APPLICATIONS OF MACHINE LEARNING IN REVENUE MANAGEMENT AND ROUTING

by

GUYVES ACHTARI

A thesis submitted to the
Smith School Of Business
in conformity with the requirements for
the degree of Doctor in Philosophy

Queen’s University
Kingston, Ontario, Canada
September 2018

Copyright © Guyves Achtari, 2018
Abstract

In this thesis, I use machine learning techniques to solve issues in revenue management and public transportation design. The first two chapters relate to problems of revenue management and online learning. The problem of sequential learning and optimization of the demand function has been an important topic in revenue management. Finding the optimal policy faces numerical complexity and is prone to the curse of dimensionality. It is mostly solved using heuristics and restrictive assumptions. In the first chapter, I use a novel non-parametric approach to solving dynamic pricing and learning problems. I develop a flexible method to approximate the optimal policy using polynomial approximation, thus reducing complexity. I make use of the Bayesian framework to update the probability model and make advances in numerical methods to solve this problem. In the second chapter, I use a machine learning heuristic called Thompson sampling. I improve the performance of the heuristic over short horizons by enforcing the decreasing nature of the demand function in the sampling algorithm. Using a stylized proof, I demonstrate the performance gains associated with this method and show the merits of ordered sampling with Thompson Sampling over short horizons. The last chapter makes use of a machine learning approach called data envelopment analysis (DEA), which I use in designing new public transportation routes in rural regions. I develop algorithms and heuristics to balance cost and equity
under multiple objectives. The solution to this project was implemented in the city of Quinte West, Ontario.
Co-Authorship

Chapters 1 and 2 are joint work with Mikhail Nediak and Yuri Levin. Chapter 3 titled “Balancing equity and cost in rural transportation management with multi-objective utility analysis and data envelopment analysis: A case of Quinte West” has been published in Transportation Research Part A: Policy and Practice 95 (2017): 148-165 and is a joint work with Chialin Chen. This study was supported by Economic Developers Council of Ontario (EDCO) and Social Sciences and Humanities Research Council of Canada (SSHRC). We would like to express our sincerely thanks to the following people who helped us complete and implement this study: Ms. Shelly Ackers, Administrator of Quinte Access, Ms. Nadine Mattis, Operations Officer of Quinte Access, Mr. John Williams, Mayor of City of Quinte West, Mr. Craig Desjardins of Economic Developers Council of Ontario, Mr. Bob Wannamaker, City Councilor of Quinte West, Mr. Charlie Murphy, Chief Administrative Officer of Quinte West, and Mr. Steve Whitehead, GIS Supervisor of Quinte West.
Acknowledgments

I would like to take this opportunity to thank my supervisors Mikhail Nediak and Yuri Levin for their support, understanding, and patience with me over the years. Without your guidance, this journey would not have been possible. Mikhail, thank you for all the time and effort you’ve invested in me. I would like to also thank Jeff McGill for his help and advice in the early days of this journey. May you rest in peace Jeff. It was an honor to have you as my adviser and teacher. Special thanks to Chialin Chen for his help and support over the years. Thank you for getting me involved with your projects, and allowing me to work with you. It was an insightful experience. I would like to thank everyone at the graduate office, past and present, for all their help and support over the years; Annette, Teresa, Nancy, Lisa, Jay, Yolande and anyone else who I’ve possibly missed on this list, thank you.

My friends and family are the ones who have supported me emotionally throughout this process. Without you, this journey would have been very different. Royah, thank you for all your support. I’m very fortunate to have you by my side. Myla, I’m sorry for my lack of availability over the last couple of months. I’m looking forward to spending more time together. I would like to thank my parents for always supporting my decisions, and for being there for me. I would like to thank Roozbeh Yousefi and Jue Wang for your friendship and generosity. The time you spent reviewing and
commenting on my work was invaluable. And finally, this section is not complete without me thanking all my friends who have supported me in some way or another during this period.
# Contents

Abstract ................................................................. i

Co-Authorship .......................................................... iii

Acknowledgments ......................................................... iv

Contents ................................................................. vi

List of Tables ........................................................... ix

List of Figures .......................................................... x

Chapter 1: Introduction ............................................... 1

Chapter 2: A Non-Parametric Approach to Dynamic Pricing with Demand Learning 4

2.1 Abstract ............................................................ 4

2.2 Introduction ....................................................... 5

2.3 Literature Review .................................................... 7

2.4 Model Formulation ................................................... 9

2.4.1 Learning framework ............................................. 11

2.5 Model heuristic ..................................................... 16

2.6 Fast approximations ............................................... 25

2.6.1 Chebyshev polynomial ........................................ 27

2.6.2 Negative binomial .............................................. 29

2.7 Simulation Setup ................................................... 29

2.7.1 Variance Reduction ............................................. 30

2.8 Numerical Results .................................................. 32

2.9 Conclusion .......................................................... 35

2.10 Summary of variables ............................................. 36
Appendix C: Balancing Equity and Cost in Rural Transportation Management with Multi-Objective Utility Analysis and Data Envelopment Analysis: A Case of Quinte West

C.1 Flow Diagrams for Data Gathering and Analysis .......................... 133
C.2 Flow Diagrams for Model Development ......................................... 135
C.3 Flow Diagrams for Route Generation and Testing .......................... 136
List of Tables

2.1 Numerical methods for different prior types . . . . . . . . . . . . . . 27
2.2 Table of variables and notation . . . . . . . . . . . . . . . . . . . . 37
3.1 Table of variables and notation . . . . . . . . . . . . . . . . . . . . 69
4.1 MOUA Regression Results . . . . . . . . . . . . . . . . . . . . . . . . 98
4.2 Summary of Research Outcomes . . . . . . . . . . . . . . . . . . . . 101
List of Figures

2.1 Non-parametric representation of demand ......................... 5
2.2 Comparison of upper bounds generated by uninformative priors using
5 prices ................................................................. 33
2.3 Simulation of the regret function with the Lagrangian relaxation heuris-
tics, and informative priors ......................................... 34
2.4 Simulation of the regret function with the Lagrangian relaxation heuris-
tics, and uninformative priors ...................................... 35

3.1 Demand learning and pricing for a single product with Thompson Sam-
pling derived from [37] ................................................. 43
3.2 Importance sampling with replacement for Thompson Sampling .... 46
3.3 Metropolis-Hastings algorithm with parallel chains ................ 49
3.4 Comparison of the [37] example with ordered Beta distribution using
Metropolis-Hastings for short horizon T=10 ......................... 50
3.5 Thompson Sampling for pricing application with uniform prior ..... 53
3.6 Comparison of sampling and integration regions for both dependent
and independent distributions ....................................... 55
3.7 Comparison of the expected number of pulls, with uniform prior ... 60
3.8 Comparison of $\text{BayesRegret}_{dep}(t, 2)$ (red) and $\text{Bayesregret}_{ind}(t, 2)$ (black) with 2 arms, and no capacity constraint. .................. 63
3.9 Comparison of the regret function, with uniform prior, and different values of $\mu_1$ ........................................ 64
3.10 Comparison of dependent (green) and independent (blue) constrained expected regret function, with uniform prior ............... 67

4.1 The Built-in GIS Data and Utility/Accessibility Scores of a Grid Cell 75
4.2 An Example of Original and Alternative Routes ..................... 90
4.3 A Summary of the Proposed Methodology ............................ 93
4.4 Old Bus Routes in Quinte West ............................ 95
4.5 Sample Questionnaire Used to Determine Utility Values in MOUA 97
4.6 Three New Routes Shown in Different Colors with Google Earth ... 100
4.7 The Efficient Frontier and the Old, New, and Shortest Routes ....... 103
Chapter 1

Introduction

My research evolves around machine learning applications that are subject to resource constraints. The first two chapters focus on solving sequential demand learning and pricing optimization problems over short horizons. The objective is to find near-optimal heuristics with low complexity, by leveraging the decreasing nature of the demand curve. In the first chapter, I introduce a dynamic program with a Bayesian updating model that simultaneously explores and exploits the demand function for a set of admissible discrete prices. The Bayesian updating model is an essential component of this research, as it sequentially updates the joint probability distribution of demand for all prices. It is worth mentioning that the numerical evaluation related to the Bayesian model is non-trivial, and that to the best of my knowledge, the methodology presented to solve it is both novel and efficient.

Finding the optimal policy to sequential learning and pricing models is a difficult task, subject to the curse of dimensionality. In this chapter, I propose a novel approach by approximating the upper-bound of the value function with a polynomial. I use simulations to show the performance of the upper-bound heuristic relative to the optimal solution.
The problem presented in the first chapter is a sequential exploration and exploitation problem, also known as the multi-armed bandit (MAB) in the computer science literature. Thompson Sampling is a computationally efficient heuristic, solving this class of problems by combining randomized sampling with Bayesian updating. The algorithm is asymptotically optimal, which doesn’t make it suitable for short horizon applications. In the second chapter, I modify the sampling procedure in Thompson Sampling by enforcing an ordering rule on the sample set. This modification reduces the number of trials and errors in the algorithm. I use the Metropolis-Hastings algorithm with a Dirichlet kernel to enforce sample ordering after each iteration. In a 2-arm stylized model, I show the performance gap between using ordered and independent samples. Short horizons with small discounts in the order of 5% to 20% benefit the most from the ordered sampling.

In the third chapter, I use Data Envelopment Analysis (DEA) in designing new public transportation routes in rural regions. These regions are under tremendous financial constraints while their public transportation systems stretch over long distances, serving an aging population. It is imperative that the design of the new routes considers the tradeoff between the costs and benefits associated with them. I conceptualize and define equity in rural transportation management with the development of new performance measures and an analytical model for decision making with multiple desirable and undesirable objectives. I also develop a heuristic procedure based on data envelopment analysis for characterizing and analyzing the route design choices on the frontier between costs and multi-objective measures of equity. The new methodology has been successfully implemented by Quinte Access, a not-for-profit organization in a rural community in Ontario, to help redesign bus routes
with significant quantitative benefits observed in multiple performance dimensions including the total ridership, the total multi-objective utility score, total population covered, and rural accessibility score per minute of bus service time, respectively.
Chapter 2

A Non-Parametric Approach to Dynamic Pricing with Demand Learning

2.1 Abstract

We consider a monopolist selling a capacity constrained product over a finite time horizon. The seller learns and earns simultaneously while assuming a non-parametric demand function and choosing from a discrete set of prices. The learning model exploits the decreasing nature of the demand function. The ordering structure of the set of prices and the associated probability of sales applied to the Bayesian updating method, generalizes the incomplete beta function, with the Euler integral as the normalizing constant. We offer numerical advances on two fronts: first, we develop tractable numerical methods to compute the value of the multi-variate Euler integral. Second, given the large state-space of the dynamic program and the obvious curse of dimensionality, we develop an efficient heuristic by approximating the upper-bound of the expected revenue with a polynomial. The heuristic performance is measured by comparing its average regret against that of the optimal policy. The approximation heuristic is numerically tractable, and converges rapidly to the optimal solution.
2.2 Introduction

We propose a non-parametric near-optimal sequential dynamic pricing and demand learning model with a capacity constraint over a short horizon. The model assumes a monopolist selling a single product over a short horizon while demand distribution remains stationary. It is applicable to flash sales of products and services also known as the deal-of-the-day model, as offered by e-commerce platforms such as Amazon, Ideeli and Rue La La. These retailers offer a single product with limited inventory over a short period of time, where the exact response to different discount levels is typically unknown. Given the unknown demand curve and the short horizon, we make use of a non-parametric demand model, flexible enough to accommodate any decreasing demand model as shown in figure 2.1. Given the limited time, separating exploration and exploitation of the demand curve is inefficient and must be performed jointly. A small set of admissible price discounts is planned ahead of time to explore
2.2. INTRODUCTION

pricing and optimize revenue as sales progress.

We aim to find a tractable solution which converges rapidly to the expected revenue derived from the optimal policy when demand distribution is known. Given the rapid convergence rate, our methodology can be extended to scenarios where demand is non-stationary. Using a non-parametric model, demand distribution is estimated by keeping count of sales statistics at different prices over time. Changes in demand distribution can be accounted for by increasing the importance of new sales information over old ones by applying exponential smoothing to the parameters that keep track of sales information.

In this paper, we propose a recursive model that keeps track of the purchasing decisions customers make for a given posted price. Using a non-parametric demand model, a Bayesian approach is used to update the retailer’s belief of the demand distribution after every customer arrival. The proposed model in this paper takes a generalized approach toward the distribution of the willingness-to-pay by exploiting the ordering structure of the set of discrete prices and the associated probability of sales; as prices increase, the expected probability of sales decreases. Our model takes into account the joint probability of sales among the entire set of prices selected by the retailer. This setup has the advantage of updating the belief on the demand distribution for all prices when update occurs. A successful sale at a higher price increases the probability of sale at lower prices. Similarly, a failure to sell at a lower price, decreases the probability of sale at higher prices. This setup leads to the generalization of the beta-binomial distribution, and the use of the Euler integral as the normalizing constant.

This paper’s contributions are as follows:
• Bayesian updating with the generalization of the beta-binomial distribution, and the use of the Euler integral as the normalizing factor.

• Two numerical approaches to compute multi-variate Euler integrals covering a large set of non-integer parameters.

• A new approach to approximate dynamic programming using polynomial approximation to construct an upper bound to the value function.

This paper is organized as follows: Section 2 contains the literature review. Section 3 introduces the discrete pricing model and the associated learning framework. In section 4, we present our heuristic, using polynomial approximation to upper bound the value function. Section 5 offers numerical methods to estimate the Euler integral for different prices. Sections 6 and 7 cover simulation setup and numerical results. Section 8 provides concluding comments.

2.3 Literature Review

Dynamic pricing with demand learning has been an important topic in revenue management with considerable literature on the subject. We refer to [19] and [31] for a general overview on this topic. An important question in the learning literature is about the value of information, how much value can learning the demand for each particular price add to total profits. There are two possible ways to address this question. First, by separating exploration (where price experimentation takes place) and exploitation (where revenue is maximized based on existing information) into two distinct phases, and optimizing how long to explore and experiment with different prices before using the acquired information to optimize revenue, see [14]. However,
this method is not suitable when demand is non-stationary as presented in [60], as exploration stops after a fixed time and does not account for changes in demand function over time. The other shortcoming is due to limitations in inventory or short selling horizons. With limited inventory, there exists a risk of exhausting all inventory before exploitation phase begins, and short selling horizons do not leave much time for exploration.

The second approach requires joint exploration-exploitation. This setting is more complex, and has been widely studied by scientists in different disciplines. A significant body of literature addresses this problem in the framework of multi-armed bandits (MAB). The upper confidence bound (UCB) based policies are popular in addressing this setup, see [7]. Simultaneous dynamic pricing and learning performs better under the simplifying assumption of a parametric model. [32] studies this problem using parametric uncertainty on the demand distribution. [12] uses a parametric family to learn demand distribution, formulates a dynamic program intractable to solve, and resorts to heuristics using state-space reductions to find approximate solutions.

The proposed methodology solves a dynamic program using a non-parametric demand model, exploiting the ordering structure of a discrete set of prices in a Bayesian setting. The ordering requires updates of the joint probability distribution of sales, resulting in updating probability of sales information for the entire set of prices after each sale. The Bayesian updating with joint probability formulation does relate to the Bayesian multi-armed bandit problem with correlated arms as formulated by [85] and [99]. The Bayesian formulation with dynamic programming methods for pricing has been more recently used by [50] studying the performance of myopic Bayesian
2.4. MODEL FORMULATION

We consider a retailer who sells a single perishable product over a finite selling horizon. The retailer chooses from a set of $Q$ ordered prices, where the set is defined by $\{p_1, \cdots, p_Q\}$, with each price indexed by $j \in Q$. The retailer does not know the demand distribution associated with each price in the price set. The objective of the retailer is to maximize the expected revenue by choosing among the set of prices, subject to an inventory constraint. The selling horizon is divided into $T$ periods with time counted down, i.e., $t = T$ represents the beginning and $t = 0$ represents the end of the selling horizon. For simplicity, we assume an arrival in each period. There are $N$ items for sale at the beginning of the horizon, with $N \leq T$, and no more than one item can be sold during each period. We assume that valuations of the consumers arriving in different periods to be independent and identically distributed (IID). Therefore, the probability that a given price results in a sale remains constant over time and sale events are independent. The sales history across the set of possible prices is recorded in vectors $m_t = (m_{t1}, \cdots, m_{tQ})$ and $n_t = (n_{t1}, \cdots, n_{tQ})$, where $m_{tj}$ stands for the total number of successful sales at price $p_j$ over $t$ periods, and $n_{tj}$ represents the total number of times the product has been offered but not purchased.
at price $p_j$. Due to the IID assumption, all sales transaction information is contained in the information state $(\mathbf{m}_t, \mathbf{n}_t)$. The quantity $m_{tj} + n_{tj}$ accounts for the number of times price $p_j$ has been offered over previous $T - t$ time periods, with $\mathbf{m}_T + \mathbf{n}_T = \mathbf{0}$. Moving forward, we drop the time index from the information state unless needed for clarification.

We define the vector of purchase probabilities $\mathbf{\theta} = (\theta_1, \cdots, \theta_Q)$, where $\theta_j$ is a non-parametric representation of the probability of sale for price $p_j$. The posterior probability of a sale at each price depends on the information vector $(\mathbf{m}, \mathbf{n})$, i.e., the distribution of $\mathbf{\theta}$ is conditional on $(\mathbf{m}, \mathbf{n})$. Inventory at the end of each period is defined as $I_t = N - \sum_{j=1}^{Q} m_j$. We define a dynamic program (DP) maximizing the expected revenue over the discrete set of prices for a given inventory and information state:

$$V_t(\mathbf{m}, \mathbf{n}, I_t) = \max_{j=1, \cdots, Q} \left\{ E[\theta_j | \mathbf{m}, \mathbf{n}](p_j + V_{t-1}(\mathbf{m} + \mathbf{e}_j, \mathbf{n}, I_t - 1)) \right\}$$

$$+ (1 - E[\theta_j | \mathbf{m}, \mathbf{n}] V_{t-1}(\mathbf{m}, \mathbf{n} + \mathbf{e}_j, I_t)$$

where $\mathbf{e}_j$ is the indicator vector with the $j^{th}$ component being 1 and zero in all other components. The terminal values of the value function are defined as $V_0(\mathbf{m}, \mathbf{n}, I_0) = 0$ and $V_t(\mathbf{m}, \mathbf{n}, 0) = 0$. The above DP suffers from the curse of dimensionality and is not tractable as the size of the state space grows proportionally to $(NT)^Q$. 
2.4. MODEL FORMULATION

2.4.1 Learning framework

The retailer does not know the probabilities of a sale across the price set, but learns about its distribution by updating the information state. A Bayesian updating framework is used to represent this learning process. This setup has been presented in the literature on the multi-armed bandit problems and more recently in [37]. The latter study presents a multi-armed bandit learning-based framework that achieves asymptotic optimality as the scale of the problem, including the selling horizon increases. Thus, the focus of the model presented in this paper is to improve learning with small inventory and short selling horizons.

To accelerate learning, we make use of the inherent ordering structure of the demand-price relation. A discrete price set may always be arranged as an ordered set, such that \( p_1 \leq \cdots \leq p_Q \), and our model exploits this structure to improve convergence time. Following nomenclature from revenue management, we define a shut-down price \( p_{Q+1} \) for which the probability of sales is zero. Given the ordered price set, it is natural for all vectors of probabilities of sales to follow the reverse order, i.e., \( 1 \geq \theta_1 \geq \cdots \geq \theta_Q > 0 \). The initial belief of the retailer can be captured by a prior distribution \( \pi(\theta) \) on sales probabilities \( \theta \in \Theta \), where \( \Theta \) is the set of all ordered purchase probability vectors.

This set-up resembles the sequential estimation of the parameters of several different Bernoulli distributions, which would typically be performed with independent Bayesian updating processes using beta-binomial conjugate distributions. However, the problem under consideration differs in that the parameters are ordered; thus, the estimation must be done jointly. Furthermore, the updating cannot be performed with the usual multinomial generalization of the binomial because the outcomes are
2.4. MODEL FORMULATION

not mutually exclusive. Nonetheless, Bayesian updating can be accomplished with the generalization of the beta-binomial estimation procedure.

The basis of the Bayesian updating procedure is the following generalization of the beta prior distribution to multiple dimensions:

$$
\pi(\theta | a, b) = C \prod_{j=1}^{Q} \theta_j^{a_j-1}(1 - \theta_j)^{b_j-1}, \theta \in \Theta (2.2)
$$

where $C$ is a normalizing constant and $a = (a_1, \cdots, a_Q), b = (b_1, \cdots, b_Q)$ are the vectors with strictly positive elements forming the retailer’s initial beliefs $(a, b)$. Historic sales information together with retailer’s intuition on expected revenue for each price may be encoded in the belief vector. Prices for which sales information is not available should be encoded with small fractional numbers, representing uninformative priors. Price certainty should be translated by using parameters with large numbers, as large pairs of $(a_j, b_j)$ tend to concentrate the demand distribution for price index $j$. The belief vector offers flexibility on encoding uninformative priors, with small fractional parameters (with components $a_j < 1, b_j < 1, \forall j \in Q$), and informative priors such as pessimistic priors (with components $a_j < 1, b_j = 1, \forall j \in Q$), optimistic priors (with components $a_j = 1, b_j < 1, \forall j \in Q$), and uniform priors (with components $a_j = b_j = 1, \forall j \in Q$) by changing the relative ratio of the belief vector components.

The expected revenue and the chosen course of actions depend on the initial belief vector the seller holds about consumer demand. Most pricing applications using beta-binomial priors with Thompson sampling or Bayesian updating frameworks use a uniform prior as their initial uninformative belief. It is important to note that a uniform prior is informative, and does not represent the retailer’s lack of information. Uninformative priors are represented by an initial belief vector with non-integer
components smaller than one. Uniform priors are commonly used with beta-binomial distributions, since non-integer values are notoriously difficult to calculate and work with.

The following proposition generalizes the truncated beta-binomial distribution.

**Proposition 1 (Preservation of Conjugacy).** Given the initial belief vector \((a, b)\), and the information vector \((m, n)\), the posterior distribution of the truncated beta-binomial model preserves conjugacy and maintains the same form as (2.2), with updated parameter vector \((a + m, b + n)\).

**Proof:** Immediate, by comparing the forms of the likelihood function for the sales outcomes captured by the information vector \((m, n)\), the prior, and the posterior distribution. The algebraic expressions are identical to those in the standard beta-binomial conjugate family, but here they also hold in the presence of the ordering constraint on the components of \(\theta\). □

**Lemma 2.4.1 (Euler Integral).** The normalizing constant in the Bayesian updating scheme of (2.2), is the reciprocal of the Euler integral:

\[
\frac{1}{C} = I(a, b) = \int_{\theta \in \Theta} \prod_{j=1}^{q} \theta_j^{a_j-1}(1 - \theta_j)^{b_j-1} d\theta_j.
\] (2.3)

**Proof:** Immediate, by a direct evaluation of the integral of (2.2) over \(\Theta\) and setting its value to one. □

The equation (2.3) is a special case of the generalized Euler integral defined in [42]. The variables of integration in this \(Q\)-dimensional integral are the probabilities of sale at each of the prices and the area of integration is determined by their order. In the absence of the ordering structure, the Euler integral would reduce to the
product of beta-functions $\frac{1}{C} = \prod_{i=1}^{Q} B(a_i, b_i)$ (where the beta-function is defined as $B(a_i, b_i) = \int_0^1 \theta^{a_i-1}(1-\theta)^{b_i-1}d\theta$). The ratio of two Euler integrals can be used to compute the expected probability of a sale at a given price index:

**Proposition 2.** The expected probability of a sale at a given price given the initial belief vector $(a, b)$ and information vector $(m, n)$ can be computed as the ratio of two Euler integrals as follows:

$$E[\theta_j | a, b, m, n] = \frac{I(a + m + e_j, b + n)}{I(a + m, b + n)}, \quad j = 1, \ldots, Q. \quad (2.4)$$

**Proof:** The posterior distribution is parametrized by the vector $(a + m, b + n)$ and its normalization constant is reciprocal to $I(a + m, b + n)$. Taking the expected value of $\theta_j$ over the posterior leads to the Euler integral with parameter vector $(a + m + e_j, b + n)$. □

This proposition highlights an important distinction between our approach and Thompson sampling, where the vector of probabilities of sales is sampled from the posterior distribution itself and used to make pricing decisions. In contrast, we employ the expected probability of sales for each price. Example 1 shows how the ordering structure affects the probability of sales.

**Example 1.** A retailer chooses between a set of 2 prices $\{p_l, p_h\}$ to sell a product. Using two independent beta-binomial distributions with uniform priors, each price has a $1/2$ probability of success. A uniform prior on the joint distribution gives a higher probability of sales of $E[\theta_l | a = b = 1] = 2/3$ to the lower price and a lower probability of sales of $E[\theta_h | a = b = 1] = 1/3$ to the higher price. After a sale at $p_h$, the probability of sale at both prices increases to $E[\theta_l | a_l = b_l = b_h = 1, a_h = 2] = 3/4$.
and \( E[\theta | \alpha_t = b_t = b_h = 1, a_h = 2] = 1/2 \) respectively. A sale at the higher price increases the chance of sales at the lower price.

Applying our learning framework to the initial DP (2.1) creates the following:

\[
V_t(a, b, I_t) = \max_{j=1, \ldots, Q} \left\{ \frac{I(a + e_j, b)}{I(a, b)}(p_j + V_{t-1}(a + e_j, b, I_t - 1)) + \frac{I(a, b + e_j)}{I(a, b)}V_{t-1}(a, b + e_j, I_t) \right\}
\]

with the terminal conditions \( V_0(a, b, I_0) = V_t(a, b, 0) = 0 \). This DP formulation “earns and learns” simultaneously by optimally balancing between the marginal value of learning and the economic profit of selling a unit of inventory at a given price. To see this balance, we define \( \Delta V_t(a + e_j, b, I_t) = V_t(a, b, I_t) - V_{t-1}(a + e_j, b, I_t - 1) \) as the opportunity cost of selling at price \( j \). Similarly, we define the opportunity cost of learning at price \( j \) as \( \Delta V_t(a, b + e_j, I_t) = V_t(a, b, I_t) - V_{t-1}(a, b + e_j, I_t) \).

By adding and subtracting \( V_t(a, b, I_t) \) within each term in (2.5) and observing that \( I(a + e_j, b) + I(a, b + e_j) = I(a, b) \), we get

\[
V_t(a, b, I_t) = \max_{j=1, \ldots, Q} \left\{ \frac{I(a + e_j, b)}{I(a, b)}(p_j - \Delta V_t(a + e_j, b, I_t)) - \frac{I(a, b + e_j)}{I(a, b)}\Delta V_t(a, b + e_j, I_t) \right\} + V_t(a, b, I_t).
\]

The first term under the maximization operation corresponds to the expected revenue adjusted by the expected opportunity cost of selling if price \( p_j \) is used. The second term is the expected opportunity cost of learning if price \( p_j \) is used. In the standard dynamic pricing problem under complete information, the dynamic program would...
not contain the second term since the opportunity cost of learning under complete information is zero.

**Lemma 2.4.2.** For any given combination of information vector and inventory, the optimal price balances the expected opportunity cost of learning and the expected economic profit. The optimal price is defined by the following equality:

$$\frac{I(a, b + e_j)}{I(a, b)} \Delta V_t(a, b + e_j, I_t) = \frac{I(a + e_j, b)}{I(a, b)} \left( p_j - \Delta V_t(a + e_j, b, I_t) \right)$$

(2.7)

The proof is immediate by rearranging the terms in equation (2.6).

2.5 Model heuristic

[101] consider approximating value functions for any practical problem an exercise of accuracy and speed. The objective of the heuristic presented in this paper is to find an accurate model solved in polynomial time. The use of upper-bounds is very common in approximating value functions. Recent work on Thompson sampling such as [37], and variations of Multi-Armed-Bandit problems make use of the Upper-Confidence-Bound (UCB). The heuristic proposed in this paper makes use of a similar upper bound related to a linear program maximizing deterministic revenue subject to the inventory constraint. However, we focus on the Lagrangian representation of this linear program. The summary of the bounding approach is given below.

In the first step, the value function $V_t(a, b, I_t)$ for a given belief vector and inventory level is bounded by its perfect-information (with respect to $\theta$) upper bound obtained as the expectation of the value functions $V_t(I_t|\theta)$ of the dynamic pricing
problems with a horizon of length $t$ and initial inventory $I_t$ and known $\theta$. The expectation is taken over the distribution $\pi(\theta|a,b)$. This leads to an easy fact that $V_t(a,b,I_t) \leq E[V_t(I_t|\theta)|a,b]$.

In the second step, each value function $V_t(I_t|\theta)$ can in turn be bounded using a Lagrangian approach. We denote the corresponding $\theta$-conditional Lagrangian bound $V^L_t(I_t,\mu|\theta)$ where $\mu$ is the Lagrangian multiplier for the capacity constraint. The level of inventory availability in this context is more appropriately measured by the ratio $\nu_t = I_t/t$. In the sequel, we will use $\nu$ as a generic functional argument where the appropriate values of $\nu_t$ for a specific $t,I_t$ can be substituted. In contrast, $\nu_t$ is indeed tied to specific $t,I_t$.

In the third step, we select $\mu = \mu(\theta,\nu)$ in a class of linear functions of $\theta,\nu$ and construct a polynomial $V^{\text{Poly}}_t(\nu_t,\mu(\theta,\nu),\theta)$ so that it is a pointwise upper bound to the Lagrangian bound $V^L_t(I_t,\mu(\theta,I_t/t)|\theta)$ for all possible values of $\theta,\nu_t$.

In the fourth step, we take the expected value of $V^{\text{Poly}}_t(\nu_t,\mu(\theta,\nu_t),\theta)$ over $\theta$ sampled according to the $\pi(\theta|a,b)$. Because of the upper-bounding or equality relations in each step of this process, we obtain the chain of upper bounds

$$V_t(a,b,I_t) \leq E[V_t(I_t|\theta)|a,b] \leq E[V^L_t(I_t,\mu(\theta,I_t/t)|\theta)|a,b] \leq E[V^{\text{Poly}}_t(I_t/t,\mu(\theta,I_t/t),\theta)|a,b].$$

Polynomial approximation-based techniques were first introduced by [10], and since have been used in dynamic programming and pricing by [30] and [29]. The following method uses Lagrangian relaxation similar to what is used in [39] to derive a polynomial upper bound. However, to our knowledge, the methodology used to derive the upper bound for $V_t(a,b,I_t)$ in this paper and its representation as the expected value
of a polynomial are novel and have not been previously documented. The coefficients used in the bound are derived using a linear program in polynomial time.

Since step one of the bounding process is straightforward, we turn our attention to the derivation of the Lagrangian bound of step two. Consider a perfect-information dynamic pricing problem with a horizon of length \( t \) and initial inventory \( I_t \). Any general randomized non-anticipating control policy \( u_\tau \), \( \tau = t, \ldots, 1 \) for this problem can be stipulated in terms of probabilities \( u_{\tau j} \) of using price index \( j \) at time \( \tau \). In other words, we have \( u_{\tau j} \geq 0, j = 1, \ldots, Q \) and \( \sum_{j=1}^{Q} u_{\tau j} = 1, \tau = t, \ldots, 1 \). We can then define binary demand random variable \( D_\tau \) at time \( \tau \) so that its conditional expected value given price index \( j \) used is \( E[D_\tau|j, \theta] = \theta_j \) and, unconditionally on \( j \), \( E[D_\tau|\theta] = \sum_{j=1}^{Q} \theta_j E[u_{\tau j}|\theta] \). Given that \( u_{\tau j} \) depends on the history of the system (prior and sale information), it is a random variable and is measured in terms of its expected value. Similarly, for the revenue \( R_\tau \) at time \( \tau \), \( E[R_\tau|j, \theta] = p_j \theta_j \) and \( E[R_\tau|\theta] = \sum_{j=1}^{Q} p_j \theta_j E[u_{\tau j}|\theta] \). The perfect-information problem and its value function can be stated as

\[
V_t(I_t|\theta) = \max_{u_\tau} \quad E \left[ \sum_{\tau=1}^{t} R_\tau | \theta \right] \tag{2.8}
\]

s.t. \( \sum_{\tau=1}^{t} D_\tau - I_t \leq 0 \) a.s., \( \tag{2.9} \)

where “a.s.” stands for almost surely in the measure-theoretic sense and indicates that the probability of violation of constraint (2.9) is equal to zero. A straightforward Lagrangian relaxation of the problem (2.8)-(2.9) is obtained by subtracting the expected value of the left-hand-side of (2.9) times a non-negative Lagrange multiplier \( \mu \) from the objective. As a side comment, we remark that (2.9) represent constraints
over non-zero probability realizations of the system dynamics and a constant $\mu$ is the simplest possible option. In fact, we could have used an arbitrary system path-dependent form of the Lagrangian multipliers. The resulting Lagrangian relaxation bound is

$$V_{LR}^t(I_t, \mu | \theta) = \max_{\tau} E \left[ \sum_{\tau=1}^{t} R_{\tau} - \mu \left( \sum_{\tau=1}^{t} D_{\tau} - I_t \right) \right] | \theta]. \tag{2.10}$$

**Proposition 3.** For any $\mu \geq 0$, we have $V_t(I_t | \theta) \leq V_{LR}^t(I_t, \mu | \theta) = tL^*(\theta, I_t/t, \mu)$, where

$$L^*(\theta, \nu, \mu) = \max_{j=1, \ldots, Q} \{(p_j - \mu)\theta_j\} + \mu \nu. \tag{2.11}$$

**Proof:** The inequality result is immediate since $\mu \geq 0$ by the standard Lagrangian relaxation reasoning. Indeed, if constraint (2.9) is not violated for any non-zero probability realization of the system dynamics, then $E \left[ \mu \left( \sum_{\tau=1}^{t} D_{\tau} - I_t \right) \right] | \theta] \leq 0$. Thus, optimizing over feasible policies will not lead to a lower value of the Lagrangian than the original objective value. On the other hand, the maximization is performed over a larger set (non-anticipating control policies that are not necessarily feasible). Thus, the maximum value of the Lagrangian is not smaller than the maximum value of the original objective.

To establish the inequality, start by evaluating the expected value in (2.10):

$$E \left[ \sum_{\tau=1}^{t} R_{\tau} - \mu \left( \sum_{\tau=1}^{t} D_{\tau} - I_t \right) \right] | \theta] = E \left[ \sum_{\tau=1}^{t} (R_{\tau} - \mu D_{\tau}) \right] | \theta] + \mu I_t$$

$$= \sum_{\tau=1}^{t} \sum_{j=1}^{Q} (p_j - \mu)\theta_j E[u_{\tau j} | \theta] + \mu I_t.$$

The stationarity of $\theta_j$ implies that one can consider control policies that result in $E[u_{\tau j} | \theta]$ that remain constant over time. Indeed, an arbitrary permutation of time
indices would lead to the same objective value as well as a convex combination of all such permutations with the equal weights. Thus, we denote \( w_j = E[u_{rj} | \theta], \ j = 1, \ldots, Q \) and consider a stationary problem:

\[
\begin{align*}
\max_{\mathbf{w}} & \quad t \left( \sum_{j=1}^{Q} (p_j - \mu) \theta_j w_j + \mu \nu_t \right) \\
\text{s.t.} & \quad \sum_{j=1}^{Q} w_j = 1, \\
& \quad w_j \geq 0, \quad j = 1, \ldots, Q,
\end{align*}
\]

where \( \nu_t = I_t/t \). Since the maximum value of this problem is attained by setting \( w_j = 1 \) for \( j \) that attains the maximum of \( (p_j - \mu) \theta_j \), the equality claim of the proposition holds. \( \square \)

At this point, it is beneficial to contemplate whether it is possible to directly use the expected value of \( V_{t}^{LR}(I_t, \mu | \theta) = tL^*(\theta, I_t/t, \mu) \) over \( \theta \) as an approximation for the value function to build a dynamic pricing heuristic. Unfortunately, this would be a very difficult endeavor due to the piece-wise structure of \( L^*(\theta, I_t/t, \mu) \) which would also become non-linear for any non-trivial dependence of \( \mu \) on \( \theta, \nu \). Therefore, we avoid a direct integration of \( V_{t}^{LR}(I_t, \mu | \theta) \) and instead construct yet another upper bound but in a tractable polynomial form. To this end, we adopt the general notation used in [64] for polynomial representation. Let us define \( \sum_{\alpha} h_{\alpha} x^{\alpha} \), where \( \alpha \) is a non-negative integer vector \( (\alpha_1, \ldots, \alpha_{Q+1}) \in \mathbb{Z}_{+}^{Q+1} \) representing the degree of a polynomial term, \( x \) is the set of variables, \( h_{\alpha} \) are the coefficients of the polynomial, and we use a shorthand notation \( x^{\alpha} = \prod_{j=1}^{Q+1} x_j^{\alpha_j} \) for the respective monomials. To account for the level of inventory over the remaining horizon, we include the variable \( \nu \) into
the polynomial. The full set of variables describing the polynomial is defined as $x = (\theta_1, \cdots, \theta_Q, \nu)$.

**Proposition 4.** 1. Given a set of discrete prices $\{p_j\}_{j=1}^Q$, a vector of sale probabilities $\theta = (\theta_1, \ldots, \theta_Q)$, and $\nu$, the inventory availability over a finite horizon, as well as the Lagrange multiplier $\mu$ that functionally depends on $\theta$ and $\nu$ as a polynomial of up to power $\bar{n}$, it is possible to represent an upper bound on $L^*(\theta, \nu, \mu)$ via a positive polynomial of power up to $\bar{n} + 1$.

2. The ordering constraints of the demand probabilities together with bounds on the inventory can be explicitly used to represent a positive polynomial in a linear program that finds the coefficients of the polynomial resulting in a lowest upper bound (in expected value sense).

**Proof:** For part one, we use the notation introduced before the statement of the proposition and construct polynomial $\sum_\alpha h_\alpha x^\alpha$ to upper bound the optimal Lagrangian $L^*(\theta, \nu, \mu)$ which is a pointwise maximum of the expression $\{(p_j - \mu)\theta_j\} + \mu\nu$ over $j$. The constructed polynomial has to dominate these quantities for every $j$ but we also allow $\mu = \mu(\theta, \nu)$ to exhibit a degree $\bar{n}$ polynomial dependence on $\theta$. This is done to capture a variable level of expected capacity constraint violation penalty represented by $\mu$, depending on the demand distribution. Thus, for all price indices, probabilities of sale, and inventory availability levels, we want to enforce:

$$ (p_j - \mu(\theta))\theta_j + \mu(\theta)\nu \leq \sum_\alpha h_\alpha x^\alpha, \quad j = 1, \ldots, Q, \forall \theta, \nu. \quad (2.12) $$

The free parameters are the polynomial coefficients $h_\alpha$ and the form of the Lagrangian
multiplier \( \mu(\theta, \nu) \). Condition (2.12) effectively stipulates that polynomial

\[
 f(x) = \sum_{\alpha} h_{\alpha} x^\alpha - (p_j - \mu(\theta)) \theta_j - \mu(\theta) \nu \geq 0, \quad \forall \ j = 1, \ldots, Q, \ \theta, \nu. \tag{2.13}
\]

Moreover, if \( \mu(\theta) \) is a \( \bar{n} \)th degree polynomial, we can guarantee that \( f(x) \) is at most \( \bar{n} + 1 \)th degree polynomial as long as \( |\alpha| = \sum_{j=1}^{Q+1} \alpha_j \leq \bar{n} \).

For part two, consider a given information vector \((a, b)\), and the polynomial \( f(x) \) representing an upper bound to the optimal Lagrangian \( L^*(\theta, \nu, \mu) \). The tightest upper bound can be found by minimizing the expectation of the polynomial over the probability of sales \( \theta \) sampled from \( \pi(\theta | a, b) \). There are \( O(Q^{\bar{n}+1}) \) coefficients in this polynomial, so there are finitely many variables and their number is bounded by a polynomial in the size of the input. However, there are uncountably many constraints in (2.13). And the set of relevant \( \theta, \nu \) needs to represent the ordering of \( \theta \) as well as reasonable bounds on \( \nu \).

Fortunately, we can take the ordering structure into account by constructing a new basis for \( f(x) \). We make use of Theorem 2.23 from [64] that uses \( \mathbb{R}[x] \) to denote the set of polynomials over \( x \in \mathbb{R}^n \) and states:

**Theorem 2.5.1.** Let \( g_k(x) \in \mathbb{R}[x] \) be affine for every \( k = 1, \ldots, \bar{m} \) and assume that \( K = \{x \in \mathbb{R}^n : g_k(x) \geq 0, k = 1, \ldots, \bar{m} \} \) is compact with a nonempty interior. If \( f \in \mathbb{R}[x] \) is strictly positive on \( K \) then \( f = \sum_{\beta \in \mathbb{N}^{\bar{m}}} c_\beta g_1^{\beta_1} \cdots g_{\bar{m}}^{\beta_{\bar{m}}} \) for finitely many nonnegative scalars \( c_\beta \).

Since constraints on \( \theta, \nu \) are all linear, this theorem provides a very clear guidance for choosing the basis. Moreover, the problem of enforcing non-negativity is reduced to non-negativity of a finitely many variables \( c_\beta \). Under an appropriate restriction on
the degree of the polynomial, the number of these new variables is also bounded by a polynomial in the size of the input.

Given that $1 \geq \theta_1 \geq \ldots \geq \theta_Q \geq 0$, we define the new basis $g = (1 - \theta_1, \theta_1 - \theta_2, \ldots, \theta_Q, \bar{\nu} - \nu, \nu - \underline{\nu})$, where each component of $g$ is affine by construction. $\bar{\nu}$ and $\underline{\nu}$ provide the upper and lower limits for the inventory constraint $\nu$. Given that $f(x)$ is strictly positive by design, we can change the basis of the polynomial as per the theorem above and impose the following equality constraint $f(x) = \sum_{\beta} c_{\beta} g^{\beta}$. This constraint reduces to a finite collection of equality constraints between polynomial terms of matching degrees (and the size of the collection is bounded by a polynomial in the size of the input).

Thus, in summary, we obtain the linear program with variables defined by the (unbounded) coefficients $h_{\alpha}$, nonnegative coefficients $c_{\beta}$, and nonnegative coefficients for a similar representation for $\mu(\theta, \nu) \geq 0$. The constraints of this linear program are equalities describing identities between coefficients of the polynomial terms of the same degree (in particular, $f(x) = \sum_{\beta} c_{\beta} g^{\beta}$ and a similar set for alternative representation of $\mu(\theta, \nu) \geq 0$ if required). The objective is to construct the tightest (expected value) upper bound $E[f(x)|a, b]$ where the expected value is taken over $\theta$ sampled from $\pi(\theta|a, b)$. In terms of our discussion at the beginning of this section $V^\text{Poly}_t(\nu_t, \mu(\theta, \nu_t), \theta) = tf(x)$ where we substitute $\nu = \nu_t$ into $f(x)$.

We provide the example of constraints corresponding to second degree polynomials in Appendix A. In practice, such constraints are best found using software packages for simplifying symbolic algebraic expressions. □

The LP formulation developed in the above proof is dependent on the order $n+1$ of the resulting polynomial. We provide numerical simulation results for a second order
polynomial. For this case, it is possible to simplify the notation for the polynomial. In particular, consider an extended vector \((1, \theta, \nu)\) in \(Q + 2\) dimensions. A general second degree polynomial in \(\theta, \nu\) is specified by a collection of coefficients \(h_{j,k}, j = 0, \ldots, Q + 1, k = j, \ldots, Q + 1\). Please note that this notation is different from \(h_\alpha\) used above as there are only two subscripts on each coefficient and they range between 0 and \(Q + 1\). We can write a polynomial in these terms as

\[
\sum_{j=0}^{Q+1} \sum_{k=j}^{Q+1} h_{j,k} x_j x_k = \nu^2 h_{Q+1,Q+1} + \sum_{j=0}^{Q} \sum_{k=j}^{Q} h_{j,k} \theta_j \theta_k + \nu \sum_{j=0}^{Q} h_{j,Q+1} \theta_j,
\]

where \(\theta_0 \equiv 1\). Taking the expectation of this polynomial over \(\theta\) sampled from \(\pi(\theta | a, b)\), we obtain

\[
\nu^2 h_{Q+1,Q+1} + \sum_{j=0}^{Q} \sum_{k=j}^{Q} \frac{I(a + e_j + e_k, b)}{I(a, b)} h_{j,k} + \nu \sum_{j=0}^{Q} \frac{I(a + e_j, b)}{I(a, b)} h_{j,Q+1},
\]

(2.14)

with \(e_0\) indicating a vector of zeros representing no sales. Minimizing (2.14) results in the tightest second-order polynomial bound on the expected revenues using the explained method.

**Proposition 5.** Given a belief vector \((a, b)\) and inventory \(I_t\), the optimal price index maximizing the upper bound on revenue using a polynomial of second degree can be
2.6. FAST APPROXIMATIONS

Data: $a, b, I_t$

while $I_t > 0$ and $t < T$ do

1. Run LP in (A.3) to get polynomial coefficients;
2. Find optimal price index $j^*$ with equation (2.15) in proposition 4;
3. Observe sales;
4. Update $a, b$, and $I_t$;
end

Algorithm 1: Optimal pricing heuristic using polynomial approximation

obtained using the following equation:

$$
 j^* = \arg \max_j \left\{ I(a + e_j, b)p_j + \frac{2I_t - 1}{t - 1} I(a, b + e_j)h_{Q+1,Q+1} + \sum_{i=0}^{Q} [I(a + e_i, b + e_j)h_{i,Q+1}] \right\}
$$

Proof: The proof is immediate by substituting the expression (2.14) instead of the value function in the RHS of the original DP presented in (2.5).

The resulting procedure is outlined algorithm 1. For every possible state space, a linear program is solved before finding the optimal pricing policy. Sales are observed and the belief vector and inventory updated before the algorithm is executed again.

2.6 Fast approximations

Numerical calculation of the Euler integral $I_Q(a, b)$ is challenging. The Euler integral is defined as the generalization of the incomplete beta function (GIBF) to multiple dimensions. Existing literature such as [33] show the difficulty of calculating the incomplete beta function (IBF) with high precision. The GIBF can be defined as the
following function:

\[ G_d(z|a, b) = \int_{0 \leq \theta_Q \leq \cdots \leq \theta_d \leq z} \prod_{j=d}^{Q} \theta_j^{a_j-1} (1 - \theta_j)^{b_j-1} d\theta_j \] (2.16)

where \( I_Q(a, b) = G_1(1|a, b) \). The GIBF can be iteratively computed using the following lemma.

**Lemma 2.6.1.** The generalized incomplete beta function may be written as: \( G_j(z|a, b) = \int_0^z \theta_j^{a_j-1} (1 - \theta_j)^{b_j-1} G_{j+1}(\theta_j|a, b) d\theta_j \)

Calculation methods for the Euler integral depend on the prior composition. With either of \( a \) or \( b \) vectors composed of integer components only, a closed form formulation of the Euler integral can be formulated using its hypergeometric representation.

**Proposition 6.** The multi-dimensional Euler integral of first type can be calculated in closed form with "b" integer components greater than 1 and "a" positive real components using the following formulation:

\[ I_Q(a, b) = \sum_{j_1=0}^{b_1-1} \cdots \sum_{j_Q=0}^{b_Q-1} \prod_{l=1}^{Q} \left( \frac{b_l - 1}{j_l} \right) \frac{(-1)^{j_l}}{1 + \sum_{k=1}^{Q} (a_k + j_k)} \] \( \forall b \in \mathbb{Z}^+, \ a \in \mathbb{R}^+ \) (2.17)

See Appendix B for proof.

Numerical calculation of the Euler integral becomes exceedingly difficult for large values of \( Q \) or real components in \((a, b)\) vector. We propose two different approaches to compute the value of the Euler integral. The first solution is based on the Direct Cosine Transforms or DCT-II. This method is efficient, however its implementation suffers from a lack of precision due to the Fast Fourier Transform (FFT) package.
implementation. The rounding error manifests itself with uninformative priors and with large values of $Q$. Our numerical experiments show that with prior components smaller than 0.1, reasonable precision can be obtained with 100 digit mantissa. This is the main shortcoming of the DCT based solution.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Prior</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a, b \in \mathbb{Z}_{++}$</td>
<td>Uniform $a = b = 1$</td>
<td>Closed form, Chebyshev (exact solution)</td>
</tr>
<tr>
<td></td>
<td>Pessimistic $a &lt; 1, b = 1$</td>
<td>Closed form, Chebyshev</td>
</tr>
<tr>
<td></td>
<td>Uninformative $a, b &lt; 1$</td>
<td>Chebyshev, Negative Binomial Series</td>
</tr>
</tbody>
</table>

Table 2.1: Numerical methods for different prior types

The second approach is based on developing the iterative calculation of the GIBF presented in lemma (2.6.1) with the negative binomial series. This method requires a convolution operation over each dimension to calculate the integral, which is certainly slower than the FFT method. However choosing the number of terms in this series together with a high precision number representation package such as Boost float 100 offers much control over the precision of the calculations. Table 2.1 offers a summary of the different methods used to calculate the Euler integral.

2.6.1 Chebyshev polynomial

DCT-II is used to compute expansions into Chebyshev polynomial series $\frac{s_0}{2} + \sum_{k=1}^{M} s_k T_k(z)$, and the inverse of DCT-II is used to efficiently evaluate the values of the expansion at $M + 1$ collocation points (roots of Chebyshev polynomial $T_{M+1}(z)$ of degree $M + 1$). The algorithm maintains Chebyshev expansion of the GIBF defined
in (2.16), with vector $\vec{g} = (\vec{g}_0, \cdots, \vec{g}_M)$ representing the values of this expansion at the collocation points $z_j = \frac{1}{2}(\cos(\phi_j) + 1), j = 0, \cdots, M$ where $\phi_j = (j + \frac{1}{2}) \frac{\pi}{M+1}$. The algorithm stated below proceeds by iteratively computing the generalized incomplete beta function presented in lemma 2.6.1. The algorithm is stated in exhibit Algorithm 2.

**input**: $a, b, M$ (the highest order in Chebyshev expression)

**output**: $I(a, b)$

1. Initialize $\tilde{g}_j := 1$ for $j = 0, \cdots, M$;

2. for $d = Q$ to 1 do
   - Let $\bar{g}_j = \tilde{g}_j z_j^{a_d-1}(1 - z_j)^{b_{d-1}}$ for $j = 0, \cdots, M$;
   - Let $s := \frac{2}{M+1}$ (DCT-II of $\bar{g}$), i.e. $s_k = \frac{2}{M+1} \sum_{j=0}^{M} \bar{g}_j \cos(k \phi_j), k = 0, \cdots, M$;
   - Find series $\sum_{k=0}^{M} s_k^I T_k(z)$ representing $\int_0^z \left(\frac{s_0}{2} + \sum_{k=1}^{M} s_k T_k(z')\right) dz'$:
     - $s_M^I := \frac{s_{M-1}}{2M}$;
     - $s_k^I := \frac{s_{k-1} - s_{k+1}}{2k}, k = 1, \cdots, M - 1$;
     - $s_0^I := \sum_{k=1}^{M} (-1)^{k-1} s_k^I$;
   - if $d > 1$ then compute inverse DCT-II of $s^I$ and assign to $\bar{g}$:
     - $\bar{g}_j := \frac{s_0^I}{2} + \sum_{k=1}^{M} s_k^I T_k(z_j), j = 0, \cdots, M$;
   - end
3. Return $\frac{s_0}{2} + \sum_{k=1}^{M} s_k^I$ as $I(a, b)$

**Algorithm 2**: Algorithm for calculating $I(a, b)$

Using this method with integer-valued $(a, b)$, the integrands remain polynomial in each step. Thus, the calculations are exact as long as $M$ exceeds the order of the polynomial in the final step, i.e. $\sum_{j=0}^{q} (a_j + b_j) - Q$. For general $r$ times differentiable functions, the guaranteed pointwise accuracy of representation by Chebyshev series is within $O(M^{-r})$ according to [9]. Since DCT-II run in $O(M \log M)$ time and there are $Q$ steps in the algorithm, $I(a, b)$ runs in $O(QM \log M)$ time.
2.6.2 Negative binomial

Any realistic numerical use of the Euler integral requires the implementation of an efficient method that works with uninformative priors (small values of $a$ and $b$). In the absence of a reasonable FFT package, we develop the GIBF presented in lemma (2.6.1) with the negative binomial series. This transformation creates a convolution which has higher complexity than the FFT based solution. However, it poses no restriction on the precision of the calculations.

Proposition 7. In a multi-dimensional Euler integral, the integral for each dimension may be represented under the form $G_{d+1}(z) = \sum_{\tilde{n} \geq 0} c_{\tilde{n}} z^{\tilde{n} + \gamma}$, where $c_{\tilde{n}}$ represents the coefficient of the series, and $\gamma$ is a fixed parameter. The update for the next dimension follows a similar form $G_{d}(z) = \sum_{\tilde{n} \geq 0} \tilde{c}_{\tilde{n}} z^{\tilde{n} + \gamma + a_d}$, where $\tilde{c}$ is obtained following a convolution $\tilde{c}_{\tilde{n}} = \frac{1}{\tilde{n} + \gamma + a_d} \sum_{l=0}^{\tilde{n}} \frac{(1-b_d)^l}{l!} c_{\tilde{n}-l}$. (See Appendix C for proof)

In the above proposition $(a)_k = a(a+1)\cdots(a+k-1)$ is the rising factorial symbol, $\tilde{n}$ represents the number of terms in the series. The individual terms in these computations are bounded by $O(2^{-\tilde{n}})$, so the series converges very fast, and does not require too many terms (large $\tilde{n}$) to compute it to high precision. The algorithm achieves accuracy of $\epsilon$ with complexity $O(Q\tilde{n}^2 P \ln(P))$, where $P = \ln(1/\epsilon)$.

Proposition 8. The negative binomial series algorithm can achieve complexity of $O(Q\tilde{n}P \ln(P))$ by only considering half the domain of $z$. (See Appendix D for proof)

2.7 Simulation Setup

Our focus remains on numerical experimentation. In order to benchmark the heuristic presented earlier, we measure the average regret using Monte-Carlo simulations. We
define the retailer’s regret given the initial inventory, belief vector and selling horizon as:

\[ Regret_T(a, b, I) = 1 - \frac{E[Rev(T, I)|a, b]}{E[Rev^*(T, I)|a, b]} \]  (2.18)

where \( E[Rev^*(T, I)|a, b] \) is the perfect information upper bound obtained by averaging the optimal revenue over a large sample set. The optimal revenue for each sample is calculated using the original DP formulation

\[ V_t(I_t|\theta^*) = \max_{j=1,\ldots,Q} \theta^*_j(p_j + V_{t-1}(I_{t-1}|\theta^*)) + (1 - \theta^*_j)V_{t-1}(I_{t-1}|\theta^*) \]  (2.19)

where \( \theta^* \) is a vector member of a set of sampled probabilities from the joint distribution. We obtain the average revenue by averaging the sampled revenue function over the sample size \( \frac{1}{L} \sum_{s=1}^{L} V_t(I_t|\theta^*) \). By the strong law of large numbers, almost surely, \( \frac{1}{L} \sum_{s=1}^{L} V_t(I_t|\theta^*) \rightarrow E[Rev^*(T, I)|a, b] \), as \( L \rightarrow \infty \).

2.7.1 Variance Reduction

Simulation of the model presented in this paper requires sampling the joint probability distribution of the generalized beta/binomial distribution, where components of each sample are correlated due to the ordering structure. The discrete maximization of the value function with correlated samples requires careful control of the variance. Given the possibility of simulating the perfect information upper bound to a reasonable precision ahead of starting simulations, we can use a control variate using the difference estimator to reduce variance.

The difference estimator takes advantage of the perfect information upper bound,
which is close in value to the expected revenue estimated via simulations. It is con-
structed so that the expected value of the difference estimator is close to that of the
estimated function, but its variance reduces to the difference between that of the
estimated function and the perfect information upper bound.

To get a proxy function close in value to that of the estimated value function
using simulations, it is possible to run a one time simulation, with a large number of
samples relative to the sample size used in simulations.

Given a set of estimates $\theta^s$, the objective is to maximize the expected revenue,
using a simulation environment. The expected revenue is defined by $\eta = \mathbb{E}_\theta[V_t(I_t|\theta^s)]$, 
where $V_t(I_t|\theta^s)$ is the estimated revenue function and $\theta^s$ represents the set of sam-
pies. Variance reduction uses a control variate $h^c(\theta)$ which is a close estimator of
the original estimated function $V_t(I_t|\theta^s)$. We take advantage of the perfect informa-
tion upper bound, as a numerically inexpensive way of constructing a proxy for the
simulated estimate. Running over $10^7$ samples takes no more than a few minutes
on any state of the art computing platform. Let $\xi = \mathbb{E}[h^c(\theta)]$, where $\theta$ corresponds
to a large set of samples. We define a difference estimator $\hat{\eta}_{diff}$, e.g., presented in
[84], as $\hat{\eta}_{diff} = \frac{1}{s}\sum_{i=1}^{s}(V(\theta_i) - h^c(\theta_i)) + \xi$, where $s$ is the number of samples in
the regular simulations. Since $\mathbb{E}[\frac{1}{s}\sum_{i=1}^{s} h^c(\theta_i)] = \xi$, the expected value of $\hat{\eta}_{diff}$ is
$\eta$. However, the variance of the difference estimator is the difference between the
two closely related functions, and therefore much smaller than the original variance
$Var(\hat{\eta}_{diff}) = \frac{1}{s}Var(V(\theta) - h^c(\theta))$, which leads to higher precision.
2.8 Numerical Results

Figure 2.2 depicts the solution to the linear program (A.3) for different values of $a$, $b$ and $\nu$. Figure 2.2(a) compares the solution of the LP between uniform and pessimistic priors, whereas figure 2.2(b) compares the solution of the LP between uniform and uninformative priors. Given the results of recent papers on Thompson Sampling using uniform priors as an uninformative prior as the starting point, see [4] [69] and [37], we believe figure 2.2 offers valuable information on the informativeness of priors.

The expected revenue is concave and increasing in $\nu$, therefore the marginal value of the expected revenue increases as $\nu$ decreases. For fixed number of periods, the marginal value of the expected revenue increases when inventory decreases. This conforms to the structure of the optimal policy of a single resource in revenue management. The uniform prior creates the maximum curvature, the most pronounced change in the marginal expected revenue among all the priors considered in this paper. For uniform and pessimistic priors, the marginal value of the expected revenue increases as $\nu$ decreases. This observation is less pronounced for uninformative priors. As the value of all components of $a$ and $b$ approach zero, the prior becomes increasingly uninformative with less change in the marginal value of the expected revenue. Pessimistic priors preserve their curvatures, however the more pessimistic the prior, the smaller the expected revenue for a given value of $\nu$. More pessimism is directly linked to a lack of learning and lower revenue. Unlike pessimistic priors, uninformative priors do not restrict learning, and under some inventory constraints have higher expected revenue than with uniform priors.

To summarize, uniform priors are informative priors, offering information on how
2.8. NUMERICAL RESULTS

revenue changes with respect to the remaining horizon and available inventory. Uninformative priors are those where the values of the components of $a$ and $b$ are both smaller than 1. The more uninformative a prior, the smaller the values of their components, resulting in smaller change in the marginal value of the expected revenue.

Figures 2.3 and 2.4 show the performance of the heuristic approach presented in this paper with a polynomial of second degree by plotting the regret function defined in (3.9). As discussed earlier, pessimistic and uninformative priors display different behaviors.

In figure 2.3, we compare the uniform prior with two pessimistic priors with different inventory levels of 2, 5 and 7 items to sell over a horizon of 100 periods. Graphs in each row differ in their inventory levels, whereas graphs in each column represent different priors. Each simulation is run with 2 and 5 prices, where prices are standardized between 0 and 1. As observed in figure 2.3, the uniform prior performs better than the pessimistic priors. As expected from our initial assessment of figure 2.2, pessimistic priors are conservative in price exploration and produce myopic policies resulting in an increasing regret function. The uniform prior maintains a
gap of less than 5% from the optimal policy after few iterations. It is important to note that similar performance is achieved using Thompson Sampling asymptotically, however our model achieves this performance when information is scarce, i.e. with small inventory and short horizon.

![Figure 2.3: Simulation of the regret function with the Lagrangian relaxation heuristics, and informative priors](image)

Uninformative priors’ performance compares to that of the uniform prior. Uninformative priors require a longer initiation period to learn. During this initial period, the performance is slightly lower than that of the uniform prior. However, as sales activity progresses, models using uninformative priors converge faster than those using uniform priors. The more uninformative the prior with smaller values of $a$ and $b$, the more pronounced this characteristic becomes.

The uninformative prior offers a better convergence in terms of regret for longer horizons, up to 100 periods. For longer horizons, both uninformative and uniform priors have similar convergence properties. The downside of uninformative priors is
2.9 Conclusion

We took a practical approach to the problem of dynamic pricing and learning over a finite horizon with limited capacity and a set of discrete prices. We introduced a non-parametric approach to demand learning using a Bayesian updating model taking advantage of the intrinsic structure existing between increasing prices and decreasing probabilities of sales. The structure generalizes the beta-binomial distribution, providing closed form solution expressing the expected probability of sales. We also
provide numerically tractable methods to calculate the Euler integral for fractional values. These methods cover the case for all types of priors, including the general case of pessimistic priors with small fractional parameters. The final contribution was the approximation of an efficient upper-bound for the expected revenue. Transforming the inventory constraint and the ordering relationship information into linear constraints to be used in a linear program required methods which, to our knowledge, are novel. The use of polynomial approximation and optimization offered a novel approach to approximating the expected revenue.

2.10 Summary of variables
### Table 2.2: Table of variables and notation

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Q$</td>
<td>Number of price components</td>
</tr>
<tr>
<td>$m_t, n_t$</td>
<td>Information state</td>
</tr>
<tr>
<td>$a, b$</td>
<td>Vector of priors</td>
</tr>
<tr>
<td>$p_j$</td>
<td>jth price in the set</td>
</tr>
<tr>
<td>$N$</td>
<td>Initial Inventory</td>
</tr>
<tr>
<td>$T$</td>
<td>Horizon</td>
</tr>
<tr>
<td>$t, \tau$</td>
<td>Time index</td>
</tr>
<tr>
<td>$\theta = (\theta_1, \cdots, \theta_q)$</td>
<td>Vector of probabilities</td>
</tr>
<tr>
<td>$I_t$</td>
<td>Inventory at time $t$</td>
</tr>
<tr>
<td>$V_t(m, n, I_t)$</td>
<td>Expected revenue</td>
</tr>
<tr>
<td>$\pi(\theta)$</td>
<td>Prior distribution</td>
</tr>
<tr>
<td>$\Theta$</td>
<td>Vector of all purchase probabilities</td>
</tr>
<tr>
<td>$C$</td>
<td>Normalizing constant</td>
</tr>
<tr>
<td>$I(a, b)$</td>
<td>Euler Integral</td>
</tr>
<tr>
<td>$B(a_i, b_i)$</td>
<td>Beta distribution</td>
</tr>
<tr>
<td>$V_t(I_t, \theta)$</td>
<td>Value function from the perfect information distribution</td>
</tr>
<tr>
<td>$\mu$</td>
<td>Lagrange multiplier</td>
</tr>
<tr>
<td>$\nu_t$</td>
<td>Ratio of inventory over time</td>
</tr>
<tr>
<td>$u_r$</td>
<td>Randomized control policy</td>
</tr>
<tr>
<td>$D_\tau$</td>
<td>Binary demand random variable</td>
</tr>
<tr>
<td>$R_\tau$</td>
<td>Revenue at time $\tau$</td>
</tr>
<tr>
<td>$L^*(\theta, \nu, \mu)$</td>
<td>Lagrangian relaxation</td>
</tr>
<tr>
<td>$w_j$</td>
<td>Convex combination of weights</td>
</tr>
<tr>
<td>$h_{\alpha, \beta}$</td>
<td>Polynomial coefficients of degree $\alpha$ and $\beta$</td>
</tr>
<tr>
<td>$\alpha, \beta$</td>
<td>Vectors representing degrees of the polynomials</td>
</tr>
<tr>
<td>$x, g$</td>
<td>Vector of polynomial variables</td>
</tr>
<tr>
<td>$\bar{n}$</td>
<td>Maximum degree of a polynomial</td>
</tr>
<tr>
<td>$\bar{m}$</td>
<td>Dimension of polynomial</td>
</tr>
<tr>
<td>$G_d(z</td>
<td>a, b)$</td>
</tr>
<tr>
<td>$M$</td>
<td>Number of roots of Chebyshev polynomial</td>
</tr>
<tr>
<td>$g = (g_0, \cdots, g_M)$</td>
<td>Expansion values of Chebyshev polynomial</td>
</tr>
<tr>
<td>$s_k, s_k^l$</td>
<td>Chebyshev polynomial coefficients</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Fixed parameter</td>
</tr>
<tr>
<td>$L$</td>
<td>Sample size</td>
</tr>
<tr>
<td>$\theta^*$</td>
<td>Vector of samples from the true distribution</td>
</tr>
<tr>
<td>$h^*(\theta)$</td>
<td>Control variate function</td>
</tr>
</tbody>
</table>
Chapter 3

Short-horizon dynamic pricing and learning with Thompson Sampling

3.1 Abstract

Thompson Sampling (TS), also called probability matching or posterior sampling, is a randomized Bayesian algorithm providing an efficient heuristic to sequential learning problems. Recent analyses of the algorithm provide regret bounds, which have been shown to be asymptotically optimal. The objective of our research is to simultaneously explore and exploit demand for an admissible set of discrete prices over a short horizon. We apply the fact that offering a product at a lower price should have at least as high of a probability of sale as offering it at a higher price to our sampling procedure. Our results with ordered samples show a 10% to 30% improvement in performance over other pricing algorithms using TS, such as the ts-update algorithm presented in [37] over short horizons. We analyze the optimal policy for a 2 arm 2 period stochastic bandit problem with both ordered and independent demand distributions and compare the results in terms of their regrets. This analysis provides the preliminary explanation supporting the use of ordered distributions for short-horizon
3.2 Introduction

Dynamic pricing with learning is an important topic in revenue management and has been extensively researched by scientists from different communities such as computer science, economics, and operations research. This interest is attributed not only to its practical applications in different industries but also to the difficulty of providing near optimal heuristics under different business constraints. Indeed, heuristics remain the only response to the numerical complexity and curse of dimensionality associated with this class of problems. The framework of multi-armed bandits (MAB) has been popular in describing and solving sequential learning and discrete pricing problems using different heuristics. The use of Thompson Sampling, a combination of Bayesian updating together with randomized sampling in solving MAB-like problems, offers an elegant and practical solution to an otherwise difficult problem. Asymptotic optimality is a problem for short-horizon applications. This chapter modifies the sampling procedure used in Thompson Sampling for pricing applications to improve performance over short horizons.

3.3 Literature review

Thompson Sampling was first introduced in [102] and later used in [52]. It has strong practical applications in sequential learning problems, known in general as the multi-armed bandit (MAB) problems. As an efficient heuristic, offering strong empirical results, this algorithm has been adapted to multi-play MAB and assortment optimization in [61] and [2], online network pricing with inventory constraint in [37],
and parametrized reinforcement learning in [47] in a long list of other applications and adaptations. For a general overview of the heuristic and a comprehensive list of its capabilities, [92] provides an excellent tutorial on the subject.

Despite Thompson Sampling’s excellent performance, its adaptation to different applications is mostly recent. A contributing factor seemed to be a lack of strong bounding analyses as explained in [92]. The first theoretical bound for MAB like problems was offered by [63]. Among the many heuristics dealing with stochastic multi-armed bandit problems, the Upper Confidence Bound (UCB) algorithm, ([6], [58]), provides theoretical bounds that were subsequently linked to Thompson Sampling ([91]). A complete analysis of Thompson Sampling bounds was first conducted in [3], [16], and later using a martingale approach in [4].

Dynamic pricing with learning has a rich literature that started with [40], with comprehensive overview papers on the most current research with [19] and [31]. Dynamic pricing and learning with Thompson Sampling makes use of Bayesian updating. This requires the use of different prior distributions as discussed in [62]. The importance of choosing priors that reflect existing data has been brought up in [92], and the link between the quality of the prior and the performance bound has been analyzed in [69]. Fractional parameters can be used with the beta distribution to encode uninformative and pessimistic beliefs in the prior. Expressing the same information with integer parameters is inefficient. Numerical methods to solve truncated multivariate beta distributions with fractional parameters have been presented in the first chapter of this thesis. To the best of our knowledge, few papers study the use of fractional parameters as a prior for the beta distribution. The algorithms presented in this paper are agnostic to this issue and can be used with fractional parameters.
3.4 Thompson Sampling

We start with Thompson Sampling for the Bernoulli-bandit problem with \( q \) independent arms, each with a probability of success \( \mu_i \), and reward of 0 or 1. We use the beta-binomial distribution instead of the beta-Bernoulli distribution to adapt TS to pricing applications, where demand for different prices is correlated. The beta-binomial distribution is a popular choice as the beta distribution is the conjugate prior to the binomial one. Consider a set of prices \( p_i \) indexed by \( i = 1 \cdots q \), each associated with a demand distribution with parameter \( \mu_i \), representing the probability that a sale occurs every time an incoming customer is presented with price \( p_i \). The price set is always ordered such that \( p_1 \leq \cdots \leq p_q \). The use of the beta-binomial distribution is natural in learning problems with this structure. The terms price and arm are used interchangeably throughout this paper.

At the start of the selling horizon, the retailer does not know the true probability of success for each price. This is represented by a beta distribution with prior parameters \((a_i, b_i)\) with \( i = 1, \cdots, q \), and mean \( \frac{a_i}{a_i+b_i} \). \((a, b)\) represents the vector of prior parameters for all admissible prices. The beta distribution for arm \( i \) is defined by

\[
f(\theta_i|a_i, b_i) = \frac{\Gamma(a_i+b_i)}{\Gamma(a_i)\Gamma(b_i)} \theta_i^{a_i-1}(1-\theta_i)^{b_i-1}.\]

The horizon is divided in \( T \) periods, with \( t = 0, \cdots, T \). Only one arm is pulled during each period. Experimenting with each arm results in success/failure information vector \((m, n)\), with \( \sum_{i=1}^q (m_i + n_i) = 1 \). As the conjugate prior of the binomial distribution, the parameters of the beta distribution for each arm \( i \) are updated by \((a_i + m_i, b_i + n_i)\). The accumulation of the information vector over the parameters of the beta distribution over time reveals the average demand for each price.
Thompson Sampling operates by using randomized samples from the posterior distribution to make optimal decisions. Consider a random sample vector \( \theta = \{\theta_1, \ldots, \theta_q\} \) from distribution \( \prod_{i=1}^{q} f(\theta_i|a_i, b_i) \). In a pricing application, the algorithm decides which price to offer by selecting the price index with the highest expected revenue = arg max, revenue(\( \theta_i \)) for a given sample vector \( \theta \). Once the arm is chosen, consumers decide to accept or reject buying the product based on the offered price. This information is added to the posterior distribution, and the next set of randomized samples allows the algorithm to continue this procedure. The procedure has shown satisfactory empirical results, and the regret bounds are asymptotically optimal as presented in [3], [4] and [37].

Pricing and revenue management adaptations for Thompson Sampling consider independent beta distributions for a set of admissible prices. We argue that pricing represents a special setting where the set of discrete prices is always ordered. The ordering structure also applies to the samples associated with the demand distribution, and therefore enforcing such structure in the algorithm reduces the time it takes to learn the true parameters of the model.

The idea behind our approach is simple. Instead of using independent beta distributions representing each price and updating individual distributions after observing the reward, we make use of a joint distribution over all demand parameters. This is necessary since sales at different prices are correlated and the sale outcome at a given price affects the demand distribution for other prices in the set. This translates into choosing ordered samples from the updated joint distribution.
3.4. THOMPSON SAMPLING

Initialize: Set the prior parameters \((a_i, b_i)\) for \(i = 1 \cdots q\) for each arm

Until end of horizon:
1- Sample each beta distribution independently for each price
2- Solve the following linear program with respect to the weights \(v_i\)

\[
\begin{align*}
\max \sum_{i=1}^{q} p_i \theta_i v_i \\
\text{subject to } \sum_{i=1}^{q} v_i &\leq 1 \\
\sum_{i=1}^{q} \theta_i v_i &\leq \frac{y_t}{t} \\
v_i &\geq 0, \quad i = 1 \cdots q
\end{align*}
\]

3- Offer price \(p_i\) with probability \(v_i\), and offer shutdown price \(p_{q+1}\) with probability \(1 - \sum_{i=1}^{q} v_i\).
4- Observe realized demand vector \((m, n)\)
5- Update the parameter vector of the priors and inventory using the realized demand with \((a + m, b + n)\) and \(y_{t-1} = y_t - \sum_{i=1}^{q} m_i\).

Figure 3.1: Demand learning and pricing for a single product with Thompson Sampling derived from [37]

[37] proposes a capacity constrained online network revenue management algorithm using Thompson Sampling. We review a simplified version (with a single product) of TS-update to illustrate how the algorithm is used and where it should be modified. The TS algorithm is extended with a linear program similar to the one presented in the first chapter of this thesis to account for the capacity constraint. \(v_i\) corresponds to a set of convex weights transforming the objective function into continuous form, and \(y_t\) represents the remaining inventory at time \(t\). The algorithm is illustrated in figure 3.1.

A few observations are relevant to this algorithm. First, the system is initialized
with independent beta distributions, and samples are chosen independently. In our analysis, we prove that considering ordering of the random samples accelerates the convergence of the algorithm. Second, the shutdown price is only relevant to network pricing, when individual items from a product group are out of stock. In situations where the capacity constraint is active and the convex weights do not sum up to 1, the excess weight is generally attributed to the shutdown price, which postpones sales. With a single product and a short horizon, there’s no disadvantage in selling out early at the highest price. The algorithm is always better off offering the highest price rather than the shutdown price.

[37] provides an example with a single product and a set of 4 prices, showing asymptotic convergence using simulations. The algorithm works over long horizons because random sampling allows the algorithm to explore the demand function and approach the optimal policy. However, the realized demand for a price does not affect the posterior distribution of demand for other prices over a short horizon. It requires a long cycle of experimentation before possible errors, due to random sampling or prior misspecification, are corrected. For example, if at the beginning of the selling horizon and given a uniform prior, the sample associated with the highest price results in successful sales over multiple periods, the posterior of the demand distribution for the highest price changes to reflect that and the algorithm becomes biased in isolation of other prices. In such cases, the sample from the posterior distribution for the highest price will have a higher value than all other samples with high probability. This kind of unordered sampling is inefficient. Although the randomized sampling prevents incomplete learning, it will take longer to correct such situations.

Sample ordering prevents this situation effectively. The sample from the demand
distribution of the highest price should maintain its ordering with respect to the samples for other prices, otherwise be eliminated or replaced. There are a few rather simple approaches to achieving this outcome. The first approach consists of sampling independent beta distributions and accepting only those that follow the ordering structure. The method is simple to implement, and does not require any special treatment as distributions are sampled independently, however the probability of acceptance diminishes rapidly, at a rate of at least \((q!)^{-1}\). It becomes excessively difficult to sample with this method as the number of prices increases.

A different approach involves importance sampling, where initially an ordered set of samples is chosen from a symmetric distribution with parameters \((\bar{a}, \bar{b})\), where \(\bar{a} = \min\{a_1, \cdots, a_q\}\) and \(\bar{b} = \min\{b_1, \cdots, b_q\}\). Symmetric distributions offer the easiest approach to generating ordered samples, where \(q\) samples are drawn from a single distribution with parameters \((\bar{a}, \bar{b})\) and arranged in increasing order. To apply importance sampling we define \(\bar{m} = \{a_1 - \bar{a}, \cdots, a_q - \bar{a}\}\) and \(\bar{n} = \{b_1 - \bar{b}, \cdots, b_q - \bar{b}\}\) and accept samples with probability \(\prod_{i=1}^{q} \theta^{\bar{m}_i}(1 - \theta)^{\bar{n}_i}\). As the sales horizon increases, the acceptance rate reduces rapidly due to the accumulation of sales results in the information vector. The algorithm is outlined in figure 3.2.

This algorithm works well with a small number of prices and short horizons, however, the algorithm becomes less effective as the number of prices or the length of the horizon increases. Both methods presented so far suffer from the curse of dimensionality. To remedy this situation, we intend to use a combination of rejection sampling to initialize the sampling procedure, together with the Metropolis-Hasting algorithm to have the sample set converge to the desired distribution.
### 3.4. THOMPSON SAMPLING

Initialize: Set the prior parameters \((a_i, b_i)\) for \(i = 1 \cdots q\) for each arm

Until end of horizon:

1- Set \(\bar{a} = \min\{a_1, \cdots, a_q\}, \bar{b} = \min\{b_1, \cdots, b_q\}\)

3- Set \(\mathbf{m} = \{a_1 - \bar{a}, \cdots, a_q - \bar{a}\}\) and \(\mathbf{n} = \{b_1 - \bar{b}, \cdots, b_q - \bar{b}\}\)

4- Sampling procedure

   **while Sample not accepted do**

   4.1- Get ordered sample from symmetric distribution

   \[
   \theta \sim \prod_{i=1}^q \theta_i^{a_i-1}(1 - \theta_i)^{b_i-1}
   \]

   4.2- Given realized demand vector \((\mathbf{m}, \mathbf{n})\) accept samples with probability

   \[
   \prod_{j=1}^q \theta_j^{m_j}(1 - \theta_j)^{n_j}.
   \]

**end**

5- Solve the following linear program with respect to the weights \(v_i\)

\[
\max \sum_{i=1}^q p_i \theta_i v_i
\]

subject to

\[
\sum_{i=1}^q v_i \leq 1
\]

\[
\sum_{i=1}^q \theta_i v_i \leq \frac{y_t}{t}
\]

\[
v_i \geq 0, \ i = 1 \cdots q
\]

6- Offer price \(p_i\) with probability \(v_i\), and offer shutdown price \(p_{q+1}\) with probability \(1 - \sum_{i=1}^q v_i\).

7- Observe realized demand vector \((\mathbf{m}, \mathbf{n})\)

8- Update the parameter vector of the priors and inventory using the realized demand with \((a = a + \mathbf{m}, b = b + \mathbf{n})\) and \(y_{t-1} = y_t - \sum_{i=1}^q m_i\).

---

Figure 3.2: Importance sampling with replacement for Thompson Sampling
3.5 Metropolis Monte-Carlo approach

The Metropolis-Hastings algorithm is a reasonable method to obtain samples from multi-dimensional distributions. It is a Markov Chain Monte-Carlo method and works by using a proposal density (transitional kernel) to sequentially decide which sample point to choose next. The choice of the proposal density has a significant impact on the algorithm’s acceptance rate and performance. To sample from the ordered multivariate beta distribution, we use the Dirichlet distribution for the proposal density. There are 3 reasons behind this choice:

- Given that the Dirichlet distribution is the multivariate generalization of the beta distribution, it is natural to use it as the proposed density in this application.

- The Dirichlet distribution is parameterized by the difference of the neighboring sample vector components. Given the ordering among the samples, these differences are always non-negative, and provide good approximation for candidate target samples.

- Ultimately, simulation results and the acceptance rate of the algorithm proved to be the deciding factors for choosing the Dirichlet distribution as proposal density.

One of the shortcomings of the Metropolis-Hastings method is that successive samples in a chain are correlated. To get independent samples, it is recommended to discard most samples and take spaced out samples by examining the auto-correlation between the samples. In this application, we use a slightly different method: instead of using a single chain to select all samples from, we run M separate chains based on the
initial sample set, and after the burn-in period, select one set of samples from each chain. Parallel chains have been studied and used for different applications, see [55] and [105]. Using multiple chains reduces the correlation between the samples, as they have different starting points and evolve over different chains. Additionally, this allows for the future parallel implementation of the algorithm. The proposed algorithm is outlined in figure 3.3.

The Metropolis-Hastings algorithm is tuned with the sensitivity parameter $S$, the length of the chain $K$, and the number of chains $M$. The sensitivity and the length of the chain are dependent on each other, as higher sensitivity requires longer chains to converge to the true distribution. Running $M$ chains allows the algorithm to run shorter chains, and reduces the problem of correlated samples. However, running a large number of chains may not be as efficient as picking lagged samples from a smaller number of chains. The focus of our simulation was not to fine tune the parameters of the sampling algorithm. We chose a small number of chains between 1 and 10, with the sampling algorithm running between 500 and 1000 iterations. These had minimal effect on our results. However, the sensitivity parameter is directly linked to the acceptance probability, with a target range between 30% and 60%. Acceptance probability in this range produced results in line with what is presented in figure 3.4. A lower acceptance probability may be due to high sensitivity and may suggest that the algorithm is stuck or slow moving. A high acceptance rate may suggest the opposite, where jumps are too wide for the algorithm to converge. A more thorough analysis of the algorithm, exploring the optimal number of chains and iterations will certainly be part of our future research.

The algorithm starts by running rejection sampling with replacement on a given
3.5. METROPOLIS MONTE-CARLO APPROACH

Initialize: Draw ordered sample vector from symmetric distribution (i.e.  
\[ a_i = a, b_i = b, i = 1, \ldots, q \]),  
\[ \theta_r \sim \prod_{i=1}^{q} \frac{\theta_i^{a_i-1}(1-\theta_i)^{b_i-1}}{Beta(a_i,b_i)}, r = 1 \cdots M; \]

Set  \( m = n = 0 \)

Until end of horizon:
1-for  \( r = 1 \) TO  \( M \) do
   for  \( k = 1 \) TO  \( K \) do
      Let  \( \bar{\theta}_r \) be the kth  \( \theta_r \) vector in the MCMC chain.
      1.1- Let  \( \alpha_j = S\gamma_j \) with  \( \gamma_j = \bar{\theta}_{r,j} - \bar{\theta}_{r,j-1} \)
      1.2- Sample from the Dirichlet distribution with  \( x^* \sim Dirichlet(\alpha) \),
          and  \( \theta^*_j = \sum_{j'=j+1}^{q+1} x^*_{j'} \) as follows:
          1.2.1- Draw samples  \( \xi_j \sim \Gamma(\alpha_j, 1) \),  
              \( j = 1, \ldots, q+1 \)
          1.2.2- Set  \( x^*_{j} = \frac{\xi_j}{\sum_{j'=1}^{q+1} \xi_{j'}} \)
      1.3- Accept  \( \theta^* \) with probability  \( \min\{1, \frac{L(\theta^*)Q(\bar{\theta}_r|\theta^*)}{L(\bar{\theta}_r)Q(\theta^*|\bar{\theta}_r)}\} \)
          where  \( L(\theta) \) is the posterior probability function defined as  
          \[ L(\theta) = \prod_{i=1}^{q} \theta_i^{a_i-1}(1-\theta_i)^{b_i-1} \]
          and  \( Q(\theta^*|\theta) \) is the proposal density defined as  
          \[ Q(\theta^*|\theta) = \left( \prod_{j=1}^{q+1} \Gamma(\alpha_j) \right)^{-1} \prod_{j=1}^{q+1} (\theta^*_j - \theta^*_{j-1})^{\alpha_j-1} \]
      1.4- if accepted  \( \bar{\theta}_{r+1} = \theta^* \) otherwise  \( \bar{\theta}_{r+1} = \bar{\theta}_r \)
   end
2- Solve the following linear program with respect to the weights  \( v_i \)

\[
\max \sum_{i=1}^{q} p_i \theta_i v_i
\]
subject to  
\[
\sum_{i=1}^{q} v_i \leq 1
\]
\[
\sum_{i=1}^{q} \theta_i v_i \leq \frac{y_t}{t}
\]
\[
v_i \geq 0, \ i = 1 \cdots q
\]

3- Offer price  \( p_i \) with probability  \( v_i \), and offer shutdown price  \( p_{q+1} \) with probability  \( 1 - \sum_{i=1}^{q} v_i \).
4- Observe realized demand vector  \( (m,n) \)
5- Update the parameter vector of the priors and inventory using the realized demand with  \( (a = a + m, b = b + n) \) and  \( y_{t-1} = y_t - \sum_{i=1}^{q} m_i \).

Figure 3.3: Metropolis-Hastings algorithm with parallel chains
3.5. METROPOLIS MONTE-CARLO APPROACH

Figure 3.4: Comparison of the [37] example with ordered Beta distribution using Metropolis-Hastings for short horizon T=10

set of M samples. At the start of the sampling algorithm, a sample set is available either from a previous run of the algorithm or from the prior distribution at initialization. The rejection sampling with replacement helps the algorithm to have randomized starting points for the parallel chains. The Dirichlet distribution is parametrized using the differences between the neighboring sample components as shown in step (1.1) of the algorithm. Step (1.2) is dedicated to sampling from the Dirichlet distribution. As shown in steps (1.2.1) and (1.2.2), the Dirichlet distribution with parameter vector \( \{\alpha_1, \cdots, \alpha_{q+1}\} \) may be sampled by first drawing \( q+1 \) random samples \( (\xi_1, \cdots, \xi_{q+1}) \) from Gamma distributions each with density \( \Gamma(\alpha_j, 1), j = 1, \cdots, q+1 \). The Dirichlet samples are constructed as follows:

\[
x_j^* = \frac{\xi_j}{\sum_{j=1}^{q+1} \xi_j}.
\]

Given the parameterization of the Dirichlet distribution with sample differences, we reconstruct the candidate state (target samples) by adding the Dirichlet samples back together by

\[
\theta_j^* = \sum_{j'=j+1}^{q+1} x_{j'}^*.
\]

\( \theta_j^* \) is an ordered sample set representing a candidate state in the Markov chain. It is accepted with the acceptance probability set by the algorithm in
3.6 PRIOR MISSPECIFICATION, AND THE USE OF FRACTIONAL PARAMETERS

step (1.3), where $Q(\theta^*|\theta)$ is the proposed distribution to choose sample $\theta^*$ given the current sample $\theta$, and $L(\theta^*)$ is the posterior probability function.

Figure 3.4 compares the example in [37] using independent beta samples and the Metropolis-Hastings algorithm providing ordered samples. We replicate the setup in [37] for a single product with the following set of discrete prices \{$29.90, $34.90, $39.90, $44.90\}. The mean demand for the prices in this set is \{0.8, 0.6, 0.3, 0.1\}. We perform our experiment on a short horizon of 10 periods, with an inventory of 3, 5 and 9, so the capacity is always constrained. Results for both algorithms are measured against the revenue achieved with true demand. Figure 3.4 also compares the results against the TS-Fixed policy, where inventory constraint is not updated as sales progress. The difference in performance is similar to what we observed with TS-update. With any level of inventory, the performance of the algorithm with the Metropolis-Hastings extension is 10% to 30% better than using only independent distributions.

The objective of this experiment was to compare the results obtained from our methodology with existing results from [37], and does certainly not constitute a comprehensive characterization. We do intend to expand on this research in the future, and fully characterize the performance of the proposed algorithm.

3.6 Prior misspecification, and the use of fractional parameters

Uniform priors are commonly used in TS applications using beta-binomial distributions. The choice is advantageous as it can be easily obtained by setting $a_i = b_i = 1, i = 1, \cdots, q$. During parameter updates, this ensures the use of integer parameters only, as numerical evaluation of the beta distribution with fractional
3.6. PRIOR MISSPECIFICATION, AND THE USE OF FRACTIONAL PARAMETERS

parameters can be difficult. In practice, data may suggest otherwise, and preference may be given to a price over another. As discussed in \cite{92}, this can affect the performance of Thompson Sampling. \cite{92} uses samples from Beta(1,50) and other similar distributions to show the effect of misspecified priors. It is also known that the beta distribution concentrates rapidly as the number of trials progresses, as used by \cite{3} to help prove convergence of TS. Although pessimistic priors Beta(1,100), Beta(0.1,10) and Beta(0.01,1) have the same mean, they convey different information and produce different sample sets. Let us consider a prior with parameters (1,100), which gives an average chance of success of approximately 1%. If the prior is misspecified, and the mean of the true distribution is considerably different, it will take many trials before converging to the true parameter, and the concentration of the distribution will slow down the process. However a fractional prior with parameters (0.01,1) gives a better chance at exploration if indeed the probability of success is higher. With the first successful trial, the prior becomes more uniform and allows for more randomized sampling. Should the prior be properly specified, then failed attempts will quickly accumulate in the parameter, and the distribution converges to its true value.

One of the problems facing researchers is the ability to calculate probabilities and get samples from the beta distribution with fractional parameters. The Metropolis-Hastings algorithm does not require any modification to use fractional numbers. Calculating the normalizing factor for both ordered and independent distributions with fractional priors has been described in the first chapter of this thesis, and can be achieved with high accuracy for both informative and uninformative priors. These calculations are beyond the scope of this paper.
3.7 Performance analysis of a 2-arm bandit problem with TS

We focus our analysis on the explicit probability associated with selecting random samples in the TS algorithm. This approach is hardly scalable, and forces the analysis into a stylized 2 arm configuration, but offers new insight on the performance of the algorithm. The objective is to study the effect of price discounts on the Bayesian regret over short horizons. The use of stylized models, similar to what is used in supply chain models is not uncommon and has been used in [13], and [11]. In this analysis, we look for performance discrepancies between the dependent and independent distributions. Let us rewrite the simplified online pricing algorithm with beta-binomial distribution and no capacity constraint in vector form:

For all arms set the prior vectors $\mathbf{a} = \mathbf{b} = \mathbf{1}$

for $t=T$ to $t=0$ do

1. Sample distribution and get randomized sample vector $\theta$ from

$$\frac{\prod_{i=1}^{A} \theta_i^a_i - 1 (1-\theta_i)^{b_i - 1} \cdot 
\int_D \prod_{i=1}^{A} \theta_i^a_i - 1 (1-\theta_i)^{b_i - 1} d\theta_i}{\int_D \prod_{i=1}^{A} \theta_i^a_i - b_i}.$$ 

2. Select arm $i$ with highest revenue $\arg \max_i p_i \theta_i$

3. Offer price associated with arm index $i$ and observe sales result.

4. If sold, $\mathbf{a} = \mathbf{a} + \mathbf{e}_i$, otherwise $\mathbf{b} = \mathbf{b} + \mathbf{e}_i$

end

Figure 3.5: Thompson Sampling for pricing application with uniform prior

$\mathbf{e}_i$ is the indicator vector with 1 on component $i$ and zero elsewhere. The normalizing factor depends on which distribution is used. Step 1 of the algorithm represents the sampling procedure presented in step 1 of figures 3.1 and 3.3.

Let us consider Thompson Sampling with two different beta-binomial distributions: First, independent distributions as presented in the current literature such as [3], [37], representing demand distribution for discrete prices. It is worth mentioning
that Thompson Sampling has also been adapted for continuous prices using parametric models such as linear systems in [1], [83] and of course [37]. The focus of this paper remains on the discrete price set. We depict the independent distribution as

\[ f_{\text{ind}}(\theta | a, b) = \prod_{i=1}^{q} \theta_{ai}^{-1}(1-\theta_{i})^{b_{i}-1} \prod_{i=1}^{q} \text{Beta}(a_{i}, b_{i}) \].

The second distribution has an ordering constraint, and is represented by

\[ f_{\text{dep}}(\theta | a, b) = \prod_{i=1}^{q} \frac{\theta_{ai}^{-1}(1-\theta_{i})^{b_{i}-1}}{I(a, b)}, \text{ where } I(a, b) = \int_{D \{ 0 \leq \theta_{2} \leq \theta_{1} \leq 1 \}} \prod_{i=1}^{q} \theta_{ai}^{a_{i}-1}(1-\theta_{i})^{b_{i}-1} d\theta_{i} \text{ is a special case of the Euler integral as defined in [42].}

To the best of our knowledge, the methods presented in chapter 1 constitute the only efficient methods for numerical computation of this integral.

In a Bayesian updating setting, the normalizing factor used for each of these distributions is different. Each independent distribution is integrated over the entire domain, which is defined as \( D_{\text{ind}} = \{ 0 \leq \theta_{1} \leq 1, 0 \leq \theta_{2} \leq 1 \} \), whereas the domain of the dependent distribution is limited by the ordering of the variables. Consider a set of prices normalized such that \( 0 \leq p_{1} \leq p_{2} \leq 1 \), the domain of integration is limited to \( D_{\text{dep}} = \{ 0 \leq \theta_{2} \leq \theta_{1} \leq 1 \} \).

TS selects arms based on revenue achieved for a given pair of randomized samples \((\theta_{1}, \theta_{2})\). The probability of higher revenue for arm 2 is expressed as \( Pr(p_{1}\theta_{1} \leq p_{2}\theta_{2} | a, b) \). Define \( z = \frac{p_{1}}{p_{2}}, \text{ with } 0 \leq z \leq 1 \). As shown in figure 3.6, the probability of \( \theta_{2} \geq z\theta_{1} \) with the ordered distribution is confined to the space between the two lines \( \theta_{2} = \theta_{1} \) and \( \theta_{2} = z\theta_{1} \) with domain \( \{ z\theta_{1} \leq \theta_{2} \leq \theta_{1}, 0 \leq \theta_{1} \leq 1 \} \), whereas with the independent distribution it is limited to \( \{ z\theta_{1} \leq \theta_{2} \leq 1, 0 \leq \theta_{1} \leq 1 \} \). Any independent sample above the \( \theta_{2} = z\theta_{1} \) line contributes to selecting arm 2, and since \( z \leq 1 \), the area contributing to selecting arm 2 is always larger than the area for arm 1. The domain of integration is biased toward arm 2. This bias is eliminated with ordered
samples as shown by the rejection zone in figure 3.6a.

(a) Sampling zone for 2 arms with ordering distributions
(b) Sampling zone for 2 arms with independent distributions

Figure 3.6: Comparison of sampling and integration regions for both dependent and independent distributions

For the given set of samples \((\theta_1, \theta_2)\), TS chooses arm 2 over arm 1 with

\[
Pr_{dep}(\theta_2 \geq z\theta_1 \mid a, b) = \frac{\int_0^1 d\theta_1 \int_0^{\theta_1} d\theta_2 \cdot h(\theta, a, b)}{\int_0^1 d\theta_1 \int_0^{\theta_1} d\theta_2 \cdot h(\theta, a, b)}
\]

with the dependent distribution, and

\[
Pr_{ind}(\theta_2 \geq z\theta_1 \mid a, b) = \frac{\int_0^1 d\theta_1 \int_0^{\theta_1} d\theta_2 \cdot h(\theta, a, b)}{\int_0^1 d\theta_1 \int_0^{\theta_1} d\theta_2 \cdot h(\theta, a, b)}
\]

with the independent distribution, where

\[
h(\theta, a, b) = \prod_{i=1}^2 \theta_i^{a_{i-1}} (1 - \theta_i)^{b_i - 1}.
\]

**Lemma 3.7.1.** For any prior vectors \((a, b)\) the probability of selecting arm 2 over arm 1 is a decreasing function of \(z\).

Proof is trivial. \(z\) decreases as the domain of integration becomes smaller.

In fact the probability of choosing an arm can be reduced to a decreasing polynomial in \(z\), computed in closed form for the 2 arm case. For comparison we provide the following equations:

\[
P_{dep}(\theta_2 > z\theta_1 \mid a + e_2, b, b_2) = \frac{a_2 A}{a_2 A - 1} P_{dep}(\theta_2 > z\theta_1 \mid a, b) + \sum_{j=1}^{b_2} \binom{b_2}{j} z^{a_2 + b_2 - j} (1 - z)^j \frac{Beta(a_1 + a_2, b_1 + b_2 - j)}{Beta(a_1 + a_2, b_1 + b_2)(a_2 A - 1)}
\]
3.7. PERFORMANCE ANALYSIS OF A 2-ARM BANDIT PROBLEM WITH TS

where $A = \frac{I(a, b)}{B(a_1 + a_2, b_1 + b_2)}$ and the ratio $\frac{B(a_1 + a_2, b_1 + b_2 - j)}{B(a_1 + a_2, b_1 + b_2)} = \prod_{l=1}^{j} (1 + \frac{a_1 + a_2}{b_1 + b_2 - l})$. With independent distributions the same recursion denoted as $P_{ind}$ is formulated as follows:

\[
P_{ind}(\theta_2 > z\theta_1 | a + e_2, b) = P_{ind}(\theta_2 > z\theta_1 | a, b) + \frac{\sum_{j=0}^{b_2} \binom{b_2}{j} z^{a_2 + b_2 - j}(1 - z)^j \text{Beta}(a_1 + a_2, b_1 + b_2 - j)}{a_2 \text{Beta}(a_1, b_1) \text{Beta}(a_2, b_2)}
\]

Let us consider the special case where $p_1 = p_2$ and $z = 1$, where the consumer should be indifferent in choosing an arm over the other. With the dependent distribution, the probability of $P_{dep}(\theta_2 > \theta_1 | a, b) = 0$ for any given prior. The dependent distribution forces the arm selection process to account for the sample order. The independent distribution does not enforce this ordering. The recursive probability of accumulating a successful pull over arm 2 is as follows:

\[
P_{ind}(\theta_2 > \theta_1 | a + e_2, b) = P_{ind}(\theta_2 > \theta_1 | a, b) + \frac{\text{Beta}(a_1 + a_2, b_1 + b_2)}{a_2 \text{Beta}(a_1, b_1) \text{Beta}(a_2, b_2)}
\]

With a uniform prior $P_{ind}(\theta_2 > \theta_1 | a = 1 + e_2, b = 1) = \frac{1}{2} + \frac{1}{6} = \frac{2}{3}$. Both distributions perform well over long horizons, however with price sensitive applications over short time horizons, independent distributions end up with considerable bias toward a specific arm.

To measure the impact of this bias on performance, we provide the following performance metrics for each distribution. First, we measure the regret for both distributions, which is defined as:

\[
Regret_d(t, y_t, \mu) = E[rev^*(t, y_t) | \mu] - E[rev_d(t, y_t) | \mu]
\] (3.1)
3.7. PERFORMANCE ANALYSIS OF A 2-ARM BANDIT PROBLEM WITH TS

where \( d = "dep" \) when using the dependent distribution and \( d = "ind" \) when using the standard independent distribution. \( \mu \) is the vector of the true parameters. Revenue is a random variable, as it depends on the samples drawn for the demand distribution.

We also use an equivalent form of the regret as proven by [63], where with no capacity constraint, regret can be written as the difference in expected revenues times the expected number of pulls. For the price set \((p_1, p_2)\), and an optimal arm 1 under the condition \( \mu_1 p_1 \geq \mu_2 p_2 \), the regret can be expressed as follows

\[
\text{Regret}_d(t, y, \mu|a, b) = (\mu_1 p_1 - \mu_2 p_2) \times \text{Expected # of pulls of arm 2}.
\]

Another useful measure of performance is the expected regret with respect to \( \mu \), which provides the overall performance of the regret for different price sets. The expected regret, also known as the Bayesian regret, is defined by

\[
\text{BayesRegret}_d(t, y, \mu|a, b) = E_\mu[\text{Regret}_d(t, y, \mu|a, b)].
\]

Under general conditions, the expected revenue with known parameter vector \( \mu \) can be summarized as the following dynamic program:

\[
E[\text{rev}^*(t, y, \mu)] = V_t(y, \mu) = \max_j \{\mu_j(p_j + V_{t-1}(y, 1, \mu)) + (1 - \mu_j)V_{t-1}(y, \mu)\}
\]

\[
= V_{t-1}(y, \mu) + \max_j \{\mu_j(p_j - \Delta V_{t-1}(y, \mu))\}
\]

\[\text{(3.2)}\]

The recursion assumes one arrival in every period. Consumers accept or reject price \( j \) with known probability \( \mu_j \) or \( 1 - \mu_j \). The marginal value is defined as \( \Delta V_{t-1}(y, \mu) = V_{t-1}(y, \mu) - V_{t-1}(y, 1, \mu) \), and the boundary conditions of the dynamic program are as follows: \( V_0(y_0, \mu) = V_t(0, \mu) = 0 \). \( j^*_t(y, \mu) \) is the optimal index, attaining the maximum revenue from (3.2), i.e.,

\[
\mu_j^*(y, \mu)(p_j^*(y, \mu) - \Delta V_{t-1}(y, \mu)) = \max_j \{\mu_j(p_j -
\]
3.7. PERFORMANCE ANALYSIS OF A 2-ARM BANDIT PROBLEM WITH TS

\[ \Delta V_{t-1}(y_t, \mu) \}\}. Given the optimal index and a vector of samples \( \theta = \{\theta_1, \ldots, \theta_q\} \) the expected revenue at time \( t \), for a given set of \( q \) prices and remaining inventory \( y \) is defined as follows:

\[
rev_d(t, a, b, \mu, y_t) = \sum_{j=1}^{q} Pr_d(j^*_t(y_t, \theta) = j|a, b)[\mu_j(p_j +
\text{rev}_d(t - 1, a + e_j, b, \mu, y_t)) + (1 - \mu_j)\text{rev}_d(t - 1, a, b + e_j, \mu, y_t)]
\]

Equation (3.3) can be numerically evaluated, but does not have a closed form solution. To gain insight into the performance of ordered sampling, we develop the closed form solution for the 2-arm, 2-period case and compare results with independent sampling.

3.7.1 Expected number of pulls

To understand the expected number of pulls, let’s take a look at the operation of Thompson Sampling. The algorithm maintains a joint distribution which is initially set based on a prior and subsequently updated based on the decision to offer a price by the retailer and the action from the consumer to accept or reject the product at that specific price. The accumulation of information within this joint distribution helps the algorithm to learn the true parameter of demand \( \mu_i \) for each price and optimize revenue. The algorithm picks a randomized sample from the distribution and chooses a specific price with probability \( Pr_d(j^*_t(y_t, \theta) = j|a, b) \). The consumer accepts the price with probability \( \mu_i \) or rejects it with probability \( (1 - \mu_i) \). The general form for the expected number of pulls of arm \( i \) is given by:
3.7. PERFORMANCE ANALYSIS OF A 2-ARM BANDIT PROBLEM WITH TS

\[ k_{i,d}(t, a, b, \mu, y_t) = \sum_{j=1}^{q} Pr_d(j_t^*(y_t, \theta) = j|a, b) \left[1_{i=j}\right] + \mu_j k_{i,d}(t-1, a + e_j, b, \mu, y_t - 1) + (1 - \mu_j) k_{i,d}(t-1, a, b + e_j, \mu, y_t) \]

For the 2-arm scenario, where arm selection reduces to \( \theta_2 \geq z\theta_1 \), the expected number of pulls on arm 2 when arm 1 is optimal is given by:

\[
k_{2,d}(t, a, b, \mu, z) = Pr_d(\theta_2 \geq z\theta_1|a, b) \left(1 + \mu_2 k_{2,d}(t-1, a + e_2, b, \mu, z) + (1 - \mu_2) k_{2,d}(t-1, a, b + e_2, \mu, z)\right) +
\]

\[
Pr_d(\theta_2 < z\theta_1|a, b) \left(\mu_1 k_{2,d}(t-1, a + e_1, b, \mu, z) + (1 - \mu_1) k_{2,d}(t-1, a, b + e_1, \mu, z)\right)
\]

with \( k_{2,d}(0, a, b, \mu, z) = 0 \). We observe \( k_{2,d}(t, a, b, \mu, 0) = t \), which suggests that if the lower price is zero, the seller draws no revenue from this price, and offers the second price all the time. Similarly, we observe \( k_{2,d}(t, a, b, \mu, 1) = 0 \), which suggests that if the price for both arms is identical, and the demand ordered, the seller will always choose the arm with the higher probability of sales, and the second arm never gets pulled.

**Lemma 3.7.2.** In a two arm, two period setting, the expected number of pulls of arm
2 is monotonic and non-increasing with respect to the ratio of price \( z \), and vice-versa, the expected number of pulls of arm 1 is monotonic and non-decreasing with respect to \( z \).

Please refer to the proof in the appendix.

![Graphs showing expected number of pulls with ordered and independent distributions](image)

(a) Expected number of pulls of arm 1 and 2 with ordered distribution  
(b) Expected number of pulls of arm 1 and 2 with independent distribution with \( \mu_1 = 1 \)

Figure 3.7: Comparison of the expected number of pulls, with uniform prior

The expected number of pulls with the ordered distribution in figure 3.7a may look counter-intuitive at first. A slight difference in price should never create such a disparity between the expected number of pulls of two arms. In fact, figure 3.7b seems like a more reasonable outcome. To see the insight, consider a retailer offering the same product at slightly different prices, with a known higher probability of sales and higher expected revenue at the lower price. Inevitably, the retailer will always choose the lower price and the expected number of pulls at the higher price will be zero. The expected number of pulls for both distributions is as follows:
3.7. PERFORMANCE ANALYSIS OF A 2-ARM BANDIT PROBLEM WITH TS

\[ k_{2,\text{dep}}(2, 1, 1, \mu, z) = 1 + Pr_{\text{dep}}(\theta_2 \geq z\theta_1|a, b) \left[ \mu_2 Pr_{\text{dep}}(\theta_2 \geq z\theta_1|a + e_2, b) + (1 - \mu_2) Pr_{\text{dep}}(\theta_2 \geq z\theta_1|a, b + e_2) \right] - z(1 - Pr_{\text{dep}}(\theta_2 \geq z\theta_1|a, b)) \]

\[ k_{2,\text{ind}}(2, 1, 1, \mu, z) = 1 + Pr_{\text{ind}}(\theta_2 \geq z\theta_1|a, b) \left[ \mu_2 Pr_{\text{ind}}(\theta_2 \geq z\theta_1|a + e_2, b) + (1 - \mu_2) Pr_{\text{ind}}(\theta_2 \geq z\theta_1|a, b + e_2) \right] - z \left( \frac{\mu_1 + 1}{3} \right) (1 - Pr_{\text{ind}}(\theta_2 \geq z\theta_1|a, b)) \]

The effect of arm 1 on the expected number of pulls is given by the last term of equation (3.5). The part in red shows the effect of the bias of selecting arm 1 during the last period. As observed, the expected number of pulls with the dependent distribution is independent of \( \mu_1 \) with a uniform prior. The effect of the probability bias during the first period can be observed by replacing the probabilities for each distribution with the corresponding polynomial:

\[ k_{2,\text{dep}}(2, 1, 1, \mu, z) = 1 + (1 - z) \left[ \left( -\frac{3}{2}z^2 + \frac{3}{2}z \right) \mu_2 + \frac{z^2}{2} - \frac{3z}{2} + 1 \right] - z^2 \]

\[ k_{2,\text{ind}}(2, 1, 1, \mu, z) = 1 + \left( 1 - \frac{z}{2} \right) \left[ \left( -\frac{2}{3}z^2 + z \right) \mu_2 + \frac{z^2}{3} - z + 1 \right] - \frac{\mu_1 + 1}{3} z^2 \]

These equations correspond to the monotonically decreasing graphs in \( z \) as displayed in figure 3.7. The expected number of pulls with the dependent distribution varies between 0 and 2, whereas the expected number of pulls with independent distributions varies between 1 and 2 and depends both on \( \mu_1 \) and \( \mu_2 \).
To measure the impact of the two distributions, we measure the regret and the expected regret. The regret changes depending on which arm is optimal:

\[
\text{regret}_d(t, \mu) = \begin{cases} 
(\mu_1 p_1 - \mu_2 p_2) k_2, d(t, a, b, \mu, z) & \text{for } \mu_2 \leq z \mu_1 \\
(\mu_2 p_2 - \mu_1 p_1) k_1, d(t, a, b, \mu, z) & \text{for } \mu_2 \geq z \mu_1
\end{cases}
\]

with \( k_{i,d}() \) representing the expected number of pulls of arm \( i \) with distribution \( d \). The Bayesian regret discussed earlier may be calculated by integrating the regret function against an unknown parameter vector \( \mu \). The parameter vector \( \mu \) is represented by a beta distribution with prior parameters \((a, b)\). The Bayesian regret for the the two arm case is defined as follows:

\[
\text{BayesRegret}_d(t, y_t) = E[\text{regret}_d(t, y_t, \mu)|z \geq \mu_2] Pr_d(\mu_2 \leq z \mu_1) \\
+ E[\text{regret}_d(t, y_t, \mu)|z < \frac{\mu_2}{\mu_1}] Pr_d(\mu_2 > z \mu_1)
\]

\[
= \int_0^1 \int_0^{\frac{\mu_1}{\mu_2}} \frac{(\mu_1 p_1 - \mu_2 p_2) k_2(2, a, b, \mu, z) \prod_{i=1}^2 \mu_i a_i - 1 (1 - \mu_i) b_i - 1 d\mu_i}{\int_0^1 \int_0^{\frac{\mu_1}{\mu_2}} \prod_{i=1}^2 \mu_i a_i - 1 (1 - \mu_i) b_i - 1 d\mu_i}
\]

We compare the expected regret of the two arm setting with multiple periods with uniform prior for both the dependent and independent probability distributions. To simplify notation we set \( p_2 = 1 \) so \( z = p_1 \). The figure below shows the expected regret for both distributions.

The objective of this paper is certainly not to offer another regret analysis of Thompson Sampling. There is a recent and significant body of literature dealing with regret bounds in it, (e.g., [58], [4] and [37]). And certainly, the expected regret with
3.7. PERFORMANCE ANALYSIS OF A 2-ARM BANDIT PROBLEM WITH TS

independent priors follows those bounds. However, we do attempt to demonstrate that with pricing applications, learning and optimization are affected by the relative order and proximity of the price set. This setup shows that when both arms offer the same revenue, the expected regret (Bayesian regret) should be close to zero, however, this can only be achieved if the ordering of the prices is taken into consideration.

In a two price setting, our model may be useful when experimenting with small discounts over short horizons. We first observe that the expected regret with independent priors is non-decreasing in $z$ or $p1$, whereas with the dependent distribution it is concave. The expected regret is minimal with extreme price settings, both when prices are very close or very far apart from each other. The biggest difference between the two distributions is observed when price discounts are small. Second, this behavior is consistent over longer periods. To get a better understanding, we focus on the regret for each distribution and make a visual comparison to better understand the fundamental differences between them.

Figure 3.9 shows the regret function for $\mu_1 = 1$. The green and red curves represent the regret with the dependent distribution, whereas the blue and yellow curves represent the regret with independent distributions. We observe the indifference line
3.7. PERFORMANCE ANALYSIS OF A 2-ARM BANDIT PROBLEM WITH TS

(a) Regret with $\mu_1 = 1$. Dependent dist. is represented by the green and red curves. Independent dist. is represented by the blue and yellow curves.

(b) Ratio of regrets of dependent over independent distributions. Arm 1 optimal zone represented by the green curve, and arm 2 optimal zone by the red curve.

Figure 3.9: Comparison of the regret function, with uniform prior, and different values of $\mu_1$

where regret is zero when revenue from both arms are identical. We observe that the regret with dependent distributions is close to zero when $p_1$ approaches 1. This makes the dependent regret symmetric around the $z\mu_1$ line. This symmetry keeps the regret function balanced over the domain of integration.

The regret ratio shows the ratio of regrets of the dependent over independent distributions. The green curve shows the ratio of the regrets when arm 1 is optimal, and the red curve corresponds to when arm 2 is optimal. The difference in regrets when $z$ is close to 1 is visible in the ratio. The red curve tends to zero when $p_1$ approaches 1. This explains the behavior of the expected regret when price differences are small.
3.7.2 Expected regret with capacity constraint

We’ve characterized the performance of the regret and the expected regret for both distribution models without any capacity constraint. The question extends naturally to the capacity constrained model, with the algorithm discussed in [37] and in the earlier section of this paper. Similarly to the previous section, we solve an exact stylized problem with a 2-arm, 2-period setting with a single item to sell. We are interested in the relative behavior of the regret over 2 periods.

We analyze the probability of sale under the constrained regime. The domain of integration changes considerably as shown below:

**Lemma 3.7.3.** In a 2 arm 2 period setting with a single item to sell, and given the true demand parameters \( \mu_1 \) and \( \mu_2 \), the probability of choosing the optimal arm is defined by the following boundary condition:

\[
\frac{z_{\mu_1}(1-\mu_1)}{1-z_{\mu_1}}.
\]

Please refer to the proof in the appendix.

Unlike the unconstrained formulation where the probability of pulling an arm depends only on the revenue \( \theta_2 \geq z\theta_1 \), the constrained formulation adjusts the boundary according to the conditions derived from the optimal value function as shown in the appendix. In the first period, with 1 item to sell, the probability of choosing arm 2 is expressed by \( Pr(\theta_2 \geq \frac{(1-\theta_1)\theta_1}{1-z\theta_1}) \), and the expected revenue from the algorithm for
3.7. PERFORMANCE ANALYSIS OF A 2-ARM BANDIT PROBLEM WITH TS

the first period is as follows:

\[ \text{rev}_d(t = 2, a, b, \mu, y_t = 1) = \]

\[
Pr_d \left( \theta_2 \geq \frac{(1 - \theta_1)z\theta_1}{1 - z\theta_1} | a, b \right) \left[ \mu_2p_2 + (1 - \mu_2)\text{rev}_d(t = 1, a, b + e_2, \mu, y_t = 1) \right] + \\
Pr_d \left( \theta_2 \leq \frac{(1 - \theta_1)z\theta_1}{1 - z\theta_1} | a, b \right) \left[ \mu_1p_1 + (1 - \mu_1)\text{rev}_d(t = 1, a, b + e_1, \mu, y_t = 1) \right]
\]

The boundary conditions and the last period value function are similar to the unconstrained case, and are provided below for reference:

\[ \text{rev}_d(t = 0, a, b, \mu, y_t) = V(t, a, b, \mu, y_t = 0) = 0 \]

\[ \text{rev}_d(t = 1, a, b, \mu, y_t = 1) = \mu_1p_1 + (\mu_2p_2 - \mu_1p_1)Pr(\theta_2 \geq z\theta_1 | a, b) \]

The regret function varies depending on the following 3 cases. As shown in the proof of lemma (3.7.3), the integration domain is bounded by the optimality condition of each arm. Arm 1 is optimal when \( \mu_2 \leq \frac{(1 - \mu_1)z\mu_1}{1 - z\mu_1} \). Arm 2 is optimal when \( \mu_2 \geq \frac{(1 - \mu_1)z\mu_1}{1 - z\mu_1} \), however the value function and the regret change when \( \mu_2 \geq z\mu_1 \).

\[
\text{regret}_d(t = 2, y_t = 1, \mu) = \\
\begin{cases} 
2z\mu_1 - z\mu_1^2 - \text{rev}_d(t = 2, a, b, \mu, y_t = 1) & \text{for } \mu_2 \leq \frac{z\mu_1(1 - \mu_1)}{1 - z\mu_1} \\
z\mu_1 + \mu_2 - z\mu_1\mu_2 - \text{rev}_d(t = 2, a, b, \mu, y_t = 1) & \text{for } z\mu_1 \geq \mu_2 \geq \frac{z\mu_1(1 - \mu_1)}{1 - z\mu_1} \\
2\mu_2 - \mu_2^2 - \text{rev}_d(t = 2, a, b, \mu, y_t = 1) & \text{for } \mu_2 \geq z\mu_1 
\end{cases}
\]
3.8 Conclusion and future work

The expected regret for the constrained and unconstrained models have similar forms as shown in figure 3.10. The main difference is in the relative scale, as the expected regret for the unconstrained model is 2 to 3 time larger than the constrained model.

![Figure 3.10: Comparison of dependent (green) and independent (blue) constrained expected regret function, with uniform prior](image)

Figure 3.10: Comparison of dependent (green) and independent (blue) constrained expected regret function, with uniform prior

### 3.8 Conclusion and future work

Over the last decade, Thompson Sampling has become an increasingly popular heuristic thanks to its asymptotic optimality and low computational complexity. In this paper, we analyzed the heuristic for pricing applications with short horizons. This has practical implications, as some retailers don’t discount products for a long time. We used the Metropolis-Hastings algorithm with a Dirichlet kernel to construct and update ordered samples from a multi-variate beta-binomial distribution. We compared our results to those from [37] for shorter horizons, and found 10% to 30% improvements using ordered samples. The intuition behind the ordered samples is simple and works well for pricing applications. The ordering structure reduces the number of trials required to learn the demand distribution.
In the 2 price setting, our results show the importance of the ordered distributions when prices are close to each other and discounts are small. The Bayesian regret with the ordered model and small discounts is negligible, whereas with independent distributions it is an increasing function as discounts decrease. This makes the use of the ordered distributions even more important, as small discounts of 5% to 10% over short horizons are common.

Future research will focus on extending the model to an arbitrary number of prices, and longer horizons. This will answer the question of what the optimal range for using the ordered samples is. It would also be interesting to expand on the Metropolis-Hasting algorithm, and compare the overall performance of the algorithm with respect to the existing parameters.
3.9 Summary of variables

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_i$</td>
<td>Price index i</td>
</tr>
<tr>
<td>q</td>
<td>Number of prices</td>
</tr>
<tr>
<td>$\mu_i$</td>
<td>Probability of success for price i</td>
</tr>
<tr>
<td>$(a, b)$</td>
<td>Vector of prior parameters</td>
</tr>
<tr>
<td>$(m, n)$</td>
<td>Information vector</td>
</tr>
<tr>
<td>t</td>
<td>Time index</td>
</tr>
<tr>
<td>T</td>
<td>Time horizon</td>
</tr>
<tr>
<td>$\theta = (\theta_1, \ldots, \theta_q)$</td>
<td>Vector of random samples</td>
</tr>
<tr>
<td>$f(\theta_i</td>
<td>a_i, b_i)$</td>
</tr>
<tr>
<td>$v_i$</td>
<td>Convex weight for arm i</td>
</tr>
<tr>
<td>$y_t$</td>
<td>Remaining inventory at time t</td>
</tr>
<tr>
<td>M</td>
<td>Number of M-H chains</td>
</tr>
<tr>
<td>K</td>
<td>Length of M-H chain</td>
</tr>
<tr>
<td>S</td>
<td>Sensitivity parameter for the M-H algorithm</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Parameter of the Dirichlet distribution</td>
</tr>
<tr>
<td>$\gamma_j$</td>
<td>Sample difference between price index j-1 and j</td>
</tr>
<tr>
<td>$x^*$</td>
<td>Sample vector from the Dirichlet distribution</td>
</tr>
<tr>
<td>$\xi$</td>
<td>Random samples from the Gamma distribution</td>
</tr>
<tr>
<td>$\theta^*$</td>
<td>Candidate state in the M-H algorithm</td>
</tr>
<tr>
<td>$L(\theta)$</td>
<td>Posterior probability function</td>
</tr>
<tr>
<td>$Q(\theta^*</td>
<td>\theta)$</td>
</tr>
<tr>
<td>$D_d$</td>
<td>Domain of integration for d={dependent,independent} distributions</td>
</tr>
<tr>
<td>$f_d(\theta</td>
<td>a, b)$</td>
</tr>
<tr>
<td>$I(a, b)$</td>
<td>Euler integral</td>
</tr>
<tr>
<td>z</td>
<td>Ratio of $p_1$ over $p_2$</td>
</tr>
</tbody>
</table>

Table 3.1: Table of variables and notation
Chapter 4

Balancing Equity and Cost in Rural Transportation Management with Multi-Objective Utility Analysis and Data Envelopment Analysis: A Case of Quinte West

Rural transportation management plays a critical role in the sustainable future of human society. Two emerging challenges faced by rural communities today are cost control and equity due to the increasing demand and limited operations resources available and the need to deal with the inevitable tradeoffs among multiple objectives and criteria. In this paper, we develop a new methodology for rural transportation management which takes into consideration of both the equity and cost factors under multiple objectives. We conceptualize and define equity in rural transportation management with the development of new performance measures and an analytical model for decision making with multiple desirable and undesirable objectives. We also develop a heuristic procedure based on data envelopment analysis for characterizing and analyzing the route design choices on the frontier between costs and multi-objective
measures of equity. A GIS-based decision support system is constructed to process the extensive data required in our analysis. The new methodology has been successfully implemented by Quinte Access, a not-for-profit organization in a rural community in Ontario, to help redesign bus routes with significant quantitative benefits observed in multiple performance dimensions. It is also expected that the new knowledge, insights, and decision support tools developed through this study for transportation planning, big data management, and transportation service operations can be transferred to other rural communities in order to deliver more sustainable transportation services in rural regions around the world.
4.1 Introduction

Rural transportation systems around the world have been severely challenged in recent years due to the limited operations resources available and the increasing demand as a result of aging population in rural regions. Traditionally, research in transportation operations and supply chain management focuses mostly on the industrial and urban settings (e.g., [27], [54], [5], [96]) with little attention given to rural communities. However, it is important to note that there is no way to separate the future of rural regions from the rest of the global economy, particularly as it relates to mobility. Two of the emerging issues in rural mobility are residents’ access to healthcare, and the fact that many elderly citizens are unable to drive. Healthcare is an essential service, and accessibility to healthcare facilities is thus of great importance to citizens in rural communities. In addition, the increasing proportion of senior population in rural communities has also created an urgent need for sustainable public transit systems as many senior citizens become unable to drive as they get older. This makes tasks such as grocery shopping, regular pharmacy visits, doctor appointments, and visiting friends and family increasingly difficult. Rural public transportation services are also important for disabled citizens and low income families. As stated by Dale Marsico before the U.S. Congress ([75], Testimony to the U.S. House Appropriations Committee), rural transportation management is always vital—but it is especially important today because of the terrific pressure on our transportation system created by the current economic crisis, the changing ways in which healthcare is being delivered, an aging population that requires additional needs to remain in the places they call home, and the potential needs of immediate relief when disasters strike. All these situations combined create a significant challenge that requires more research.
in sustainable rural transportation management.

Today, two biggest challenges faced by rural communities in transportation service management are cost control and equity ([36]). For example, it is reported that the operating expenses for demand responsive transport services in the United States have more than doubled in the past ten years due to the increasing demand and operations costs. The soaring expenses, coupled with the fact that many public rural transit systems have been significantly underfunded due to the recent economic downturn, have thus forced decision makers in rural communities to prioritize their services because the existing resources are not sufficient to satisfy the demand of every citizen in need of transportation services. The limited transportation service capacity in rural regions thus inevitably creates an emerging issue of equity concerning how to operate and prioritize rural transportation services, such as regular buses and demand responsive transportation (dial-a-ride), to satisfy only a subset of the total demand based on some measures of multiple objectives and criteria ([36]).

In this paper, we develop a new methodology for designing and managing rural transportation services which takes into consideration of both the equity and cost factors under multiple objectives. We conceptualize and define equity in rural transportation management with the development of new performance measures and an analytical model for decision making with multiple desirable and undesirable objectives. We also develop a heuristic procedure based on data envelopment analysis for characterizing and analyzing the route design choices on the frontier between costs and multi-objective measures of equity. To process the extensive data required in our analysis, we utilize GIS (Geographic Information System) as a big data platform
4.1. INTRODUCTION

to develop a decision support system for compiling, exporting, importing, and synchronizing data and analytical results. Figure 4.1 presents an example of geographic information (coordinates, population, numbers of different points of interest, etc.) retrieved from GIS and the utility and accessibility scores derived from multi-objective utility analysis for a particular grid cell built in the decision support system. All these data can be quickly obtained and displayed by a simple click on the grid cell of the Google Earth file in KML format. (Refer to the geographic, demographic, and utility data for all the 3,359 grid cells in Quinte West in the submitted Google Earth file entitled “GIS Decision Support System.”)

The new methodology has been successfully implemented by Quinte Access, a not-for-profit organization in a rural community in Ontario, to help redesign bus routes with significant quantitative benefits observed in multiple performance dimensions including the total ridership, the total multi-objective utility score, total population covered, and rural accessibility score per minute of bus service time, respectively. Data envelopment analysis is also used to further verify the quantitative benefits with multiple objectives by showing that two of the three new routes are non-dominated “efficient designs.” The remainder of the paper is organized as follows. Relevant literature will be reviewed in Section 2. In Section 3, we will first conceptualize and define transport equity, and then present an analytical model and a heuristic procedure for solving the multi-objective problem with both desirable and undesirable factors. The case study will be presented in Section 4, and concluding remarks are given in Section 5.
4.2. RELATED LITERATURE

Various management science methods have been proposed to analyze different issues related to transportation service design and management (e.g., [106], [17], [95], [48], [87]). Most analytical models in the existing literature, however, focus on a single objective, such as transportation distance or time, without jointly taking resource scarcity and equity into consideration. A number of algorithms and heuristic procedures for analyzing vehicle routing problems with multiple objectives have been developed with the scalar methods, which transfer a multi-objective problem into a
single objective problem (e.g., [15], [65]), and the Pareto methods, which search for solutions along the optimal “Pareto efficient frontier” of multiple objectives ([104], [86]). An excellent review is given in [56]. Objectives which have been considered in the existing studies include cost, distance, makespan, time, satisfaction, risk, and emissions (e.g., [88], [44], [51], [26], [5]). [53] evaluate “the spread in service level across nodes in a solution” as a technical measurement of equity. [103] develop two-stage data envelopment analysis models to explore determinants of technical, allocative, and cost efficiency with data of twenty international urban rail systems. Unlike our approach, however, they do not consider the heterogeneity across different regions with different attributes as the multi-objective measurement of equity for rural transportation management.

The issue of balancing cost and equity, the main topic of this paper, has been subject to much research in the wider context of urban service delivery. [71] takes an in-depth look at the role of transporation and its effect on social exclusion. It highlights the wider issue of equity in transportation services. In a follow up paper [72] reflects on the progress and evolution of policy and the state of the agenda around social exclusion, and concludes that “Metrics are needed to establish the minimum level and standards of public transport which are necessary for social inclusion given certain distances, densities, level of services, etc. and local target set to achieve these within given timewframes”. [82] evaluates and compares different measure of equity by evaluating the properties of place-based and people-based accessibility measures. [41] takes a closer look at the availability of public transit systems, and concludes that much work is needed to provide transportation services where it is needed.
4.2. RELATED LITERATURE

Two major constraints faced by researchers in jointly analyzing the scarcity and equity issues are information availability and complexity since the required data such as population density, traveling distance, geographical features of different regions, and stakeholders’ interests are typically difficult to obtain and synchronize, especially for rural communities ([36]). However, as will be shown in our analysis, the recent advance of GIS, which provides a unique platform to integrate the big data required to perform analyses under multiple objectives, has created new opportunities for researchers and practitioners to tackle the emerging issues of resource scarcity and equity in rural transportation management ([90]). Another major issue faced by researchers and practitioners is performance evaluation under multiple objectives. Equity in transportation research is generally defined as the “equitable levels of access to basic amenities” ([36]), which will be incorporated in our proposed methodology for evaluating the relative values of multiple amenities or points of interests to assess equity on a consistent basis for rural transportation planning and management.

Solving multi-objective problems typically involves generating non-inferior sets first and then applying different techniques to identify the best compromise solution(s) (e.g., [20], [35], [80]). Various methods have been proposed to obtain non-inferior solutions, which include the weighting method (e.g., [108], [73]), the epsilon-constraint method (e.g., [74], [22]), the noninferior set estimation method ([21]), and multi-objective linear programming (simplex) method (e.g., [89], [109]). The weighted sum technique, which converts a multi-objective problem into a single-objective problem, is perhaps the most straightforward way to obtain the compromise solution(s). How to obtain the weights to correctly articulate the preferences of decision makers, however, represents a major challenge. Different techniques for objective weight assessment
with multi-attribute utility functions, such as portfolio analysis and multi-objective utility analysis, have been proposed by researchers (e.g., [45], [59], [20]). [49] develop the surrogate worth tradeoff method based on the assumption that decision makers are able to exhibit an implicit multi-attribute utility function obtained from systematically comparing two attributes at a time. The goal programming method proposed by [18] allows decision makers to specify a target for each objective function to identify a preferred solution that minimizes the sum of the deviations from the prescribed set of target value. Similarly, [107] and [109] develop methods to identify the best-compromise solution that minimizes the distance from a pre-determined idea solution.

Data envelopment analysis (DEA), the approach adopted in this paper, can be viewed as an extension of the above-mentioned multi-objective optimization methods/techniques. Specifically, searching the efficient frontier in a DEA model can be formulated as a multi-objective linear programming problem ([18]). The procedure for identifying the set of weights which maximize the efficiency of each decision making unit in DEA may be characterized as a method of objective weight assessment ([38]). In recent years, DEA has been applied to performance evaluation for various transportation problems, such as airport operations management, airline efficiency evaluation, and rail systems performance measurement (e.g., [78], [79], [77], [103], [76]). To the authors’ best knowledge, however, the current paper proposes the first integrated approach with a novel DEA-based heuristics procedure for balancing the equity and cost objectives in rural transportation management. In the next section, we will present our integrated methodology with multi-objective utility analysis and data envelopment analysis.
4.3 Methodology

To introduce our proposed methodology, we will first conceptualize transport equity with multi-objective performance measures. We will then present the analytical framework and heuristic procedure for modeling and solving the multi-objective problem for our study.

4.3.1 Measuring Equity with Multiple Objectives

Based on the general conceptualization of equity for transit and land use planning given in [36], we define equity in rural transportation operations management as “the equitable levels of accessibility to major points of interest in a rural community.” This definition is consistent with those in the existing literature of public transport planning which generally include two major dimensions: (i) need (measured with region-specific factors or variables) and (ii) access (measured with the proportion of population covered) ([81]). While the assessment of access can be done with population data from GIS, need assessment in our project requires a consistent measurement of the relative values of different points of interest (amenities) in a rural community. This is accomplished with Multi-Objective Utility Analysis (MOUA) or Conjoint Analysis, a statistical technique commonly used for market research which involves tradeoffs among multiple objectives and criteria ([67], [59], [68]). The main objective is to determine the importance, or utility, of different levels of multiple objectives (factors) considered in a decision. For example, factors considered for rural transport accessibility could include different points of interest or amenities, such as hospitals,
4.3. METHODOLOGY

retirement residences, schools, and grocery stores, in a given area. Since each factor in our study can be viewed as a continuous variable with which preference will increase as the quantity of the factor (i.e., the total number of units of a particular point of interest) increases, the Vector Model is adopted in our analysis ([67]). MOUA consists of generating different combinations (product profiles) of factors at different levels and having a respondent assign a unique score to each combination. Statistical techniques are then used to estimate the specific utility value for each factor based on the overall respondent preference.

To structure the MOUA problem, the entire region of a community is first divided into multiple grid cells of the same size (e.g., 400 meters by 400 meters, which is typically considered as an acceptable walking distance in public transport planning). Information about the “coordinates,” population, and attributes (number of schools, number of hospitals, number of stores, etc.) of each of the grid cells can then be obtained from GIS to construct an “attribute database.” Each of the grid cells in the community is considered as a unique “product profile” with different combinations of factors (attributes) at different levels. The multi-objective utility score (MOUS) for a grid cell \( k \), denoted by \( \text{MOUS}_k \), is defined as

\[
\text{MOUS}_k = \sum_{b=1}^{B} u_b A_b^k
\]  

(4.1)

where \( b = \) index of factors (points of interest), \( b = 1, \ldots, B \),

\( B = \) total number of factors,

\( u_b = \) utility value of each unit of factor \( b \),

\( A_b^k = \) the total number of units of factor \( b \) in grid cell \( k \).
4.3. METHODOLOGY

To obtain the values of \( u_b \) in (4.1), a questionnaire needs to be completed by key decision makers and stakeholders in a rural community, and the results are used in the following regression:

\[
U_p = \sum_{b=1}^{B} u_b A_b^p + \epsilon_p
\] (4.2)

where \( p \) represents a particular combination (product profile) of selected factors and levels for a “virtual grid cell” generated by the system, and \( U_p \) is the score provided by a respondent for \( p \) in utility assessment while \( \epsilon_p \) is the error terms with normal distribution, zero mean, and variance equal to the square of the standard deviation for all \( p \) ([67], [68]).

In the existing literature of rural service planning, the gravity-type methods, which define accessibility as an increasing function in population and attributes and a decreasing function in distance, have been widely used as a performance measure ([70], [57], [90]). We thus define the rural accessibility score (RAS) for grid cell \( k \) as the combination of MOUS and population as follows:

\[
RAS_k = \left[ \sum_{b=1}^{B} u_b A_b^k \right] \times POP_k
\] (4.3)

where \( POP_k \) is the population in grid cell \( k \). The rural accessibility score provides
4.3. METHODOLOGY

a basic measurement of equity in rural transportation operations management, which is defined earlier as the equitable levels of accessibility (measured by population) to major points of interest (measured by MOUS). It is noted that our analysis can be easily modified with $POP_k$ representing only a particular subgroup of population, such as senior residents with disabilities, employers in need of public transport, and low-income families, for communities with different needs and priorities ([100], [93]).

For a route that connects $Q$ different grid cells, indexed by $q = 1 \cdots Q$, the multi-objective accessibility score (MOAS) is defined as

$$MOAS = \sum_{q=1}^{Q} \frac{RAS_q}{DIS}$$

(4.4)

where $DIS$ is the total distance travelled to cover all the $Q$ grid cells within the road network. The gravity-based MOAS defined in (4.4) provides a simple integrated measurement for decision makers to compare the potential benefits (utility and accessibility scores) and costs (distances) of different routes. It should be noted, however, that there exists a potential issue associated with this gravity-type performance measure because the resulting scores may sometimes be skewed by measurement scales and units. Therefore, while we will still report the multi-objective rural accessibility scores of different routes to allow meaningful comparison between our research outcomes and those derived from the existing methods, our proposed heuristic procedure will primarily be based on data envelopment analysis, which can be used with different input-output measurements, to avoid the potential
scaling issue. The specific analytical framework used to analyze the multi-objective optimization problem for our study will be presented next.

### 4.3.2 Analytical Framework

Consider a multi-objective optimization problem where $s$ objectives are desirable (e.g., population covered, utility scores, etc.) while $m$ objectives are undesirable (e.g., travel time, distance, environmental impacts, etc.)\(^1\). To be consistent with the problem we study, assume $m \geq 1$ and $s \geq 1$. Group all the objective functions into undesirable ones ($i = 1, \cdots, m$) and desirable ones ($j = 1, \cdots, s$). The multi-objective optimization problem, modified from [56], for our study can be stated as follows:

$$
\min F(x) = [f_1(x), f_2(x), \cdots, f_m(x), -g_1(x), -g_2(x), \cdots, -g_s(x)]
$$

subject to $x \in D$

where $f_i(x) = \text{objective function for an undesirable objective, } i = 1, 2, \ldots, m,$

$g_j(x) = \text{objective function for a desirable objective, } j = 1, 2, \ldots, s,$

$x = (x_1, x_2, \cdots, x_r) = \text{the decision variable vector with } r \text{ as the total number of route segments under consideration}$

$D = \text{the feasible solution space},$

---

\(^1\)The total service time may not be proportional to distance due to different speed limits on different roads and the fact that a route can cover multiple grids in various ways with different numbers of intermediate stops depending on the road network.
4.3. METHODOLOGY

\( F(\mathbf{x}) \) = the objective vector.

Let \( O = V(D) \) denote the set that corresponds to the feasible solutions in the objective space, and \( \mathbf{y} = (y_1, y_2, \cdots, y_{m+s}) \) denote a solution of the objective space, where \( y_k = f(\mathbf{x}), k = 1, \ldots, m \) and \( y_k = g_{k-m}(\mathbf{x}), k = m + 1, \ldots, m + s \). The solution to the above multi-objective problem is the set of non-dominated solutions called the “Pareto set,” where dominance is defined as:

Definition. A solution \( \mathbf{y} = (y_1, y_2, \cdots, y_{m+s}) \) dominates another solution \( \mathbf{z} = (z_1, z_2, \cdots, z_{m+s}) \) if and only if (i) \( \forall k \in \{1, \cdots, m+s\}, y_k \leq z_k \) for \( k = 1, \cdots, m \) and \( y_k \geq z_k \) for \( k = m + 1, \cdots, m + s \), and (ii) \( \exists k \in \{1 \cdots, m + s\}, y_k < z_k \) for \( k = 1, \cdots, m \) or \( y_k > z_k \) for \( k = m + 1, \cdots, m + s \).

Based on the theory of decision with multiple objectives ([59]), all the non-dominated solutions form a “Pareto efficient frontier” along which no solution is dominated according to the above definition of dominance.

As mentioned previously, the scalar methods and Pareto methods are two commonly adopted optimization algorithms for solving multi-objective problems [56]. The scalar methods, which typically use weighted linear aggregation to transfer the multi-objective problem into a single objective problem, is rather simple to implement. However, there are several disadvantages associated with the scalar techniques.
First, weights must be arbitrarily selected, which can be a rather challenging task. It has also been shown that the scalar techniques only consider solutions on the convex hull of the optimal Pareto set as opposed to all the Pareto optimal solutions ([43]).

Given the shortcomings of the scalar techniques, the Pareto methods, which apply the notion of Pareto dominance directly, have become more popular in recent years ([56]). As a Pareto method, our proposed heuristic procedure is based on the “Pareto Local Search” concept ([104], [86]) through data envelopment analysis (DEA). DEA is a nonparametric method for establishing the Pareto efficient frontier among all the possible solutions (alternatives) for a multi-objective problem. As a nonparametric method, DEA has the advantage of uncovering the relationships among multiple objectives without the use of any explicit mathematical function of the Pareto efficient frontier. Each possible solution \( y = (y_1, y_2, \cdots, y_{m+s}) = (f_1(x), \cdots, f_m(x), g_1(x), \cdots, g_s(x)) \) in the objective space is considered as a decision making unit (DMU) in DEA where the desirable objectives \( g_j(x), j = 1, \ldots, s \) are considered as outputs and the undesirable objectives \( f_i(x), i = 1, \ldots, m \) are considered as inputs ([5]). Suppose that there are a total of \( N \) possible solutions under evaluation, denoted by \( y_n = (y_{1,n}, y_{2,n}, \cdots, y_{m+s,n}) = (f_{1,n}, \cdots, f_{m,n}, g_{1,n}, \cdots, g_{s,n}) \), \( n = 1, \ldots, N \). Various forms of DEA models have been developed over the years. To evaluate the relative efficiency of a possible solution \( y_o, o \in \{1, \cdots, N\} \), the input-oriented constant return to scale (CRS) model [18] can be formulated as follows:
Based on [24] and [23], the above CRS model can be converted to a variable return to scale (VRS) model [8] by adding the following convexity condition:

\[
\sum_i \eta_i + \sum_j \mu_j = 1
\] (4.5)

The efficiency of each possible solution in DEA is defined as a ratio (i.e., the weighted sum of desirable objectives divided by the weighted sum of undesirable objectives), adjusted to be a number between 0 and 1, where \(\mu_j\) and \(\eta_i\) are the weights of multiple objectives. The less undesirable objectives and the more desirable
objectives, the more efficient is a possible solution. It is important to note that
the weights used in DEA are not arbitrarily determined in advance. Rather, the
weights for each solution are obtained from the above model formulation. This is
the unique feature of DEA and introduces the required element of objectivity which
is missing when applying arbitrary weights. The efficiency scores of all the decision
making units can then be obtained by solving a series of linear programs. Those
“efficient solutions” which attain an efficiency of 1 or 100% form a mathematical
space (the Pareto efficient frontier) that “envelops” all the other solution points.
Although difficult to visualize in the multi-objective space, the Pareto efficient
frontier is very precisely defined in DEA to allow differentiation between efficient
and non-efficient solutions. The efficiency score for each possible solution (which
is less than or equal to 1 or 100% ) indicates how far the solution is away from
the Pareto efficient frontier along the solution vector in the objective space. This
provides a useful way to measure and characterize the suboptimality of different
solutions for a multi-objective optimization problem against the Pareto efficient
frontier ([5]). Another advantage of DEA is the capability to synchronize different
input/output measurements on a consistent basis to avoid the previously mentioned
issues associated with scaling and different measurement units ([34]).

It is noted that the above problem can also be formulated as an output-oriented
model ([24], [23]). To choose between the input- and output-oriented models, a general
rule of thumb is that, if decision makers have more control or higher influence on the
inputs, an input-oriented model should be used, and vice versa ([79]). It should
also be noted that the envelopment frontier will differ depending on different scale
assumptions - constant returns to scale (CRS) or variable returns to scale (VRS). CRS reflects the fact that output will change by the same proportion as inputs are changed; VRS reflects the fact that production technology may exhibit increasing, constant, and decreasing returns to scale ([94], [24]). Another potential issue is whether the efficiency scores obtained from the traditional DEA are biased as various bootstrapping methods have been proposed to derive the bias-corrected efficiency scores ([97], [98], [79], [76]). The specific DEA-based heuristic procedure will be discussed next.

### 4.3.3 Heuristic Procedure

Determining the optimal solution(s) of a vehicle routing problem is known as an NP-complete problem in combinatorial optimization which requires extensive computational effort ([46]). Unlike most vehicle routing problems which optimize only the travel distance, our analysis further requires the optimization of multiple objectives, including distance as the traditional performance measure and the total multi-objective utility score and population covered, without exceeding the total service time constraint $T$. We thus propose a heuristic procedure which starts with distance minimization as the single objective, and then applies a DEA-based process to continuously improve a route design with multiple objectives.

As the starting point, the main stops which are strategically important for a rural community (e.g., city hall, Wal-Mart store, industrial park) of an initial new route design can be specified by the decision and policy makers. Consider all the main stops in a particular route specified in the initial route plan as the required nodes of

\[^2\text{As discussed in [94], the test results obtained from input- and output-based efficiency measures are equivalent under the assumption of constant returns to scale.}\]
4.3. METHODOLOGY

A weighted graph $G = (V, E)$ with vertex (node) set $V = \{v_1, \cdots, v_S\}$ and edge set $E$, where $S$ is the total number of main stops. Each edge $v_hv_l \in E$ is characterized by the distance $d_{hl}$, the average service time $t_{hl}$, as well as $POP_{hl}$, $MOUS_{hl}$, and $MOAS_{hl}$, representing the total population and multi-objective utility and accessibility scores, respectively, between two vertices $v_h$ and $v_l$, where $v_h, v_l \in V$. In the proposed heuristic procedure, we first use Dijkstra algorithm ([110], [28]) to find the original solution (path) on each edge that minimizes the travel distance as the single objective within the road network of a community.

While the distance between each pair of nodes (main stops) in the original solution is minimized with Dijkstra algorithm, it is possible to find an alternative path between the nodes which covers a different set of grid cells with higher total MOUS, MOAS, and/or population covered but longer travel distance. Specifically, the route between two required nodes covers a number of grid cells depending on the chosen path. Each grid cell with non-zero utility score or population on an alternative path between two required nodes adds to the total utility score or population of the edge by the amounts equal to the MOUS or population density of the grid cell as an intermediate stop. The total MOUS, population, and MOAS between two required nodes can be calculated as $MOUS_{hl} = \sum_k MOUS_k$, $POP_{hl} = \sum_k POP_k$, and $MOAS_{hl} = \sum_k \frac{RAS_k}{d_{hl}}$, where $k$ represents grid cells located on a path between required nodes $h$ and $l$.

Figure 4.2 presents an example of swapping the original path (solid line) with an alternative path (dash line) in a particular area of Quinte West in our case study, which will increase the total travel distance from 1,450 m to 3,560 m. However, the swapping will also increase the total MOUS from 57.7 to 168, the total population from 365 to 1,738, and total MOAS from 5,264 to 32,131. The average service time will
4.3. METHODOLOGY

also increase from 5.2 minutes to 11.4 minutes due to additional traveling time and the average loading/unloading time at each intermediate stop. If the additional service time is within the total service time constraint, it is possible to use the alternative route that would result in a better combination of multiple objectives taking into account the total travel distance (the original criterion), MOUS, and population (the new criteria), subject to other physical constraints such as traffic restrictions and street conditions.

![Figure 4.2: An Example of Original and Alternative Routes](image)

To identify the best path to swap, we utilize DEA based on the model defined in (4.5) and in the process of improving the route design with multiple objectives. Each feasible path is considered as a decision making unit (DMU) with MOUS and population as the outputs and travel distance and service time as the inputs. The population covered and multi-objective utility score, as the measures of equitable levels of access to different points of interest in a rural community, are reasonable outputs for DEA.
since they are the two major factors which the community would like to maximize. The distance and service time are reasonable measures of inputs for DEA because they are directly related to the costs of fuel, labor (drivers), vehicle depreciation, and other relevant operational expenses. DEA can thus be used to measure and characterize how the limited resources (inputs) are utilized to achieve multiple objectives (outputs) on the cost/equity frontier. The detailed heuristic procedure used in our study is described as follows:

1. Within the road network, for each pair of required nodes $h$ and $l$, use Dijkstra algorithm to identify the shortest path that minimizes the travel distance. Let $\{ d_{hl}^0, MOUS_{hl}^0, POP_{hl}^0, t_{hl}^0 \}$ represents the set of distance, MOUS, population covered, and average service time for the shortest path. Calculate the remaining idle time on the route: $T - \sum_{h} \sum_{l} t_{hl}^0$, where $T$ is the specified maximum service time.

2. The path between each pair of required nodes $i$ and $j$ may be represented by a limited set of $w + 1$ feasible routes denoted by

$$v_{hi}v_{il} = \{ \{ d_{hi}^0, MOUS_{hi}^0, POP_{hi}^0, t_{hi}^0 \}, \{ d_{hi}^1, MOUS_{hi}^1, POP_{hi}^1, t_{hi}^1 \}, \ldots, \{ d_{hi}^w, MOUS_{hi}^w, POP_{hi}^w, t_{hi}^w \} \}$$

where $w$ is the number of alternative routes other than the original shortest route.

3. Use $MOUS_{hl}^k$ and $POP_{hl}^k$ as the outputs and $d_{hl}^k$ and $t_{hl}^k$ as the inputs for each decision making unit $DMU_{hl}^k, \forall k \in \{0, \cdots, w\}$. Based on all the decision
4.3. METHODOLOGY

making units (all the feasible routes) for all pairs of required nodes, use DEA to obtain the efficiency scores of all the feasible routes.

4. Rank all the feasible routes in non-increasing order of DEA efficiency scores. For routes with the same efficiency scores, rank them in non-increasing order of MOAS. Shorten the list by retaining only the feasible route with the highest rank for each pair of required nodes while removing all the other routes with lower ranks.

5. Swap the feasible route with the highest rank with the respective original route derived in Step 1. (No swap is needed if the original shortest route is also the highest ranked route.) Update the remaining idle time. Then remove the swapped route as well as all the feasible routes whose service times exceed the remaining idle time from the list of feasible routes.

6. Repeat Step 5 until the total traveling time limit $T$ is reached or until all the remaining feasible routes on the list are assigned.

It should be noted that, like many of the existing methods for solving vehicle routing problems, the proposed heuristic procedure is based on the concept of greedy search [46]. The use of DEA efficiency scores, however, allows the integrated evaluation of solutions against the cost/equity frontier on a consistent basis in the process to continuously improve a route design. Figure 4.3 provides a summary of the proposed methodology with a four-step process that includes (1) Data Gathering and Analysis, (2) Model Development, (3) Route Generating
and Testing, and (4) Implementation and Continuous Improvement. Detailed flow diagrams for the first three steps are also given in the Appendix.

Figure 4.3: A Summary of the Proposed Methodology

4.4 Case Study: Quinte Access - Challenges and Opportunities

4.4.1 Background

Quinte West is a municipality in Southeastern Ontario, Canada. According to the latest census by Statistics Canada, Quinte West has a population of 43,086 over 494.15 square kilometers. Like many of the rural communities in North America, Quinte West faces the problem of fast aging population as the median age of residents in the region increased by 5.2 years from 38.2 in 2001 to 43.5 in 2011. Like many of

---

3The data of population densities, multi-objective utility scores, and accessibility scores of all the grid cells in the region as well as the IDs of grid cells covered by each route derived from the heuristic procedure in Step 3 are exported from GIS to Google Earth to generate an interactive map in KML format.
the rural communities with similar geographical and demographical characteristics in North America, developing an affordable and sustainable transportation system is a challenging and critical task for Quinte West to serve its aging population as well as to retain and attract qualified professionals/workers. The community was thus selected to demonstrate the proposed methodology for balancing equity and cost in rural transportation management.

With the goal of “keeping our community on the move,” Quinte Access is the region’s public transportation service operator. The not-for-profit organization operates a specialized service and a regular public transit system. The public transit system provides regular bus services covering a relatively small portion of the region. Three buses are currently available for service on the daily basis. Each bus, which is wheelchair lift-equipped, has a seating capacity of 16 passengers plus 2 wheelchairs and 6-7 standing passengers. The specialized service, which is more costly to operate, is primarily available for persons with disabilities and seniors in areas not covered by the public transit service. As the city expands through amalgamation, responding to different needs within the community given the scarce and underfunded resources poses a great challenge to the operator of the public transit system. It requires the bus routes to be strategically designed in order to achieve a balance among multiple objectives such as accessibility, equity, and operating costs.

The old bus routes in the public transit system of Quinte West consisted of two 90-minute routes continuously operated from 6:00 A.M. to 7:30 P.M. on weekdays and from 9:00 A.M. to 4:30 P.M. on Saturday. (The average service time for one of the old routes was actually 80 minutes.) A map showing the two old routes is provided in Figure 4.4. To improve transport accessibility in the rural community and to
reduce the increasing loads on the specialized service, Quinte Access made a number of operational improvements in the public transit system in recent years. However, the transit time of 90 minutes posed a major problem. In October 2013, Quinte Access started to plan to transition to a new route plan, consisting of three 60-minute routes, to provide increased flexibility in serving more strategically important areas in the community. (With three 60-minute routes, the total bus service time would be approximately the same as that with two 90-minute routes in the old system.) Implementing the expanded route plan requires a new scientific methodology which make use of the extensive demographic and geographic data from the community’s GIS to help design and evaluate different route alternatives under multiple objectives and criteria. In the subsection that follows, we will discuss how our proposed methodology was implemented to redesign the public transit system in Quinte West.

Figure 4.4: Old Bus Routes in Quinte West
4.4. CASE STUDY: QUINTE ACCESS - CHALLENGES AND OPPORTUNITIES

4.4.2 Research Procedure

In the case study, interviews were first conducted to identify ten factors considered in the multi-objective utility analysis, which included Apartment Building Cluster, Community Park, Hospital, Grocery/Department Store, Industrial/Business Park, Cluster with Medical Clinics and Doctor/Dentist/Optometrist Offices, Pharmacy, Retirement/Nursing Home Residence, Recreational/Entertainment Facility, and School. An initial questionnaire was then designed for utility analysis. However, the pretest based on the initial questionnaire indicated that, due to the relatively large number of factors considered in the study, respondents had trouble assigning scores when multiple units of each of the factors were present. The questionnaire was thus modified with the maximum of one unit of each factor given in a particular virtual grid cell, with a sample questionnaire shown in Figure 4.5.\(^4\) This modification also made it easier for respondents to relate the virtual grid cells to the actual situation in the community where most grid cells have no more than a single unit of each point of interest. The questionnaire was then completed by the mayor, a senior councilor, the chief administrative officer, the chief financial officer, the manager of public works, and the administrators of Quinte Access of Quinte West to obtain the relative values (weights) of all the factors and the utility scores of all the grid cells. It should be noted that the application of MOUA typically requires the participation of only a limited number (fewer than 10) of “expert” decision makers, as shown in various real-world applications presented in [59], because the purpose is to help decision makers in decision-making with multiple objectives as opposed to seeking views and opinions from the general public.

\(^4\)Since some of the key decision were not technically savvy, a paper-based questionnaire was used as an alternative to conduct the survey.
### 4.4. CASE STUDY: QUINTE ACCESS - CHALLENGES AND OPPORTUNITIES

<table>
<thead>
<tr>
<th>Area Profile Number</th>
<th>Apartment Building Cluster</th>
<th>Community Park</th>
<th>Hospital</th>
<th>Grocery/Department Store</th>
<th>Industrial/Business Park</th>
<th>Cluster of Medical Clinics &amp; Offices</th>
<th>Pharmacy</th>
<th>Retirement/Nursing Home Residence</th>
<th>Recreational/Entertainment Facility</th>
<th>School</th>
<th>Total Utility Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0 0 1 1 1 1 0 0 0 1 1 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1 1 0 0 0 0 0 0 1 1 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0 0 1 1 1 0 0 0 1 1 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0 0 1 1 1 0 0 0 1 1 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1 1 0 1 1 0 0 1 1 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0 1 1 1 0 0 0 0 0 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0 1 0 0 1 0 1 0 0 0 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0 1 1 1 0 0 0 1 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>1 1 0 1 1 1 1 1 0 0 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0 1 1 0 0 0 0 1 0 1 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0 1 1 1 0 1 1 0 0 0 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>1 1 0 0 1 1 0 0 0 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>1 1 0 0 0 0 0 0 1 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>1 1 0 1 1 1 0 0 0 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0 1 0 1 0 1 0 0 0 0 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>0 1 1 1 1 1 0 0 0 0 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAX</td>
<td>1 1 1 1 1 1 1 1 1 1 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100</td>
</tr>
<tr>
<td>MIN</td>
<td>0 0 0 0 0 0 0 0 0 0 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

**Instructions**

1. Rank all 16 area profiles and assign them a utility score between 1 and 100.
2. The Area Profile MAX shows that a score of 100 can only be attained from having all the points of interest.
3. The Area Profile MIN shows that a score of 1 should be given to an area with none of the points of interest.
4. Place score in highlighted column with "?"
5. Example: Area Profile #1 corresponds to an area in Quinte West that has no apartment building cluster, no community park, one hospital, one grocery/department store, one industrial/business park, one cluster with medical clinics/doctor offices, no pharmacy, no retirement/nursing home, one recreational/entertainment facility, and one school.

**Figure 4.5:** Sample Questionnaire Used to Determine Utility Values in MOUA

Based on the results of utility assessment, a regression analysis was performed according to the Vector Model specified in (4.1) and (4.2) to derive the utility value of each of the ten factors, as shown in Table 4.1. According to the table, the P-values of the top-four factors (Hospital, Grocery Store, Medical Facility, and Apartment Building) are all within 0.01, and the P-values of seven out of the ten factors are within 0.1. The only factor with a high P-value is Community Park. One option is to drop this factor from further consideration. However, since this was a factor which Quinte Access would like to include in the study, and the utility value associated with this factor is rather low (with likely minimal effects), it was decided that this factor
4.4. CASE STUDY: QUINTE ACCESS - CHALLENGES AND OPPORTUNITIES

<table>
<thead>
<tr>
<th>Factors</th>
<th>Utility Value</th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P -value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospital</td>
<td>20.35</td>
<td>4.14</td>
<td>4.91</td>
<td>0.00</td>
</tr>
<tr>
<td>Grocery Store</td>
<td>13.56</td>
<td>3.61</td>
<td>3.76</td>
<td>0.00</td>
</tr>
<tr>
<td>Medical Facility</td>
<td>13.34</td>
<td>4.76</td>
<td>2.80</td>
<td>0.01</td>
</tr>
<tr>
<td>Apartment Building</td>
<td>12.93</td>
<td>4.68</td>
<td>2.76</td>
<td>0.01</td>
</tr>
<tr>
<td>School</td>
<td>9.68</td>
<td>7.45</td>
<td>1.30</td>
<td>0.20</td>
</tr>
<tr>
<td>Pharmacy</td>
<td>7.89</td>
<td>4.15</td>
<td>1.90</td>
<td>0.06</td>
</tr>
<tr>
<td>Industrial Park</td>
<td>7.62</td>
<td>4.99</td>
<td>1.53</td>
<td>0.13</td>
</tr>
<tr>
<td>Retirement Residence</td>
<td>7.33</td>
<td>4.46</td>
<td>1.64</td>
<td>0.10</td>
</tr>
<tr>
<td>Recreational Facility</td>
<td>6.57</td>
<td>3.84</td>
<td>1.71</td>
<td>0.09</td>
</tr>
<tr>
<td>Community Park</td>
<td>2.67</td>
<td>5.85</td>
<td>0.46</td>
<td>0.65</td>
</tr>
</tbody>
</table>

Table 4.1: MOUA Regression Results

would still be considered in our successive analyses.

The regression results are used to calculate the utility and accessibility scores of each grid cell, and the resulting data are exported to GIS. A large-scale Google Earth file is then created with the map of the entire community in which a decision maker can quickly retrieve information about utility/accessibility scores and population density by clicking on a particular grid cell, as previously shown in Figure 4.1. Note that the population data are estimated from the number of households in each grid cell available in the community’s GIS and the average number of residents per household in the community. Distance between different nodes can be directly obtained using Google Earth. With data of population, distance, service time, and utility/accessibility scores retrieved from the system, the main stops are first specified, and the initial route plan is identified with three new routes of the shortest distances. The DEA-based heuristic rule is then applied to evaluate alternative routes. Since there is a priori belief that all the routes will be operated at the optimal scale, the
constant returns to scale model is used ([24], [34]). In addition, we also perform a bootstrapping analysis by using the “dea.boot” function in R with 2,000 iterations to calculate the bias-corrected efficiency scores to allow meaningful comparison ([79], [76]). The final route plan is then obtained according to the analytical process outlined in Section 3.3.3. A zoomed-in version of the Google Earth file with the proposed new routes and the tick marks of the top-50 grid cells with the highest rural accessibility scores is given in Figure 4.6.

4.4.3 Research Results and Discussion

Table 4.2 presents a summary of the old and proposed new route designs on key performance dimensions including population covered, multi-objective utility score (MOUS), distance, rural accessibility score (RAS), service time, and multi-objective accessibility score (MOAS). As shown in the table, the new route design will increase the total population covered from 29,918 to 37,637 (25.80% increase), total MOUS from 1,855.77 to 2,398 (29.22% increase), and total RAS from 431,338 to 538,091 (24.75% increase), respectively. Since the total service time of the three new bus routes is slightly longer than that of the two old routes, we also compute the results based on “per minute of bus service time” in order to compare the old and new route designs on a more consistent basis. According to Table 4.2, the new route design will increase the population covered per minute from 175.99 to 209.09 (18.81% increase), MOUS per minute from 10.92 to 13.32 (22.04% increase), and RAS per minute from 2,537.28 to 2,989.40 (17.82% increase), respectively. Although these improvements are at the expense of longer travel distance with the new route design (to reach some

\[^5\]According to the protocol of the DEA scale assumption given in [34], VRS model should be used only when there is not a priori reason and when scale effects can be demonstrated.
4.4. CASE STUDY: QUINTE ACCESS - CHALLENGES AND OPPORTUNITIES

grid cells with high population and MOUS), it is noted that the new route design also increases MOAS (i.e., the ratio of the total RAS to distance) per minute of bus service time from 81.24 to 116.37 (43.25% increase). The significantly higher MOAS implies that the new route design presents a better combination of multiple objectives and criteria.

![Figure 4.6: Three New Routes Shown in Different Colors with Google Earth](image)

To provide a quantitative benchmark of our solutions, we also compare the objective values of the proposed new routes with those of the shortest routes derived from Dijkstra algorithm, a common benchmark approach in the literatures of OR, geographic science, and urban/regional planning. As shown in Table 4.2, our proposed method leads to higher values of desirable objectives (population covered and MOUS score) and higher values of undesirable objectives (travel time and distance).
4.4. CASE STUDY: QUINTE ACCESS - CHALLENGES AND OPPORTUNITIES

Table 4.2: Summary of Research Outcomes

<table>
<thead>
<tr>
<th>Route Type</th>
<th>Population Covered</th>
<th>MOUS Score</th>
<th>Distance (km)</th>
<th>RAS Score</th>
<th>Service Time (min)</th>
<th>MOAS Score</th>
<th>DEA Score (original)</th>
<th>DEA Score (bootstrap)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old Route A</td>
<td>15,022</td>
<td>932.19</td>
<td>32.0</td>
<td>216,051</td>
<td>80</td>
<td>6,752</td>
<td>0.8505</td>
<td>0.8073</td>
</tr>
<tr>
<td>Old Route B</td>
<td>14,897</td>
<td>923.58</td>
<td>30.5</td>
<td>215,287</td>
<td>90</td>
<td>7,059</td>
<td>0.8496</td>
<td>0.8076</td>
</tr>
<tr>
<td>New Route A</td>
<td>13,104</td>
<td>826.23</td>
<td>22.8</td>
<td>193,755</td>
<td>60</td>
<td>8,498</td>
<td>1.0000</td>
<td>0.9332</td>
</tr>
<tr>
<td>New Route B</td>
<td>13,956</td>
<td>814.48</td>
<td>30.0</td>
<td>198,721</td>
<td>60</td>
<td>6,624</td>
<td>1.0000</td>
<td>0.9394</td>
</tr>
<tr>
<td>New Route C</td>
<td>10,577</td>
<td>757.28</td>
<td>25.0</td>
<td>145,616</td>
<td>60</td>
<td>5,825</td>
<td>0.9166</td>
<td>0.8923</td>
</tr>
<tr>
<td>Shortest Route A</td>
<td>8,453</td>
<td>556.77</td>
<td>18.0</td>
<td>143,755</td>
<td>43</td>
<td>7,945</td>
<td>0.9403</td>
<td>0.9096</td>
</tr>
<tr>
<td>Shortest Route B</td>
<td>7,555</td>
<td>549.00</td>
<td>17.8</td>
<td>137,824</td>
<td>43</td>
<td>7,765</td>
<td>0.9272</td>
<td>0.9010</td>
</tr>
<tr>
<td>Shortest Route C</td>
<td>7,834</td>
<td>492.22</td>
<td>19.5</td>
<td>125,553</td>
<td>44</td>
<td>6,453</td>
<td>0.8129</td>
<td>0.7841</td>
</tr>
<tr>
<td>Old Routes (total)</td>
<td>29,918</td>
<td>1,855.77</td>
<td>62.5</td>
<td>431,338</td>
<td>130</td>
<td>13,810</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>New Routes (total)</td>
<td>37,637</td>
<td>2,398.01</td>
<td>77.9</td>
<td>538,091</td>
<td>180</td>
<td>20,947</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Shortest Routes (total)</td>
<td>23,842</td>
<td>1,597.99</td>
<td>55.2</td>
<td>406,457</td>
<td>130</td>
<td>22,169</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Old Routes (per min)</td>
<td>175.99</td>
<td>10.93</td>
<td>0.37</td>
<td>2,537.28</td>
<td>-</td>
<td>81.24</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>New Routes (per min)</td>
<td>209.09</td>
<td>13.32</td>
<td>0.43</td>
<td>2,989.40</td>
<td>-</td>
<td>116.37</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Shortest Route (per min)</td>
<td>183.40</td>
<td>12.29</td>
<td>0.42</td>
<td>3,126.59</td>
<td>-</td>
<td>170.53</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

For the two gravity-type measurements, our method has an advantage on RAS score but a disadvantage on MOAS score. These results are reasonable given the fact that Dijkstra algorithm minimizes the total travel distance with fewer grid cells covered. The real challenge, however, is how to evaluate different solutions with advantages on different objectives on a consistent basis. The accomplish this goal, we also calculate the DEA efficiency scores for the two old routes, three new routes, and three shortest routes with population and MOUS treated as outputs and distance and service time treated as inputs, as shown in Table 4.2. The biased-corrected efficiency scores with bootstrapping are also provided to allow meaningful comparison. According to Table 4.2, new routes A and B are the efficient designs (with 100% efficiencies) based on DEA while new route C also has a higher efficiency (91.66%) than old routes A and...
B (with efficiencies equal to 85.05% and 84.98%, respectively). Each of the three new routes also has a higher efficiency score than the corresponding shortest route (i.e., the shortest routes A, B, and C with efficiencies equal to 94.03%, 92.72%, and 81.29%). The DEA efficiency scores can also be used to quantitatively characterize the suboptimality of our solutions based on the mathematical definition of “Pareto dominance” presented earlier. Specifically, the efficiency scores of 100% attained by new routes A and B and 91.66% attained by new route C in Table 4.2 show that the newly designed routes are either on or close to the Pareto efficient frontier established with DEA, an indication of good multi-objective solutions for addressing the cost/equity tradeoff in public transport planning. As depicted in Figure 4.7 with the DEA efficient frontier based on the CRS assumption, the proposed DEA-based heuristic procedure improves the initially dominated solutions derived from Dijkstra algorithm toward the Pareto efficient frontier in the multi-objective space (i.e., improving the efficiencies of routes A, B, and C from 94.03% to 100%, 92.72% to 100%, and 81.29% to 91.66%, respectively).

As mentioned previously, the efficiency scores obtained from the traditional DEA might be biased ([97], [98]). Table 4.2 presents the bias-corrected efficiency scores derived from bootstrapping. According the table, the bias-corrected efficiency scores are generally lower than the original DEA efficiency scores, which implies that, consistent with the results in the existing literature (e.g.,[76], [79]), the traditional DEA without bootstrapping may overestimate the efficiencies of DMUs. However, the bias-corrected efficiency scores of the new routes A, B, and C are still higher than those of the corresponding shortest routes A, B, and C as well as the two old routes, which again shows the effectiveness of the proposed methodology in improving the original
and shortest route designs toward the efficient frontier of multiple objectives.

![Graph showing the efficient frontier and old, new, and shortest routes.](image)

Figure 4.7: The Efficient Frontier and the Old, New, and Shortest Routes

The new public transit system was officially launched by Quinte Access in November 2014. The increase in total ridership in the first year was equal to 23% from the same period of months in the previous years. Quinte Access has also received overwhelmingly positive feedback from residents in the rural community. It is important to note that, since the total bus service time with three 60-minute routes under the new system is approximately the same as that with two 90-minute routes under the old system, the performance improvements were largely accomplished with similar levels of operating resources. Therefore, the strategic deployment of bus routes in the newly designed system appears to be the major driving force of the improved transportation service achieved by Quinte Access. The community is also planning
to make the necessary adjustments with different solution approaches (e.g., different multi-objective optimization methods, DEA with different scale assumptions and bootstrapping procedures, etc.) and updated geographical and demographical data on a regular basis (every two to three years) to continuously improve the overall route design for the public transit system in the future.

4.5 Conclusion

Rural transportation management is of crucial importance to today’s society due to the increasing demand and the limited operations resources available. It is also an area which has received little attention from the research community. In this paper, we develop a new methodology for analyzing the resource scarcity and equity issues in rural transportation management. It is shown with a case study that the proposed methodology can lead to significant quantitative benefits, including increased total ridership and population covered and better resource utilization to achieve multiple objectives, on different performance measures for a rural community. Our analysis can also be modified and applied to other fields such as disaster relief where vehicles need to be quickly sent to relieve only a limited number of areas in a region based on some measures of the relative importance of each grid cell. In every challenge, there lies opportunity. It is expected that the new knowledge, insights, and tools derived from this study for rural transportation planning, big data management, and vehicle routing and operations can be transferred to other communities in the ongoing process of developing sustainable transportation services in rural regions around the world.


[75] DJ Marsico. Transportation challenges of rural america: Testimony to the house appropriations 19 committee, subcommittee on transportation, housing and urban development.


Appendix A

A Non-Parametric Approach to Dynamic Pricing with Demand Learning

A.1 Example with polynomial of degree 2

We use a more compact notation of $c_{i,j,k}^j$ for the polynomial coefficients, where $j$ and $k$ indicate the position of the variables in the basis, and $i$ corresponds to the price index. $c_{0,0}^i$ refers to the constant, and $c_{0,k}^i$ represents all variables of degree one. We assume all $j \leq k$. Starting with $\sum_{\beta} c_{\beta} g^{\beta}$, the second degree polynomial is represented
A.1. EXAMPLE WITH POLYNOMIAL OF DEGREE 2

by the following equation:

$$\sum_{\beta} c_{\beta} g^{\beta} = c_{0,0}^i + c_{0,1}^i (1 - \theta_i) + \ldots + \sum_{k=2}^{Q} c_{0,k}^i (\theta_{j-1} - \theta_j) + c_{0,Q+3}^i (\nu - \nu)$$  \hspace{1cm} (A.1)

$$+ \sum_{j=2}^{Q} \sum_{j \leq k} c_{j,k}^i (\theta_{j-1} - \theta_j) (\theta_{k-1} - \theta_k) + \ldots + c_{Q+3,Q+3}^i (\nu - \nu)^2$$

(A.1) represents the right side of the equality $f(x) = \sum_{\beta} c_{\beta} g^{\beta}$. The upper bound polynomial derived earlier from the Lagrangian relaxation with degree 2 can be represented under a similar form, and the coefficients from both polynomials can be matched into equality constraints. Polynomial $f(x)$ contains also a Lagrangian multiplier which has to be estimated in this process. In this example the multiplier is projected on the polynomial basis and is written as a linear combination of $\mu = \mu_0 + \sum_{j=1}^{Q} \mu_j \theta_j + \nu \mu_{Q+1}$. Expanding (A.2) with this approximation creates a polynomial with cross-terms that are different for each price index:

$$f(x) = \sum_{\alpha} h_{\alpha} x^{\alpha} - (p_i - \mu_0) \theta_i + \mu_{Q+1} \nu \theta_i - \nu \sum_{j=1}^{Q} \mu_j \theta_j - \mu_{Q+1} \nu^2 - \mu_0 \nu + \sum_{j=1}^{Q} (\mu_j \theta_j) \theta_i \geq 0$$  \hspace{1cm} (A.2)

Using (A.2) and (A.1) all coefficients in the equations below are matched together to get the ordering constraint for the set of all prices:
A.1. EXAMPLE WITH POLYNOMIAL OF DEGREE 2 121

\[ h_{0,0} = c_{0,0}^i + c_{0,1}^i + c_{1,1}^i + \bar{v}c_{0,Q+2}^i + \bar{v}c_{1,Q+2}^i + \bar{v}^2c_{Q+2,Q+2}^i \]

\[ - \nu c_{0,Q+3}^i - \nu c_{1,Q+3}^i - \nu \bar{v}c_{Q+2,Q+3}^i + \nu^2 c_{Q+3,Q+3}^i \]

\[ h_{0,1} = -2c_{1,1}^i - c_{0,1}^i + c_{1,2}^i + c_{0,2}^i + \bar{v}c_{2,Q+2}^i - \nu c_{1,Q+2}^i - \nu c_{2,Q+3}^i \]

\[ + \nu c_{1,Q+3}^i - (p_i - \mu_0)\delta_{ij} \]

\[ h_{0,j} = -c_{1,j}^i - c_{0,j}^i + c_{1,j+1}^i + c_{0,j+1}^i + \bar{v}c_{j+1,Q+2}^i - \bar{v}c_{j,Q+2}^i \]

\[ - \nu c_{j+1,Q+3}^i + \nu c_{j,Q+3}^i - (p_i - \mu_0)\delta_{ij} \quad \forall j = 2, \ldots, Q \]

\[ h_{0,Q+1} = -c_{1,Q+2}^i - c_{0,Q+2}^i + c_{1,Q+3}^i + c_{0,Q+3}^i - 2\bar{v}c_{Q+2,Q+2}^i \]

\[ - 2\nu c_{Q+3,Q+3}^i + \bar{v}c_{Q+2,Q+3}^i + \nu c_{Q+2,Q+3}^i - \mu_0 \]

\[ h_{j,k} = -c_{j+1,k}^i + c_{j,k}^i + c_{j+1,k+1}^i - c_{j,k+1}^i + \mu_k\delta_{ij} \]

\[ \forall j + 1 < k < Q + 1, j = 1, \ldots, Q - 1, i \leq j \]

\[ h_{j,k} = -c_{j+1,k}^i + c_{j,k}^i + c_{j+1,k+1}^i - c_{j,k+1}^i + \mu_j\delta_{ik} \]

\[ \forall j + 1 < k < Q + 1, j = 1, \ldots, Q - 1, i > j \]

\[ h_{j,j+1} = -2c_{j+1,j+1}^i + c_{j,j+1}^i + c_{j+1,j+2}^i - c_{j,j+2}^i + \mu_{j+1}\delta_{ij} \]

\[ \forall j = 1, \ldots, Q - 1, i \leq j \]

\[ h_{j,j+1} = -2c_{j+1,j+1}^i + c_{j,j+1}^i + c_{j+1,j+2}^i - c_{j,j+2}^i + \mu_j\delta_{i,j+1} \]

\[ \forall j = 1, \ldots, Q - 1, i > j \]

\[ h_{j,j} = c_{j,j}^i + c_{j+1,j+1}^i - c_{j,j+1}^i + \mu_j\delta_{ij} \quad \forall j = 1, \ldots, Q \]

\[ h_{j,Q+1} = -c_{j+1,Q+2}^i + c_{j,Q+2}^i + c_{j+1,Q+3}^i - c_{j,Q+3}^i + \mu_{Q+1}\delta_{ij} - \mu_j \]

\[ \forall j = 1, \ldots, Q \]

\[ h_{Q+1,Q+1} = c_{Q+2,Q+2}^i + c_{Q+3,Q+3}^i - c_{Q+2,Q+3}^i - \mu_{Q+1} \]

\[ c_{j,k}^i \geq 0 ; i = 1, \ldots, Q \]
A.1. EXAMPLE WITH POLYNOMIAL OF DEGREE 2

The above equality constraints hold for each and every price index $i$. These equality constraints capture the inventory and ordering constraints of the initial model expressed in terms of the polynomial coefficients. The above set of constraints are designated by $\mathcal{H}$. Given the positivity constraint on $\mu$, the linear combination of all the variables defining $\mu$ should be positive as well. This leads to a new set of constraints defined below:

\begin{align*}
\mu_0 + \bar{\nu}\mu_{Q+1} & \geq 0 \\
\mu_0 + \nu\mu_{Q+1} & \geq 0 \\
\mu_0 + \sum_{k=1}^j \mu_k + \bar{\nu}\mu_{Q+1} & \geq 0 \quad \forall j = 1, \ldots, Q \\
\mu_0 + \sum_{k=1}^j \mu_k + \nu\mu_{q+1} & \geq 0 \quad \forall j = 1, \ldots, Q
\end{align*}

These constraints are designated as $\mathcal{M}$. The combination of the constraint sets $\mathcal{M}$ and $\mathcal{H}$ define all the linear constraints necessary to minimize the expected return of the polynomial:

\begin{equation}
\begin{aligned}
&\text{minimize } \mathbb{E}\left[\sum_{\alpha} h_{\alpha} x^\alpha \right] \mid a, b \\
&\text{subject to } \mathcal{H}, \mathcal{M}
\end{aligned}
\end{equation}

(A.3)

The polynomial coefficients are used within the learning DP (2.5) to find the optimal value function and pricing policy for a given belief vector and inventory constraint.
A.2. CLOSED FORM EULER INTEGRAL

A.2 Closed form Euler Integral

We proceed from lemma (2.6.1). Starting with $G_Q(\theta_Q - 1 | a_Q, b_Q)$ with $G_Q+1() = 1$, $G_Q()$ transforms into the regular incomplete beta function, which may be written using a Hypergeometric function representation with $B(z; a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} z^a (1-z)^{b-a}$. Therefore every recursion of $G_j()$ can be written as an incomplete beta function such as:

$$G_{Q-1}(\theta_{Q-2} | a_{Q-1}, b_{Q-1}) = \sum_{j_{Q-2}=0}^{+\infty} (1-b_{Q-1})_{j_{Q-2}} \int_{\theta_{Q-2}=0}^{\theta_{Q-1}} \theta_{Q-1}^{a_{Q-1}+a_{Q}-1+j_{Q}-1}(1-\theta_{Q-1})^{b_{Q}-1} d\theta_{Q-1}$$

(A.4)

The last term $G_1(\theta_1 = 1 | a_1, b_1)$ is a complete beta function. The closed form expression for $I_Q(a + e_1, b)$ where $a$ and $b$ are vectors of positive real components can be written as follows:

$$I_Q(a + e_1, b) = \sum_{j_{1,\ldots,Q}=0}^{+\infty} \prod_{l=1}^{Q} \frac{(1-b_l)_{j_l}}{j_l!} (1_{\{l\leq i\}} + \sum_{k=1}^{Q}(a_k + j_k))$$

(A.5)

The latter expression holds when $a$ and $b \in \mathbb{R}^{++}$. For the cases where $b \in \mathbb{Z}^{++}$ and all components of $b$ are greater or equal than $1$, the rising factorial is limited to $b_l - 1$ components. The rising factorial can be used to express the binomial coefficient as $\frac{(1-b_l)_{j_l}}{j_l!} = (-1)^{j_l} (b_l - 1)_{j_l}$, and the summations are limited by $b_l - 1$ numbers in the rising factorial. The equation reduces to

$$I_Q(a + e_1, b) = \sum_{j_{1,\ldots,Q}=0}^{b_{1-1}} \prod_{j_{Q-2}=0}^{Q} \frac{(b_l - 1)_{j_l}}{j_l!} (1_{\{l\leq i\}} + \sum_{k=1}^{Q}(a_k + j_k)) \forall b \in \mathbb{Z}^{++}, a \in \mathbb{R}^{++} (A.6)$$

with exact calculations of the integral.
A.3 Euler Integral with Negative Binomial

We claim that function $G_j(z|a,b)$ denoted thereafter as $G_j(z)$ is an analytic function in the complex plain $\mathbb{C}$ with two cuts along the intervals $(-\infty, 0]$ and $[1, \infty)$. Next, it is easy to obtain a series representation for $G_j(z)$ in the circle $|z| < 1$. This will be done by induction. Assume that

$$G_{j+1}(z) = \sum_{n \geq 0} c_n z^{n+\gamma}, \quad (A.7)$$

for some $\gamma \geq 0$ and some real coefficients $|c_n|$. Then, using the Binomial series

$$(1 - x)^{b_j - 1} = \sum_{k \geq 0} (1 - b_j)_k x^k,$$

where $(a)_k = a(a + 1) \ldots (a + k - 1)$ is the rising factorial symbol, we obtain

$$G_j(z) = \int_0^z x^{a_j - 1}(1 - x)^{b_j - 1}G_{j+1}(x)dx$$

$$= \int_0^z x^{a_j - 1} \left[ \sum_{k \geq 0} \frac{(1 - b_j)_k}{k!} x^k \right] \times \left[ \sum_{n \geq 0} c_n x^{n+\gamma} \right] dx = \sum_{n \geq 0} \tilde{c}_n z^{n+\gamma+a_j}, \quad (A.8)$$

where

$$\tilde{c}_n = \frac{1}{n+\gamma+a_j} \sum_{l=0}^n \frac{(1 - b_j)_l}{l!} c_{n-l}. \quad (A.9)$$

Now, using formula (A.9) we can obtain the series expansion of $G_1(z)$ near $z = 0$. This series expansion converges for all $z \in [0, 1)$, and in fact one can prove by induction that the coefficients are bounded by one.
In the same way, one can obtain a series expansion of $G_1(z)$ near $z = 0$, though that will be a bit more complicated. Again, this is done by induction. Assume that

$$G_{j+1}(1 - z) = \sum_{i=1}^{j} \sum_{n \geq 0} c_{i,n} z^{n+\gamma_i}. \quad (A.10)$$

Define

$$A_j = \int_0^1 x^{a_j-1}(1 - x)^{b_j-1} G_{j+1}(x) \, dx. \quad (A.11)$$

This number can be computed very efficiently by writing

$$A_j = \int_0^{1/2} x^{a_j-1}(1 - x)^{b_j-1} G_{j+1}(x) \, dx + \int_{1/2}^{1} x^{b_j-1}(1 - x)^{a_j-1} G_{j+1}(1 - x) \, dx$$

$$= \int_0^{1/2} x^{a_j-1}(1 - x)^{b_j-1} \left[ \sum_{n \geq 0} c_n x^{n+\gamma} \right] \, dx +$$

$$\int_0^{1/2} x^{b_j-1}(1 - x)^{a_j-1} \left[ \sum_{i=1}^{j} \sum_{n \geq 0} c_{i,n} z^{n+\gamma_i} \right] \, dx.$$ 

Next, we expand $(1 - x)^{b_j-1}$ and $(1 - x)^{a_j-1}$ in power series and we compute the above two integrals as infinite sums (as in formula (A.8) above, evaluated at $z = 1/2$). The individual terms in these infinite series are bounded by $O(2^{-n})$, so these series converge very fast and we do not need too many terms to compute it to high precision.

Going back to computing expansion of $G_j(1-z)$, we write $\int_0^{1-z} = \int_0^1 - \int_{1-z}$ and
obtain

\[ G_j(1 - z) = A_j - \int_0^z x^{b_j-1}(1 - x)^{a_j-1}G_{j+1}(1 - x)dx \]

\[ = A_j - \int_0^z x^{b_j-1}(1 - x)^{a_j-1} \left[ \sum_{i=1}^j \sum_{n \geq 0} c_{i,n} x^{n+\gamma_i} \right] dx. \]

Now, we deal with each series in the above equation as we did above in formulas (A.8) and (A.9). So if we do this recursively, we obtain a series expansion of \( G_j(1 - z) \) at \( z = 0 \) (it will consist of \( j \) different series with terms \( x^{n+\gamma_i+a_j} \) and a constant term \( A_j \)).

Note that the original integral that we wanted to compute is given by \( G_1(1) = A_1 \). So, once we compute \( A_1 \) – we may stop the algorithm.

Also, note that the sum in (A.9) is a convolution, therefore at each step of the algorithm we can compute all new coefficients in \( O(N \ln(N)) \) computations. Overall, to achieve accuracy \( \epsilon \) this algorithm will require \( O(n^2N \ln(N)) \) computations with \( N = \ln(1/\epsilon) \). It would also need \( O(nN) \) memory.
A.4 Optimizing Euler integral calculations

Consider

\[ G_j^q(a, b; z) = \int_{D_j(z)} x_1^{a_1-1} \cdots x_j^{a_j-1} (1 - x_1)^{b_1-1} \cdots (1 - x_Q)^{b_Q-1} dx_1 \cdots dx_Q, \]

where \( z \in (0, 1) \) and the domain is

\[
D_j(z) = \{0 \leq x_Q \leq \cdots \leq x_{j+1} \leq z \leq x_j \leq \cdots \leq x_1 \leq 1\}, \quad j = 1, \ldots, Q - 1,
\]
\[
D_Q(z) = \{z \leq x_Q \leq \cdots \leq x_1 \leq 1\},
\]
\[
D_1(z) = \{0 \leq x_Q \leq \cdots \leq x_1 \leq z\} \quad \text{(i.e., the original} \ D).\]

For each fixed \( z \in (0, 1) \), the value of \( f_Q(a_Q, b_Q; 1) \), as originally defined by (A.7), can be computed by enumerating all possible \( Q + 1 \) splits of the variables in two groups by \( z \) and adding up corresponding integrals:

\[
G_Q(a_Q, b_Q; 1) = \sum_{j=0}^{Q} G_j^Q(a_Q, b_Q; z). \tag{A.12}
\]

Since \( z \) is a constant, the subvectors \((x_1, \ldots, x_j)\) and \((x_{j+1}, \ldots, x_Q)\) vary independently from each other and

\[
G_j^Q(a_Q, b_Q; z) = \int_{1 \geq x_1 \geq \cdots \geq x_j \geq z} x_1^{a_1-1} \cdots x_j^{a_j-1} (1 - x_1)^{b_1-1} \cdots (1 - x_Q)^{b_Q-1} dx_1 \cdots dx_j
\]
\[
\times \int_{z \geq x_{j+1} \geq \cdots \geq x_Q \geq 0} x_{j+1}^{a_{j+1}-1} \cdots x_Q^{a_Q-1} (1 - x_{j+1})^{b_{j+1}-1} \cdots (1 - x_Q)^{b_Q-1} dx_{j+1} \cdots dx_Q.\]
The first integral in the above product is the original $G_j(a_j, b_j; z)$, while the second integral can be transformed by a variable change $\bar{x}_1 = 1 - x_Q, \ldots, \bar{x}_{Q-j} = 1 - x_{j+1}$ (note that the determinant of the Jacobian is 1 in absolute value) to

$$
\int_{1-z \leq 1-\bar{x}_{Q-j} \leq \ldots \leq 1-\bar{x}_1 \leq 1} (1 - \bar{x}_{n-j})^{a_{j+1}-1} \ldots (1 - \bar{x}_1)^{a_Q-1} \bar{x}_{Q-j}^{b_{j+1}-1} \ldots \bar{x}_1^{b_Q-1} d\bar{x}_{Q-j} \ldots d\bar{x}_1
$$

$$
= G_{Q-j}((b_Q, \ldots, b_{j+1}), (a_Q, \ldots, a_{j+1}); 1 - z).
$$

When $z \geq \frac{1}{2}$, expansion (A.8) can be used to evaluate $G_{Q-j}((b_Q, \ldots, b_{j+1}), (a_Q, \ldots, a_{j+1}); 1 - z)$. Thus, we let $z = \frac{1}{2}$ in (A.12) leading to

$$
f_Q(a_Q, b_Q; 1) = \sum_{j=0}^{Q} f_j(a_j, b_j; \frac{1}{2}) f_{Q-j}((b_Q, \ldots, b_{j+1}), (a_Q, \ldots, a_{j+1}); \frac{1}{2}).
$$
Appendix B

Short horizon dynamic pricing and learning with Thompson sampling

B.0.1 Proof of Lemma 7.2

\( k_2(1, a, b, \mu_1, \mu_2, z) = Pr(\theta_2 \geq z\theta_1|a, b) \). The domain of integration for the probability is decreasing in \( z \), and the first derivative is negative. To show that \( k_2(2, a, b, \mu, z) \) is monotonically non-increasing, we show that the first derivative with respect to \( z \) is
negative for all $z \in [0..1]$.

$$\frac{\partial k_2(2, a, b, \ldots)}{\partial z} = \frac{\partial Pr(\theta_2 \geq z\theta_1|a, b)}{\partial z} \left[1 + \mu_2 k_2(1, a + e_2, b, \mu, z) + (1 - \mu_2) k_2(1, a, b + e_2, \mu, z) - \mu_1 k_2(1, a + e_1, b, \mu, z)\right]$$

$$(1 - \mu_2) k_2(1, a, b + e_2, \mu, z) - \mu_1 k_2(1, a + e_1, b, \mu, z) -$$

$$(1 - \mu_1) k_2(1, a, b + e_1, \mu, z) + Pr(\theta_2 \geq z\theta_1|a, b) \left[\mu_2 \frac{\partial k_2(1, a + e_2, b, \mu, z)}{\partial z} + (1 - Pr(\theta_2 \geq z\theta_1|a, b)) \left[\mu_1 \frac{\partial k_2(1, a + e_1, b, \mu, z)}{\partial z}\right]\right]$$

Second and third brackets are negative. The content of the first bracket can be replaced by minimums and maximums of the value function:

$$1 + \min\{k_2(1, a + e_2, b, \mu, z), k_2(1, a, b + e_2, \mu, z)\} - \max\{k_2(1, a + e_1, b, \mu, z), k_2(1, a, b + e_1, \mu, z)\} =$$

$$1 + \min\{Pr(\theta_2 \geq z\theta_1|a + e_2, b), Pr(\theta_2 \geq z\theta_1|a + e_2, b)\} - \max\{Pr(\theta_2 \geq z\theta_1|a + e_1, b), Pr(\theta_2 \geq z\theta_1|a + e_1, b)\}$$

Content of first bracket is bigger or equal to zero, which makes the entire first derivative of the expected pulls of arm 2 with respect to $z$ negative.

We compare the expected number of pulls on the second arm using independent
Beta distributions with ordered joint distribution.

B.0.2 Proof of Lemma 7.3

We look at the optimal policy under known parameters \( \mu_i, i = 1, \ldots, q \):

\[
V_t(y, \mu) = \max_j \left\{ \mu_j (p_j + V_{t-1}(y-1, \mu)) + (1 - \mu_j) V_{t-1}(y, \mu) \right\} \tag{B.1}
\]

\[
= V_{t-1}(y, \mu_j) + \max_j \left\{ \mu_j (p_j + V_{t-1}(y-1, \mu) - V_{t-1}(y, \mu)) \right\} \tag{B.2}
\]

We look at the 2 period, 2 price case with 1 item to sell. The value function for the last period is simply \( V_1(1, \mu) = \max_j \{ \mu_j p_j \} = \max_j \{ z \mu_1, \mu_2 \} \). The first period is slightly more involved, considering the inequality between \( z \mu_1 \) and \( \mu_2 \). When \( z \mu_1 \geq \mu_2 \), then the value function reduces to:

\[
V_2(1, \mu) = z \mu_1 + \max \{ \mu_1 (z - \mu_1), \mu_2 (1 - z \mu_1) \} \tag{B.3}
\]

The value function is dependent on the following inequality \( \mu_2 \geq \frac{z \mu_1 (1 - \mu_2)}{1 - z \mu_1} \). When this inequality holds arm 2 is optimal the value function is as follow \( V_2(1, \mu) = z \mu_1 + \mu_2 (1 - z \mu_1) \). With the inequality not holding arm 1 is optimal and the value function reduces to \( V_2(1, \mu) = z \mu_1 + \mu_1 (z - z \mu_1) \).

In the case where \( \mu_2 \geq z \mu_1 \), the value function is as follows:

\[
V_2(1, \mu) = \mu_2 + \max \{ \mu_1 (z - \mu_2), \mu_2 (1 - \mu_2) \} \tag{B.4}
\]
In this case $\mu_2(1 - \mu_2)$ is always greater or equal $\mu_1(z - \mu_2)$, and the value function reduces to $V_2(1, \mu) = \mu_2 + \mu_2(1 - \mu_2)$. Getting both conditions together, arm 2 is optimal when $\mu_2 \geq \frac{z\mu_1(1-\mu_1)}{1-z\mu_1}$.
Appendix C

Balancing Equity and Cost in Rural Transportation Management with Multi-Objective Utility Analysis and Data Envelopment Analysis: A Case of Quinte West

C.1 Flow Diagrams for Data Gathering and Analysis
C.1. FLOW DIAGRAMS FOR DATA GATHERING AND ANALYSIS

Data Gathering and Analysis

Define areas, factors, options

Define requirements

Define inputs and outputs

Create survey (factors, options)

Respondents complete survey

Check survey

Complete?

Yes

Model Development

No

Acceptable?

Yes

GIS

Export data to spreadsheets

Coding with GIS

Create layers using KML

No
C.2 Flow Diagrams for Model Development

1. Model Development

2. Define variables for calculating weights

3. Calculation of weights process

4. Utility Survey (utility scores)

5. Weights (spreadsheet or output file)

6. Define performance measure and prepare data sheet

7. Create programs to calculate performance measures

8. Input

   - Yes
     - Route Generation and Testing
   - No
     - Compile performance data
     - Plot points in GIS
     - KML or SHAPE file with plotted scores
C.3 Flow Diagrams for Route Generation and Testing

1. Route Generation & Testing
   - KML or SHAPE file with plotted scores
2. Existing Routes?
   - Yes
     - Calculate value of existing routes
     - Total Route Scores
     - Plan for Implementation
   - No
     - KML or SHAPE file with plotted scores
     - Use DEA-based heuristic procedure to identify routes
     - Maps and details for potential routes
     - Final Selection