m-HVRP

We run our computational exercises on 21 node and 30 node networks using formulations given in Appendix D.3. For this study, we assume that the network is served by at most two identical vehicles with a capacity and a route-length limit. The demand characteristics are assumed to be same as that given in Table 5.1. Due to the complexity of the problem, we were unable to run the computational exercises to optimality for 30 nodes network with the 2–Relaxed Priority Rule. Our focus in this research is to study the relevance of various models to humanitarian relief operations and the authors believe that future research can be done to reduce computational time for solving bigger networks.

Recall that as discussed in Section 5.4.2, we can enforce the d–Relaxed Priority Rule locally or globally. Tables 5.5 shows the optimal results of using Two-stage optimization with Local Timing Rule on a 21 node network. For 21 node network, we fix the route-length limit at 18 units and vary the capacity. The network can be serviced by 2 vehicles with a capacity of 50 units each or with a capacity of 75 units each. As one can notice from Table 5.5, route length is never binding for

either of the situations and the vehicle capacity constraint is binding only when the total capacity is 100 units. First, let us discuss the different performance measures for total capacity of 100 units. For different d values, the demand satisfied is the same as the total vehicle capacity, thus all the models are the equivalent on this performance measure. The total distance traversed is 24.72, 21.43, 21.43 as the d value increases. Thus, one can see that $d=1$ and $d=2$ are equivalent with respect to distance traversed. We will combine the discussion of service times distribution using the Local Timing Rule with the discussion of optimal results when we apply the Global Timing Rule.

Next, let's discuss the optimal results for total capacity of 150 units. For the different d values, all of demand is satisfied, hence all the models are the equivalent with respect to this performance measure. The total distance traversed is 29.51, 25.04, and 23.44, corresponding to increasing d values. The two vehicles traverse a longer distance as all of the demand is satisfied with 150 units of total capacity. The vehicles do not always distribute the load evenly but rather it is possible that one vehicle entirely satisfies demand of a priority class and the second vehicle satisfies the demand of another priority class. Though it is beneficial in terms of meeting demand, the drawback is that if there are many Priority 1 nodes compared to Priority 3 nodes then that last Priority 1 node may be serviced at a much later time compared to a Priority 3 node. Later in this section, we discuss how the Global Timing Rule addresses this issue.

Table 5.6 shows the optimal results for Two-stage optimization on 30 nodes with the Local Timing Rule. We do not discuss this situation in detail as one can

Table 5.5: Optimal Results for m-HVRP using Two-stage Optimization and Local Timing Rule on 21 nodes network

d (Stage)	Total dis- tance	Distance by Vehicle (#1, #2)	CAP used by vehicle (#1, #2)	Demand met by Priority for Veh 1	Demand met by Priority for Veh 2	Time
Two vehicles with $CAP = 50$, $L^{max} = 18$						
0 (I)	30.69	13.04, 17.65	50, 50	35, 15, 0	9, 34, 7	2.80s
0 (II)	24.72	13.94, 10.78	50, 50	21, 29, 0	23, 20, 0	8.26s
1 (I)	30.79	17.35, 13.44	50, 50	32, 11, 7	12, 38, 0	3.49s
1 (II)	21.43	12.44, 8.99	50, 50	23, 27, 0	21, 22, 7	0.32h
2 (I)	30.45	12.56, 17.89	50, 50	26, 17, 7	18, 32, 0	2.28s
2 (II)	21.43	8.99, 12.44	50, 50	21, 22, 7	23, 27, 0	9.03m
Two vehicles with $CAP = 75$, $L^{max} = 18$						
0 (I)	34.01	17.97, 16.04	69, 63	0, 41, 28	44, 8, 11	0.49m
0 (II)	29.51	11.77, 17.74	62, 70	0, 41, 21	44, 8, 18	9.25s
1 (I)	35.29	17.70, 17.59	71, 61	21, 15, 35	23, 34, 4	0.17m
1 (II)	25.04	15.85, 9.19	70, 62	23, 29, 18	21, 20, 21	9.58h
2 (I)	34.99	17.40, 17.59	58, 74	9, 21, 28	35, 28, 11	2.97s
2 (II) ^a	23.44	14.25, 9.19	70, 62	23, 29, 18	21, 20, 21	3.39h

h:hours, m:min, s:sec, ^a5.81% gap

make similar observations. However, the computational time grows rapidly as the LP-bound used in the branch-and-bound tree for CPLEX grows slowly.

Similar to the 1-HVRP, we attempted to solve this two stage procedure in a single stage by developing a weighted combination of distance and demand. The optimal results for this Weighted Objective approach are shown in Table 5.7 for 21 nodes with 2 vehicles of capacity 50 units each and a route-length limit of 18 units each. Similar to the 1-HVRP, we observe that one needs to design the weights aptly to obtain Pareto-optimal solutions that maximize demand and yet travel efficiently in terms of distance.

Next, we apply Two-Stage optimization procedure using Global Timing Rule. The optimal results on 21 nodes are shown in Table 5.8. The network is assumed to

Table 5.6: Optimal Results for m-HVRP using Two-stage Optimization and Local Timing Rule on 30 nodes network

d (Stage)	Total dis- tance	Distance by Vehicle (#1, #2)	CAP used by vehicle (#1, #2)	Demand met by Priority for Veh 1	Demand met by Priority for Veh 2	Time
0 (I)	68.44	34.96, 33.48	121, 77	27, 66, 28	36, 10, 31	70.75m
0 (II)	57.71	27.12, 30.59	121, 77	21, 66, 21	42, 10, 38	0.85m
1 (I)	69.34	34.43, 34.91	121, 77	36, 54, 2	27, 22, 57	16.00m
1 (II) ^a	48.40	30.41, 17.99	133, 65	43, 43, 47	20, 33, 12	6.98d
2 (I)	65.69	33.31, 32.28	121, 77	40, 59, 26	23, 17, 33	0.92m
2 (II) ^b	42.53	33.34, 9.19	148, 50	54, 56, 38	9, 20, 21	11.41h

d:days, h:hours, m:min, s:sec, ^a4.97% gap, ^b11.31% gap

Table 5.7: Optimal Results for m-HVRP using Weighted Objective Approach and Local Timing Rule on 21 nodes network

$\alpha(\beta)$	Total Distance	Distance by vehicle (#1, #2)	Total Demand Met	Demand met by vehicle (#1, #2)	Demand met by Priority (#1, #2, #3)
1.0 (0.0)	34.62	16.66, 17.96	100	50, 50	44, 49, 7
0.9 (0.1)	24.72	13.94, 10.78	100	50, 50	44, 49, 7
...
0.3 (0.7)	24.72	13.94, 10.78	100	50, 50	44, 49, 7
0.2 (0.8)	19.47	16.66, 17.96	84	41, 43	44, 40, 0
0.1 (0.9)	3.92	1.40, 2.52	12	0, 12	12, 0, 0

be served by 2 vehicles with capacity of 75 units each and route-length limit of 18 units. Similar to the Local Timing Rule, all of the demand is satisfied for both these models, hence both models are equivalent in terms of demand satisfied. Recall from Figure 5.3 that this is not necessarily true in all situations. The two vehicles together travel a total distance of 30.87 units for $d=0$, and 25.55 for $d=1$ with Global Timing Rule. Compared to Local Timing Rule model, the total distance is 4.6% and 2.0% higher and this might be reasonable, depending on how these models perform with respect to other performance measures. The load distribution for the two vehicles

between the Local and Global model is very similar.

Finally, let us compare the Local and Global Timing Rule models on the distribution of service times. Table 5.9 gives us the comparison of service times for Local Timing Rule and Global Timing rule models on this network. Figure 5.4 illustrates the Two-stage optimization using Local Timing Rule for $d=0$ and $d=1$ and Figure 5.5 illustrates the Two-stage optimization using Global Timing Rule for $d=0$ and $d=1$. Observe that the Stage I route for both the rules is not optimized in terms of distance, but notice how the routes are efficient distance wise, after Stage II optimization. In case of the Local Timing Rule model with $d=0$, notice that vehicle 1 services *all* the Priority 1 (red) nodes and a few Priority 2, 3 nodes. On the other hand, observe that vehicle 2 starts servicing with Priority 2 nodes and finishes with Priority 3 nodes. This means that with the Local Timing Rule, it is possible that a higher priority node will have to wait since the other vehicle is attending to a lower priority node. If the nodes are badly affected by a disaster, this can prove fatal and hence may not be an acceptable solution as resources can, alternatively, be pooled to service higher priority nodes. From Table 5.9, with Local Timing Rule for $d=0$, the earliest and latest service times for Priority 1 nodes using vehicle 2 are 1.26 to 10.49 (vehicle 1 does not service Priority 1 nodes). The earliest and latest service times for Priority 2 nodes using vehicle 1 are 1.10 to 7.90 and using vehicle 2, they are 11.27. Comparing the average service times, we see that it is 5.73 for Priority 1, where as for Priority 2 is 4.70 and 11.27. That is, some of the Priority 2 nodes are serviced earlier than Priority 1 nodes! On the other hand, with Global Timing Rule, the service times reflect the urgency levels of the nodes, that is, all vehicles cater to

Priority 1 nodes first and so on. In the Local Timing Rule model with $d=1$ we see that some of the Priority 3 nodes get serviced earlier compared to Priority 1 nodes! On the other hand, with Global Timing Rule model with $d=1$, the latest service times for Priority 1 nodes are (6.96, 5.66) which is lower than the earliest service times for Priority 3 nodes. Thus, we observe that the Global Timing Rule model gives us an ordered distribution of service times based on urgency levels compared to Local Timing Rule model. But it should also be emphasized that Local Timing Rule model has its benefits as discussed in Section 5.4.2 when the vehicle capacity or route-length is limited.

Table 5.8: Optimal Results for m-HVRP using Global Timing Rule on 21 nodes network

d (Stage)	Total dis- tance	Distance by Vehicle (#1, #2)	CAP used by vehicle (#1, #2)	Demand met by Priority for Veh 1	Demand met by Priority for Veh 2	Time
Two vehicles with $CAP = 75$, $L^{max} = 18$						
0 (I)	34.80	16.92, 17.88	57, 75	17, 29, 11	27, 20, 28	13.67s
0 (II)	30.87	14.13, 16.74	70, 62	23, 26, 21	21, 23, 18	70.50s
1 (I)	34.70	17.76, 16.94	75, 57	24, 19, 32	20, 30, 7	22.08m
1 (II) ^a	25.55	15.56, 9.99	66, 66	23, 25, 18	21, 24, 21	18.46h

h:hours, s:sec, m:min, ^a4.14% gap

5.6. Conclusions

In this research we discussed the challenges that are faced in the last mile distribution of the aid to disaster victims. We developed two routing models 1-HVRP and m-HVRP. For both these models, the vehicle(s) are constrained by capacity and route-length restrictions. Apart from the d —Relaxed Priority Rule, we enforce the

Table 5.9: Service Times for m-HVRP using Local & Global Timing Rules (after Stage II)

d	Service Times for Veh 1 by Priority	Service Times for Veh 2 by Priority
Local Timing Rule		
0	Min: N/A, 1.10, 9.12	Min: 1.26, 11.27, 11.96
	Max: N/A, 7.90, 11.07	Max: 10.49, 11.27, 14.69
	Avg: N/A, 4.70, 9.94	Avg: 5.73, 11.27, 13.29
1	Min: 4.53, 3.66, 9.48	Min: 1.26, 3.99, 6.04
	Max: 8.56, 14.75, 12.21	Max: 4.86, 5.32, 8.49
	Avg: 6.43, 9.80, 10.88	Avg: 3.08, 4.56, 7.02
Global Timing Rule		
0	Min: 1.26, 7.33, 11.48	Min: 3.44, 6.57, 11.17
	Max: 6.46, 10.26, 13.43	Max: 5.79, 10.65, 13.90
	Avg: 3.93, 8.96, 12.30	Avg: 4.62, 8.27, 12.57
1 ^a	Min: 2.93, 1.29, 9.99	Min: 2.06, 1.10, 7.34
	Max: 6.96, 12.03, 12.72	Max: 5.66, 6.12, 9.29
	Avg: 5.06, 7.66, 11.39	Avg: 3.88, 3.86, 8.16

^aStage II optimality gap is 4.14%

Order of Demand Fulfillment rule for these routing models. The routing models are developed to cater to four performance metrics: Distance, Demand Satisfied, Response Time and Number of Customers Served. We develop MIP formulations for the capacitated vehicle routing problem with single (1-HVRP) and multiple vehicles (m-HVRP). Inherent conflicts in the performance measures in the case of single and multiple vehicles, are addressed by multi-stage optimization or adding appropriate constraints. Multiple vehicles adds on to the complexity of the problem in terms of whether all the vehicles need to service all the nodes simultaneously or sequentially. The routing models are successfully run on small test problems with 21, and 30 nodes network to optimality (near-optimal) using optimization software. Then, we discuss the performance of the routing models on four different performance metrics – distance traversed, demand satisfied, response time and number of customers sat-

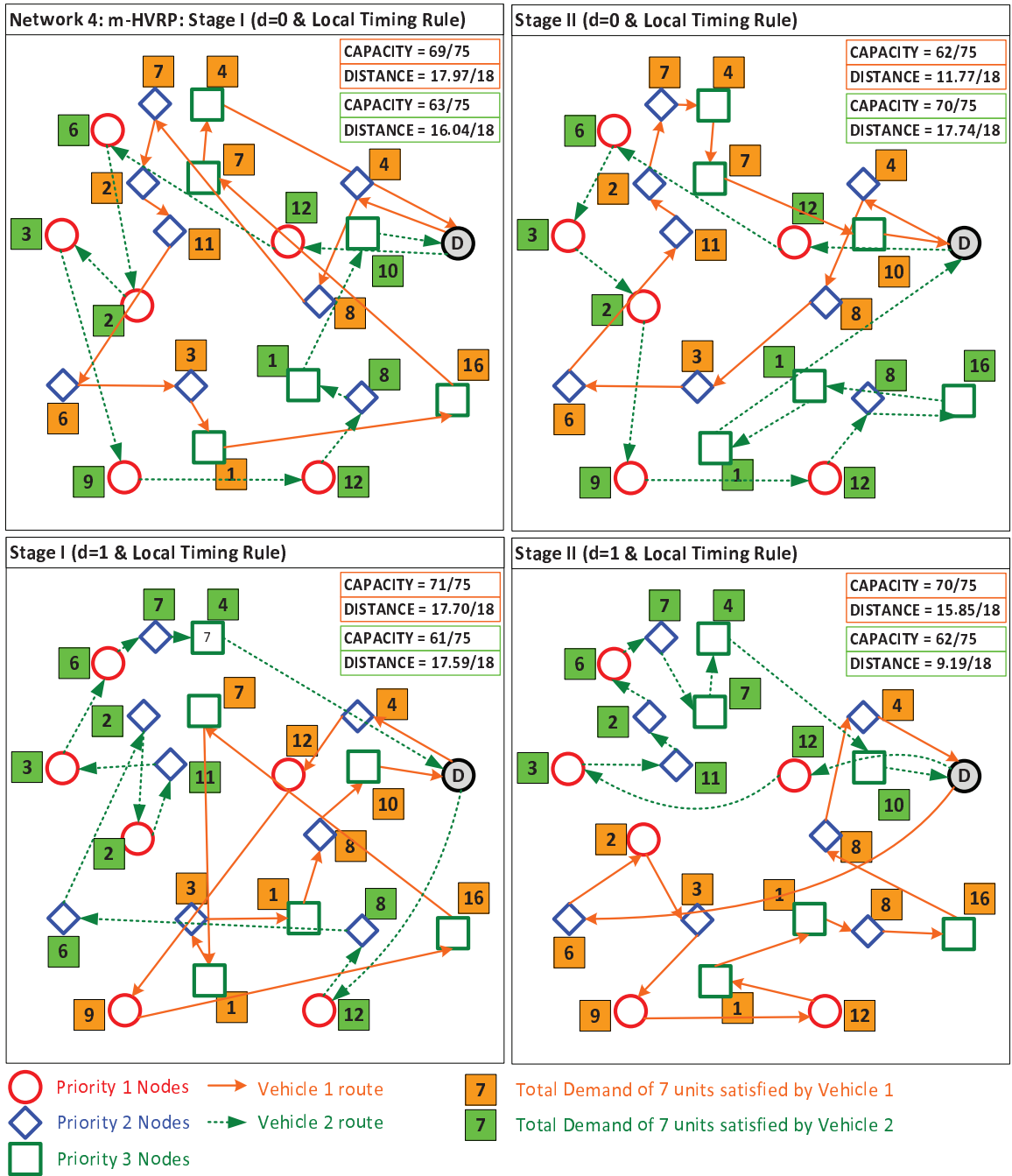


Figure 5.4: Optimal results of m-HVRP on 21 nodes using the Local Timing Rule

ified. Finally, we show that traditional routing models fail as they do not capture the multiple performance measures relevant for humanitarian relief operations.

Delivery of aid concerns not only just relief agencies but businesses as well

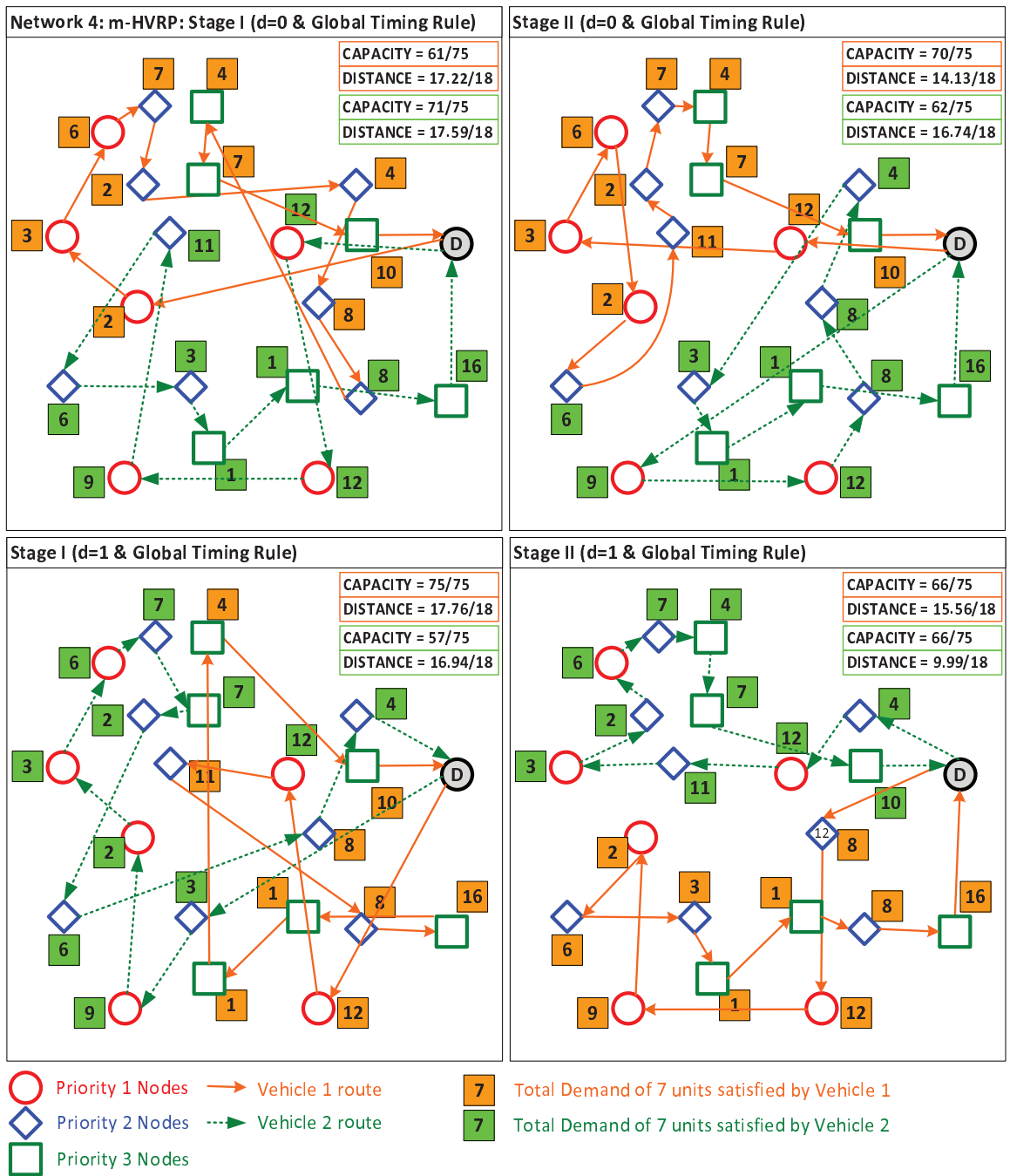


Figure 5.5: Optimal results of m-HVRP on 21 nodes using the Global Timing Rule since their global supply chain is affected. During hurricane Katrina, businesses like Waffle House, Lowes, Home Depot, and Wal-Mart did an outstanding job of restoring their stores to help serve the needs of local community. This helped them

earn tremendous social value, and in the process helped their businesses to thrive as their stores were up and ready to serve the community. In a recent panel discussion (Humanitarian Logistics Conference, Feb 2009), director of World Food Programme – an organization known for its logistics and distribution, mentioned that cost and efficiency of humanitarian relief operations as important as the humanitarian aspect of the operations to the agency and as well as to the donors. With involvement of numerous parties in the relief operations (e.g., businesses, international and local relief agencies), it is of vital importance that distribution of relief aid is efficient. Our research designs efficient and humanitarian delivery routes for distribution of relief goods which provides a better understanding of relief operations for businesses and relief agencies.

There are many possible extensions to this work. Most of the relief operations cannot be completed in one time-period and hence one can extend this work to multi-period vehicle routing problem. In this work, we ignored the computational complexity of these routing models; hence future research can be geared towards to solving these problems in real-time, efficiently to optimality. Our work is of tremendous research value as it contributes to the vehicle routing problem (VRP) literature and has defined a new VRP variant.

5.7. Summary of Insights

In this chapter, we extended the uncapacitated VRP model for humanitarian relief situations (that was discussed in Chapter 4) to consider single and multiple

vehicle(s) with capacity and route-length restrictions. The key insights from this chapter are:

- Developed two routing models 1-HVRP and m-HVRP corresponding to routing with single vehicle and multiple vehicles, respectively, with capacity and route-length restrictions. Also, introduced new constraint - Order of Demand Fulfillment that is relevant for dealing with nodes that have demand and priorities. This says that all high priority demand must be satisfied if a vehicle has enough available capacity.
- Discussed the additional challenges in the distribution of aid that are posed when one has single and multiple vehicles with capacity and route-length restrictions. Primary objective is to satisfy as much demand as possible. But, can we achieve this using the least distance? Does maximum demand mean, we can satisfy as many victims as possible? With multiple vehicles, should relief organization do sequential or simultaneous delivery for a priority class? For example, one can imagine two routes which satisfy as much demand as possible but these routes may differ on the response time for the nodes. Based on these research questions, we develop models that capture the four performance metrics: (i) Distance, (ii) Response Time (Earliest, Latest, Average), (iii) Fill Rate (or % Demand Satisfied), (iv) Number of Customers Serviced.
- Discussed two optimization models for solving 1-HVRP: (i) Two-stage optimization and (ii) Weighted objective function.
- The two-stage optimization model consists of two stages. In first stage, the

routes that deliver maximum demand are identified. This stage is not trivial to solve as the vehicle has route-length restriction. In second stage, such a route is cleaned up for efficiency in terms of distance and number of customers serviced. First stage caters to the humanitarian aspect and the second stage caters to the efficiency of the operations.

- MIP formulations are developed for the two-stage optimization model. First stage problem is formulated as an Orienteering Problem, taking into account the node priorities. Prize is the demand at each node. Second stage problem is modeled as (i) u-HVRP for *the same* subset of nodes that are visited in first stage or (ii) u-HVRP for *any* subset of nodes that meet the maximum demand objective in the first stage.
- In the other approach - weighted objective function is formed by taking a convex combination of total demand satisfied and total distance traversed. Depending on the weights of the objectives, one objective dominates the other. If distance is highly weighted, then the vehicle starts delivering at lower levels, which is unacceptable for humanitarian relief operations. Thus, this approach is discarded for 1-HVRP.
- Computationally, 1-HVRP formulations are solved to optimality for 21 node and 30 node networks. Results indicate that the routes that deliver maximum demand do not necessarily travel least distance. Weighted objective function approach is not guaranteed to obtain such desired solutions as the objectives do not have inverse relationship.

- With m-HVRP, one important issue that arises is the sequential or simultaneous service of a priority class. These two models arise as a result of how the d -Relaxed Priority Rule is applied – for individual vehicle separately or for all vehicles. In sequential servicing (we call it local timing model), a high priority node can be serviced at a later time than a low priority node if the two nodes are visited by two different vehicles. In simultaneous service (we call it global timing model), a high priority node is always serviced earlier than a lower priority node.
- Similar to 1-HVRP, we have two-stage optimization and weighted objective approach. We discard the weighted objective approach for the same issues as in 1-HVRP. MIP formulations for two-stage optimization are developed. First stage is modeled as a Team-Orienting Problem with d -Relaxed Priority Rule enforced locally or globally. In the second stage, the routes from first stage are cleaned up for efficiency for distance and number of customers satisfied. Second stage problem is modeled as (i) multiple vehicle extension of u-HVRP for *the same* subset of nodes that are visited in first stage or (ii) multiple vehicle extension of u-HVRP for *any* subset of nodes that meet the maximum demand objective in the first stage.
- Computationally, m-HVRP formulations are successfully run on test problems with 21 nodes network to optimality using CPLEX optimization software. Local timing rule results in some of the high priority nodes serviced later than a low priority node, where as with global timing rule, high priority nodes have

lower service times compared to low priority nodes.

- Our work is of practical value as we developed comprehensive routing models for single and multiple vehicles that consider different performance metrics. Our models not only meet the primary goal of satisfying as much demand as possible but also cater to other performance metrics like distance, response time and number of customers satisfied.
- Developed a new orienteering or team-orienteering model that not only has prize, but also places importance on the order of visit to collect these prizes. Thus, this work is of value to the vehicle routing problem (VRP) literature as we have defined a new class of VRPs.

Chapter 6

Conclusions and Future Work

In this dissertation, we introduced and studied research problems in Retail Operations and Humanitarian Logistics. We summarize our work and discuss future research directions below.

In Chapters 2 and 3, we focused on Retail Operations. Matching supply with demand at the right time is a big challenge for retailers, especially for fashion apparel goods due to long procurement and production lead times and relatively short sales horizon. Retailers have to order their goods long before a sales season and face market uncertainty at a macro level and demand uncertainty at a micro level. In Retail Operations, the inventory that ends up as unsaleable at primary markets can be significant (up to 20% of the retail product). Retailers of short life cycle products look for strategies like selling in secondary markets at a discounted price to increase revenues. Apart from this strategy, retailers also enter into contractual agreements with suppliers to share the cost of excess (unsaleable) inventory. The effect of such contracts on the retailer's order quantities and the retailer's channel strategies is not trivial.

In Chapter 2, we focus on decisions that are prevalent primarily for high-end fashion retail: Many retailers make use of their own stores (brands) as well as outlets. It is typical in practice to sell the same goods in the stores and in the outlets, but the

retailers choose to sell the goods first in their primary stores and then at the outlets with a delay. We build a model to answer the following questions: Should the retailer selling a seasonal product offer it only in the primary market? If not, then what is the best time to start offering the product in the secondary market? Our model is novel that it captures the temporal (sequential) aspect of the timing decision and also analyzes the optimal order quantity for the retailer that faces both market and demand uncertainty. Critical to our model and its structural properties is a demand model where the demand in each market is proportional to the length of time the product is offered in that market. By establishing analytical properties of the retailer's expected revenue and profits and conducting computational experiments, we find that higher optimal order quantities in general result shortens the sales horizon in the primary market while extending it in the secondary market. Steeper discounts in the secondary market do not necessarily result in higher order quantities as the revenue potential in both markets plays an important role in the retailer's total expected revenues. Most importantly, it may or may not be optimal for the retailer to use a sequential, dual channel approach: The retailer can maximize his expected profits by selling only in the primary market or only in the secondary market; this depends on problem parameters, including demand and prices.

In Chapter 3, we develop a model to study the effect of contract types on a retailer's use of primary and secondary markets to sell his product. We introduce the MM contract, which is commonly used in the industry, but has seldom been studied in the research literature on supply chains. In this chapter, we not only study retailer's ordering and timing decisions but also the supplier's choice of wholesale

price given contractual agreements between the retailer and the supplier. We show that for a given order quantity, the time to transfer the goods are sold for longest time in the primary market when BB contract is in place; the second longest sales horizon in the primary market is observed under WP. MM, on the other hand, increases the length of the sales horizon for the secondary market. Considering the optimal order quantities for the retailer, the retailer places the highest amount of orders under the MM contract. While this is advantageous for the supplier, increasing his own sales, it also poses challenges because the potential for paybacks from the suppliers to the retailer are also higher when the order quantity is high. In general, retailer benefits more from the MM contract but both parties can benefit from MM.

Our work contributes to the literature on Retail Operations, especially with sequential management of channels. There are many extensions possible to the problem discussed in this dissertation. One can look at menu of contracts that can result in equitable distribution of inventory-related costs. We have considered costless transfer and in future work, this assumption can be relaxed. With the advent of technology, customers are strategic in that they wait for sales to make the purchase. For example, a customer walking into store with iPhone has access to review of the product and feedback from the customers of the product without incurring additional search costs. Another potential extension is when the retailer is carrying multiple products, across multiple channels. The issue of substitution needs to be considered to accurately model the demand. Do we time the product changes at the same time? Channel conflict issues need to be examined carefully. Yet, another extension is managing sales channels sequentially when the sales channels

are owned by different parties.

The next two chapters – Chapter 4 and 5, we discuss one particular aspect of humanitarian relief chain: The distribution of relief aid for non-profit organizations, commonly known as “Last Mile Distribution”. This is a complex problem as the goal of the relief operation is humanitarian yet at the same time, due to limited resources, the operations have to be managed efficiently.

In Chapter 4, we consider the vehicle routing problem without capacity and route-length restrictions for humanitarian relief operations. We discuss vehicle routing problem for such relief operations whose primary goal is to deliver as much aid as possible but that also is efficient on other performance metrics like response time. In this paper we discuss routing for relief operations using one vehicle without capacity restrictions. Contrary to the existing vehicle routing models, the key property of our routing models is that the nodes have priorities along with humanitarian needs. Nodes with higher priority represent places that urgently need service; hence they need to be serviced before a lower priority node. Mixed integer programs (MIP) are used to formulate routing models with strict and relaxed forms of priority restrictions for a single vehicle with no capacity restrictions. We derive bounds for this problem and show that this bound is attained in limiting condition for a worst-case example. Finally, we test the formulations on a set of small problems (up to 40 nodes) and evaluate the optimal solutions with respect to the performance metrics like response time and distance and show that u-HVRP, our vehicle routing model with priorities and d-Relaxed Priority rule capture the trade-off between distance and response time effectively compared to VRP models without priorities.

In Chapter 5, we discuss two vehicle routing models for distribution of a single relief good (e.g., water, blankets, medicines) to disaster victims using single and multiple vehicles consisting of homogeneous fleet with capacity and route length restrictions. We call these problems as 1-HVRP and m-HVRP respectively. The goal of a humanitarian relief operation is to deliver as much aid as possible but at the same time, due to limited resources, the operations have to be efficient on other metrics as well. This imposes new challenges for the routing models and in this work, we consider metrics like fill rate (demand satisfied), distance traversed, response time and number of victims satisfied to capture performance metrics relevant for humanitarian relief operations. The key aspect of our routing models is that the nodes have priorities and all the routes have to satisfy two important constraints: d-Relaxed Priority Rule and Order of Demand Fulfillment. Nodes with higher priority represent places that urgently need service; hence they need to be serviced first compared to a lower priority node. Different routing models are formulated as Mixed Integer Programs and are solved to optimality for small test problems (up to 30 nodes). Finally, we show that our routing models show the trade-off between these four metrics as the former constraint caters to operational efficiency and the later constraint caters to the humanitarian nature of this routing problem.

Our work is of tremendous research value as it contributes to the vehicle routing problem (VRP) literature and has defined a new VRP variant. We have studied the VRP for humanitarian relief situations and shown that traditional models fail to capture the relevant performance metrics. There are many possible extensions to this work. In this work we assumed demand is deterministic; if the demand is

random, then we have Stochastic VRPs. Most of the relief operations cannot be completed in one time-period and hence one can extend this work to multi-period vehicle routing problem. In this work, we ignored the computational complexity of these routing models; hence future research can be geared towards solving these problems in real-time, efficiently to optimality.

Appendix A

Appendix for Essay 1

A.1. Proof for Componentwise Concavity of the Sample Path Revenue

We will show that the expected revenue function is concave in sample-path sense. For a sample-path, $X_1 = k_1$ and $X_2 = k_2$, we see that $R(\beta|Q)$ can consist of possibly three slopes as derived in Equation (2.12). For this sample path, if the first order derivative with respect to β decreases in β then our proof is complete.

First, we identify the feasible range of β for which $h_i(\beta, Q)$ for $i = 1, 2, 3$ are defined, in Table A.1. Analyzing different cases based on ‘Condition Set 1’ and ‘Condition Set 2’ we arrive at feasible ranges of β for $h_i(\beta, Q)$ functions. Each function has a different partial derivative, hence a distinct ‘slope,’ with respect to β . Using this table and based on the sign of three terms, $Q - k_1$, $Q - k_2$, and $k_1 - k_2$, we list the possible sequence of slopes for $\beta \in [0, 1]$ in Table A.2. The last column in Table A.2 refers to the cases that are listed in Table A.1.

In summary, the retailer’s timing problem can be simplified to understanding these eight possibilities. We illustrate each feasible case in Figure A.1: In Figure A.1(a), $h_2(\beta, Q)$ is followed by $h_3(\beta, Q)$ and it is never possible for $p_1 k_1 - p_2 k_2 > (p_1 - p_2)k_1$; hence $R(\beta|Q)$ is concave in β . In Figure A.1(b), only $h_3(\beta, Q)$ exists; hence $R(\beta, Q)$ is concave in β . In Figure A.1(c), there are two possibilities: either $h_2(\beta, Q)$ is followed by $h_1(\beta, Q)$ or $h_3(\beta, Q)$ precedes $h_2(\beta, Q)$ which precedes $h_1(\beta, Q)$. Again, for this parameter combination, it is never possible that $p_1 k_1 - p_2 k_2 < (p_1 - p_2)k_1$. Therefore $R(\beta, Q)$ is componentwise concave in β for a given Q .

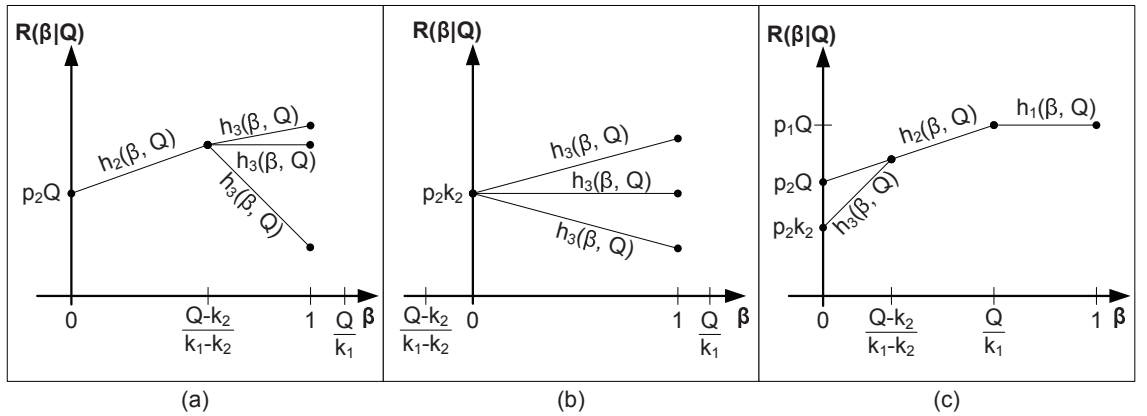


Figure A.1: Sample path revenue as a function of β

We continue by referring to the conditions in Table A.2. In No.1 and No.2, the slopes are decreasing in β since $p_1 \geq p_2$ and thus concave. Consider No.3. Notice that the slopes are decreasing since $p_1 k_1 - p_2 k_2 \geq (p_1 - p_2)k_1 \geq 0$. Otherwise we

Table A.1: Feasible region for first order derivatives (in sample-path) for the Linear Demand Model

Case	Condition Set 1	Condition Set 2	Feasible region β
$h_1(\beta, Q) = p_1 Q$			
1.1	$Q < \beta k_1 \leq k_1$		$\frac{Q}{k_1} < \beta < 1$
$h_2(\beta, Q) = (p_1 - p_2)\beta k_1 + p_2 Q$			
2.1		$Q \leq k_2, k_1 < k_2$	$0 \leq \beta \leq \frac{Q}{k_1} \leq 1 \leq \frac{Q-k_2}{k_1-k_2}$
2.2	$\beta k_1 \leq Q \leq k_1,$	$Q \leq k_2, k_1 > k_2$	$\frac{Q-k_2}{k_1-k_2} \leq 0 \leq \beta \leq \frac{Q}{k_1} \leq 1$
2.3	$Q-k_2 < \beta(k_1-k_2)$	$Q \geq k_2, k_1 > k_2$	$0 \leq \frac{Q-k_2}{k_1-k_2} \leq \beta \leq \frac{Q}{k_1} \leq 1$
2.4		$Q \geq k_2, k_1 < k_2$	Infeasible as $\beta < \frac{Q-k_2}{k_1-k_2} \leq 0$
3.1		$Q \leq k_2, k_1 < k_2$	$0 \leq \beta \leq \frac{Q-k_2}{k_1-k_2} \leq 1 \leq \frac{Q}{k_1}$
3.2	$\beta k_1 \leq k_1 \leq Q,$	$Q \leq k_2, k_1 > k_2$	Infeasible as $Q \geq k_1$ & $k_1 > k_2 \geq Q$
3.3	$Q-k_2 < \beta(k_1-k_2)$	$Q \geq k_2, k_1 > k_2$	Infeasible as $\beta > \frac{Q-k_2}{k_1-k_2} \geq 1$
3.4		$Q \geq k_2, k_1 < k_2$	Infeasible as $\beta < \frac{Q-k_2}{k_1-k_2} \leq 0$
$h_3(\beta, Q) = p_1 \beta k_1 + p_2(1 - \beta)k_2$			
4.1		$Q \leq k_2, k_1 < k_2$	Infeasible as $\beta \geq \frac{Q-k_2}{k_1-k_2} \geq 1$
4.2	$\beta k_1 \leq Q \leq k_1,$	$Q \leq k_2, k_1 > k_2$	Infeasible as $\beta \leq \frac{Q-k_2}{k_1-k_2} \leq 0$
4.3	$Q-k_2 \geq \beta(k_1-k_2)$	$Q \geq k_2, k_1 > k_2$	$0 \leq \beta \leq \frac{Q-k_2}{k_1-k_2} \leq \frac{Q}{k_1} \leq 1$
4.4		$Q \geq k_2, k_1 < k_2$	Infeasible as $Q \leq k_1$ & $Q \geq k_2 > k_1$
5.1		$Q \leq k_2, k_1 < k_2$	$0 \leq \frac{Q-k_2}{k_1-k_2} \leq \beta \leq 1 \leq \frac{Q}{k_1}$
5.2	$\beta k_1 \leq k_1 \leq Q,$	$Q \leq k_2, k_1 > k_2$	Infeasible as $\beta \leq \frac{Q-k_2}{k_1-k_2} \leq 0$
5.3	$Q-k_2 \geq \beta(k_1-k_2)$	$Q \geq k_2, k_1 > k_2$	$0 \leq \beta \leq 1 \leq \frac{Q}{k_1} \leq \frac{Q-k_2}{k_1-k_2}$
5.4		$Q \geq k_2, k_1 < k_2$	$\frac{Q-k_2}{k_1-k_2} \leq 0 \leq \beta \leq 1 \leq \frac{Q}{k_1}$

will violate the assumptions of this sample-path. No.4 is not feasible. For No.5, the slopes are decreasing as, $(p_1 - p_2)k_1 \geq p_1 k_1 - p_2 k_2 \geq 0$ to satisfy the conditions for this sample-path. No.6 is again not feasible. In No.7 and No.8, there is only one slope and thus it is linear, which implies it is concave. Thus, we see that the expected revenue function is concave in sample-path sense. In the above table, we implicitly assumed that $k_1 > 0, k_2 > 0, k_1 \neq k_2$. For technical completeness, let us look at the special situations when these restrictions are violated. If $k_1 = 0, k_2 = 0$, then it is a trivial case with constant revenue of 0 for all β . If $k_1 = 0, k_2 > 0$ then it is equivalent to selling only in the secondary market. Then, the slopes are: 0 for $\beta \leq 1 - \frac{Q}{k_2}$ and $-p_2 k_2$ for $\beta > 1 - \frac{Q}{k_2}$. Slopes are decreasing and thus it is concave. If $k_1 = k_2 > 0$, then we have the following slopes: $(p_1 - p_2)k_1$ for $\beta \leq \frac{Q}{k_1}$ and 0 for $\beta > \frac{Q}{k_1}$. We see that slopes are decreasing and thus the sample path revenue function is concave.

Table A.2: First order derivative wrt β (in sample-path) for the Linear Demand Model

No.	Conditions	Sequence of Slopes for β	Feasible β
1	$Q \leq k_1, Q \leq k_2, k_1 < k_2$	$(p_1 - p_2)k_1, 0$	Cases 2.1, 1.1
2	$Q \leq k_1, Q \leq k_2, k_1 > k_2$	$(p_1 - p_2)k_1, 0$	Cases 2.2, 1.1
3	$Q \leq k_1, Q \geq k_2, k_1 > k_2$	$p_1k_1 - p_2k_2, (p_1 - p_2)k_1, 0$	Cases 4.3, 2.3, 1.1
4	$Q \leq k_1, Q \geq k_2, k_1 < k_2$	–	Not feasible
5	$Q \geq k_1, Q \leq k_2, k_1 < k_2$	$(p_1 - p_2)k_1, p_1k_1 - p_2k_2, 0$	Cases 3.1, 5.1
6	$Q \geq k_1, Q \leq k_2, k_1 > k_2$	–	Not feasible
7	$Q \geq k_1, Q \geq k_2, k_1 > k_2$	$p_1k_1 - p_2k_2$	Case 5.3
8	$Q \geq k_1, Q \geq k_2, k_1 < k_2$	$p_1k_1 - p_2k_2$	Case 5.4

A.2. Proof of Joint Concavity of the Sample Path Revenue Function

We will prove the joint concavity of β, Q using the definition of concavity. Consider two points (Q_1, β_1) and (Q_2, β_2) and $0 \leq \alpha \leq 1$. Define $Q_0 = \alpha Q_1 + (1 - \alpha)Q_2$ and $\beta_0 = \alpha\beta_1 + (1 - \alpha)\beta_2$. We will show that the expected revenue function is concave in β, Q in a sample-path sense. That is, for any sample path, $X_1 = k_1$, and $X_2 = k_2$, we will show that $R(\beta_0|Q_0) \geq \alpha R(\beta_1|Q_1) + (1 - \alpha)R(\beta_2|Q_2)$. There are many cases depending on the relation between the elements in each of the sets: $\{Q_1 - \beta_1k_1, (1 - \beta_1)k_2\}$, $\{Q_2 - \beta_2k_1, (1 - \beta_2)k_2\}$, $\{Q_0 - \beta_0k_1, (1 - \beta_0)k_2\}$, $\{Q_1, Q_2\}$ and $\{\beta_1, \beta_2\}$. To simplify the notation, let us denote $I_1 = Q_1 - \beta_1k_1, D_1 = (1 - \beta_1)k_2$, $I_2 = Q_2 - \beta_2k_1, D_2 = (1 - \beta_2)k_2$, and $I_0 = Q_0 - \beta_0k_1, D_0 = (1 - \beta_0)k_2$. Table A.3 summarizes the possible conditions and lists where the analysis is provided. Please refer to Lemma 1 through 11 below to complete the proof.

Let us define the following for this section: $LHS \equiv R(\beta_0|Q_0)$ and $RHS \equiv \alpha R(\beta_1|Q_1) + (1 - \alpha)R(\beta_2|Q_2)$. To prove concavity, we need to show $LHS \geq RHS$ for all the cases listed in Table A.3.

Lemma 1 *If $I_1 \leq 0 \leq D_1$ or $0 \leq I_1 \leq D_1$ or $0 \leq D_1 \leq I_1, I_2 \leq 0 \leq D_2, I_0 \leq 0 \leq D_0$ then $LHS = RHS$*

Proof If $I_1 \leq 0$ then $LHS = p_1Q_0, RHS = p_1(\alpha Q_1 + (1 - \alpha)Q_2) = p_1Q_0 = LHS$. If $0 \leq I_1 \leq D_1$ then $LHS = p_1Q_0$ and $RHS = \alpha(p_1\beta_1k_1 + p_2(Q_1 - \beta_1k_1)) + (1 - \alpha)(p_1Q_2)$. Since $p_2 \leq p_1$, we have $RHS \leq p_1(\alpha\beta_1k_1 + \alpha Q_1 - \alpha\beta_1k_1) + (1 - \alpha)(p_1Q_2) = p_1Q_0 = LHS$. If $0 \leq D_1 \leq I_1$ then $LHS = p_1Q_0$ and $RHS = \alpha(p_1\beta_1k_1 + p_2(1 - \beta_1)k_2) + (1 - \alpha)(p_1Q_2)$. As $p_2 \leq p_1$, $RHS \leq \alpha(p_1\beta_1k_1 + p_1Q_1 - p_1\beta_1k_1) + (1 - \alpha)(p_1Q_2) = p_1Q_0 = LHS$. •

Lemma 2 *If $I_1 \leq 0 \leq D_1, 0 \leq I_2 \leq D_2$ or $0 \leq D_2 \leq I_2, I_0 \leq 0 \leq D_0$ then $LHS \geq RHS$*

Proof If $0 \leq I_2 \leq D_2$, then $LHS = p_1Q_0, RHS = \alpha(p_1Q_1) + (1 - \alpha)(p_1\beta_2k_1 + p_2(Q_2 - \beta_2k_1))$. Since $p_2 \leq p_1$, then $RHS \leq \alpha(p_1Q_1) + (1 - \alpha)(p_1\beta_2k_1 + p_1(Q_2 - \beta_2k_1)) = p_1Q_0 = LHS$.

Table A.3: Proof of joint concavity in sample-path for the Linear Demand Model

No.	Conditions	Proof
$Q_2 \leq Q_1, \beta_2 \leq \beta_1$		
1	$I_1 \leq 0 \leq D_1, I_2 \leq 0 \leq D_2, I_0 \leq 0 \leq D_0$	Lemma 1
2	$I_1 \leq 0 \leq D_1, 0 \leq I_2 \leq D_2, I_0 \leq 0 \leq D_0$	Lemma 2
3	$I_1 \leq 0 \leq D_1, 0 \leq D_2 \leq I_2, I_0 \leq 0 \leq D_0$	Lemma 2
4	$I_1 \leq 0 \leq D_1, 0 \leq I_2 \leq D_2, 0 \leq I_0 \leq D_0$	Lemma 3
5	$I_1 \leq 0 \leq D_1, 0 \leq I_2 \leq D_2, 0 \leq D_0 \leq I_0$	Lemma 4
6	$I_1 \leq 0 \leq D_1, 0 \leq D_2 \leq I_2, 0 \leq I_0 \leq D_0$	Lemma 3
7	$I_1 \leq 0 \leq D_1, 0 \leq D_2 \leq I_2, 0 \leq D_0 \leq I_0$	Lemma 4
8	$0 \leq I_1 \leq D_1, I_2 \leq 0 \leq D_2, I_0 \leq 0 \leq D_0$	Lemma 1
9	$0 \leq I_1 \leq D_1, I_2 \leq 0 \leq D_2, 0 \leq I_0 \leq D_0$	Lemma 5
10	$0 \leq I_1 \leq D_1, I_2 \leq 0 \leq D_2, 0 \leq D_0 \leq I_0$	Lemma 6
11	$0 \leq I_1 \leq D_1, 0 \leq I_2 \leq D_2, 0 \leq I_0 \leq D_0$	Lemma 7
12	$0 \leq I_1 \leq D_1, 0 \leq I_2 \leq D_2, 0 \leq D_0 \leq I_0$	Lemma 8
13	$0 \leq I_1 \leq D_1, 0 \leq D_2 \leq I_2, 0 \leq I_0 \leq D_0$	Lemma 9
14	$0 \leq I_1 \leq D_1, 0 \leq D_2 \leq I_2, 0 \leq D_0 \leq I_0$	Lemma 8
15	$0 \leq D_1 \leq I_1, I_2 \leq 0 \leq D_2, I_0 \leq 0 \leq D_0$	Lemma 1
16	$0 \leq D_1 \leq I_1, I_2 \leq 0 \leq D_2, 0 \leq I_0 \leq D_0$	Lemma 5
17	$0 \leq D_1 \leq I_1, I_2 \leq 0 \leq D_2, 0 \leq D_0 \leq I_0$	Lemma 10
18	$0 \leq D_1 \leq I_1, 0 \leq I_2 \leq D_2, 0 \leq I_0 \leq D_0$	Lemma 9
19	$0 \leq D_1 \leq I_1, 0 \leq I_2 \leq D_2, 0 \leq D_0 \leq I_0$	Lemma 8
20	$0 \leq D_1 \leq I_1, 0 \leq D_2 \leq I_2, 0 \leq I_0 \leq D_0$	Lemma 9
21	$0 \leq D_1 \leq I_1, 0 \leq D_2 \leq I_2, 0 \leq D_0 \leq I_0$	Lemma 11
$Q_2 \leq Q_1, \beta_2 \geq \beta_1$		
22	$I_1 \leq 0 \leq D_1, I_2 \leq 0 \leq D_2, I_0 \leq 0 \leq D_0$	See No.1
23	$0 \leq I_1 \leq D_1, I_2 \leq 0 \leq D_2, I_0 \leq 0 \leq D_0$	See No.8
24	$0 \leq I_1 \leq D_1, I_2 \leq 0 \leq D_2, 0 \leq I_0 \leq D_0$	See No.9
25	$0 \leq I_1 \leq D_1, I_2 \leq 0 \leq D_2, 0 \leq D_0 \leq I_0$	See No.10
26	$0 \leq I_1 \leq D_1, 0 \leq I_2 \leq D_2, 0 \leq I_0 \leq D_0$	See No.11
27	$0 \leq I_1 \leq D_1, 0 \leq I_2 \leq D_2, 0 \leq D_0 \leq I_0$	See No.12
28	$0 \leq I_1 \leq D_1, 0 \leq D_2 \leq I_2, 0 \leq I_0 \leq D_0$	See No.13
29	$0 \leq I_1 \leq D_1, 0 \leq D_2 \leq I_2, 0 \leq D_0 \leq I_0$	See No.14
30	$0 \leq D_1 \leq I_1, I_2 \leq 0 \leq D_2, I_0 \leq 0 \leq D_0$	See No.15
31	$0 \leq D_1 \leq I_1, I_2 \leq 0 \leq D_2, 0 \leq I_0 \leq D_0$	See No.16
32	$0 \leq D_1 \leq I_1, I_2 \leq 0 \leq D_2, 0 \leq D_0 \leq I_0$	See No.17
33	$0 \leq D_1 \leq I_1, I_2 \leq 0 \leq D_2, 0 \leq I_0 \leq D_0$	See No.18
34	$0 \leq D_1 \leq I_1, 0 \leq I_2 \leq D_2, 0 \leq D_0 \leq I_0$	See No.19
35	$0 \leq D_1 \leq I_1, 0 \leq I_2 \leq D_2, 0 \leq I_0 \leq D_0$	See No.20
36	$0 \leq D_1 \leq I_1, 0 \leq D_2 \leq I_2, 0 \leq D_0 \leq I_0$	See No.21

$\beta_2 k_1)) = p_1 Q_0 = LHS$. If $I_2 \geq D_2$, then $LHS = p_1 Q_0, RHS = \alpha(p_1 Q_1) + (1 - \alpha)(p_1 \beta_2 k_1 + p_2(1 - \beta_2)k_2)$. Since $p_2 \leq p_1$ and $D_2 < I_2$, this implies that $RHS \leq \alpha(p_1 Q_1) + (1 - \alpha)(p_1 \beta_2 k_1 + p_1(Q_2 - \beta_2 k_1)) = p_1 Q_0 = LHS$. •

Lemma 3 *If $I_1 \leq 0 \leq D_1, 0 \leq I_2 \leq D_2$ or $0 \leq D_2 \leq I_2, 0 \leq I_0 \leq D_0$ then $LHS \geq RHS$*

Proof If $0 \leq I_2 \leq D_2$, then

$$\begin{aligned} LHS &= p_1\beta_0k_1 + p_2(Q_0 - \beta_0k_1), RHS = \alpha(p_1Q_1) + (1 - \alpha)(p_1\beta_2k_1 + p_2(Q_2 - \beta_2k_1)) \\ LHS - RHS &= p_1(\beta_0k_1 - \alpha Q_1 - (1 - \alpha)\beta_2k_1) + p_2(Q_0 - \beta_0k_1 - (1 - \alpha)(Q_2 - \beta_2k_1)) \\ &= p_1(\alpha\beta_1k_1 - \alpha Q_1) + p_2(\alpha Q_1 - \alpha\beta_1k_1) = (p_1 - p_2)\alpha(\beta_1k_1 - Q_1) \geq 0 \end{aligned}$$

If $D_2 \leq I_2$, then

$$\begin{aligned} LHS &= p_1\beta_0k_1 + p_2(Q_0 - \beta_0k_1), RHS = \alpha(p_1Q_1) + (1 - \alpha)(p_1\beta_2k_1 + p_2(1 - \beta_2)k_2) \\ LHS - RHS &= p_1(\beta_0k_1 - \alpha Q_1 - (1 - \alpha)\beta_2k_1) + p_2(Q_0 - \beta_0k_1 - (1 - \alpha)(1 - \beta_2)k_2) \\ &\geq p_1(\alpha\beta_1k_1 - \alpha Q_1) + p_2(Q_0 - \beta_0k_1 - (1 - \alpha)(Q_2 - \beta_2k_1)) = (p_1 - p_2)\alpha(\beta_1k_1 - Q_1) \geq 0 \end{aligned}$$

•

Lemma 4 *If $I_1 \leq 0 \leq D_1, 0 \leq I_2 \leq D_2$ or $0 \leq D_2 \leq I_2, 0 \leq I_0 \geq D_0$ then $LHS \geq RHS$*

Proof If $0 \leq I_2 \leq D_2$, then

$$\begin{aligned} LHS &= p_1\beta_0k_1 + p_2(1 - \beta_0)k_2, RHS = \alpha(p_1Q_1) + (1 - \alpha)(p_1\beta_2k_1 + p_2(Q_2 - \beta_2k_1)) \\ LHS - RHS &= p_1(\beta_0k_1 - \alpha Q_1 - (1 - \alpha)\beta_2k_1) + p_2((1 - \beta_0)k_2 - (1 - \alpha)(Q_2 - \beta_2k_1)) \\ &\geq p_1(\alpha\beta_1k_1 - \alpha Q_1) + p_2((1 - \beta_0)k_2 - (1 - \alpha)(1 - \beta_2)k_2) \\ &= p_1(\alpha\beta_1k_1 - \alpha Q_1) + p_2k_2\alpha(1 - \beta_1) \geq 0 \end{aligned}$$

If $I_2 \geq D_2$, then

$$\begin{aligned} LHS &= p_1\beta_0k_1 + p_2(1 - \beta_0)k_2, RHS = \alpha(p_1Q_1) + (1 - \alpha)(p_1\beta_2k_1 + p_2(1 - \beta_2)k_2) \\ LHS - RHS &= p_1(\beta_0k_1 - \alpha Q_1 - (1 - \alpha)\beta_2k_1) + p_2((1 - \beta_0)k_2 - (1 - \alpha)(1 - \beta_2)k_2) \\ &= p_1(\alpha\beta_1k_1 - \alpha Q_1) + p_2\alpha k_2(1 - \beta_1) \geq 0 \end{aligned}$$

•

Lemma 5 *If $0 \leq I_1 \leq D_1$ or $0 \leq D_1 \leq I_1, I_2 \leq 0 \leq D_2, 0 \leq I_0 \leq D_0$ then $LHS \geq RHS$*

Proof If $0 \leq I_1 \leq D_1$ then

$$\begin{aligned} LHS &= p_1\beta_0k_1 + p_2(Q_0 - \beta_0k_1), RHS = \alpha(p_1\beta_1k_1 + p_2(Q_1 - \beta_1k_1)) + (1 - \alpha)(p_1Q_2) \\ LHS - RHS &= p_1(\beta_0k_1 - \alpha\beta_1k_1 - (1 - \alpha)Q_2) + p_2(Q_0 - \beta_0k_1 - \alpha Q_1 + \alpha\beta_1k_1) \\ &= (1 - \alpha)(\beta_2k_1 - Q_2)(p_1 - p_2) \geq 0 \end{aligned}$$

If $0 \leq D_1 \leq I_1$ then

$$\begin{aligned} LHS &= p_1\beta_0k_1 + p_2(Q_0 - \beta_0k_1), RHS = \alpha(p_1\beta_1k_1 + p_2(1 - \beta_1)k_2) + (1 - \alpha)(p_1Q_2) \\ LHS - RHS &\geq p_1((1 - \alpha)\beta_2k_1 - (1 - \alpha)Q_2) + p_2(Q_0 - \beta_0k_1 - \alpha(Q_1 - \beta_1k_1)) \\ &= (1 - \alpha)(\beta_2k_1 - Q_2)(p_1 - p_2) \geq 0 \end{aligned}$$

•

Lemma 6 If $0 \leq I_1 \leq D_1, I_2 \leq 0 \leq D_2, 0 \leq D_0 \leq I_0$ then $LHS \geq RHS$

Proof

$$\begin{aligned} LHS &= p_1\beta_0k_1 + p_2(1 - \beta_0)k_2, RHS = \alpha(p_1\beta_1k_1 + p_2(Q_1 - \beta_1k_1)) + (1 - \alpha)(p_1Q_2) \\ LHS - RHS &= p_1(\beta_0k_1 - \alpha\beta_1k_1 - (1 - \alpha)Q_2) + p_2((1 - \beta_0)k_2 - \alpha(Q_1 - \beta_1k_1)) \\ &\geq p_1(1 - \alpha)(\beta_2k_1 - Q_2) + p_2((1 - \beta_0)k_2 - \alpha(1 - \beta_1)k_2) \\ &= p_1(1 - \alpha)(\beta_2k_1 - Q_2) + p_2k_2(1 - \alpha)(1 - \beta_2) \geq 0 \end{aligned}$$

•

Lemma 7 If $0 \leq I_1 \leq D_1, 0 \leq I_2 \leq D_2, 0 \leq I_0 \leq D_0$ then $LHS \geq RHS$

Proof

$$\begin{aligned} LHS &= p_1\beta_0k_1 + p_2(Q_0 - \beta_0k_1) \\ RHS &= \alpha(p_1\beta_1k_1 + p_2(Q_1 - \beta_1k_1)) + (1 - \alpha)(p_1\beta_2k_1 + p_2(Q_2 - \beta_2k_1)), LHS - RHS = 0 \end{aligned}$$

•

Lemma 8 If $0 \leq I_1 \leq D_1$ or $0 \leq D_1 \leq I_1, 0 \leq I_2 \leq D_2$ or $0 \leq D_2 \leq I_2, 0 \leq D_0 \leq I_0$ then $LHS \geq RHS$

Proof If $0 \leq I_1 \leq D_1, 0 \leq I_2 \leq D_2$ then

$$\begin{aligned} LHS &= p_1\beta_0k_1 + p_2(1 - \beta_0)k_2 \\ RHS &= \alpha(p_1\beta_1k_1 + p_2(Q_1 - \beta_1k_1)) + (1 - \alpha)(p_1\beta_2k_1 + p_2(Q_2 - \beta_2k_1)) \\ LHS - RHS &\geq p_2k_2(1 - \beta_0 - \alpha(1 - \beta_1) - (1 - \alpha)(1 - \beta_2)) = 0 \end{aligned}$$

If $0 \leq I_1 \leq D_1, 0 \leq D_2 \leq I_2$ then

$$\begin{aligned} LHS &= p_1\beta_0k_1 + p_2(1 - \beta_0)k_2 \\ RHS &= \alpha(p_1\beta_1k_1 + p_2(Q_1 - \beta_1k_1)) + (1 - \alpha)(p_1\beta_2k_1 + p_2(1 - \beta_2)k_2) \\ LHS - RHS &\geq p_2k_2(1 - \beta_0 - \alpha(1 - \beta_1) - (1 - \alpha)(1 - \beta_2)) = 0 \end{aligned}$$

If $0 \leq D_1 \leq I_1, 0 \leq I_2 \leq D_2$ then

$$LHS = p_1\beta_0k_1 + p_2(1 - \beta_0)k_2$$

$$RHS = \alpha(p_1\beta_1k_1 + p_2(1 - \beta_1)k_2) + (1 - \alpha)(p_1\beta_2k_1 + p_2(Q_2 - \beta_2k_1))$$

$$LHS - RHS \geq p_2k_2(1 - \beta_0 - \alpha(1 - \beta_1) - (1 - \alpha)(1 - \beta_2)) = 0$$

•

Lemma 9 *If $0 \leq I_1 \leq D_1$ or $0 \leq D_1 \leq I_1, 0 \leq I_2 \leq D_2$ or $0 \leq D_2 \leq I_2, 0 \leq I_0 \leq D_0$ then $LHS \geq RHS$*

Proof If $0 \leq I_1 \leq D_1$ and $0 \leq D_2 \leq I_2$, then,

$$LHS = p_1\beta_0k_1 + p_2(Q_0 - \beta_0k_1)$$

$$RHS = \alpha(p_1\beta_1k_1 + p_2(Q_1 - \beta_1k_1)) + (1 - \alpha)(p_1\beta_2k_1 + p_2(1 - \beta_2)k_2)$$

$$LHS - RHS \geq p_2(Q_0 - \beta_0k_1 - \alpha(Q_1 - \beta_1k_1) - (1 - \alpha)(Q_2 - \beta_2k_1)) = 0$$

If $0 \leq D_1 \leq I_1$ and $0 \leq I_2 \leq D_2$, then,

$$LHS = p_1\beta_0k_1 + p_2(Q_0 - \beta_0k_1)$$

$$RHS = \alpha(p_1\beta_1k_1 + p_2(1 - \beta_1)k_2) + (1 - \alpha)(p_1\beta_2k_1 + p_2(Q_2 - \beta_2k_1))$$

$$LHS - RHS \geq p_2(Q_0 - \beta_0k_1 - \alpha(Q - \beta_1k_1) - (1 - \alpha)(Q_2 - \beta_2k_1)) = 0$$

If $0 \leq D_1 \leq I_1$ and $0 \leq D_2 \leq I_2$, then,

$$LHS = p_1\beta_0k_1 + p_2(Q_0 - \beta_0k_1)$$

$$RHS = \alpha(p_1\beta_1k_1 + p_2(1 - \beta_1)k_2) + (1 - \alpha)(p_1\beta_2k_1 + p_2(1 - \beta_2)k_2)$$

$$LHS - RHS \geq p_2(Q_0 - \beta_0k_1 - \alpha(Q_1 - \beta_1k_1) - (1 - \alpha)(Q_2 - \beta_2k_1)) = 0$$

•

Lemma 10 *If $0 \leq D_1 \leq I_1, I_2 \leq 0 \leq D_2, 0 \leq D_0 \leq I_0$ then $LHS \geq RHS$*

Proof

$$LHS = p_1\beta_0k_1 + p_2(1 - \beta_0)k_2, RHS = \alpha(p_1\beta_1k_1 + p_2(1 - \beta_1)k_2) + (1 - \alpha)(p_1Q_2)$$

$$LHS - RHS = p_1(1 - \alpha)(\beta_2k_1 - Q_2) + p_2k_2(1 - \alpha)(\beta_1) \geq 0$$

•

Lemma 11 *If $0 \leq D_1 \leq I_1, 0 \leq D_2 \leq I_2, 0 \leq D_0 \leq I_0$ then $LHS \geq RHS$*

Proof

$$LHS = p_1\beta_0k_1 + p_2(1 - \beta_0)k_2$$

$$RHS = \alpha(p_1\beta_1k_1 + p_2(1 - \beta_1)k_2) + (1 - \alpha)(p_1\beta_2k_1 + p_2(1 - \beta_2)k_2)$$

$$LHS - RHS = 0$$

•

Appendix B

Appendix for Essay 2

B.1. Proofs for Retailer's Timing Problem with WP Contract

B.1.1 Proof for Proposition 1

We use a simplified notation here: We drop the contract type WP from the parameters.

(a) We derive the first order partial derivative of $E[R(\beta|Q)]$. Let us define the following:

$$V(j, k) = p_1 \min(Q, j) + p_2 \min(k, (Q - j)^+) \quad (\text{B.1})$$

For convenience in this proof, let us define the following notation: $N_1(\beta) = N_1^\beta$ and $N_2(\beta) = N_2^\beta$. Then, we can re-write the optimization problem as:

$$R^*(Q) = \max_{0 \leq \beta \leq 1} E[R(\beta|Q)] = \max_{0 \leq \beta \leq 1} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} V(j, k) P(N_1^\beta = j) P(N_2^\beta = k) \quad (\text{B.2})$$

From (B.2) we have the first order derivative in terms of V function as follows:

$$\frac{\partial E[R(\beta|Q)]}{\partial \beta} = \frac{\partial E[V(N_1^\beta, N_2^\beta)]}{\partial \beta} \quad (\text{B.3})$$

We shall evaluate (B.3), the first order derivative from basic principles:

$$\frac{\partial E[R(\beta|Q)]}{\partial \beta} = \lim_{\varepsilon \rightarrow 0} \frac{E[V(N_1^{\beta+\varepsilon}, N_2^{\beta+\varepsilon})] - E[V(N_1^\beta, N_2^\beta)]}{\varepsilon} = \lim_{\varepsilon \rightarrow 0} \frac{VR}{\varepsilon}. \quad (\text{B.4})$$

Note that when $\varepsilon \rightarrow 0$, there is at most one arrival in the interval $[\beta, \beta + \varepsilon]$ following the Poisson model. Conditioning on the N_1^β and N_2^β , we can write the following:

$$N_1^{\beta+\varepsilon} = \begin{cases} N_1^\beta & \text{with probability } (1 - \lambda_1 \varepsilon) \\ N_1^\beta + 1 & \text{with probability } \lambda_1 \varepsilon \end{cases} \quad (\text{B.5})$$

Similarly we have:

$$N_2^{\beta+\varepsilon} = \begin{cases} N_2^\beta & \text{with probability } (1 - \lambda_2 \varepsilon) \\ N_2^\beta - 1 & \text{with probability } \lambda_2 \varepsilon \end{cases} \quad (\text{B.6})$$

Using (B.5) and (B.6) to expand the term $E[V(N_1^{\beta+\varepsilon}, N_2^{\beta+\varepsilon})]$, we get:

$$\begin{aligned} E[V(N_1^{\beta+\varepsilon}, N_2^{\beta+\varepsilon})] &= (1 - \lambda_1\varepsilon)(1 - \lambda_2\varepsilon)E[V(N_1^\beta, N_2^\beta)] \\ &\quad + (1 - \lambda_1\varepsilon)(\lambda_2\varepsilon)E[V(N_1^\beta, N_2^\beta - 1)] \\ &\quad + (\lambda_1\varepsilon)(1 - \lambda_2\varepsilon)E[V(N_1^\beta + 1, N_2^\beta)] \end{aligned} \quad (\text{B.7})$$

Substituting (B.7) in Equation (B.4) and evaluating the numerator, we get,

$$\begin{aligned} VR &= E[V(N_1^{\beta+\varepsilon}, N_2^{\beta+\varepsilon})] - E[V(N_1^\beta, N_2^\beta)] \\ &= (1 - \lambda_1\varepsilon - \lambda_2\varepsilon + \lambda_1\lambda_2\varepsilon^2 - 1) E[V(N_1^\beta, N_2^\beta)] \\ &\quad + (\lambda_2\varepsilon - \lambda_1\lambda_2\varepsilon^2) E[V(N_1^\beta, N_2^\beta - 1)] \\ &\quad + (\lambda_1\varepsilon - \lambda_1\lambda_2\varepsilon^2) E[V(N_1^\beta + 1, N_2^\beta)] \end{aligned} \quad (\text{B.8})$$

Replacing the numerator in (B.4), and re-arranging the terms, we finally obtain,

$$\begin{aligned} \frac{\partial E[R(\beta|Q)]}{\partial \beta} &= \lambda_1 \left(E[V(N_1^\beta + 1, N_2^\beta)] - E[V(N_1^\beta, N_2^\beta)] \right) \\ &\quad - \lambda_2 \left(E[V(N_1^\beta, N_2^\beta)] - E[V(N_1^\beta, N_2^\beta - 1)] \right) \end{aligned} \quad (\text{B.9})$$

Using the definition for the expectation, we have,

$$\begin{aligned} \frac{\partial E[R(\beta|Q)]}{\partial \beta} &= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \left\{ \lambda_1 [V(j+1, k) - V(j, k)] \right. \\ &\quad \left. - \lambda_2 [V(j, k+1) - V(j, k)] \right\} P(N_1^\beta = j) P(N_2^\beta = k) \end{aligned} \quad (\text{B.10})$$

The difference terms required in Equation (B.10) are provided in Table B.1. Substituting the values from Table B.1 into Equation (B.10) we obtain the first order derivative with respect to β as stated in this proposition.

Table B.1: Evaluation of the difference functions $V(j+1, k) - V(j, k)$ and $V(j, k+1) - V(j, k)$

Case	$V(j+1, k) - V(j, k)$	$V(j, k+1) - V(j, k)$
$j > Q$	0	0
$j \leq Q$ and $j+k \leq Q$	p_1	p_2
$j \leq Q$ and $j+k > Q$	$p_1 - p_2$	0

(b) First, consider the interesting case of $\lambda_1 \leq \lambda_2$. In this case, the second order derivative wrt β is given in Lemma 12 below. Let us consider the sign of $\frac{\partial^2 E[R(\beta|Q)]}{\partial \beta^2}$ for different inventory levels. For a given inventory level that satisfies Lemmas 13

- 15, the proof for the proposition follows trivially as $\frac{\partial^2 E[R(\beta|Q)]}{\partial \beta^2} \leq 0$. In case of high inventory levels, as mentioned in Lemma 17, $\frac{\partial^2 E[R(\beta|Q)]}{\partial \beta^2} \rightarrow 0$ and hence the expected revenue function exhibits affine behavior which implies that it is concave in β . Consider the case when $\lambda_1 > \lambda_2$. In this case, the first order derivative with respect to β is always positive. This means that the expected revenue function is always strictly increasing in β . Thus, by definition of quasiconcavity, the expected revenue function is quasiconcave in β when $\lambda_1 > \lambda_2$.

B.1.2 Lemmas for Retailer's Timing Problem with Poisson Demand Model

Lemma 12 (a) *The second order derivative with respect to β is given as follows:*

$$\begin{aligned} \frac{\partial^2 E[R(\beta|Q)]}{\partial \beta^2} &= p_2(\lambda_2 - \lambda_1) \left[\lambda_1 P(N_1^\beta + N_2^\beta = Q) - \lambda_2 P(N_1^\beta + N_2^\beta = Q + 1) \right] \\ &\quad - \lambda_1^2(p_1 - p_2) P(N_1^\beta = Q) \end{aligned} \quad (\text{B.11})$$

(b) *If $\gamma = \frac{\beta\lambda_1 + (1-\beta)\lambda_2}{Q+1}$, then we have:*

$$\frac{\partial^2 E[R(\beta|Q)]}{\partial \beta^2} = p_2(\lambda_2 - \lambda_1)(\lambda_1 - \gamma \lambda_2) P(N_1^\beta + N_2^\beta = Q) - \lambda_1^2(p_1 - p_2) P(N_1^\beta = Q) \quad (\text{B.12})$$

Proof (a) We can write the first order derivative, given in Equation (3.11) as follows:

$$\begin{aligned} \frac{\partial E[R(\beta|Q)]}{\partial \beta} &= \left[p_1 \lambda_1 - p_2 \lambda_2 \right] \left[1 - P(N_1^\beta + N_2^\beta > Q) \right] \\ &\quad + \left[p_1 \lambda_1 - p_2 \lambda_1 \right] \left[P(N_1^\beta + N_2^\beta > Q) - P(N_1^\beta > Q) \right] \end{aligned} \quad (\text{B.13})$$

In order to evaluate the second order derivative, it is sufficient if we evaluate the following expressions: $\frac{\partial}{\partial \beta} P(N_1^\beta + N_2^\beta > Q)$ and $\frac{\partial}{\partial \beta} P(N_1^\beta > Q)$. Writing down the derivative from first principles, we have:

$$\frac{\partial}{\partial \beta} P(N_1^\beta + N_2^\beta > Q) = \lim_{\varepsilon \rightarrow 0} \frac{P(N_1^{\beta+\varepsilon} + N_2^{\beta+\varepsilon} > Q) - P(N_1^\beta + N_2^\beta > Q)}{\varepsilon} \quad (\text{B.14})$$

where ε is so chosen that there is at most one arrival in this interval. Using Equations (B.5) and (B.6) we can evaluate the numerator in (B.14) as follows:

$$\begin{aligned} P(N_1^{\beta+\varepsilon} + N_2^{1-\beta-\varepsilon} > Q) &= (1 - \lambda_1 \varepsilon) (1 - \lambda_2 \varepsilon) P(N_1^\beta + N_2^\beta > Q) \\ &\quad + (\lambda_1 \varepsilon) (1 - \lambda_2 \varepsilon) P(N_1^\beta + 1 + N_2^\beta > Q) \\ &\quad + (1 - \lambda_1 \varepsilon) (\lambda_2 \varepsilon) P(N_1^\beta + N_2^\beta - 1 > Q) \end{aligned} \quad (\text{B.15})$$

Substituting (B.15) into Equation (B.14), and taking limits, we have,

$$\begin{aligned}\frac{\partial}{\partial\beta}P(N_1^\beta + N_2^\beta > Q) &= (-\lambda_1 - \lambda_2) P(N_1^\beta + N_2^\beta > Q) \\ &\quad + (\lambda_1) P(N_1^\beta + N_2^\beta + 1 > Q) \\ &\quad + (\lambda_2) P(N_1^\beta + N_2^\beta - 1 > Q)\end{aligned}\tag{B.16}$$

Note that the probabilities in the above expression are related as follows:

$$P(N_1^\beta + N_2^\beta + 1 > Q) = P(N_1^\beta + N_2^\beta > Q) + P(N_1^\beta + N_2^\beta = Q)\tag{B.17}$$

$$P(N_1^\beta + N_2^\beta - 1 > Q) = P(N_1^\beta + N_2^\beta > Q) - P(N_1^\beta + N_2^\beta = Q + 1)\tag{B.18}$$

Using the above relationship, we thus finally arrive at:

$$\frac{\partial}{\partial\beta}P(N_1^\beta + N_2^\beta > Q) = \lambda_1 P(N_1^\beta + N_2^\beta = Q) - \lambda_2 P(N_1^\beta + N_2^\beta = Q + 1)\tag{B.19}$$

Proceeding on similar lines, we can derive the following derivatives as:

$$\frac{\partial}{\partial\beta}P(N_1^\beta + N_2^\beta \geq Q) = \lambda_1 P(N_1^\beta + N_2^\beta = Q - 1) - \lambda_2 P(N_1^\beta + N_2^\beta = Q)\tag{B.20}$$

$$\frac{\partial}{\partial\beta}P(N_1^\beta > Q) = \lambda_1 P(N_1^\beta = Q)\tag{B.21}$$

$$\frac{\partial}{\partial\beta}P(N_1^\beta \geq Q) = \lambda_1 P(N_1^\beta = Q - 1)\tag{B.22}$$

Differentiating Equation (B.13) and using appropriate relations from (B.19) - (B.22), we have:

$$\begin{aligned}\frac{\partial^2 E[R(\beta)]}{\partial\beta^2} &= \left[\lambda_1 p_1 - \lambda_2 p_2 \right] \left[-\lambda_1 P(N_1^\beta + N_2^\beta = Q) + \lambda_2 P(N_1^\beta + N_2^\beta = Q + 1) \right] \\ &\quad + \left[\lambda_1 p_1 - \lambda_1 p_2 \right] \left[\lambda_1 P(N_1^\beta + N_2^\beta = Q) - \lambda_2 P(N_1^\beta + N_2^\beta = Q + 1) \right] \\ &\quad - \lambda_1^2 \left[p_1 - p_2 \right] P(N_1^\beta = Q)\end{aligned}\tag{B.23}$$

Re-arranging the terms, we obtain the second order derivative as stated in the lemma. (b) We know from the Poisson process that the probabilities are related as:

$$P(N_1^\beta + N_2^\beta = Q + 1) = \frac{\beta\lambda_1 + (1 - \beta)\lambda_2}{Q + 1} P(N_1^\beta + N_2^\beta = Q)\tag{B.24}$$

Using the relation between the probabilities and definition of γ we get the required expression. •

Lemma 13 *If $Q + 1 \leq \lambda_1$, then the second order derivative is non-positive.*

Proof Note that the expression $\beta\lambda_1 + (1 - \beta)\lambda_2$ lies between λ_1 and λ_2 as it is a convex combination of λ_1 and λ_2 . In this case, we see that:

$$\gamma = \frac{\beta\lambda_1 + (1 - \beta)\lambda_2}{Q + 1} \geq \frac{\beta\lambda_1 + (1 - \beta)\lambda_2}{\lambda_1} \geq 1$$

Now, consider the second order derivative mentioned in Equation (B.12):

$$\frac{\partial^2 E[R(\beta)]}{\partial \beta^2} \leq p_2(\lambda_2 - \lambda_1)(\lambda_1 - \lambda_2)P(N_1^\beta + N_2^\beta = Q) - \lambda_1^2(p_1 - p_2)P(N_1^\beta = Q) \leq 0 \quad (\text{B.25})$$

Equation (B.25) follows as $\lambda_1 \leq \lambda_2$ and that $\gamma \geq 1$. Thus, we can see that the second order derivative as mentioned in Equation (B.12) is non-positive. •

Lemma 14 *If $\lambda_1 \leq Q + 1 \leq \lambda_2$, then the second order derivative is non-positive.*

Proof Consider the following expression:

$$\begin{aligned} \lambda_1 - \gamma \lambda_2 &= \frac{1}{Q + 1} [\lambda_1(Q + 1) - (\beta\lambda_1 + (1 - \beta)\lambda_2)\lambda_2] \\ &= \frac{1}{Q + 1} [\lambda_1(Q + 1)(\beta + 1 - \beta) - (\beta\lambda_1 + (1 - \beta)\lambda_2)] \\ &= \frac{1}{Q + 1} [\beta\lambda_1(Q + 1 - \lambda_2) + (1 - \beta)(\lambda_1(Q + 1) - \lambda_2^2)] \\ &\leq \frac{1}{Q + 1} [\beta\lambda_1(Q + 1 - \lambda_2) + (1 - \beta)(\lambda_2(Q + 1) - \lambda_2^2)] \\ &= \frac{1}{Q + 1} [\beta\lambda_1(Q + 1 - \lambda_2) + (1 - \beta)\lambda_2(Q + 1 - \lambda_2)] \\ &\leq 0 \end{aligned} \quad (\text{B.26})$$

Thus in this case, using Equation (B.26) in the second order derivative mentioned in Equation (B.12), we can see that it consists of all non-positive terms. •

Lemma 15 *If $\lambda_2 \leq Q + 1 \leq \lambda_2 \left(\frac{\lambda_2}{\lambda_1}\right)$, then the second order derivative is non-positive.*

Proof Using the assumption of lemma and by definition of γ we have:

$$\gamma = \left(\frac{\beta\lambda_1 + (1 - \beta)\lambda_2}{Q + 1} \right) \geq \frac{\lambda_1}{Q + 1} \geq \frac{\lambda_1}{\lambda_2}$$

Using the above inequality we derive:

$$\begin{aligned}\lambda_1 - \gamma\lambda_2 &= \lambda_1 - \left(\frac{\beta\lambda_1 + (1-\beta)\lambda_2}{Q+1} \right) \lambda_2 \leq \lambda_1 - \left(\frac{\lambda_1}{\lambda_2} \right) \lambda_2 \\ &= 0\end{aligned}\tag{B.27}$$

Using Equation (B.27) we see that in this case also, the second order derivative wrt β as mentioned in Equation (B.12) is non-positive. •

We need to prove the following additional lemma 16, before we can prove the lemma 17 which is regarding the affine nature of the expected revenue function.

Lemma 16 *The following probabilities share the relationship:*

$$P(N_1^\beta + N_2^\beta = Q) = P(N_1^\beta = Q) e^{-\lambda_2(1-\beta)} \left(1 + \frac{\lambda_2(1-\beta)}{\lambda_1\beta} \right)^Q \tag{B.28}$$

Proof The proof for the above relation is as follows:

$$\begin{aligned}P(N_1^\beta + N_2^\beta = Q) &= \sum_{i=0}^Q P(N_1^\beta = i) P(N_2^\beta = Q - i) \\ &= \sum_{i=0}^Q \frac{e^{-\lambda_1\beta}(\lambda_1\beta)^i}{i!} \frac{e^{-\lambda_2(1-\beta)}(\lambda_2(1-\beta))^{Q-i}}{Q-i!} \\ &= \frac{e^{-\lambda_1\beta}(\lambda_1\beta)^Q}{Q!} e^{-\lambda_2(1-\beta)} \sum_{i=0}^Q \left(\frac{\lambda_2(1-\beta)}{\lambda_1\beta} \right)^{Q-i} \frac{Q!}{i!Q-i!} \\ &= P(N_1^\beta = Q) e^{-\lambda_2(1-\beta)} \sum_{i=0}^Q \left(\frac{\lambda_2(1-\beta)}{\lambda_1\beta} \right)^{Q-i} 1^i \frac{Q!}{i!Q-i!} \\ &= P(N_1^\beta = Q) e^{-\lambda_2(1-\beta)} \left(1 + \frac{\lambda_2(1-\beta)}{\lambda_1\beta} \right)^Q\end{aligned}$$

•

Lemma 17 *If $\lambda_2 \left(\frac{\lambda_2}{\lambda_1} \right) \leq Q + 1$, then the second order derivative approaches zero.*

Proof From Equation (B.28) we can see that for low Q , $P(N_1^\beta + N_2^\beta = Q) \leq P(N_1^\beta = Q)$. For sufficiently high Q , we can say that, $P(N_1^\beta + N_2^\beta = Q) \geq P(N_1^\beta = Q)$. Let us re-write the second order derivative using Equation (B.28):

$$\begin{aligned}\frac{\partial^2 E[R(\beta)]}{\partial \beta^2} &= \left\{ p_2(\lambda_2 - \lambda_1)(\lambda_1 - \gamma\lambda_2) e^{-\lambda_2(1-\beta)} \left(1 + \frac{\lambda_2(1-\beta)}{\lambda_1\beta} \right)^Q \right. \\ &\quad \left. - \lambda_1^2(p_1 - p_2) \right\} P(N_1^\beta = Q)\end{aligned}\tag{B.29}$$

As Q is increased, γ goes to 0 and the first term approaches very high positive value as $\left(1 + \frac{\lambda_2(1-\beta)}{\lambda_1\beta}\right)^Q$ increases in Q , compared to the negative second term which does not grow with Q . But as we know, Poisson probabilities, $P(N_1^\beta = Q)$, drop rapidly after $Q \geq \lambda_1$. Since $\lambda_2 \left(\frac{\lambda_2}{\lambda_1}\right) \leq Q + 1$, the second order derivative approaches 0. To illustrate, consider the limiting case when $Q \rightarrow \infty$, the first term is highly positive compared to the second term, hence the difference is positive. But as we are multiplying the whole expression by $P(N_1^\beta = Q) \rightarrow 0$, the second order derivative approaches 0. In short, for Q mentioned in this case, the second order derivative approaches zero in the limiting case. In fact, the first order derivative is $p_1\lambda_1 - p_2\lambda_2$ and thus there is no change in the first order derivative wrt β . •

B.2. Proofs for Retailer's Timing Problem with BB Contract

B.2.1 Proof for Proposition 2

Proof For notational convenience, we write, $N_1(\beta)$ as N_1^β and $N_2(\beta)$ as N_2^β .

(a) From Equation (3.13), we have the first order derivative as:

$$\frac{\partial E[R^{BB}(\beta)]}{\partial \beta} = \frac{\partial E[V^{BB}(N_1^\beta, N_2^\beta)]}{\partial \beta} \quad (\text{B.30})$$

Performing similar analysis to the WP contract as in Chapter 2, we can write the following:

$$\begin{aligned} \frac{\partial E[R^{BB}(\beta)]}{\partial \beta} = & \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \left\{ \lambda_1 [V^{BB}(j+1, k) - V^{BB}(j, k)] \right. \\ & \left. - \lambda_2 [V^{BB}(j, k+1) - V^{BB}(j, k)] \right\} P_j P_k \end{aligned} \quad (\text{B.31})$$

Evaluating difference terms in Equation (B.31) we obtain Table (B.2). Substituting

Table B.2: Difference Function for BB contract

Case	$V^{BB}(j+1, k) - V^{BB}(j, k)$	$V^{BB}(j, k+1) - V^{BB}(j, k)$
$j > Q$	0	0
$j \leq Q$ and $j+k \leq Q$	$p_1 - b$	$p_2 - b$
$j \leq Q$ and $j+k > Q$	$p_1 - p_2$	0

the values from Table (B.2) into Equation (B.31) we obtain the partial differential

as:

$$\begin{aligned} \frac{\partial E[R^{BB}(\beta)]}{\partial \beta} &= \sum_{j=0}^Q \sum_{k=0}^{Q-j} [(p_1 - b)\lambda_1 - (p_2 - b)\lambda_2] \mathbf{P}_j^1(\beta) \mathbf{P}_j^2(\beta) \\ &\quad + \sum_{j=0}^Q \sum_{k=Q-j}^{\infty} [p_1 - p_2]\lambda_1 \mathbf{P}_j^1(\beta) \mathbf{P}_j^2(\beta) \end{aligned} \quad (\text{B.32})$$

We simplify Equation (B.32) and thus arrive at the first order derivative of expected revenue function with respect to β .

(b) We will prove this using the second order derivative wrt β . Just like WP contract, this can be derived from Equation (3.14). That is,

$$\frac{\partial^2 E[R^{BB}(\beta)]}{\partial \beta^2} = \frac{\partial^2 E[R^{WP}(\beta)]}{\partial \beta^2} + b[\lambda_2 - \lambda_1] \frac{\partial}{\partial \beta} P(N_1^\beta + N_2^\beta \leq Q) \quad (\text{B.33})$$

Substituting the second order derivative for WP contract from Equation (B.11) in Equation (B.33), we arrive at the second order derivative with respect to β as follows:

$$\begin{aligned} \frac{\partial^2 E[R^{BB}(\beta)]}{\partial \beta^2} &= (p_2 - b)(\lambda_2 - \lambda_1) \left[\lambda_1 P(N_1^\beta + N_2^\beta = Q) - \lambda_2 P(N_1^\beta + N_2^\beta = Q + 1) \right] \\ &\quad - \lambda_1^2 (p_1 - p_2) P(N_1^\beta = Q) \end{aligned} \quad (\text{B.34})$$

Let $\gamma = \frac{\beta\lambda_1 + (1-\beta)\lambda_2}{Q+1}$, as defined in Lemma 12. Then, we can re-write as follows:

$$\begin{aligned} \frac{\partial^2 E[R^{BB}(\beta|Q)]}{\partial \beta^2} &= (p_2 - b)(\lambda_2 - \lambda_1)(\lambda_1 - \gamma\lambda_2) P(N_1^\beta + N_2^\beta = Q) \\ &\quad - \lambda_1^2 (p_1 - p_2) P(N_1^\beta = Q) \end{aligned} \quad (\text{B.35})$$

This looks similar to the second order derivative wrt β for WP contract, except for the term $p_2 - b$. The proof that the expected revenue function for BB contract is concave follows from the proof for WP contract discussed in Section B.1.1. The lemmas 13 through 15 and 17 still hold true when we substitute $p_2 - b$ for p_2 and thus the expected revenue function is concave in β . \bullet

B.3. Proofs for Retailer's Timing Problem with MM Contract

B.3.1 Proof for Proposition 3

Proof (a) We will use the analysis from WP contract to obtain first order derivative wrt β for MM contract as shown below.

- Two-market Rebate

In the analysis from Section B.1.1 we substitute $p_1 = p_2 = w(1 + g)$ but still

use p_1 and p_2 for effective rate calculations for Poisson process in both the markets. Substituting these values we get:

$$\begin{aligned} \frac{\partial E[R^{MM}(\beta|Q)]}{\partial \beta} &= w(1+g)[\lambda_1 - \lambda_2] P(N_1^\beta + N_2^\beta \leq Q) \\ &\quad + [w(1+g) - w(1+g)] P(N_1^\beta \leq Q < N_1^\beta + N_2^\beta) \end{aligned}$$

The second term vanishes and simplifying the above expression we arrive at the first order derivative wrt β .

- One-market Rebate

In this case the primary market situation is unchanged compared to WP contract, that is, revenue per unit sold is p_1 and use p_1 for effective rate calculations. In secondary market we substitute $p_2 = w(1+g)$ but use p_2 for calculation of effective rate for Poisson process. Substituting corresponding values into analysis from WP contract, we arrive at the required expression.

(b) We will prove using the second order derivative wrt β . Using the second order derivative wrt β for WP contract, we can write the same for MM contract:

- Two-market Rebate

$$\begin{aligned} \frac{\partial^2 E[R^{MM}(\beta|Q)]}{\partial \beta^2} &= w(1+g)(\lambda_2 - \lambda_1) \left[\lambda_1 P(N_1^\beta + N_2^\beta = Q) \right. \\ &\quad \left. - \lambda_2 P(N_1^\beta + N_2^\beta = Q + 1) \right] \end{aligned} \quad (\text{B.36})$$

- One-market Rebate

$$\begin{aligned} \frac{\partial^2 E[R^{MM}(\beta|Q)]}{\partial \beta^2} &= w(1+g)(\lambda_2 - \lambda_1) \left[\lambda_1 P(N_1^\beta + N_2^\beta = Q) \right. \\ &\quad \left. - \lambda_2 P(N_1^\beta + N_2^\beta = Q + 1) \right] \\ &\quad - \lambda_1^2 (p_1 - w(1+g)) P(N_1^\beta = Q) \end{aligned} \quad (\text{B.37})$$

The above derivation can be trivially obtained. From Table (3.2), we see that MM contract is a special instance of WP contract. Thus, concavity follows directly from the WP contract. •

B.4. Description of Algorithm for Computational Experiments

In this research, we considered a game theoretic model between the supplier and the retailer in three stages. In stage 1, the supplier decides on the optimal wholesale price. In stage 2, the retailer decides on the optimal order quantity and in stage 3, the retailer decides on the optimal timing of the discount based on the

state of the market. We discuss the algorithms (coded in C++) used in each of this stages below:

- The supplier's expected profit function is not concave and hence we do discrete search with an initial step size of w_{step}^0 to find out the optimal wholesale price for this step size, w_{step0}^* . The idea is that this w_{step0}^* represents the approximate location of optimal wholesale price and hence we do further refined search, using a smaller step size w_{step} in the neighborhood of w_{step0}^* to get closer to w^* . We stop when the percentage change in supplier expected profit is not more pre-specified limit, say $\delta\%$. Specifically we use $w_{step}^0 = 1$, decrease step size by $1/2$ and $\delta = 1$.
- Q is discrete and hence we use discrete search with step size of 1 to identify the Q^* . If the retailer's expected profit function were concave in Q for any contract, then we make use of this property in finding the stopping criterion and thus make the algorithm efficient in identifying the optimal order quantity, Q^* .
- In stage 3, the retailer decides on the optimal timing of the markdown. We showed that the retailer's expected revenue function is quasi-concave in β and hence we use Fibonacci search algorithm (see Bazaraa et al. (2006)) to identify the β^* . We identify the number of steps required based on the convergence criteria.

Appendix C

Appendix for Essay 3

C.1. Notation

We define the common notation¹ that will be used for MIP formulations as follows:

Sets:

N = Set of nodes i in the network. E.g. $\{1, 2, \dots, N\}$

$\bar{N}_i = N - \{i\}$. That is, set of nodes in the network except i .

E = Set of edges $i - j$ in the network. E.g. $\{1 - 2, \dots, E\}$

P = Set of priorities p for all nodes. E.g. $\{1, 2, \dots, P\}$

N_p = Set of nodes i in the network with priority p . E.g. $\{5, 7, 8, 10\}$

Parameters:

L_{ij} = Shortest path distance between nodes i and j

T_{ij} = Travel time corresponding to shortest path between nodes i and j

UT_i = Time of unloading required at node i

L_i^{eq} = Distance equivalent of unloading time at node i

P_i = Priority for node i

Decision Variables:

$$x_{ij} = \begin{cases} 1 & \text{if vehicle uses shortest path from node } i \text{ to node } j \\ 0 & \text{otherwise} \end{cases}$$

s_i = Time at which service is rendered by vehicle at node i

s^{max} = Latest service time for the entire network

C.2. Formulation MinDist(d) for u-HVRP

The objective function denoted by equation (C.1) is to minimize total cost of travel by the vehicle while satisfying **d**-Relaxed Rule of Priority. Constraint (C.2) says the vehicle enters a node only once. Similarly constraint (C.3) says that vehicle has to leave the node exactly once. Note that even though these constraints say that the vehicle visits a node only once in this modified complete graph, it does not necessarily mean that the vehicle visits that node only once in the original graph. All these assignment constraints say is that the vehicle has to visit node exactly once

¹In this chapter, please note that we use different typeface to differentiate parameters or given data and *decision variables* in the MIP formulations

to satisfy demand. Constraint (C.4) says that the service time for the depot is 0. Constraints (C.5) and (C.6) together calculate the time of service at a node j . If the vehicle traverses to node j from node i , then $x_{ij} = 1$ and thus the two constraints become binding. Constraint (C.7) enforces the \mathbf{d} -Relaxed Rule of Priority. It says that service time for nodes of priority p and $p + 1 + \mathbf{d}$ are appropriately related. For e.g., if 0-Relaxed Rule of Priority, then it says that the service time of all nodes with priority p is earlier than that of all nodes with priority $p + 1$. Constraints (C.8) corresponds to the integrality and non-negativity constraints for the decision variables.

$$\text{Min} \quad \sum_{i \in \mathbf{N}} \sum_{j \in \mathbf{N}} (\mathbf{L}_{ij} x_{ij}) \quad (\text{C.1})$$

$$\sum_{j \in \mathbf{N}} x_{ij} = 1 \quad \forall i \in \mathbf{N} \quad (\text{C.2})$$

$$\sum_{i \in \mathbf{N}} x_{ij} = 1 \quad \forall j \in \mathbf{N} \quad (\text{C.3})$$

$$s_1 = 0 \quad (\text{C.4})$$

$$s_i + \mathbf{T}_{ij} + \mathbf{UT}_i - M(1 - x_{ij}) \leq s_j \quad \forall i \in \mathbf{N}, j \in \bar{\mathbf{N}}_1 \quad (\text{C.5})$$

$$s_i + \mathbf{T}_{ij} + \mathbf{UT}_i + M(1 - x_{ij}) \geq s_j \quad \forall i \in \mathbf{N}, j \in \bar{\mathbf{N}}_1 \quad (\text{C.6})$$

$$s_i + \mathbf{T}_{ij} + \mathbf{UT}_i \leq s_j \quad \forall p \in \mathbf{P}, i \in \mathbf{N}_p, j \in \mathbf{N}_{p+1+\mathbf{d}} \quad (\text{C.7})$$

$$x_{ij} \in \{0, 1\}, s_i \geq 0 \quad \forall i, j \in \mathbf{N} \quad (\text{C.8})$$

C.2.1 Proof for Proposition 1

Proof We will prove this by contradiction. Assume that there are $n_s > 1$ sub-tours in the optimal solution to Single Uncapacitated VRP with Priorities. This translates into n_s directed cycles in the graph. One sub-tour should contain the depot and hence that sub-tour is valid. Let us consider any sub-tour without the depot. Call it $x_1, x_2, \dots, x_k, x_1$. Assume that the unloading times are zero (this result holds even when the unloading times assume positive values) and let us write down the binding service time constraints for any node x_r in this sub-tour.

$$s_{x_r} = s_{x_{r-1}} + \mathbf{T}_{x_{r-1}x_r} = s_{x_1} + \mathbf{T}_{x_1x_2} + \sum_{i=2}^{r-1} (s_{x_i} + \mathbf{T}_{x_i x_{i+1}})$$

Writing down the constraint for x_1 , we obtain the following:

$$\begin{aligned} s_{x_1} &= s_{x_k} + \mathbf{T}_{x_k x_1} = s_{x_1} + \mathbf{T}_{x_1 x_2} + \sum_{i=2}^{k-1} (s_{x_i} + \mathbf{T}_{x_i x_{i+1}}) \\ \Rightarrow \mathbf{T}_{x_1 x_2} + \sum_{i=2}^{k-1} (s_{x_i} + \mathbf{T}_{x_i x_{i+1}}) &= 0 \end{aligned}$$

The above expression on the left hand side consists of positive and non-negative terms; thus it can never vanish. Hence it is not possible to have sub-tours in the optimal solution when we have service time constraints. •

C.2.2 Proof for Proposition 2

Proof We will prove this by contradiction. Assume triangle inequality holds true for T_{ij} and consider any three nodes a, b, c on the optimal route as shown in Figure C.1. To enforce 0–Relaxed Priority Rule, all we need the constraint to say is that the service time for node c is higher than a . However, we would like to strengthen the inequality by saying that service time at node c is not just higher but specifically higher by amount $T_{ac} + UT_a$. Let us see if this will alter the optimal route $a \rightarrow b \rightarrow c$. If we had 0–Relaxed Priority Rule for this network, then we

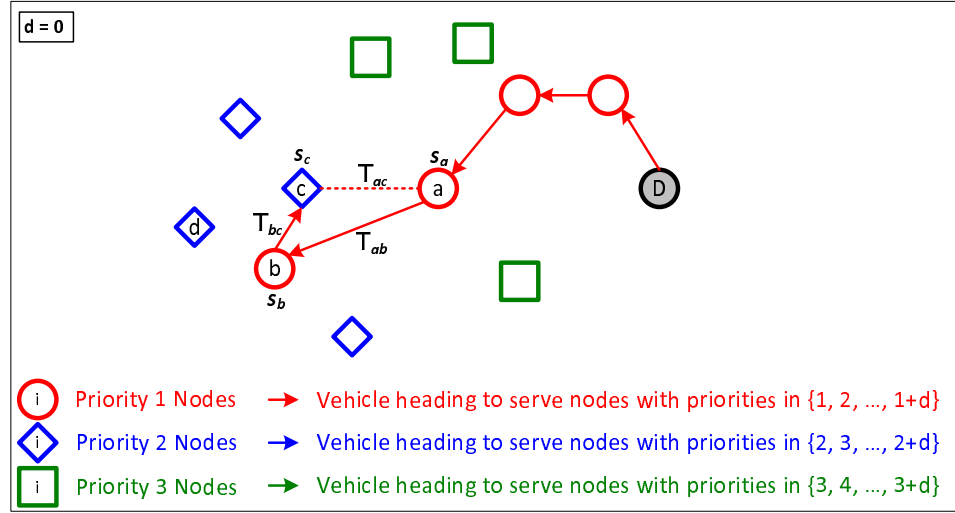


Figure C.1: Proof for d–Relaxed Priority Rule Constraint

write the following constraint for nodes a , b , and c :

$$s_a + T_{ac} + UT_a \leq s_c \text{ and } s_b + T_{bc} + UT_b \leq s_c$$

Since $a \rightarrow b \rightarrow c$ is the optimal route, the second constraint is binding. The first constraint says that s_c is higher than $s_a + T_{ac} + UT_a$. If this constraint forces the vehicle to travel to another node d from b , then we are altering the optimal route. If this must happen then s_c using path $a \rightarrow c$ is *higher than* s_c using optimal path $a \rightarrow b \rightarrow c$. That is,

$$\begin{aligned} s_a + T_{ac} + UT_a &> s_b + T_{bc} + UT_b \\ s_a + T_{ac} + UT_a &> (s_a + T_{ab} + UT_a) + T_{bc} + UT_b \\ (T_{ac}) + UT_a &> (T_{ab} + T_{bc}) + UT_a + UT_b \\ \Rightarrow T_{ac} &> T_{ab} + T_{bc} \quad \because 0 \leq UT_a \leq UT_a + UT_b \end{aligned}$$

This clearly violates the triangle inequality, which is stated as: $T_{ac} \leq T_{ab} + T_{bc}$. Hence this constraint does not alter the route in optimality as long as triangle inequality holds true for travel times. Note that this constraint is a stronger form of $s_i \leq s_j$. •

C.2.3 Valid Inequalities

We add constraints (C.9) - (C.12) so that running time might be reduced. Constraint (C.9) says that the vehicle can never travel directly from a priority q node j to a priority p node i , when $q \geq p + 1 + d$. For example, when $d = 0$, then it says a vehicle can never travel directly from nodes of priorities 2, 3, and so on, to any priority 1 node. Constraint (C.10) says that vehicle will travel at most once from a priority p node i to a priority q node j when $q = p + 1 + d$. For example, when $d = 0$, the vehicle will travel at most once from a priority 1 node to a priority 2 node. Constraint (C.11) extends the constraint (C.10). It says that a vehicle can never travel directly from a priority p node i to node j of priority strictly greater than $p + d + 1$ because otherwise the vehicle would violate the d -Relaxed Priority Rule. If $d = 0$, the vehicle will not travel from any priority 1 node to any node of priority 3, 4, and so on. Constraint (C.12) says that *all* priority p nodes are serviced before *any* node in priority $p + d + 1$ can be serviced.

$$\sum_{i \in N_p} \sum_{j \in N_q} x_{ji} = 0 \quad \forall (p, q) \in P^2 \ni q - p \geq d + 1 \quad (C.9)$$

$$\sum_{i \in N_p} \sum_{j \in N_q} x_{ij} \leq 1 \quad \forall (p, q) \in P^2 \ni q - p = d + 1 \quad (C.10)$$

$$\sum_{i \in N_p} \sum_{j \in N_q} x_{ij} = 0 \quad \forall (p, q) \in P^2 \ni q - p \geq d + 2 \quad (C.11)$$

$$\sum_{i \in N} x_{ij} \geq y_p, \sum_{i \in N} x_{ik} \leq y_p, y_p \in \{0, 1\} \quad \forall j \in N_p, k \in N_{p+d+1}, p \in \{1, \dots, P - d - 1\} \quad (C.12)$$

C.3. Formulation MinMax and MinSum for u-HVRP

For MinMax and MinSum objectives as mentioned in Campbell et al. (2008) paper, we write down MIP formulations MinMax and MinSum. Notice that there are no priorities for nodes in these two formulations. These are just standard VRP

and thus the formulations are self-explanatory.

$$MinMax : \quad \text{Min} \quad s^{max} \quad (C.13)$$

$$\sum_{j \in \mathbf{N}} x_{ij} = 1 \quad \forall i \in \mathbf{N} \quad (C.14)$$

$$\sum_{i \in \mathbf{N}} x_{ij} = 1 \quad \forall j \in \mathbf{N} \quad (C.15)$$

$$s_1 = 0 \quad (C.16)$$

$$s_i + \mathbf{T}_{ij} + \mathbf{UT}_i - M(1 - x_{ij}) \leq s_j \quad \forall i \in \mathbf{N}, j \in \bar{\mathbf{N}}_1 \quad (C.17)$$

$$s_i + \mathbf{T}_{ij} + \mathbf{UT}_i + M(1 - x_{ij}) \geq s_j \quad \forall i \in \mathbf{N}, j \in \bar{\mathbf{N}}_1 \quad (C.18)$$

$$s_i \leq s^{max} \quad \forall i \in \mathbf{N}_1, \quad (C.19)$$

$$x_{ij} \in \{0, 1\}, s^{max} \geq 0, s_i \geq 0 \quad \forall i, j \in \mathbf{N} \quad (C.20)$$

$$MinSum : \quad \text{Min} \quad \sum_{i \in \mathbf{N}} (s_i) \quad (C.21)$$

$$\sum_{j \in \mathbf{N}} x_{ij} = 1 \quad \forall i \in \mathbf{N} \quad (C.22)$$

$$\sum_{i \in \mathbf{N}} x_{ij} = 1 \quad \forall j \in \mathbf{N} \quad (C.23)$$

$$s_1 = 0 \quad (C.24)$$

$$s_i + \mathbf{T}_{ij} + \mathbf{UT}_i - M(1 - x_{ij}) \leq s_j \quad \forall i \in \mathbf{N}, j \in \bar{\mathbf{N}}_1 \quad (C.25)$$

$$s_i + \mathbf{T}_{ij} + \mathbf{UT}_i + M(1 - x_{ij}) \geq s_j \quad \forall i \in \mathbf{N}, j \in \bar{\mathbf{N}}_1 \quad (C.26)$$

$$x_{ij} \in \{0, 1\}, s_i \geq 0 \quad \forall i, j \in \mathbf{N} \quad (C.27)$$

Appendix D

Appendix for Essay 4

D.1. Notation for all models

We define the common notation¹ that will be used for MIP formulations as follows:

Sets:

N = Set of nodes i in the network. E.g. $\{1, 2, \dots, N\}$

$\bar{N}_i = N - \{i\}$. That is, set of nodes in the network except i .

E = Set of edges $i - j$ in the network. E.g. $\{1 - 2, \dots, E\}$

P = Set of priorities p for all nodes. E.g. $\{1, 2, \dots, P\}$

N_p = Set of nodes i in the network with priority p . E.g. $\{5, 7, 8, 10\}$

Parameters:

L_{ij} = Shortest path distance between nodes i and j

T_{ij} = Travel time corresponding to shortest path between nodes i and j

UT_i = Time of unloading required at node i

L_i^{eq} = Distance equivalent of unloading time at node i

P_i = Priority for node i

D_i = Demand of product at node i

D.2. Appendix for 1-HVRP

Before we write the MIP formulations, we define the following notation, in addition to the definitions in Appendix D.1.

Parameters:

CAP = Capacity of the vehicle

L^{max} = Route-length restriction for the vehicle

Decision Variables:

$$v_i = \begin{cases} 1 & \text{if vehicle delivers demand to node } i \\ 0 & \text{otherwise} \end{cases}$$

$$z_p = \begin{cases} 1 & \text{if vehicle is able to meet } all \text{ of demand for nodes in } N_p \\ 0 & \text{otherwise} \end{cases}$$

¹In this paper, please note that we use different typeface to differentiate parameters or given data and *decision variables* in the MIP formulations

Decision Variables (contd.):

$$v_i^{1*} = \begin{cases} 1 & \text{if vehicle delivers demand to node } i \text{ in Stage I} \\ 0 & \text{otherwise} \end{cases}$$

$$x_{ij} = \begin{cases} 1 & \text{if vehicle uses shortest path from node } i \text{ to node } j \\ 0 & \text{otherwise} \end{cases}$$

$$s_i = \text{Time at which service is rendered by vehicle at node } i$$

D.2.1 Formulation 1-HVRP-MaxDemand

$$\text{Max} \quad \sum_{i \in \bar{N}_1} (D_i v_i) \quad (\text{D.1})$$

$$\sum_{j \in \bar{N}_1} x_{1j} = 1, \quad \sum_{i \in \bar{N}_1} x_{i1} = 1 \quad (\text{D.2})$$

$$\sum_{j \in \bar{N}_i} x_{ij} \leq 1 \quad \forall i \in \bar{N}_1 \quad (\text{D.3})$$

$$\sum_{i \in \bar{N}_j} x_{ij} \leq 1 \quad \forall j \in \bar{N}_1 \quad (\text{D.4})$$

$$\sum_{j \in \bar{N}_i} x_{ji} = \sum_{j \in \bar{N}_i} x_{ij} \quad \forall i \in \bar{N}_1 \quad (\text{D.5})$$

$$s_1 = 0 \quad (\text{D.6})$$

$$s_i + T_{ij} + UT_i - M(1 - x_{ij}) \leq s_j \quad \forall i \in N, j \in \bar{N}_1 \quad (\text{D.7})$$

$$s_i + T_{ij} + UT_i + M(1 - x_{ij}) \geq s_j \quad \forall i \in N, j \in \bar{N}_1 \quad (\text{D.8})$$

$$s_i \leq s_j \quad \forall p \in P, i \in N_p, j \in N_{p+1+d} \quad (\text{D.9})$$

$$\sum_{i \in N_1} D_i v_i \leq \text{CAP} \quad (\text{D.10})$$

$$\sum_{i \in N_p} D_i v_i \leq \max \left(0, \text{CAP} - \sum_{r=2}^{p-1} \sum_{i \in N_r} D_i \right) \quad \forall p \in P - \{1\} \quad (\text{D.11})$$

$$v_i = \sum_{j \in N} x_{ji} \quad \forall i \in \bar{N}_1 \quad (\text{D.12})$$

$$\sum_{i \in N_p} D_i v_i \geq z_p \sum_{i \in N_p} D_i \quad \forall p \in P \quad (\text{D.13})$$

$$\sum_{i \in N_p} D_i v_i \leq z_{p-1} \sum_{i \in N_p} D_i \quad \forall p \in P - \{1\} \quad (\text{D.14})$$

$$\sum_{i-j \in E} \left((L_{ij} + L_i^{eq}) x_{ij} \right) \leq L^{max} \quad (\text{D.15})$$

$$x_{ij} \in \{0, 1\}, s_i \geq 0 \quad \forall i, j \in N \quad (\text{D.16})$$

$$v_i \in \{0, 1\} \quad \forall i \in \bar{N}_1 \quad (\text{D.17})$$

$$z_p \in \{0, 1\} \quad \forall p \in P \quad (\text{D.18})$$

The objective function denoted by equation (D.1) is to maximize the total demand delivered while adhering to **d**–Relaxed Rule of Priority, capacity and order of demand constraints. Constraint (D.2) says the vehicle leaves depot only once and returns back to depot exactly once. Constraint (D.3) and (D.4) say that vehicle has to enter and leave any node (other than depot) at most once. These constraints do not assume that the vehicle has to visit all nodes as we have demand, capacity and route-length constraints in this formulation. Constraint (D.5) says that if the vehicle enters a node then it also has to leave that node. We do not need flow conservation constraints for the depot as it is already enforced via constraint (D.2). Constraints (D.6) through (D.8) are the service time calculation constraints just like in the uncapacitated version of this problem. Constraint (D.9) enforces the **d**–Relaxed Rule of Priority, similar to the constraint we had in the uncapacitated version of this problem. Essentially constraints (D.10) and (D.11) talk about the available vehicle’s capacity for servicing nodes with priority p . Constraint (D.10) says that the total capacity available to meet demand for all nodes with priority 1 is just the capacity of the vehicle. Similarly, constraint (D.11) says that the capacity available for nodes with priority $p > 1$ is what is leftover (calculated using the $\max()$ function in the equation) after deducting demand for all nodes with higher priorities. Constraints (D.12) through (D.14) refer to the partial demand fulfillment by the vehicle for different goods. Constraint (D.12) says that a node cannot have its demand satisfied until the vehicle arrives to that node. It also says that a node cannot receive more than its demand requirement. Constraints (D.13) and (D.14) enforce the order in which demand is satisfied and thus eliminate the possibility that higher priority nodes are partially fulfilled compared to the lower priority nodes. The binary variable z_p takes on a value of 1 if vehicle is able to meet all of demand of product for nodes with priority p . For e.g. if a vehicle is planning to satisfy partially or fully demand of priority p then it implies that z_{p-1} is 1. This means that total demand of priority p is completely satisfied. Constraint (D.15) says the total time traversed by the vehicle cannot exceed the total route-length restriction for the vehicle. Constraints (D.16) through (D.18) corresponds to the integrality and non-negativity constraints for the decision variables. Similar to the u-HVRP, we do not need to write down sub-tour elimination constraints as the service time constraints rule out the possibility of sub-tours in the optimal solution (shown in Proposition 1 from the previous Chapter).

D.2.2 Formulation 1-HVRP-MinDist-i

$$\text{Min} \quad \sum_{i \in \mathbf{N}} \sum_{j \in \mathbf{N}} (\mathbf{L}_{ij} x_{ij}) \quad (\text{D.19})$$

$$\sum_{j \in \overline{\mathbf{N}}_1} x_{1j} = 1, \quad \sum_{i \in \overline{\mathbf{N}}_1} x_{i1} = 1 \quad (\text{D.20})$$

$$\sum_{j \in \mathbf{N}} x_{ij} = v_i^{1*} \quad \forall i \in \overline{\mathbf{N}}_1 \quad (\text{D.21})$$

$$\sum_{i \in \mathbf{N}} x_{ij} = v_j^{1*} \quad \forall j \in \overline{\mathbf{N}}_1 \quad (\text{D.22})$$

$$\sum_{j \in \mathbf{N}} x_{ji} = \sum_{j \in \mathbf{N}} x_{ij} \quad \forall i \in \overline{\mathbf{N}}_1 \quad (\text{D.23})$$

$$s_1 = 0 \quad (\text{D.24})$$

$$s_i + \mathbf{T}_{ij} + \mathbf{UT}_i - M(1 - x_{ij}) \leq s_j \quad \forall i \in \mathbf{N}, j \in \overline{\mathbf{N}}_1 \quad (\text{D.25})$$

$$s_i + \mathbf{T}_{ij} + \mathbf{UT}_i + M(1 - x_{ij}) \geq s_j \quad \forall i \in \mathbf{N}, j \in \overline{\mathbf{N}}_1 \quad (\text{D.26})$$

$$s_i \leq s_j \quad \forall p \in \mathbf{P}, i \in \mathbf{N}_p, j \in \mathbf{N}_{p+1+d} \quad (\text{D.27})$$

$$x_{ij} \in \{0, 1\}, s_i \geq 0 \quad \forall i, j \in \mathbf{N} \quad (\text{D.28})$$

The objective function given by equation (D.19) minimizes the total distance traversed. Constraint (D.20) enforces the restrictions on the travel in and out of depot. Constraints (D.21) and (D.22) say that those nodes that have been served in Stage I have to be served. Constraint (D.23) enforces the flow conservation, that is, if the vehicle enters a node, then it has to leave the node, except for depot. Constraints (D.24) - (D.26) help in calculation of service time for the nodes served. Constraint (D.27) enforces the \mathbf{d} -Relaxed Rule of Priority. Finally, constraint (D.28) is the non-negativity and integrality of the variables. We do not need to write down demand constraints in this formulation as we indirectly enforced it with the modified assignment constraints. Also, as this solution is obtained from Stage I, we do not need vehicle capacity restrictions.

D.2.3 Formulation 1-HVRP-MinDist-a

The formulation is similar to the 1-HVRP-MinDist-i except that we use assignment constraints (as in 1-HVRP-MinDist) as we are not enforcing demand for each node serviced. Instead, we enforce it, for each priority level using constraint (D.39).

$$\text{Min} \quad \sum_{i \in \mathbf{N}} \sum_{j \in \mathbf{N}} (\mathbf{L}_{ij} x_{ij}) \quad (\text{D.29})$$

$$\sum_{j \in \overline{\mathbf{N}}_1} x_{1j} = 1, \quad \sum_{i \in \overline{\mathbf{N}}_1} x_{i1} = 1 \quad (\text{D.30})$$

$$\sum_{j \in \mathbf{N}} x_{ij} \leq 1 \quad \forall i \in \overline{\mathbf{N}}_1 \quad (\text{D.31})$$

$$\sum_{i \in \mathbf{N}} x_{ij} \leq 1 \quad \forall j \in \overline{\mathbf{N}}_1 \quad (\text{D.32})$$

$$\sum_{j \in \mathbf{N}} x_{ji} = \sum_{j \in \mathbf{N}} x_{ij} \quad \forall i \in \overline{\mathbf{N}}_1 \quad (\text{D.33})$$

$$s_1 = 0 \quad (\text{D.34})$$

$$s_i + \mathbf{T}_{ij} + \mathbf{UT}_i - M(1 - x_{ij}) \leq s_j \quad \forall i \in \mathbf{N}, j \in \overline{\mathbf{N}}_1 \quad (\text{D.35})$$

$$s_i + \mathbf{T}_{ij} + \mathbf{UT}_i + M(1 - x_{ij}) \geq s_j \quad \forall i \in \mathbf{N}, j \in \overline{\mathbf{N}}_1 \quad (\text{D.36})$$

$$s_i \leq s_j \quad \forall p \in \mathbf{P}, i \in \mathbf{N}_p, j \in \mathbf{N}_{p+1+d} \quad (\text{D.37})$$

$$v_i \leq \sum_{j \in \mathbf{N}} x_{ji} \quad \forall i \in \overline{\mathbf{N}}_1 \quad (\text{D.38})$$

$$\sum_{i \in \mathbf{N}_p} \mathbf{D}_i v_i \geq \sum_{i \in \mathbf{N}_p} \mathbf{D}_i v_i^{1*} \quad \forall p \in \mathbf{P} \quad (\text{D.39})$$

$$\sum_{i-j \in \mathbf{E}} \left((\mathbf{L}_{ij} + \mathbf{L}_i^{eq}) x_{ij} \right) \leq \mathbf{L}^{max} \quad (\text{D.40})$$

$$x_{ij} \in \{0, 1\}, s_i \geq 0 \quad \forall i, j \in \mathbf{N} \quad (\text{D.41})$$

$$v_i \in \{0, 1\} \quad \forall i \in \overline{\mathbf{N}}_1 \quad (\text{D.42})$$

D.2.4 Formulation 1-HVRP-WeightedObj

The feasible region is same as to that of formulation 1-HVRP-MaxDemand but the objective function is changed as follows:

$$\text{Max} \quad \alpha \sum_{i \in \overline{\mathbf{N}}_1} (\mathbf{D}_i v_i) - \beta \sum_{i \in \mathbf{N}} \sum_{j \in \mathbf{N}} (\mathbf{L}_{ij} x_{ij}) \quad (\text{D.43})$$

where, $0 \leq \{\alpha, \beta\} \leq 1$.

D.3. Appendix for m-HVRP

We define additional notation apart from the sets and parameters already defined in Appendix D.1.

Sets/Parameters:

K = Set of vehicles k available. E.g. $\{1, 2, \dots, K\}$

CAP^k = Capacity of the vehicle k

$L^{max\ k}$ = Route-length limit for vehicle k

Decision Variables:

$$x_{ij}^k = \begin{cases} 1 & \text{if vehicle } k \text{ uses shortest path from node } i \text{ to node } j \\ 0 & \text{otherwise} \end{cases}$$

s_i^k = Time at which service is rendered by vehicle k at node i

Decision Variables (contd.):

$$v_i^k = \begin{cases} 1 & \text{if vehicle } k \text{ delivers demand to node } i \\ 0 & \text{otherwise} \end{cases}$$

$$z_p = \begin{cases} 1 & \text{if all vehicles are able to meet all of demand for nodes in } N_p \\ 0 & \text{otherwise} \end{cases}$$

$$v_i^{k1*} = \begin{cases} 1 & \text{if vehicle } k \text{ delivers demand to node } i \text{ in Stage I} \\ 0 & \text{otherwise} \end{cases}$$

D.3.1 Formulation for Stage I with Local Timing Rule

In this MIP formulation, m-HVRP-local-MaxDemand, we enforce the Local Timing Rule as well as other restrictions like travel in and out of depot for each vehicle, capacity, Route-Length, Order of Demand Fulfillment etc.

$$\text{Max} \quad \sum_{k \in K} \sum_{i \in \bar{N}_1} (D_i v_i^k) \quad (D.44)$$

$$\sum_{j \in \bar{N}_1} x_{1j}^k = 1, \quad \sum_{i \in \bar{N}_1} x_{i1}^k = 1 \quad \forall k \in K \quad (D.45)$$

$$\sum_{j \in \bar{N}_i} x_{ij}^k \leq 1 \quad \forall i \in \bar{N}_1, k \in K \quad (D.46)$$

$$\sum_{i \in \bar{N}_j} x_{ij}^k \leq 1 \quad \forall j \in \bar{N}_1, k \in K \quad (D.47)$$

$$\sum_{j \in \bar{N}_i} x_{ji}^k = \sum_{j \in \bar{N}_i} x_{ij}^k \quad \forall i \in \bar{N}_1, k \in K \quad (D.48)$$

$$s_1^k = 0 \quad \forall k \in K \quad (D.49)$$

$$s_i^k + T_{ij} + UT_i - M(1 - x_{ij}^k) \leq s_j^k \quad \forall i \in N, j \in \bar{N}_1, k \in K \quad (D.50)$$

$$s_i^k + T_{ij} + UT_i + M(1 - x_{ij}^k) \geq s_j^k \quad \forall i \in N, j \in \bar{N}_1, k \in K \quad (D.51)$$

$$s_i^k \leq s_j^k \quad \forall p \in P, i \in N_p, j \in N_{p+1+d}, k \in K \quad (D.52)$$

$$\sum_{i \in \bar{N}_1} D_i v_i^k \leq CAP^k \quad k \in K \quad (D.53)$$

$$\sum_{k \in K} \sum_{i \in N_1} D_i v_i^k \leq \min \left(\sum_{i \in N_1} D_i, \sum_{k \in K} CAP^k \right) \quad (D.54)$$

$$\sum_{k \in K} \sum_{i \in N_p} D_i v_i^k \leq \max \left(0, \sum_{k \in K} (CAP^k) - \sum_{r=2}^{p-1} \sum_{i \in N_r} D_i \right) \quad \forall p \in P - \{1\} \quad (D.55)$$

$$v_i^k = \sum_{j \in N} x_{ji}^k \quad \forall i \in \bar{N}_1, k \in K \quad (D.56)$$

$$\sum_{k \in K} v_i^k \leq 1 \quad \forall i \in \bar{N}_1 \quad (D.57)$$

$$\sum_{k \in K} \sum_{i \in N_p} D_i v_i^k \geq z_p \sum_{i \in N_p} D_i \quad \forall p \in P \quad (D.58)$$

$$\sum_{k \in K} \sum_{i \in N_p} D_i v_i^k \leq z_{p-1} \sum_{i \in N_p} D_i \quad \forall p \in P - \{1\} \quad (D.59)$$

$$\sum_{i-j \in E} \left((L_{ij} + L_i^{eq \ k}) x_{ij}^k \right) \leq L^{max \ k} \quad \forall k \in K \quad (D.60)$$

$$x_{ij}^k \in \{0, 1\}, s_i^k \geq 0 \quad \forall i, j \in N, k \in K \quad (D.61)$$

$$v_i^k \in \{0, 1\} \quad \forall i \in \bar{N}_1, k \in K \quad (D.62)$$

$$z_p \in \{0, 1\} \quad \forall p \in P, k \in K \quad (D.63)$$

The objective function given by equation (D.44) is to maximize the demand delivered by all vehicles. Constraints (D.45) - (D.47) are the assignment constraints for the depot and general node but enforced for each vehicle separately. Constraint (D.48) is the flow conservation, that is, vehicle k that enters a node has to leave that node (except for the depot). Constraints (D.49) - (D.51) help in calculation of service time for vehicle k at nodes it services. Constraint (D.52) says that for vehicle k , it should service higher priority nodes first compared to lower priority nodes subject to d -Relaxed Rule of Priority. This constraint is local in nature, that is, it enforces this d -Relaxed Rule of Priority for each vehicle separately. Hence, it is possible that vehicle 1's service time for a node with priority p maybe higher than vehicle 2's service time for a node with priority $p + d + 1$ node. Constraint (D.53) says that the total capacity of the vehicle k cannot be exceeded. Constraint (D.54) provides an upper limit on the available capacity for priority 1 nodes. It says the upper limit is the minimum of total capacity or the total demand. Similarly, constraint (D.55)

gives the upper limit on available capacity for that particular priority p . It says that the limit is either the leftover capacity after fully satisfying demand for higher priority nodes or if it cannot, then there is no capacity available for this priority p nodes. Constraint (D.56) says that a node can be serviced only if the vehicle visits that node. Constraint (D.57) says that a particular node can be at most serviced only by one vehicle, that is, there is no split delivery. Constraints (D.58) and (D.59) enforce the Order of Demand Fulfillment restriction for each priority, across all vehicles. Constraint (D.60) is the route-length restriction for each vehicle. Finally, constraints (D.61) - (D.63) enforce the non-negativity and integrality of the decision variables respectively.

D.3.2 Formulations for Stage II with Local Timing Rule

Similar to the Single VRP with Priorities as discussed in Section 5.4.1.1, the route obtained in Stage I may not be efficient in terms of distance. Thus, we write two MIP formulations: **m-HVRP-local-MinDist-i** and **m-HVRP-local-MinDist-a** for individual and aggregate assignment respectively.

Formulation m-HVRP-local-MinDist-i:

$$\text{Min} \quad \sum_{i \in \mathbf{N}} \sum_{j \in \mathbf{N}} (L_{ij} x_{ij}^k) \quad (\text{D.64})$$

$$\sum_{j \in \bar{\mathbf{N}}_1} x_{1j}^k = 1, \quad \sum_{i \in \bar{\mathbf{N}}_1} x_{i1}^k = 1 \quad (\text{D.65})$$

$$\sum_{j \in \mathbf{N}} x_{ij}^k = v_i^{k1*} \quad \forall i \in \bar{\mathbf{N}}_1 \quad (\text{D.66})$$

$$\sum_{i \in \mathbf{N}} x_{ij}^k = v_j^{k1*} \quad \forall j \in \bar{\mathbf{N}}_1 \quad (\text{D.67})$$

$$\sum_{j \in \mathbf{N}} x_{ji}^k = \sum_{j \in \mathbf{N}} x_{ij}^k \quad \forall i \in \bar{\mathbf{N}}_1 \quad (\text{D.68})$$

$$s_1^k = 0 \quad (\text{D.69})$$

$$s_i^k + T_{ij} + UT_i - M(1 - x_{ij}^k) \leq s_j^k \quad \forall i \in \mathbf{N}, j \in \bar{\mathbf{N}}_1 \quad (\text{D.70})$$

$$s_i^k + T_{ij} + UT_i + M(1 - x_{ij}^k) \geq s_j^k \quad \forall i \in \mathbf{N}, j \in \bar{\mathbf{N}}_1 \quad (\text{D.71})$$

$$s_i^k \leq s_j^k \quad \forall p \in \mathbf{P}, i \in \mathbf{N}_p, j \in \mathbf{N}_{p+1+d} \quad (\text{D.72})$$

$$x_{ij}^k \in \{0, 1\}, s_i^k \geq 0 \quad \forall i, j \in \mathbf{N} \quad (\text{D.73})$$

Formulation m-HVRP-local-MinDist-a:

$$\text{Min} \quad \sum_{i \in \mathbf{N}} \sum_{j \in \mathbf{N}} (\mathbf{L}_{ij} x_{ij}^k) \quad (\text{D.74})$$

$$\sum_{j \in \bar{\mathbf{N}}_1} x_{1j}^k = 1, \quad \sum_{i \in \bar{\mathbf{N}}_1} x_{i1} = 1 \quad (\text{D.75})$$

$$\sum_{j \in \mathbf{N}} x_{ij}^k \leq 1 \quad \forall i \in \bar{\mathbf{N}}_1 \quad (\text{D.76})$$

$$\sum_{i \in \mathbf{N}} x_{ij}^k \leq 1 \quad \forall j \in \bar{\mathbf{N}}_1 \quad (\text{D.77})$$

$$\sum_{j \in \mathbf{N}} x_{ji}^k = \sum_{j \in \mathbf{N}} x_{ij}^k \quad \forall i \in \bar{\mathbf{N}}_1 \quad (\text{D.78})$$

$$s_1^k = 0 \quad (\text{D.79})$$

$$s_i^k + \mathbf{T}_{ij} + \mathbf{UT}_i - M(1 - x_{ij}^k) \leq s_j^k \quad \forall i \in \mathbf{N}, j \in \bar{\mathbf{N}}_1 \quad (\text{D.80})$$

$$s_i^k + \mathbf{T}_{ij} + \mathbf{UT}_i + M(1 - x_{ij}^k) \geq s_j^k \quad \forall i \in \mathbf{N}, j \in \bar{\mathbf{N}}_1 \quad (\text{D.81})$$

$$s_i^k \leq s_j^k \quad \forall p \in \mathbf{P}, i \in \mathbf{N}_p, j \in \mathbf{N}_{p+1+d} \quad (\text{D.82})$$

$$v_i^k \leq \sum_{j \in \mathbf{N}} x_{ji}^k \quad \forall i \in \bar{\mathbf{N}}_1 \quad (\text{D.83})$$

$$\sum_{i \in \mathbf{N}_p} \mathbf{D}_i v_i^k \geq \sum_{i \in \mathbf{N}_p} \mathbf{D}_i v_i^{k1*} \quad \forall p \in \mathbf{P} \quad (\text{D.84})$$

$$\sum_{i-j \in \mathbf{E}} ((\mathbf{L}_{ij} + \mathbf{L}_i^{eq}) x_{ij}^k) \leq \mathbf{L}^{max} \quad (\text{D.85})$$

$$x_{ij}^k \in \{0, 1\}, s_i^k \geq 0 \quad \forall i, j \in \mathbf{N} \quad (\text{D.86})$$

$$v_i^k \in \{0, 1\} \quad \forall i \in \bar{\mathbf{N}}_1 \quad (\text{D.87})$$

D.3.3 Formulation for Stage I with Global Timing Rule

In this MIP formulation, m-HVRP-global-MaxDemand, we enforce the Global Timing Rule version of **d**–Relaxed Priority Rule and other restrictions as mentioned in m-HVRP-local-MaxDemand.

$$\text{Max} \quad \sum_{k \in K} \sum_{i \in \bar{N}_1} (D_i v_i^k) \quad (\text{D.88})$$

$$\sum_{j \in \bar{N}_1} x_{1j}^k = 1, \quad \sum_{i \in \bar{N}_1} x_{i1}^k = 1 \quad \forall k \in K \quad (\text{D.89})$$

$$\sum_{j \in \bar{N}_i} x_{ij}^k \leq 1 \quad \forall i \in \bar{N}_1, k \in K \quad (\text{D.90})$$

$$\sum_{i \in \bar{N}_j} x_{ij}^k \leq 1 \quad \forall j \in \bar{N}_1, k \in K \quad (\text{D.91})$$

$$\sum_{j \in \bar{N}_i} x_{ji}^k = \sum_{j \in \bar{N}_i} x_{ij}^k \quad \forall i \in \bar{N}_1, k \in K \quad (\text{D.92})$$

$$s_1^k = 0 \quad \forall k \in K \quad (\text{D.93})$$

$$s_i^k + T_{ij} + UT_i - M(1 - x_{ij}^k) \leq s_j^k \quad \forall i \in N, j \in \bar{N}_1, k \in K \quad (\text{D.94})$$

$$s_i^k + T_{ij} + UT_i + M(1 - x_{ij}^k) \geq s_j^k \quad \forall i \in N, j \in \bar{N}_1, k \in K \quad (\text{D.95})$$

$$s_i^k \leq s_j^k \quad \forall p \in P, i \in N_p, j \in N_{p+1+d}, k \in K \quad (\text{D.96})$$

$$s_i^k \leq s_p^{max} \quad \forall p \in P, i \in N_p, k \in K \quad (\text{D.97})$$

$$s_p^{max} \leq s_j^k \quad \forall p \in P, j \in N_{p+1+d}, k \in K \quad (\text{D.98})$$

$$\sum_{i \in \bar{N}_1} D_i v_i^k \leq CAP^k \quad k \in K \quad (\text{D.99})$$

$$\sum_{k \in K} \sum_{i \in N_1} D_i v_i^k \leq \min \left(\sum_{i \in N_1} D_i, \sum_{k \in K} CAP^k \right) \quad (\text{D.100})$$

$$\sum_{k \in K} \sum_{i \in N_p} D_i v_i^k \leq \max \left(0, \sum_{k \in K} (CAP^k) - \sum_{r=2}^{p-1} \sum_{i \in N_r} D_i \right) \quad \forall p \in P - \{1\} \quad (\text{D.101})$$

$$v_i^k = \sum_{j \in N} x_{ji}^k \quad \forall i \in \bar{N}_1, k \in K \quad (\text{D.102})$$

$$\sum_{k \in K} v_i^k \leq 1 \quad \forall i \in \bar{N}_1 \quad (\text{D.103})$$

$$\sum_{k \in K} \sum_{i \in N_p} D_i v_i^k \geq z_p \sum_{i \in N_p} D_i \quad \forall p \in P \quad (\text{D.104})$$

$$\sum_{k \in K} \sum_{i \in N_p} D_i v_i^k \leq z_{p-1} \sum_{i \in N_p} D_i \quad \forall p \in P - \{1\} \quad (\text{D.105})$$

$$\sum_{i-j \in E} \left((L_{ij} + L_i^{eq \ k}) x_{ij}^k \right) \leq L^{max \ k} \quad \forall k \in K \quad (\text{D.106})$$

$$x_{ij}^k \in \{0, 1\}, s_i^k \geq 0 \quad \forall i, j \in \mathbf{N}, k \in \mathbf{K} \quad (\text{D.107})$$

$$v_i^k \in \{0, 1\} \quad \forall i \in \bar{\mathbf{N}}_1, k \in \mathbf{K} \quad (\text{D.108})$$

$$z_p \in \{0, 1\} \quad \forall p \in \mathbf{P}, k \in \mathbf{K} \quad (\text{D.109})$$

D.3.4 Formulations for Stage II with Global Timing Rule

Similar to the Local Timing Rule, we write two MIP formulations: **m-HVRP-global-MinDist-i** and **m-HVRP-global-MinDist-a** for individual and aggregate assignment respectively.

Formulation m-HVRP-global-MinDist-i:

$$\text{Min} \quad \sum_{i \in \mathbf{N}} \sum_{j \in \mathbf{N}} (\mathbf{L}_{ij} x_{ij}^k) \quad (\text{D.110})$$

$$\sum_{j \in \bar{\mathbf{N}}_1} x_{1j}^k = 1, \quad \sum_{i \in \bar{\mathbf{N}}_1} x_{i1}^k = 1 \quad (\text{D.111})$$

$$\sum_{j \in \mathbf{N}} x_{ij}^k = v_i^{k1*} \quad \forall i \in \bar{\mathbf{N}}_1 \quad (\text{D.112})$$

$$\sum_{i \in \mathbf{N}} x_{ij}^k = v_j^{k1*} \quad \forall j \in \bar{\mathbf{N}}_1 \quad (\text{D.113})$$

$$\sum_{j \in \mathbf{N}} x_{ji}^k = \sum_{j \in \mathbf{N}} x_{ij}^k \quad \forall i \in \bar{\mathbf{N}}_1 \quad (\text{D.114})$$

$$s_1^k = 0 \quad (\text{D.115})$$

$$s_i^k + \mathbf{T}_{ij} + \mathbf{UT}_i - M(1 - x_{ij}^k) \leq s_j^k \quad \forall i \in \mathbf{N}, j \in \bar{\mathbf{N}}_1 \quad (\text{D.116})$$

$$s_i^k + \mathbf{T}_{ij} + \mathbf{UT}_i + M(1 - x_{ij}^k) \geq s_j^k \quad \forall i \in \mathbf{N}, j \in \bar{\mathbf{N}}_1 \quad (\text{D.117})$$

$$s_i^k \leq s_j^k \quad \forall p \in \mathbf{P}, i \in \mathbf{N}_p, j \in \mathbf{N}_{p+1+d}, k \in \mathbf{K} \quad (\text{D.118})$$

$$s_i^k \leq s_p^{max} \quad \forall p \in \mathbf{P}, i \in \mathbf{N}_p, k \in \mathbf{K} \quad (\text{D.119})$$

$$s_p^{max} \leq s_j^k \quad \forall p \in \mathbf{P}, j \in \mathbf{N}_{p+1+d}, k \in \mathbf{K} \quad (\text{D.120})$$

$$x_{ij}^k \in \{0, 1\}, s_i^k \geq 0 \quad \forall i, j \in \mathbf{N} \quad (\text{D.121})$$

Formulation m-HVRP-global-MinDist-a:

$$\text{Min} \quad \sum_{i \in \mathbf{N}} \sum_{j \in \mathbf{N}} (\mathbf{L}_{ij} x_{ij}^k) \quad (\text{D.122})$$

$$\sum_{j \in \bar{\mathbf{N}}_1} x_{1j}^k = 1, \quad \sum_{i \in \bar{\mathbf{N}}_1} x_{i1} = 1 \quad (\text{D.123})$$

$$\sum_{j \in \mathbf{N}} x_{ij}^k \leq 1 \quad \forall i \in \bar{\mathbf{N}}_1 \quad (\text{D.124})$$

$$\sum_{i \in \mathbf{N}} x_{ij}^k \leq 1 \quad \forall j \in \bar{\mathbf{N}}_1 \quad (\text{D.125})$$

$$\sum_{j \in \mathbf{N}} x_{ji}^k = \sum_{j \in \mathbf{N}} x_{ij}^k \quad \forall i \in \bar{\mathbf{N}}_1 \quad (\text{D.126})$$

$$s_1^k = 0 \quad (\text{D.127})$$

$$s_i^k + \mathbf{T}_{ij} + \mathbf{UT}_i - M(1 - x_{ij}^k) \leq s_j^k \quad \forall i \in \mathbf{N}, j \in \bar{\mathbf{N}}_1 \quad (\text{D.128})$$

$$s_i^k + \mathbf{T}_{ij} + \mathbf{UT}_i + M(1 - x_{ij}^k) \geq s_j^k \quad \forall i \in \mathbf{N}, j \in \bar{\mathbf{N}}_1 \quad (\text{D.129})$$

$$s_i^k \leq s_j^k \quad \forall p \in \mathbf{P}, i \in \mathbf{N}_p, j \in \mathbf{N}_{p+1+d}, k \in \mathbf{K} \quad (\text{D.130})$$

$$s_i^k \leq s_p^{max} \quad \forall p \in \mathbf{P}, i \in \mathbf{N}_p, k \in \mathbf{K} \quad (\text{D.131})$$

$$s_p^{max} \leq s_j^k \quad \forall p \in \mathbf{P}, j \in \mathbf{N}_{p+1+d}, k \in \mathbf{K} \quad (\text{D.132})$$

$$v_i^k \leq \sum_{j \in \mathbf{N}} x_{ji}^k \quad \forall i \in \bar{\mathbf{N}}_1 \quad (\text{D.133})$$

$$\sum_{i \in \mathbf{N}_p} \mathbf{D}_i v_i^k \geq \sum_{i \in \mathbf{N}_p} \mathbf{D}_i v_i^{k1*} \quad \forall p \in \mathbf{P} \quad (\text{D.134})$$

$$\sum_{i-j \in \mathbf{E}} (\mathbf{L}_{ij} + \mathbf{L}_i^{eq}) x_{ij}^k \leq \mathbf{L}^{max} \quad (\text{D.135})$$

$$x_{ij}^k \in \{0, 1\}, s_i^k \geq 0 \quad \forall i, j \in \mathbf{N} \quad (\text{D.136})$$

$$v_i^k \in \{0, 1\} \quad \forall i \in \bar{\mathbf{N}}_1 \quad (\text{D.137})$$

Appendix E
Networks used in Essays 3 and 4

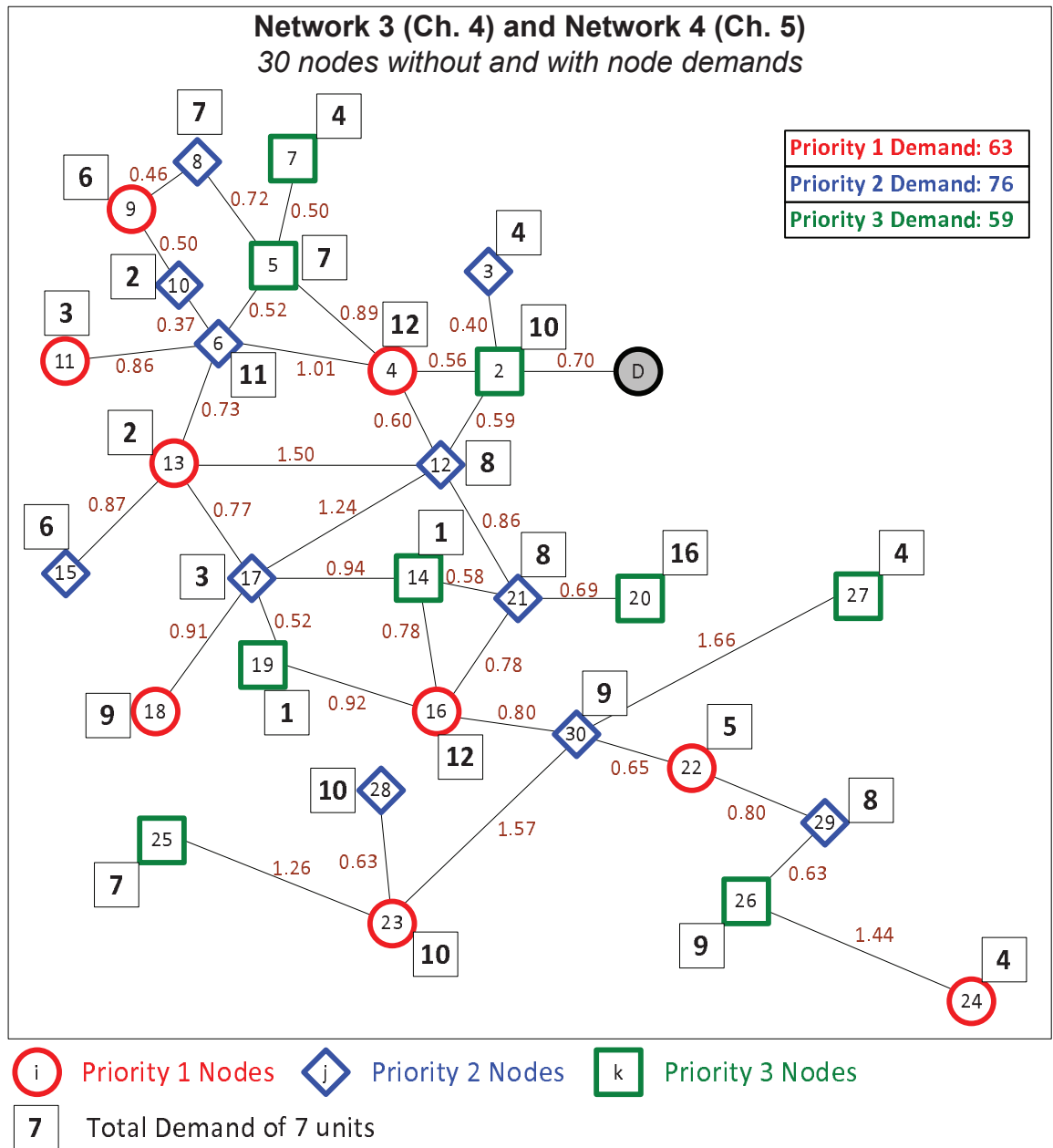


Figure E.1: 30-node network used in Chapters 4 and 5

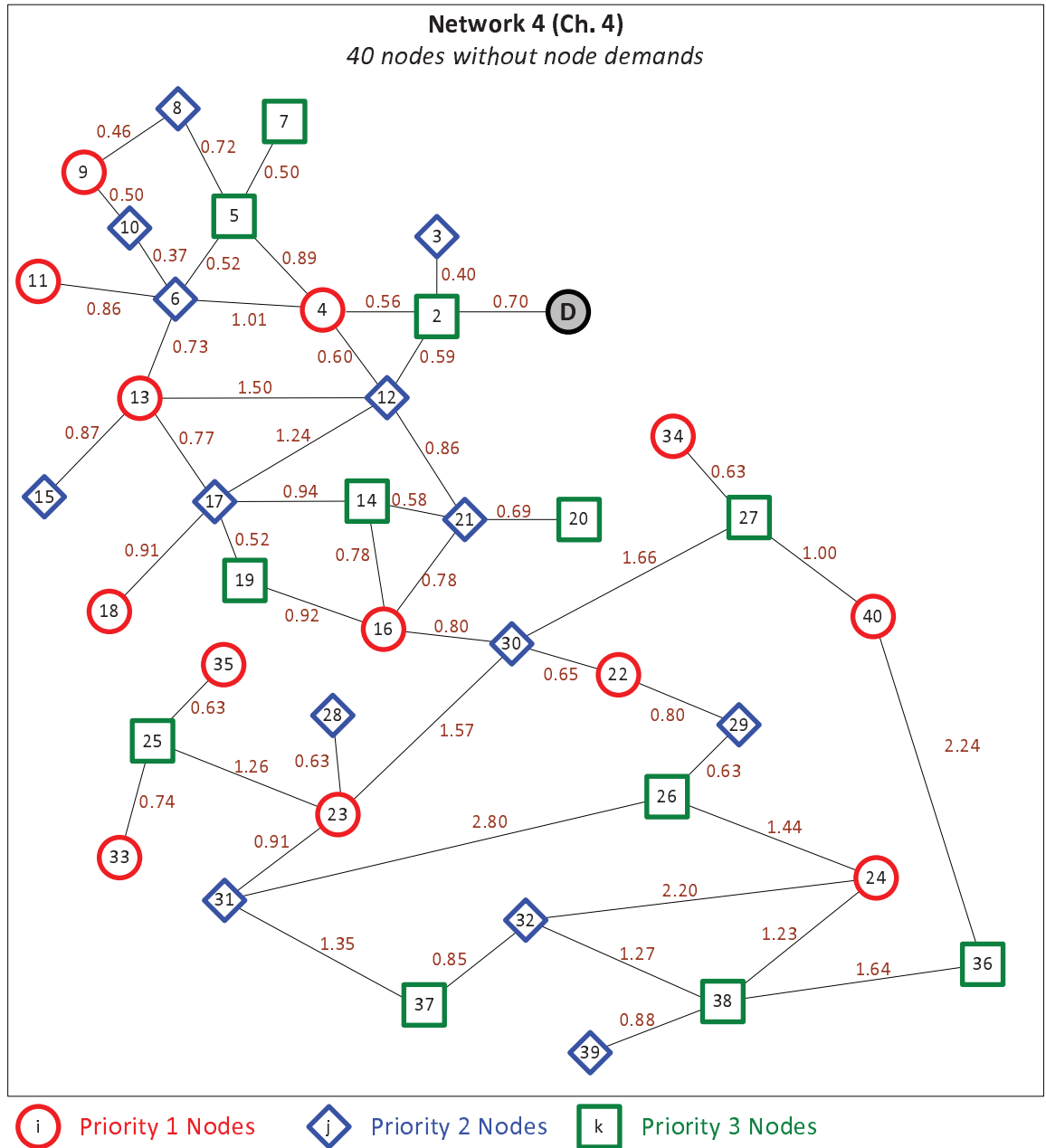


Figure E.2: 40-node network used in Chapter 4

Bibliography

- Agins, T. (2006), ‘Boss Talk: Fashion Redesigner; Liz Claiborne CEO Charron Says the Apparel World Needs Less ‘Art’, More ‘Business’’, *Wall Street Journal* .
- Balcik, B. & Beamon, B. M. (2008), ‘Facility location in humanitarian relief’, *International Journal of Logistics* **11**(2), 101–121.
- Balcik, B., Beamon, B. M. & Smilowitz, K. (2008), ‘Last Mile Distribution in Humanitarian Relief’, *Journal of Intelligent Transportation Systems* **12**(2), 51–63.
- Bazaraa, M. S., Sherali, H. D. & Shetty, C. M. (2006), *Nonlinear Programming: Theory and Algorithms*, 3rd edn, John Wiley & Sons, Inc., New Jersey, USA.
- Beamon, B. M. & Balcik, B. (2008), ‘Performance measurement in humanitarian relief chains’, *International Journal of Public Sector Management* **21**(1), 4–25.
- Birchall, J. (2010), ‘Low-cost stores expand as bigger shops falter’, <http://www.ft.com/cms/s/0/3c785144-f732-11df-9b06-00144feab49a.html#axzz16PPCIV1L>. [Online; accessed 26-November-2010].
- Bitran, G. & Caldentey, R. (2003), ‘Overview of pricing models for revenue management’, *msom* **5**(3), 203.
- Bitran, G., Caldentey, R. & Mondschein, S. (1998), ‘Coordinating clearance markdown sales of seasonal products in retail chains’, *Operations Research* **46**(5), 609–624.
- Bitran, G. R. & Mondschein, S. V. (1997), ‘Periodic Pricing of Seasonal Products in Retailing’, *Management Science* **43**(1), 64–79.
- Boussier, S., Feillet, D. & Gendreau, M. (2007), ‘An Exact Algorithm for Team Orienteering Problems’, *Operations Research* **5**, 211–230.
- Bräysy, O. & Gendreau, M. (2005a), ‘Vehicle Routing Problem with Time Windows, Part I: Route Construction and Local Search Algorithms’, *Transportation Science* **39**(1), 104–118.
- Bräysy, O. & Gendreau, M. (2005b), ‘Vehicle Routing Problem with Time Windows, Part II: Metaheuristics’, *Transportation Science* **39**(1), 104–118.
- Buffa, E. S. (1961), *Modern Production Management*, Wiley, New York, NY.
- Byron, E. (2005), ‘Saks Sued Again For Withholding Markdown Money’, *Wall Street Journal* .
- Cachon, G. P. (2003), Chapter 6: Supply Chain Coordination with Contracts, in ‘Handbooks in Operations Research and Management Science: Supply Chain Management’, North-Holland.

- Campbell, A. M., Vandenbussche, D. & Hermann, W. (2008), ‘Routing for Relief Efforts’, *Transportation Science* **42**(2), 127–145.
- Chiu, Y.-C. & Zheng, H. (2007), ‘Real-time mobilization decisions for multi-priority emergency response resources and evacuation groups: Model formulation and solution’, *Transportation Research Part E: Logistics and Transportation Review* **43**(6), 710 – 736.
- Coughlan, A. T. & Soberman, D. A. (2004), ‘A Survey of Outlet Mall Retailing: Past, Present, and Future.’, *INSEAD Working Paper Series* .
- Cuadra, G. Q. (2001), *Humanitarian Supply Management and Logistics in the Health Sector*, Pan American Health Organization and World Health Organization.
- Davidson, A. L. (2006), Key performance indicators in humanitarian logistics, Master’s thesis, Massachusetts Institute of Technology.
- Desrochers, M. & Laporte, G. (1991), ‘Improvements to the millertuckerzemlin sub-tour elimination constraints’, *Operations Research Letter* **10**, 27–369.
- Drake, M. J., Duran, S., Griffin, P. M. & Swann, J. L. (2008), ‘Optimal Timing of Switches Between Product Sales for Sports and Entertainment Tickets’, *Naval Research Logistics* **55**(1), 59–75.
- Economist, T. (Jan 2005), ‘Quality over quantity’, *Economist* .
- Elmaghraby, W. J. & Keskinocak, P. (2003), ‘Dynamic Pricing in the Presence of Inventory Considerations’, *Managment Science* **49**(10), 1287–1309.
- Erhun, F., Gonçalves, P. & Hopman, J. (2007), ‘The art of managing new product transitions’, *MIT Sloan Management Review* .
- Feng, Y. & Gallego, G. (1995), ‘Optimal Starting Times for End-of-Season Sales and Optimal Stopping Times for Promotional Fares’, *Managment Science* **41**(8), 1371–1391.
- Feng, Y. & Gallego, G. (2000), ‘Perishable Asset Revenue Management with Markovian Time dependent demand intensities’, *Managment Science* **46**.
- Fritz Institute, . (2006), ‘Hurricane Katrina: Perceptions of the Affected’, Survey Report.
- Ghose, A., Teland, R. & Krishnan, R. (2005), ‘Effect of Electronic Secondary Markets on the Supply Chain’, *Journal of Management Information Systems* **22**(2), 91–120.
- Glasserman, P. & Tayur, S. (1995), ‘Sensitivity analysis for base-stock levels in multiechelon production-inventory systems’, *Managment Science* **41**, 263–281.
- Gottlieb, R. (2005), ‘Markdown Money: It’s just a bad idea’, *Playthings* .

- Gouveia, L. & Pires, J. M. (1999), ‘The asymmetric travelling salesman problem and a reformulation of the miller-tucker-zemlin constraints’, *European Journal of Operations Research* **112**(1), 134 – 146.
- Grewal, D. & Levy, M. (2007), ‘Retailing research: Past, present, and future’, *Journal of Retailing* **83**(4), 447–464.
- Grewal, D. & Levy, M. (2009), ‘Emerging issues in retailing research’, *Journal of Retailing* **85**(4), 522–526.
- Hentenryck, P. V., Bent, R. & Coffrin, C. (2010), *Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems*, Vol. 6140 of *Lecture Notes in Computer Science*, Springer Berlin/Heidelberg, chapter Strategic Planning for Disaster Recovery with Stochastic Last Mile Distribution, pp. 318–333.
- Hugos, M. H. & Thomas, C. (2005), The Retail Distribution Channel, in ‘Supply Chain Management in the Retail Industry’, Wiley, chapter 2, pp. 31–42.
- Insighter (2006), ‘Welfare for the Retail Class’, *Insighter*.
- Jotshi, A., Gong, Q. & Batta, R. (2009), ‘Dispatching and routing of emergency vehicles in disaster mitigation using data fusion’, *Socio-Economic Planning Sciences* **43**(3), 1–24.
- Khouja, M. & Robbins, S. S. (2005), ‘Optimal pricing and quantity of products with two offerings’, *European Journal of Operational Research* **163**(2), 530 – 544.
- Kleywegt, A. J. & Shapiro, A. (2001), *Handbook of Industrial Engineering: Technology and Operations Management, Third Edition*, 3rd edn, John Wiley & Sons, Ltd, New York, chapter 102-Stochastic Optimization, pp. 2625 – 2650.
- Kök, A. G., Fisher, M. L. & Vaidyanathan, R. (2009), Assortment planning: Review of literature and industry practice, in N. Agrawal & S. A. Smith, eds, ‘Retail Supply Chain Management’, Vol. 122 of *International Series in Operations Research Management Science*, Springer US, pp. 1–55.
- Kovács, G. & Spens, K. M. (2007), ‘Humanitarian logistics in disaster relief operations’, *International Journal of Physical Distribution & Logistics Management* **37**(2), 99–114.
- Krishnan, T. V. & Soni, H. (1997), ‘Guaranteed Profit Margins: Demonstration of Retailer Power’, *International Journal of Research in Marketing* **14**, 35–56.
- Krishnan, V. & Ulrich, K. T. (2001), ‘Product development decisions: A review of the literature’, *Management Science* **47**(1), 1–21.
- Lariviere, M. A. (1998), Supply Chain Contracting and Coordination with Stochastic Demand, in S. Tayur, H. Magazine & R. Ganeshan, eds, ‘Quantitative Models of Supply Chain Management’, Kluwer Academic Publishers, pp. 233–268.

- Lee, C. H. & Rhee, B.-D. (2008), ‘Optimal guaranteed profit margins for both vendors and retailers in the fashion apparel industry’, *Journal of Retailing* **84**(3), 325–333.
- Lee, H. & Whang, S. (2002), ‘The Impact of the Secondary Market on the Supply Chain’, *Management Science* **48**(6), 719–731.
- Lehmann, D. R. & Weinberg, C. B. (2000), ‘Sales through sequential distribution channels: An application to movies and videos’, *Journal of Marketing* **64**, 18–33.
- Liu, D., Han, J. & Zhu, J. (2007), *Optimization and Systems Biology*, Vol. 7 of *Lecture Notes in Operations Research*, World Publishing Corporation China, chapter Vehicle Routing for Medical Supplies in Large-Scale Emergencies, pp. 412–419.
- Mantrala, M. K., Basuroy, S. & Gajanan, S. (2005), ‘Do Style-Goods Retailers’ Demands for Guaranteed Profit Margins Unfairly Exploit Vendors?’, *Marketing Letters* **16**(1), 53–66.
- Miller, C. E., Tucker, A. W. & Zemlin, R. A. (1960), ‘Integer programming formulation of travelling salesman problems’, *Journal of ACM* **7**, 326–329.
- National Retail Federation (2010), ‘Large retail sales and profitability’, http://www.nrf.com/modules.php?name=Pages&sp_id=1256. [Online; accessed 26-November-2010].
- Ngueveu, S. U., Prins, C. & Calvo, R. W. (2010), ‘An effective memetic algorithm for the cumulative capacitated vehicle routing problem’, *Computers & Operations Research* **37**(11), 1877 – 1885.
- Nocke, V. & Peitz, M. (2003), ‘Hyperbolic discounting and secondary markets’, *Games and Economic Behavior* **44**(1), 77–97.
- Nolz, P. C., Doerner, K. F., Gutjahr, W. J. & Hartl, R. F. (2010), *Advances in Multi-Objective Nature Inspired Computing*, Vol. 272 of *Studies in Computational Intelligence*, Springer Berlin/Heidelberg, chapter A Bi-objective Metaheuristic for Disaster Relief Operation Planning, pp. 167–187.
- Pannirselvam, G. P., Ferguson, L. A., Ash, R. C. & Siferd, S. P. (1999), ‘Operations management research: an update for the 1990s’, *Journal of Operations Management* **18**(1), 95–112.
- Pasternack, B. A. (1985), ‘Optimal Pricing and Return Policies for Perishable Commodities’, *Marketing Science* **4**(2), 166–176.
- Petruzzi, N. C. & Dada, M. (1999), ‘Pricing and the Newsvendor Problem: Review with Extensions’, *Operations Research* **47**(2), 183–194.
- Petruzzi, N. C. & Monahan, G. E. (2003), ‘Managing Fashion Goods Inventories: Dynamic Recourse for Retailers with Outlet Stores’, *IIE Transactions* **35**, 1033–1047.

- Russell, T. E. (2005), The humanitarian relief supply chain: Analysis of the 2004 south east asia earthquake and tsunami, Master's thesis, Massachusetts Institute of Technology.
- Sa Vinhas, A., Chatterjee, S., Dutta, S., Fein, A., Lajos, J., Neslin, S., Scheer, L., Ross, W. & Wang, Q. (2010), 'Channel design, coordination, and performance: Future research directions', *Marketing Letters* **21**, 223–237.
- Sanders, R. L. (2010), 'Nordstrom rack opens at arrowhead crossing in peoria', <http://www.azcentral.com/community/peoria/articles/2010/11/11/20101111peoria-nordstrom-rack-opens-arrowhead-crossing.html#ixzz16Pt83UZq>. [Online; accessed 26-November-2010].
- SAS (2007), Competing as a 21st century retailer: Maximizing profitability in a consumer-centric world, Technical report, SAS Institute Inc.
- Schmitz, H. & Niemann, S. (2009), *Metaheuristics in the Service Industry*, Vol. 624 of *Lecture Notes in Economics and Mathematical Systems*, Springer Berlin/Heidelberg, chapter A Bicriteria Traveling Salesman Problem with Sequence Priorities, pp. 1–14.
- Shen, Z., Dessouky, M. M. & Ordóñez, F. (2009), 'A two-stage vehicle routing model for large-scale bioterrorism emergencies', *Networks* **54**(4), 255–269.
- Shen, Z., Ordóñez, F. & Dessouky, M. M. (2009), *Optimization and Logistics Challenges in the Enterprise*, Vol. 30 of *Springer Optimization and Its Applications*, Springer US, chapter The Stochastic Vehicle Routing Problem for Minimum Unmet Demand, pp. 349–371.
- Sherali, H. D. & Driscoll, P. J. (2002), 'On tightening the relaxations of miller-tucker-zemlin formulations for asymmetric traveling salesman problems', *Operations Research* **50**(4), 656–669.
- Shetty, V. K., Sudit, M. & Nagi, R. (2008), 'Priority-based Assignment and Routing of a Fleet of Unmanned Combat Aerial Vehicles', *Computers & Operations Research* **35**(6), 1813–1828.
- Sheu, J.-B. (2007a), 'Challenges of emergency logistics management', *Transportation Research Part E: Logistics and Transportation Review* **43**(6), 655 – 659.
- Sheu, J.-B. (2007b), 'An emergency logistics distribution approach for quick response to urgent relief demand in disasters', *Transportation Research Part E: Logistics and Transportation Review* **43**(6), 687 – 709.
- Sheu, J.-B. (2010), 'Dynamic relief-demand management for emergency logistics operations under large-scale disasters', *Transportation Research Part E* **46**(1), 1–17.

- Singhal, K., Singhal, J. & Starr, M. K. (2007), ‘The domain of production and operations management and the role of elwood buffa in its delineation’, *Journal of Operations Management* **25**, 310–327.
- Smith, S. L., Pavone, M., Bullo, F. & Frazzoli, E. (2010), ‘Dynamic vehicle routing with priority classes of stochastic demands’, *SIAM Journal on Control and Optimization* **48**, 3224–3245.
- Sundaram, R. K. (1996), *A First Course in Optimization Theory*, Cambridge University Press, Cambridge, United Kingdom.
- Talluri, K. T. & van Ryzin, G. J. (2004), *The Theory and Practice of Revenue Management*, Kluwer Academic Publishers, Boston.
- Thomas, A. S. & Fritz, L. (2006), ‘Disaster Relief, Inc.’, *Harvard Business Review* .
- Thomas, A. S. & Kopczak, L. R. (2005), From Logistics to Supply Chain Management: The Path Forward in the Humanitarian Sector, Technical report, Fritz Institute.
- Tibben-Lembke, R. S. (2004), ‘Strategic Use of the Secondary Market for Retail Consumer Goods.’, *California Management Review* **46**(2), 90–104.
- Tomasini, R. & van Wassenhove, L. (2009), *Humanitarian Logistics*, Palgrave Macmillan.
- Tsay, A. (2001), ‘Managing Retail Channel Overstock: Markdown Money and Return Policies’, *Journal of Retailing* **47**(4), 418–420.
- Tzeng, G.-H., Cheng, H.-J. & Huang, T. D. (2007), ‘Multi-objective optimal planning for designing relief delivery systems’, *Transportation Research Part E: Logistics and Transportation Review* **43**(6), 673 – 686.
- U. S. Census Bureau (2010), Annual Retail Trade survey - 2008, Technical report, U. S. Census Bureau.
- Van Zandt, T. (2002), ‘An introduction to monotone comparative statics’, <http://faculty.insead.edu/vanzandt/teaching/CompStatics.pdf>. [Online; accessed 26-November-2010].
- Vives, X. (2000), *Oligopoly Pricing: Old Ideas and New Tools*, MIT Press, Cambridge, MA.
- Wahba, P. (2010), ‘TJX move upscale may blur lines in high-end retail’, <http://uk.reuters.com/article/idUKTRE61066A20100201>. [Online; accessed 26-November-2010].
- Yang, J., Xia, Y. & Qi, X. (2007), A nonatomic-game model for timing clearance sales under competition. Working paper.

Yi, W. & Kumar, A. (2007), ‘Ant colony optimization for disaster relief operations’,
Transportation Research Part E: Logistics and Transportation Review **43**(6), 660
– 672.