STRUCTURAL RESULTS FOR AVERAGE-COST INVENTORY MODELS WITH PARTIALLY OBSERVED MARKOV-MODULATED DEMAND

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE

OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF

MASTER OF SCIENCE

IN

INDUSTRIAL ENGINEERING

By

Harun Avcı

May 2018

STRUCTURAL RESULTS FOR AVERAGE-COST INVENTORY MODELS WITH PARTIALLY OBSERVED MARKOV-MODULATED DEMAND

By Harun Avcı

May 2018

We certify that we have read this thesis and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Kağan Gökbayrak(Advisor)

Emre Nadar(Co-Advisor)

Çağın Ararat

Zeynep Pelin Bayındır

Approved for the Graduate School of Engineering and Science:

Ezhan Karaşan
Director of the Graduate School

ABSTRACT

STRUCTURAL RESULTS FOR AVERAGE-COST INVENTORY MODELS WITH PARTIALLY OBSERVED MARKOV-MODULATED DEMAND

Harun Avcı
M.S. in Industrial Engineering
Advisor: Kağan Gökbayrak
Co-Advisor: Emre Nadar
May 2018

We consider a discrete-time infinite-horizon inventory system with full backlogging, deterministic replenishment lead time, and Markov-modulated demand. The actual state of demand can only be imperfectly estimated based on past demand data. We model the inventory replenishment problem as a Markov decision process with an *uncountable* state space consisting of both the inventory position and the most recent *belief* about the actual state of demand. When the demand state evolves according to an ergodic Markov chain, using the vanishing discount method along with a coupling argument, we prove the existence of an optimal average cost that is independent of the initial system state. With this result, we establish the average-cost optimality of a belief-dependent base-stock policy. We then discretize the belief space into a regular grid. The average cost under our discretization converges to the optimal average cost as the number of grid points grows large. Finally, we conduct numerical experiments to evaluate the use of a myopic belief-dependent base-stock policy as a heuristic. On a test bed of 108 instances, the average cost under the myopic policy deviates by no more than a few percent from the best lower bound on the optimal average cost obtained from our discretization.

Keywords: Inventory control, Markov-modulated demand, partial observations, long-run average cost, base-stock policy.

ÖZET

SAKLI MARKOV SÜRECİYLE DEĞİŞEN TALEP DAĞILIMLI ORTALAMA MALİYET ENVANTER MODELLERİNDE YAPISAL SONUÇLAR

Harun Avcı
Endüstri Mühendisliği, Yüksek Lisans
Tez Danışmanı: Kağan Gökbayrak
İkinci Tez Danışmanı: Emre Nadar
Mayıs 2018

Dönemsel gözden geçirilen, sabit tedarik süreli, karşılanamayan taleplerin kaybolmadığı ve talep durumunun sonlu bir Markov zincirine bağlı değiştiği tek ürünlük envanter sistemlerinde sonsuz ufuk ortalama maliyet problemi ele alınmıştır. Herhangi bir dönemdeki talep durumu doğrudan bilinmemekte ve talep geçmişi yardımıyla tahmin edilebilmektedir. Envanter problemi, envanter pozisyonundan ve koşullu durum dağılımından oluşan sonsuz durum uzaylı Markov karar süreci olarak modellenmiştir. Talep durumları ölçümkal (ergodik) bir durum zincirine göre değiştiğinde yok olan indirim yöntemi ile başlangıç durumdan bağımsız en iyi ortalama maliyetin varlığı ispatlanmıştır. Bu sonuçla koşullu durum dağılımına bağlı taban stok politikasının en iyi politika olduğu belirlenmiştir. Sonrasında koşullu durum dağılımı uzayı düzgün örgüye ayrıklaştırılmıştır. Düzgün örgülü durum uzayı altındaki ortalama maliyet, örgü nokta sayısı arttıkça en iyi ortalama maliyete yakınsamaktadır. Son olarak sezgisel bir yöntem olan durum dağılımına bağlı miyop taban stok politikasının performansını değerlendirmek için sayısal deney yapılmıştır. 108 örnek için miyop politika altındaki ortalama maliyetin ayrıklaştırma yöntemi ile elde ettiğmiz en iyi alt sınırdan sadece yüzde birkaç saptığı gözlemlenmiştir.

Anahtar sözcükler: Envanter sistemleri, Markov modüllü talep, kısmi gözlem, uzun vadede ortalama maliyet, taban stok politikası.

Acknowledgement

First and foremost, I would like to express my sincere gratitude to my advisors Asst. Prof. Kağan Gökbayrak and Asst. Prof. Emre Nadar for their invaluable support, encouragement, and guidance in my lifelong academic journey. I feel extremely lucky to have the opportunity to work under their supervision.

I am also grateful to Asst. Prof. Çağın Ararat and Assoc. Prof. Zeynep Pelin Bayındır for devoting their valuable time to read and review my thesis and their substantial comments.

I cannot thank enough my mother Nurgül Avcı and my father İbrahim Avcı, who have supported me in every way. Despite the actual distance during my education years, I have always felt them next to me. I would like to give a special thanks to my brother Nurullah Avcı, who has encouraged me to reach my full potential. He also has been my guide in my life.

My dearest friend Tolga Fazıloğlu, with whom we have been friends since childhood, deserves my sincere thanks for being a close friend and helping me shape my goals.

I would like to extend my sincere thanks to my high school friends - particularly Mahmut Altınay, Numan Atalay, Melih Başlopçu, Ahmet Doğukan Dağdaş, Doğancan Eser, Hasan Kürşad Gezer, Mehmet Koç, and Abdullah Uysal - who have been with me through thick and thin since the first day we met.

My university friends - especially Anıl Erdem Derinöz, Umay Kabayel, and Yavuz Mert Sarısakal - with whom I had enjoyed close friendship, also deserve thanks for supporting me on my academic journey.

I am deeply grateful to my dearest friends Merve Bolat, Hale Erkan, Utku Karaca, and Yücel Naz Yetimoğlu, who have been with me during my graduate studies for providing a supportive and joyful environment. I am also thankful to my officemates and those whom I failed to mention personally.

Lastly, I would like to thank TÜBİTAK for their support on this ARDEB 1001 Project (grant 214M243) and for the scholarship under BIDEB 2210-A.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Companies often face non-stationary demand that is driven by dynamic environmental factors, such as fluctuating economic and/or market conditions [1, 2, 3]. Associating a demand distribution with each state, Markov chains provide an elegant mathematical framework for modeling non-stationary demand. In this framework, the probability distribution of demand evolves over time according to a Markov chain whose state variable captures all the relevant information about environmental factors to represent the *demand state*. The Markov chain approach enables researchers to extend the optimal policy structures available in classical inventory models with stationary demand to their counterparts with Markov-modulated demand, by reasonably allowing the policy parameters to depend on the demand state. We refer the reader to Beyer et al. [4] for a comprehensive discussion on inventory models with Markov-modulated demand.

Although inventory models with Markov-modulated demand facilitate analytical treatment of non-stationarity, their practical applicability often suffers from the assumed *perfect* knowledge of demand state [5, 6]. Only a few researchers have addressed this issue by considering *partially* observed Markov-modulated demand. And those researchers have focused only on finite-horizon total-cost and infinite-horizon discounted-cost inventory systems. However, to our knowledge no one has studied the infinite-horizon average-cost inventory systems with

partial observations. Part of the reason for this is the notorious difficulty of the resulting partially observed Markov decision processes (POMDPs) under the average cost criterion (see [7] and Chapter 5 in [8]). In this thesis we study the average-cost inventory replenishment problem with partially observed Markov-modulated demand. We contribute to the inventory literature by establishing structural results for this problem.

Specifically, we consider a single-item discrete-time inventory system with full backlogging and non-stationary demand that arrives according to one of a finite number of probability distributions in each time period. The probability distributions undergo Markovian transitions between time periods. The state of the underlying Markov chain, i.e., the demand state, is only partially observable based on past demand data. Replenishment lead times are constant and there is no fixed replenishment order cost.

The infinite-horizon discounted-cost problem for this inventory system can be formulated as a POMDP with an information vector that contains all past demand observations and the belief about the initial demand state. The demand state belief in any period can be specified as a probability distribution over the set of demand states that forms sufficient statistics for the entire history of the process and possesses the Markovian property. The belief evolves over time, as new demand observations become available, according to the Bayes' formula. To leverage this Markovian structure of the belief, we formulate the infinite-horizon discounted-cost problem as a Markov decision process (MDP) with a state space consisting of the inventory position and the belief about the current demand state, leading to an *uncountable* space. (See [9] for more details on reduction of a POMDP to an MDP.)

Bayesian updating mechanisms were exploited in many inventory papers that consider stationary demand with unknown parameters (e.g., [10], [11], [12], [13], and [14]) and non-stationary demand with partially observed demand states (e.g., [15], [16], and [5]). A greatly simplified alternative to MDPs with Bayesian updating is to formulate and solve an MDP with *perfectly* observed demand states,

to forecast the demand state in each time period based on the so-called "maximum a posteriori" (MAP) estimation, and to take the optimal action obtained from the MDP for the forecasted demand state. (See Chapter 9 in [17] for a detailed discussion on MAP estimation.) But this alternative method leads to a significant loss of optimality according to our numerical experiments on our inventory problem (see Chapter 6).

For our infinite-horizon problem with Bayesian updating, first, we establish that the optimal inventory replenishment policy is a belief-dependent base-stock policy in the discounted-cost case (Proposition 1). Then, assuming the underlying Markov chain is ergodic (Assumption 1), we employ the vanishing discount method along with a coupling argument to prove that (i) there exists an optimal average cost independent of the initial system state, (ii) the average-cost optimality equation holds, and (iii) the belief-dependent base-stock policy is optimal in the average-cost case (Theorem 1).

Because the state space is uncountable, finding an exact solution for the average-cost optimality equation (and thus calculating the optimal average cost and base-stock levels) is a computational challenge [18, 19]. As an approximation, we discretize our belief space via the regular grid approach proposed by Lovejoy [20] and the discretization scheme proposed by Yu and Bertsekas [21]. The average cost under this approximation is a lower bound on the optimal average cost. This lower bound converges to the optimal average cost as the number of grid points goes to infinity. We then evaluate the use of a *myopic* belief-dependent base-stock policy as a heuristic replenishment policy for our average-cost problem with uncountable state space. Myopic base-stock policies can be easily implemented in practice. Myopic base-stock policies were also shown to be optimal for several inventory models in the case of stationary demand [22, 23] and in the case of non-stationary demand under certain conditions [24, 25]. Our numerical experiments reveal the practicality of the myopic policy in our problem: On our test bed of 108 instances, the average cost under the myopic policy is no more than a few percent worse than the best lower bound on the optimal average cost that can be obtained from our approximation. In addition, computations for the myopic solution are instantaneous.

The rest of this thesis is organized as follows: Chapter 2 reviews the related literature. Chapter 3 formulates our problem. Chapter 4 presents our structural results for both discounted-cost and average-cost problems. Chapter 5 offers a discretization scheme for calculation of a lower bound on the optimal average cost. Chapter 6 presents our numerical results and Chapter 7 concludes.

Chapter 2

Literature Review

In the literature most classical inventory models assume that demand in each period is independent of environmental factors other than time (Chapter 1 in [4]). There is also a growing body of literature that models non-stationary demand (due to environmental factors) as a Markov-modulated process: Song and Zipkin [26] consider an inventory system with demand modeled as a Markov-modulated Poisson process, full backlogging, ordered stochastic replenishment lead times, linear holding and shortage costs, and fixed and linear variable ordering costs. The objective is to minimize the expected discounted cost over a finite or an infinite horizon. They establish the optimality of a state-dependent (s, S) policy for this system and propose a modified value-iteration algorithm to compute the optimal policy parameters. Under the assumption of zero replenishment lead time, Sethi and Cheng [27] generalize the optimality of state-dependent (s, S) policies to inventory systems with Markov-modulated demand, full backlogging, state-dependent convex holding and shortage costs, and fixed and linear variable ordering costs. Applying the vanishing discount method to the infinite-horizon discounted-cost problem in [27], Beyer and Sethi [28] extend the optimality of state-dependent (s, S) policies to the infinite-horizon average-cost problem. Using the vanishing discount method, Huh et al. [29] partially characterize the optimal policy structures for several different single-stage inventory models with Markov-modulated demand and capacity.

There are also papers that adopt Markov-modulated demand in multi-echelon inventory systems: Chen and Song [30] prove the optimality of an echelon base-stock policy with order-up-to levels dependent on the state of the underlying Markov chain. Muharremoğlu and Tsitsiklis [31] obtain a similar result for an uncapacitated inventory model with Markov-modulated stochastic lead times under the assumption of no order crossing. Chen et al. [32] study inventory control of serial supply chains with continuous demand and a constant lead time.

All of the above papers assume that the current state of the Markov-modulated process is perfectly observed by the controller and thus the true demand distribution is always known. Several other papers have significantly relaxed this assumption: Treharne and Sox [15] consider discrete-time inventory systems in which the demand state can only be partially observed through the past demand data. They study the finite-horizon total-cost problem with deterministic replenishment lead times, linear holding and shortage costs, linear variable ordering costs, and zero fixed ordering cost. They establish the optimality of a base-stock policy with the base-stock levels that depend on the most recent belief about the actual demand state. They also propose heuristic solution methods for calculation of the base-stock levels. Arifoğlu and Özekici [16] consider discrete-time inventory systems with random yield and finite capacity. The demand state is partially revealed via some observation process that is different than the past demand data. The observation process takes values from a finite set whereas the demand is non-negative real valued. They prove the optimality of belief-dependent (s, S) policies in finite-horizon and infinite-horizon discounted-cost problems. Bayraktar Ludkovski [33] consider continuous-time inventory systems with Markov-modulated Poisson demand with intensities and discrete jump increments conditioned on the demand state. The demand state is partially observed through the past demand data. They characterize the optimal policy structure in both cases of backlogging and lost sales. All these papers incorporate partial observations into their inventory models via Bayesian updating mechanisms (and thus uncountable state spaces) in finite-horizon total-cost or infinite-horizon discounted-cost problems. In this study, however, we focus on the infinite-horizon average-cost problem.

For the infinite-horizon average-cost problems with uncountable state spaces,

the optimal average cost may depend on the initial state. And when it is independent of the initial state, an optimal stationary policy need not exist (Chapter 5 in [8]). The vanishing discount method, which was originally developed by [34], can be used to show the existence of a constant optimal average cost that is independent of the initial state. Following this method, Ross [35] shows that the uniform boundedness and equicontinuity of the discounted cost function ensures the existence of an optimal average cost. Platzman [36] proves that the uniform boundedness of the optimal differential discounted cost function is a necessary and sufficient condition for a bounded optimal average cost. Beyer and Sethi [28] establish the uniform boundedness and equicontinuity of the discounted cost function for inventory models in which the *perfectly* observed demand state evolves over time according to an *irreducible* Markov chain. Using a coupling argument to obtain certain bounds on the discounted cost function, Borkar [37] proves the uniform boundedness and equicontinuity of the discounted cost function for controlled Markov chains with partial observations when the underlying Markov chain is ergodic. We refer the reader to Arapostathis et al. [38] for a detailed review on average-cost problems. In this study, we extend the coupling argument in [37] to an inventory system with partially observed Markov-modulated demand, which enables us to show the existence of an optimal average cost.

Because solving the average-cost optimality equation on an uncountable state space is infeasible, previous work has developed discretization schemes for approximate solutions. Lovejoy [20] discretizes the uncountable state space into a regular grid with the concept of "triangulation." Zhou and Hansen [18] improve Lovejoy's result by introducing a variable-resolution regular grid. Both papers establish a lower bound for discounted-cost problems modeled as POMDPs. Yu and Bertsekas [21] present a lower approximation approach for both discounted-cost and average-cost problems modeled as POMDPs. There are also papers that approximate the average cost for MDPs with uncountable state space; see, for instance, [39] and [19]. In this study, we adopt the discretization schemes developed by Lovejoy [20] and Yu and Bertsekas [21], which enable us to obtain a lower bound on the optimal average cost that is sufficiently tight according to our numerical experiments.

Chapter 3

Problem Formulation

In this chapter, we formulate our inventory replenishment problem for a single-item system with non-stationary demand distribution. Demand in each period arrives according to a conditional distribution on the state of economy or market that undergoes Markovian transitions over time. The demand state in period t, d_t, takes a value from a finite set $\mathcal{N} := \{1, 2, ..., N\}$, $\forall t \in \mathbb{Z}_+ := \{1, 2, ...\}$. We thus model the demand state process $\{d_t\}_{t\in\mathbb{Z}_+}$ as a finite-state Markov chain with an $N \times N$ transition matrix $P = \{p_{ij}\}$ where $p_{ij} := \mathbb{P}\{d_{t+1} = j | d_t = i\}$, $\forall t \in \mathbb{Z}_+$. The demand realization in period t, w_t, takes a value from a finite set $\mathcal{M} := \{0, \ldots, M\}$, $\forall t \in \mathbb{Z}_+$. We denote by $r_i(\cdot)$ the conditional probability mass function of w_t for a given $d_t = i$, i.e., $r_i(k) := \mathbb{P}\{w_t = k | d_t = i\}$. We assume that there exists an $i \in \mathcal{N}$ and $k \in \mathcal{M}_+ := \{1, \ldots, M\}$ such that $r_i(k) > 0$. This assumption is violated if and only if the demand is always zero.

The demand state d_t, $t \in \mathbb{Z}_+$, is partially observable through the realized demand values prior to period t and the initial state belief $\pi^1 = [\pi_1^1, \ldots, \pi_N^1]$, where $\pi_i^1 := \mathbb{P}\{d_1 = i\}$, $i \in \mathcal{N}$. We define the state belief in any period, which is also known as the "conditional state distribution" in the literature (see [40]), as an N-dimensional vector consisting of the apriori probabilities of being in each demand state conditioned on the history composed of the initial state belief and all past demand observations. Therefore, the belief in period $t > 1$, $\pi^t = [\pi_1^t, \ldots, \pi_N^t]$, can be formulated as $\pi_i^t(\pi, \omega) := \mathbb{P}\{d_t = i | \pi^1 = \pi, \omega^{t-1} = \omega\}$, $i \in \mathcal{N}$, where

$\omega^{t-1} = (w_1, ..., w_{t-1})$ is the demand history up to period $t-1$. For a given initial belief $\pi^1 = \pi$, a given demand history $\omega^{t-1} = \omega$, and a given demand realization $w_t = w$, the belief π^{t+1} can be calculated as follows:

$$\begin{aligned}
&\pi_i^{t+1}(\pi, (\omega, w)) = \mathbb{P}\{d_{t+1} = i | \pi^1 = \pi, \omega^t = (\omega, w)\} \\
&= \mathbb{P}\{d_{t+1} = i | \pi^1 = \pi, \omega^{t-1} = \omega, w_t = w\} \\
&= \sum_{j \in \mathcal{N}} \mathbb{P}\{d_{t+1} = i | d_t = j, \pi^1 = \pi, \omega^{t-1} = \omega, w_t = w\} \mathbb{P}\{d_t = j | \pi^1 = \pi, \omega^{t-1} = \omega, w_t = w\} \\
&= \sum_{j \in \mathcal{N}} p_{ji} \frac{\mathbb{P}\{d_t = j, w_t = w | \pi^1 = \pi, \omega^{t-1} = \omega\}}{\mathbb{P}\{w_t = w | \pi^1 = \pi, \omega^{t-1} = \omega\}} \\
&= \frac{\sum_{j \in \mathcal{N}} p_{ji} \mathbb{P}\{w_t = w | d_t = j, \pi^1 = \pi, \omega^{t-1} = \omega\} \mathbb{P}\{d_t = j | \pi^1 = \pi, \omega^{t-1} = \omega\}}{\sum_{j' \in \mathcal{N}} \mathbb{P}\{w_t = w | d_t = j', \pi^1 = \pi, \omega^{t-1} = \omega\} \mathbb{P}\{d_t = j' | \pi^1 = \pi, \omega^{t-1} = \omega\}} \\
&= \frac{\sum_{j \in \mathcal{N}} p_{ji}\ r_j(w)\ \pi_j^t(\pi, \omega)}{\sum_{j' \in \mathcal{N}} r_{j'}(w)\ \pi_{j'}^t(\pi, \omega)} \qquad (3.1) \\
&:= T_i(\pi^t(\pi, \omega), w), \quad \forall t \in \mathbb{Z}_+, \forall i \in \mathcal{N}.
\end{aligned}$$

For notational convenience, we express $\pi^t(\pi, \omega)$ as π^t in the rest of the thesis. Let $\Pi := \left\{\pi \in [0,1]^N \ : \ \sum_{i \in \mathcal{N}} \pi_i = 1\right\}$ be the continuous space of all possible beliefs. We define $T : \Pi \times \mathcal{M} \to \Pi$ as the one-period belief update function given by $T(\cdot, \cdot) = [T_1(\cdot, \cdot), \ldots, T_N(\cdot, \cdot)] \in \Pi$ (see [41], [15], and Chapter 4 in [42] for similar belief updates). The conditional probability of w_t for a given $\pi^t = \pi$ can be written as

$$\begin{aligned}
\hat{r}_\pi(k) &= \mathbb{P}\{w_t = k | \pi^t = \pi\} \\
&= \sum_{i \in \mathcal{N}} \mathbb{P}\{w_t = k | \pi^t = \pi, d_t = i\} \mathbb{P}\{d_t = i | \pi^t = \pi\} \\
&= \sum_{i \in \mathcal{N}} r_i(k) \pi_i.
\end{aligned}$$

We assume that the planning horizon is infinite and all unmet demand is backlogged. The order placed at the beginning of period t is received at the beginning of period $t + l$, where $l \in \{0, 1, \ldots\}$ is constant, $t \in \mathbb{Z}_+$. As we allow for non-zero replenishment lead times, we define the inventory position as the number of items on hand plus the number of items on order minus the number

of backlogged demands, including it in the state description of our MDP. As the belief π^t summarizes the demand history up to period t and the initial belief, it forms a sufficient statistic for the information collected up to period t [43]. Hence we also include π^t in the state description.

We denote the inventory position at the beginning of period t by $y_t \in \mathbb{Z}$, and the quantity of the order placed at the beginning of period t by $u_t \in \mathbb{Z}_+ \cup \{0\}$, $t \in \mathbb{Z}_+$. For an initial inventory position y_1, the inventory position evolves over time as follows.

$$y_{t+1} = y_t + u_t - w_t, \quad t \in \mathbb{Z}_+. \tag{3.2}$$

There are two types of costs in our inventory model: The ordering cost in period t is linear in the order quantity and is given by cu_t, where c is the unit ordering cost. The single-period expected inventory cost in period $t+l$ is piecewise linear and is given by

$$g(\pi^t, y_t + u_t) = \mathbb{E}\left[\max\left\{h\left(y_t + u_t - \sum_{n=0}^{l} w_{t+n}\right), b\left(-y_t - u_t + \sum_{n=0}^{l} w_{t+n}\right)\right\}\middle|\pi^t\right],$$

where b and h are the unit shortage and holding costs per period, respectively. Note that

$$\mathbb{P}\left\{\sum_{n=0}^{l} w_{t+n} = k\middle|\pi^t\right\} = \sum_{k_1=0}^{k} \hat{r}_{\pi^t}(k_1)\mathbb{P}\left\{\sum_{n=1}^{l} w_{t+n} = k - k_1\middle|\pi^{t+1} = T(\pi^t, k_1)\right\}.$$

Using the above recursion, the conditional $(l+1)$-period demand distribution can be calculated as follows:

$$\mathbb{P}\left\{\sum_{n=0}^{l} w_{t+n} = k\middle|\pi^t\right\}$$
$$= \sum_{k_1=0}^{k}\sum_{k_2=0}^{k-k_1}\cdots\sum_{k_l=0}^{k-\sum_{j=1}^{l-1}k_j} \hat{r}_{\pi^t}(k_1)\hat{r}_{\pi^{t+1}}(k_2)\cdots\hat{r}_{\pi^{t+l-1}}(k_l)\hat{r}_{\pi^{t+l}}\left(k - \sum_{j=1}^{l} k_j\right)$$

where $\pi^{t+1} = T(\pi^t, k_1), \ldots, \pi^{t+l} = T(\pi^{t+l-1}, k_l)$. We summarize our notation in Table 3.1.

Table 3.1: Summary of our notation.

Symbol	Description
N	Number of demand states.
$\mathcal{N}$	Set of possible demand states, i.e., $\{1, 2, ..., N\}$.
d_t	Demand state in period t, $d_t \in \mathcal{N}$, $t \in \mathbb{Z}_+$.
$\{d_t\}$	Markov chain with transition matrix $P = \{p_{ij}\}$.
Π	Uncountable set of beliefs, i.e., $\left\{\pi \in [0,1]^N \ : \ \sum_{i\in\mathcal{N}} \pi_i = 1\right\}$.
π^t	Apriori probability distribution of the demand state in period t.
π^1	Initial state belief, i.e., $[\mathbb{P}\{d_1 = 1\}, \ldots, \mathbb{P}\{d_1 = N\}]$.
M	Maximum possible demand value across all demand states.
$\mathcal{M}$	Set of possible demand values, i.e., $\{0, \ldots, M\}$.
w_t	Demand realization in period t.
ω^t	Demand history up to period t, i.e., $\{w_1, \ldots, w_t\}$.
$r_i(\cdot)$	Conditional probability mass function of w_t given $d_t = i$.
$\hat{r}_\pi(\cdot)$	Conditional probability mass function of w_t given $\pi^t = \pi$.
y_t	Inventory position at the beginning of period t, $y_t \in \mathbb{Z}$, $t \in \mathbb{Z}_+$.
u_t	Order quantity at the beginning of period t, $u_t \in \mathbb{Z}_+ \cup \{0\}$, $t \in \mathbb{Z}_+$.
l	Replenishment lead time.
c	Unit variable ordering cost.
h	Unit holding cost per period.
b	Unit shortage cost per period.

For an initial belief π and an initial inventory position y, the expected long-run average cost per period under a replenishment policy with order quantities $U = (u_1, u_2, ...)$, $u_t \geq 0$, $t = 1, 2, ...$, can be written as

$$J^U(\pi, y) = \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T}[cu_t + g(\pi^t, y_t + u_t)]\middle|\pi^1 = \pi, y_1 = y\right] \text{ s.t. (3.1) and (3.2).}$$

The objective is to determine the replenishment policy that minimizes the expected long-run average cost per period. In this formulation we allow the order quantity to depend on the state of the system in every period. For notational convenience, however, we suppress the dependency of u_t on (π^t, y_t). In Chapter 4, using the vanishing discount approach, we prove that there exists a replenishment policy with order quantities $U^* = (u_1^*, u_2^*, \ldots) \in \mathcal{U}$, where

$\mathcal{U} = \{(u_1, u_2, \ldots) | u_t \in \mathbb{Z}_+ \cup \{0\}\}$, and a constant λ^*, which is independent of π and y, such that $\lambda^* = J^{U^*}(\pi, y) \leq J^U(\pi, y),\ \forall U \in \mathcal{U}$.

Chapter 4

Analytical Results

In this chapter, first, we provide several structural results for the discounted-cost version of our problem that we will utilize in our average-cost analysis (Chapter 4.1). Then, we employ the vanishing discount method to prove that the average-cost optimality equation holds, and use this optimality equation to characterize the optimal policy structure for our average-cost problem (Chapter 4.2). We refer the reader to Chapter 5 in [4], Chapter 5 in [8], and Chapter 8 in [44] for detailed descriptions of the vanishing discount method.

4.1 The Discounted-Cost Problem

For any discount factor $\alpha \in (0,1)$ and any initial state $(\pi, y) \in \Pi \times \mathbb{Z}$, the optimal expected total discounted cost over an infinite horizon can be defined as

$$v_\alpha(\pi, y) = \inf_{U \in \mathcal{U}} J_\alpha^U(\pi, y)$$

where $J_\alpha^U(\pi, y)$ is the expected total discounted cost for the initial state (π, y) under a replenishment policy with order quantities $U = (u_1, u_2, \ldots)$, i.e.,

$$J_\alpha^U(\pi, y) = \lim_{T\to\infty} \mathbb{E}\left[\sum_{t=1}^{T} \alpha^{t-1}[cu_t + \alpha^l g(\pi^t, y_t + u_t)]\middle| \pi^1 = \pi, y_1 = y\right].$$

Following Proposition 4.1.9 in [8], we verify that the optimal cost function v_α satisfies

$$v_\alpha(\pi, y) = \min_{u\geq 0}\left\{cu + \alpha^l g(\pi, y+u) + \alpha\sum_{w=0}^{M} v_\alpha(T(\pi, w), y+u-w)\hat{r}_\pi(w)\right\}, \quad \forall \pi \in \Pi,\ \forall y \in \mathbb{Z} \tag{4.1}$$

(see [15] for a similar formulation on a finite-horizon total-cost problem). We assume that $\alpha^l b > c$. This assumption is standard in the inventory literature; see, for instance, [27], [16], and Chapter 3 in [42]. Note that if $\alpha^l b$ were less than c, it would never be optimal to place an order in an $(l+1)$-period problem.

For the discounted-cost problem in (4.1), Proposition 1 shows that a belief-dependent base-stock policy is optimal and the optimal belief-dependent base-stock levels S_α^π are bounded between 0 and $(l+1)M$, $\forall \pi \in \Pi$, $\forall \alpha \in (0, 1)$.

Proposition 1. *For any $\alpha \in (0, 1)$, the optimal stationary inventory replenishment policy is a belief-dependent base-stock policy with base-stock levels S_α^π such that the optimal order quantity in state (π, y) is $S_\alpha^\pi - y$ if $y < S_\alpha^\pi$, and zero otherwise. Furthermore, the optimal belief-dependent base-stock levels S_α^π, $\forall \pi \in \Pi$, satisfy (i) $S_\alpha^\pi \leq (l+1)M$ and (ii) $S_\alpha^\pi \geq 0$.*

Proof. We will prove that $v_\alpha(\pi, y+1) - v_\alpha(\pi, y) \geq v_\alpha(\pi, y) - v_\alpha(\pi, y-1)$, $\forall \pi \in \Pi$, $\forall y \in \mathbb{Z}$. We call this property "discrete-convexity in y" in our proof. With this property, we are able to characterize the optimal policy structure. We consider the value iteration algorithm that can be used to calculate $v_\alpha(\cdot, \cdot)$: Let $v_\alpha^t(\cdot, \cdot)$ denote the value function at the tth iteration step of the value iteration algorithm. Letting $z = y + u$, we obtain $v_\alpha^{t+1}(\pi, y) = -cy + \min_{z\geq y}\{G_\alpha^t(\pi, z)\}$ where

$$G_\alpha^t(\pi, z) = cz + \alpha^l g(\pi, z) + \alpha\sum_{w\in\mathcal{M}} v_\alpha^t(T(\pi, w), z-w)\hat{r}_\pi(w).$$

Following Proposition 4.1.9 in [8], we verify that $\lim_{t\to\infty} v_\alpha^t(\pi, y) = v_\alpha(\pi, y)$ if $v_\alpha^0(\cdot,\cdot)$ is the zero function. We thus assume that $v_\alpha^0(\cdot,\cdot)$ is the zero function in our value iteration algorithm.

Assuming that $v_\alpha^t(\pi, y)$ is discrete-convex in y, we will show that $v_\alpha^{t+1}(\pi, y)$ is discrete-convex in y, $\forall\pi \in \Pi$. It is easy to verify that $cz + \alpha^l g(\pi, z)$ is discrete-convex in z. Hence, because we assume $v_\alpha^t(\pi, y)$ is discrete-convex in y, $\forall\pi \in \Pi$, and because $G_\alpha^t(\cdot,\cdot)$ is a sum of discrete-convex functions, $G_\alpha^t(\pi, z)$ is discrete convex in z, $\forall\pi \in \Pi$. Also, note that $\lim_{z\to+\infty} g(\pi, z) = \infty$ and $\lim_{z\to-\infty} g(\pi, z) = \infty$, $\forall\pi \in \Pi$. Because $G_\alpha^t(\pi, z) \geq cz + \alpha^l g(\pi, z)$, $\lim_{z\to+\infty} G_\alpha^t(\pi, z) \geq \lim_{z\to+\infty}\{cz + \alpha^l g(\pi, z)\} = \infty$ and $\lim_{z\to-\infty} G_\alpha^t(\pi, z) \geq \lim_{z\to-\infty}\{cz + \alpha^l g(\pi, z)\} \geq \lim_{z\to-\infty}\{(c - \alpha^l b)z\} = \infty$ (recall $\alpha^l b > c$). Therefore there exists a global minima $S_\alpha^{\pi,t}$, a value of z, which minimizes $G_\alpha^t(\pi, z)$, i.e., $S_\alpha^{\pi,t} = \arg\min_{z\in\mathbb{Z}}\{G_\alpha^t(\pi, z)\}$, $\forall\pi \in \Pi$. It is thus optimal to order $\max\{0, S_\alpha^{\pi,t} - y\}$ items in state (π, y) at the $(t+1)$th iteration step of the value iteration algorithm. This implies that

$$v_\alpha^{t+1}(\pi, y) = \begin{cases} -cy + G_\alpha^t(\pi, S_\alpha^{\pi,t}) & \text{if } y < S_\alpha^{\pi,t}, \\ -cy + G_\alpha^t(\pi, y) & \text{if } y \geq S_\alpha^{\pi,t}. \end{cases}$$

In order to show that $v_\alpha^{t+1}(\pi, y)$ is discrete convex in y, we need to consider three different cases depending on the relationship between $S_\alpha^{\pi,t}$ and y:

(1) If $y + 1 \leq S_\alpha^{\pi,t}$, we have

$$v_\alpha^{t+1}(\pi, y+1) = -c(y+1) + \min_{z\geq y+1}\left\{G_\alpha^t(\pi, z)\right\} = -c(y+1) + G_\alpha^t(\pi, S_\alpha^{\pi,t}),$$

$$v_\alpha^{t+1}(\pi, y) = -cy + \min_{z\geq y}\left\{G_\alpha^t(\pi, z)\right\} = -cy + G_\alpha^t(\pi, S_\alpha^{\pi,t}), \text{ and}$$

$$v_\alpha^{t+1}(\pi, y-1) = -c(y-1) + \min_{z\geq y-1}\left\{G_\alpha^t(\pi, z)\right\} = -c(y-1) + G_\alpha^t(\pi, S_\alpha^{\pi,t}).$$

Hence $v_\alpha^{t+1}(\pi, y+1) - v_\alpha^{t+1}(\pi, y) = -c = v_\alpha^{t+1}(\pi, y) - v_\alpha^{t+1}(\pi, y-1)$.

(2) If $y = S_\alpha^{\pi,t}$, we have

$$v_\alpha^{t+1}(\pi, y+1) = -c(y+1) + G_\alpha^t(\pi, y+1),$$
$$v_\alpha^{t+1}(\pi, y) = -cy + G_\alpha^t(\pi, S_\alpha^{\pi,t}), \text{ and}$$
$$v_\alpha^{t+1}(\pi, y-1) = -c(y-1) + G_\alpha^t(\pi, S_\alpha^{\pi,t}).$$

Since $S_\alpha^{\pi,t}$ is the global minima, $v_\alpha^{t+1}(\pi, y+1) - v_\alpha^{t+1}(\pi, y) = -c + G_\alpha^t(\pi, y+1) - G_\alpha^t(\pi, S_\alpha^{\pi,t}) \geq -c = v_\alpha^{t+1}(\pi, y) - v_\alpha^{t+1}(\pi, y-1)$.

(3) If $y - 1 \geq S_\alpha^{\pi,t}$, we have

$$v_\alpha^{t+1}(\pi, y+1) = -c(y+1) + G_\alpha^t(\pi, y+1),$$
$$v_\alpha^{t+1}(\pi, y) = -cy + G_\alpha^t(\pi, y), \text{ and}$$
$$v_\alpha^{t+1}(\pi, y-1) = -c(y-1) + G_\alpha^t(\pi, y-1).$$

By discrete-convexity of $G_\alpha^t(\pi, \cdot)$, $v_\alpha^{t+1}(\pi, y+1) - v_\alpha^{t+1}(\pi, y) = -c + G_\alpha^t(\pi, y+1) - G_\alpha^t(\pi, y) \geq -c + G_\alpha^t(\pi, y) - G_\alpha^t(\pi, y-1) = v_\alpha^{t+1}(\pi, y) - v_\alpha^{t+1}(\pi, y-1)$.

Hence $v_\alpha^{t+1}(\pi, y)$ is discrete-convex in y.

Consequently, because $v_\alpha^0(\pi, y)$ is discrete-convex in y, $\forall \pi \in \Pi$, $v_\alpha^t(\pi, y)$ is discrete-convex in y, $\forall \pi \in \Pi$, $\forall t \in \mathbb{Z}_+$. Let $G_\alpha(\pi, z) = cz + \alpha^l g(\pi, z) + \alpha \sum_{w \in \mathcal{M}} v_\alpha(T(\pi, w), z - w)\hat{r}_\pi(w)$. Because $\lim_{t\to\infty} v_\alpha^t(\pi, y) = v_\alpha(\pi, y)$, $v_\alpha(\pi, y)$ is discrete-convex in y, and thus $G_\alpha(\pi, z)$ is discrete-convex in z. Also, note that $\lim_{z\to+\infty} G_\alpha(\pi, z) = \infty$ and $\lim_{z\to-\infty} G_\alpha(\pi, z) = \infty$, $\forall \pi \in \Pi$. Therefore a belief-dependent base-stock policy with base-stock levels S_α^π is optimal. We next prove (i) and (ii):

(i) For any $\alpha \in (0, 1)$, let $U = (u_1, u_2, \ldots)$ represent the order quantities under the optimal belief-dependent base-stock levels S_α^π, $\forall \pi \in \Pi$. Suppose that $\exists \pi \in \Pi$ such that $S_\alpha^\pi > (l+1)M$. Now consider all sample paths that start with $y_1 = S_\alpha^\pi - 1$ and $\pi^1 = \pi$ where $S_\alpha^\pi > (l+1)M$. The base-stock policy implies that the optimal order quantity in the first period is $u_1 = 1$ in all

these sample paths. We now construct an alternative policy with order quantities $\widetilde{U} = (\widetilde{u}_1, \widetilde{u}_2, \ldots)$ such that

$$\widetilde{u}_t = \begin{cases} u_1 - 1 & \text{if } t = 1, \\ u_2 + 1 & \text{if } t = 2, \\ u_t & \text{otherwise.} \end{cases}$$

The inventory position plus the order quantity in period t under the alternative policy is

$$\widetilde{y}_t + \widetilde{u}_t = \begin{cases} y_1 + u_1 - 1 & \text{if } t = 1, \\ y_t + u_t & \text{if } t > 1. \end{cases}$$

Since $S_\alpha^\pi > (l+1)M$, $y_1 = S_\alpha^\pi - 1 \geq (l+1)M$. Consequently, $y_1 + u_1 - \sum_{t=1}^{l+1} w_t \geq 1$ and $\widetilde{y}_1 + \widetilde{u}_1 - \sum_{t=1}^{l+1} w_t \geq 0$. Hence:

$$\begin{aligned} & J_\alpha^{\widetilde{U}}(\pi, S_\alpha^\pi - 1) - J_\alpha^U(\pi, S_\alpha^\pi - 1) \\ &= \mathbb{E}\left[\sum_{t=1}^{\infty} \alpha^{t-1}[c\widetilde{u}_t + \alpha^l g(\pi^t, \widetilde{y}_t + \widetilde{u}_t) - cu_t - \alpha^l g(\pi^t, y_t + u_t)] \right. \\ &\qquad \left. \Big| \pi^1 = \pi, y_1 = \widetilde{y}_1 = S_\alpha^\pi - 1 \right] \\ &= \alpha^l h(\widetilde{y}_1 + \widetilde{u}_1 - y_1 - u_1) + c((\widetilde{u}_1 - u_1) + \alpha(\widetilde{u}_2 - u_2)) \\ &= -\alpha^l h - (1-\alpha)c < 0. \end{aligned}$$

We have a contradiction because the expected total discounted cost under the alternative policy cannot be smaller than the expected total discounted cost under the optimal policy. We thus conclude that any policy with $S_\alpha^\pi > (l+1)M$ for some $\pi \in \Pi$ cannot be optimal.

(ii) For any $\alpha \in (0,1)$, let $U = (u_1, u_2, \ldots)$ represent the order quantities under the optimal belief-dependent base-stock levels S_α^π, $\forall \pi \in \Pi$. Suppose that $\exists \pi \in \Pi$ such that $S_\alpha^\pi < 0$. Now consider all sample paths that start with $y_1 = S_\alpha^\pi$ and $\pi^1 = \pi$ where $S_\alpha^\pi < 0$. The base-stock policy implies that the optimal order quantity in the first period is $u_1 = 0$ in all these sample

paths. Let K be the first period with an order, i.e., $K = \min_{n\in\mathbb{Z}_+}\{n : u_n > 0|\pi^1 = \pi, y_1 = S^\pi_\alpha\}$. For a given sample path, if $K = k$, we construct an alternative policy with order quantities $\widetilde{U} = (\widetilde{u}_1, \widetilde{u}_2, \ldots)$ such that

$$\widetilde{u}_t = \begin{cases} u_1 + 1 & \text{if } t = 1, \\ u_t & \text{if } 1 < t < k, \\ u_k - 1 & \text{if } t = k, \\ u_t & \text{otherwise.} \end{cases}$$

The inventory position plus the order quantity in period t under the alternative policy is

$$\widetilde{y}_t + \widetilde{u}_t = \begin{cases} y_t + u_t + 1 & \text{if } 1 \le t \le k-1, \\ y_t + u_t & \text{if } t \ge k. \end{cases}$$

Note that $y_t + u_t < 0$ and $\widetilde{y}_t + \widetilde{u}_t \le 0$ for $t < k$. Hence:

$$\begin{aligned}
&J^{\widetilde{U}}_\alpha(\pi, S^\pi_\alpha) - J^U_\alpha(\pi, S^\pi_\alpha) \\
&= \mathbb{E}\left[\sum_{t=1}^{\infty} \alpha^{t-1}[c\widetilde{u}_t + \alpha^l g(\pi^t, \widetilde{y}_t + \widetilde{u}_t) - cu_t - \alpha^l g(\pi^t, y_t + u_t)] \right. \\
&\qquad \left. \Big| \pi^1 = \pi, y_1 = \widetilde{y}_1 = S^\pi_\alpha \right] \\
&= \sum_{k=2}^{\infty} \mathbb{E}\left[\sum_{t=1}^{\infty} \alpha^{t-1}[c\widetilde{u}_t + \alpha^l g(\pi^t, \widetilde{y}_t + \widetilde{u}_t) - cu_t - \alpha^l g(\pi^t, y_t + u_t)] \right. \\
&\qquad \left. \Big| K = k, \pi^1 = \pi, y_1 = \widetilde{y}_1 = S^\pi_\alpha \right] \mathbb{P}\{K = k\} \\
&= \sum_{k=2}^{\infty} \left[c - \alpha^{k-1}c - \alpha^l b \sum_{t=1}^{k-1} \alpha^{t-1}(\widetilde{y}_t + \widetilde{u}_t - y_t - u_t)\right] \mathbb{P}\{K = k\} \\
&= \sum_{k=2}^{\infty} [(1-\alpha)c - \alpha^l b] \frac{1-\alpha^{k-1}}{1-\alpha} \mathbb{P}\{K = k\}.
\end{aligned}$$

Because $\frac{1-\alpha^{k-1}}{1-\alpha} > 0$, $\mathbb{P}\{K = k\} > 0$ for some $k \ge 2$, and $(1-\alpha)c - \alpha^l b < 0$ (recall $\alpha^l b > c$), we have $J^{\widetilde{U}}_\alpha(\pi, S^\pi_\alpha) - J^U_\alpha(\pi, S^\pi_\alpha) < 0$. We have a contradiction

because the expected total discounted cost under the alternative policy cannot be smaller than the expected total discounted cost under the optimal policy. We thus conclude that any policy with $S_\alpha^\pi < 0$ for some $\pi \in \Pi$ cannot be optimal.

■

Similar threshold policies are also available in the extant literature: Treharne and Sox [15] establish the optimality of a belief-dependent base-stock policy for a finite-horizon total-cost inventory system with partial observation. In their study, similar to ours, the demand in each period takes a value from a finite set and the actual demand state is partially revealed through the past demand data. We thus extend the optimal policy structure in [15] to an infinite-horizon discounted-cost inventory system. Arifoğlu and Özekici [16] establish the optimality of a belief-dependent (s, S) policy for an infinite-horizon discounted-cost inventory system with partial observation. In their study, unlike ours, the demand is non-negative real-valued and the actual demand state is partially revealed via a finite observation set that is different from the past demand data.

4.2 The Average-Cost Problem

We next consider the vanishing discount method for our analysis of the average-cost problem: For a fixed $\bar{\pi} \in \Pi$, we define $\delta_\alpha(\pi, y) := v_\alpha(\pi, y) - v_\alpha(\bar{\pi}, 0)$ as the differential discounted value function, $\forall \pi \in \Pi,\ \forall y \in \mathbb{Z}$. For any $\alpha \in (0, 1)$, the equation in (4.1) implies that

$$\begin{aligned}&\delta_\alpha(\pi, y) + (1-\alpha) v_\alpha(\bar{\pi}, 0) \\ &= \min_{u \geq 0} \left\{ cu + \alpha^l g(\pi, y+u) + \alpha \sum_{w=0}^{M} \delta_\alpha(T(\pi, w), y+u-w) \hat{r}_\pi(w) \right\}. \end{aligned} \tag{4.2}$$

We will show (in Theorem 1) that there exists a constant λ^* and a locally Lipschitz continuous function $\delta^*(\cdot, \cdot)$ that together satisfy the average-cost optimality

equation:

$$\delta^*(\pi, y) + \lambda^* = \min_{u \geq 0} \left\{ cu + g(\pi, y+u) + \sum_{w=0}^{M} \delta^*(T(\pi, w), y+u-w)\hat{r}_\pi(w) \right\},$$
$$\forall \pi \in \Pi, \ \forall y \in \mathbb{Z},$$

such that $(1-\alpha)v_\alpha(\bar{\pi}, 0) \to \lambda^*$ and $\delta_\alpha(\pi, y) \to \delta^*(\pi, y)$ as α goes to 1. In order to obtain this analytical result, we establish that $(1-\alpha)v_\alpha(\bar{\pi}, 0)$ is bounded with respect to $\alpha \in (0, 1)$ (see Lemma 1), and that $\delta_\alpha(\cdot, \cdot)$ is locally Lipschitz continuous for $\alpha \in (0, 1)$ and uniformly bounded with respect to $\alpha \in (0, 1)$ (see Lemma 2). We will also show (in Theorem 1) that the optimal replenishment policy is a belief-dependent base-stock policy in our average-cost problem.

Lemma 1. *$(1-\alpha)v_\alpha(\bar{\pi}, 0)$ is bounded with respect to $\alpha \in (0, 1)$. Furthermore, there exists a sequence $(\alpha_t)_{t=1}^{\infty}$ converging to 1 and a constant λ^* such that $(1-\alpha_t)v_{\alpha_t}(\bar{\pi}, 0) \to \lambda^*$ as t goes to infinity.*

Proof. Proof. For any $\alpha \in (0, 1)$ and the initial inventory position $\widetilde{y}_1 = 0$, consider a replenishment policy with order quantities $\widetilde{U} = (\widetilde{u}_1, \widetilde{u}_2, \ldots)$ such that

$$\widetilde{u}_t = \begin{cases} 0 & \text{if } t = 1, \\ w_{t-1} & \text{if } t > 1. \end{cases}$$

Note that the above policy corresponds to a zero base-stock level policy. The inventory position plus the order quantity in period t is

$$\widetilde{y}_t + \widetilde{u}_t = \begin{cases} 0 & \text{if } t = 1, \\ -w_{t-1} & \text{if } t > 1. \end{cases}$$

Then the following hold:

$$
\begin{aligned}
(1-\alpha)v_\alpha(\bar{\pi},0) &\leq (1-\alpha)J_\alpha^{\widetilde{U}}(\bar{\pi},0) \\
&= (1-\alpha)\mathbb{E}\left[\sum_{t=1}^{\infty}\alpha^{t-1}[c\widetilde{u}_t+\alpha^l g(\pi^t,\widetilde{y}_t+\widetilde{u}_t)]\middle|\pi^1=\bar{\pi},\widetilde{y}_1=0\right] \\
&= (1-\alpha)\mathbb{E}\left[\sum_{t=1}^{\infty}\alpha^{t-1}\left[c\widetilde{u}_t-\alpha^l b\left(\widetilde{y}_t+\widetilde{u}_t-\sum_{n=0}^{l}w_{t+n}\right)\right]\middle|\pi^1=\bar{\pi},\widetilde{y}_1=0\right] \\
&= (1-\alpha)\mathbb{E}\left[\alpha^l b\sum_{n=0}^{l}w_{1+n}+\sum_{t=2}^{\infty}\alpha^{t-1}\left(cw_{t-1}+\alpha^l b\sum_{n=-1}^{l}w_{t+n}\right)\middle|\pi^1=\bar{\pi}\right] \\
&\leq (1-\alpha)\left[\sum_{t=1}^{\infty}\alpha^{t-1}\right][c+b(l+2)]M \\
&= [c+b(l+2)]M.
\end{aligned}
$$

Thus $(1-\alpha)v_\alpha(\bar{\pi},0)$ is bounded with respect to $\alpha\in(0,1)$. By the Bolzano-Weierstrass Theorem, there exists a subsequence $\alpha_t\uparrow 1$ and a constant λ^* such that $(1-\alpha_t)v_{\alpha_t}(\bar{\pi},0)\to\lambda^*$. ■

In order to obtain further analytical results, we assume that the Markov chain governing the demand state process is ergodic. Previous work has required the *irreducibility* of the underlying Markov chain for optimal policy characterization in average-cost inventory models with *perfectly* observed Markov-modulated demand (see [28] and [29]). In this study, in addition to irreducibility, we also require the *aperiodicity* of the underlying Markov chain.

Assumption 1. *The Markov chain with transition matrix P is ergodic.*

We now consider two demand state processes $\{d_t\}_{t\in\mathbb{Z}_+}$ and $\{\tilde{d}_t\}_{t\in\mathbb{Z}_+}$, both evolving according to Markov chains with the same transition matrix. Let $\nu(i,j) := \mathbb{P}\{d=i,\tilde{d}=j\}$ denote an arbitrary joint probability mass function for demand states d and $\tilde{d}$. Also, let $V_{\pi,\widetilde{\pi}} := \left\{\nu : \sum_{j\in\mathcal{N}}\nu(i,j)=\pi_i,\forall i\in\mathcal{N},\text{ and }\sum_{i\in\mathcal{N}}\nu(i,j)=\widetilde{\pi}_j,\forall j\in\mathcal{N}\right\}$. Following [37], we define the Wasserstein distance between two beliefs π and $\widetilde{\pi}$ that correspond

to d and $\tilde{d}$, respectively:

$$\Delta(\pi, \widetilde{\pi}) := \inf_{\nu \in V_{\pi,\widetilde{\pi}}} \{\mathbb{E}_\nu[|d - \tilde{d}|]\} = \inf_{\nu \in V_{\pi,\widetilde{\pi}}} \left\{ \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}} |i - j| \nu(i, j) \right\}.$$

Using the above definition and our structural results in Proposition 1, under Assumption 1, Lemma 2 proves that $\delta_\alpha(\cdot, \cdot)$ is locally Lipschitz continuous for $\alpha \in (0, 1)$ and uniformly bounded with respect to $\alpha \in (0, 1)$.

Lemma 2. *Under Assumption 1, for $\alpha \in (0, 1)$, $\delta_\alpha(\cdot, \cdot)$ is locally Lipschitz continuous and uniformly bounded with respect to α. Furthermore, there exists a sequence $(\alpha_t)_{t=1}^\infty$ converging to 1 and a locally Lipschitz continuous function δ^* such that $\delta_{\alpha_t}(\pi, y) \to \delta^*(\pi, y)$ locally uniformly for any finite $y \in \mathbb{Z}$ and $\pi \in \Pi$ as t goes to infinity.*

Proof. Proof. Let $y_1, y_2, \ldots$ be the inventory positions of a system with beliefs $\pi^1, \pi^2, \ldots$ under the optimal belief-dependent base-stock policy with order quantities $U = (u_1, u_2, \ldots)$. Similarly, let $\widetilde{y}_1, \widetilde{y}_2, \ldots$ be the inventory positions of another system with beliefs $\widetilde{\pi}^1, \widetilde{\pi}^2, \ldots$ under an alternative policy with order quantities $\widetilde{U} = (\widetilde{u}_1, \widetilde{u}_2, \ldots)$ such that $\widetilde{u}_t = \max\{(y_t + u_t) - \widetilde{y}_t, 0\}, \ \forall t \in \mathbb{Z}_+$. For any finite $y, \widetilde{y} \in \mathbb{Z}$ and $\pi, \widetilde{\pi} \in \Pi$, assuming $\delta_\alpha(\widetilde{\pi}, \widetilde{y}) - \delta_\alpha(\pi, y) \geq 0$ without loss of generality, the following holds.

$$\begin{aligned}
\delta_\alpha(\widetilde{\pi}, \widetilde{y}) - \delta_\alpha(\pi, y) &= v_\alpha(\widetilde{\pi}, \widetilde{y}) - v_\alpha(\bar{\pi}, 0) - v_\alpha(\pi, y) + v_\alpha(\bar{\pi}, 0) \\
&= v_\alpha(\widetilde{\pi}, \widetilde{y}) - v_\alpha(\pi, y) \\
&\leq J_\alpha^{\widetilde{U}}(\widetilde{\pi}, \widetilde{y}) - v_\alpha(\pi, y) \\
&= \mathbb{E}\left[\sum_{t=1}^{\infty} \alpha^{t-1} [c\widetilde{u}_t + \alpha^l g(\widetilde{\pi}^t, \widetilde{y}_t + \widetilde{u}_t) - cu_t - \alpha^l g(\pi^t, y_t + u_t)] \right. \\
&\qquad \left. \Big| \pi^1 = \pi, \widetilde{\pi}^1 = \widetilde{\pi}, y_1 = y, \widetilde{y}_1 = \widetilde{y} \right]. \qquad (4.3)
\end{aligned}$$

Let $\eta_{ij} := \min\{n \in \mathbb{Z}_+ : d_n = \tilde{d}_n | d_1 = i, \tilde{d}_1 = j\}$ be the first period that the two demand state processes coincide, given that one process starts in state i and the other process starts in state j. Let $\tilde{K}_{ij} := \min_{k \in \mathbb{Z}_+}\{k \geq n : y_k + u_k =$

$\widetilde{y}_k + \widetilde{u}_k | \pi^n, \widetilde{\pi}^n, y_n, \widetilde{y}_n, \eta_{ij} = n\}$. Following the coupling argument in [37], we verify that the same demand values are observed in the systems starting with initial beliefs π and $\widetilde{\pi}$ once the demand states d_t and $\tilde{d}_t$ become equal to each other. Hence, if the inventory positions of these two systems become equal to each other as well in a certain period, they will remain equal to each other in all future periods, i.e., if $\tilde{K}_{ij} = k$, then $y_t = \widetilde{y}_t$ and $u_t = \widetilde{u}_t$, $\forall t \geq k+1$. The inequality in (4.3) can be rewritten as

$$\begin{aligned}
&\delta_\alpha(\widetilde{\pi}, \widetilde{y}) - \delta_\alpha(\pi, y) \\
&\leq \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}} \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} \mathbb{E}\left[\sum_{t=1}^{k} \alpha^{t-1} [\alpha^l [g(\widetilde{\pi}^t, \widetilde{y}_t + \widetilde{u}_t) - g(\pi^t, y_t + u_t)] + c(\widetilde{u}_t - u_t)] \right. \\
&\qquad\qquad \left. \Big| d_1 = i, \tilde{d}_1 = j, y_1 = y, \widetilde{y}_1 = \widetilde{y} \right] \mathbb{P}\{\tilde{K}_{ij} = k\} \mathbb{P}\{\eta_{ij} = n\} \pi_i \widetilde{\pi}_j.
\end{aligned} \tag{4.4}$$

By the alternative policy structure and Proposition 1, we have $0 \leq y_t + u_t \leq \widetilde{y}_t + \widetilde{u}_t$, $\forall t \in \mathbb{Z}_+$. Since the largest demand amount is M, $y_t + u_t - \sum_{n=0}^{l} w_{t+n} \geq -(l+1)M$ and $\widetilde{y}_t + \widetilde{u}_t - \sum_{n=0}^{l} \widetilde{w}_{t+n} \geq -(l+1)M$, $\forall t \in \mathbb{Z}_+$. Again by Proposition 1, the base-stock levels are no greater than $(l+1)M$. We thus have $-(l+1)M \leq y_t + u_t - \sum_{n=0}^{l} w_{t+n} \leq \max\{y, (l+1)M\}$ and $-(l+1)M \leq \widetilde{y}_t + \widetilde{u}_t - \sum_{n=0}^{l} \widetilde{w}_{t+n} \leq \max\{\widetilde{y}, (l+1)M\}$, $\forall t \in \mathbb{Z}_+$. Hence:

$$\sum_{t=1}^{k} \alpha^{t+l-1} [g(\widetilde{\pi}^t, \widetilde{y}_t + \widetilde{u}_t) - g(\pi^t, y_t + u_t)] \leq (k-1) \max\{\widetilde{y}h, (l+1)Mh, (l+1)Mb\}. \tag{4.5}$$

Recall that $y_{k+1} = \widetilde{y}_{k+1}$ and $w_t = \widetilde{w}_t$, $\forall t \geq n$. Also, recall that $w_t, \widetilde{w}_t \in \mathcal{M}$,

$\forall t \in \mathbb{Z}_+$. Thus:

$$\begin{aligned}
&\sum_{t=1}^{k} \alpha^{t-1}(\widetilde{u}_t - u_t) \\
&= \sum_{t=1}^{k} \alpha^{t-1}(\widetilde{y}_{t+1} - \widetilde{y}_t + \widetilde{w}_t - y_{t+1} + y_t - w_t) \\
&= \alpha^{k-1}(\widetilde{y}_{k+1} - y_{k+1}) + y_1 - \widetilde{y}_1 + (1-\alpha)\sum_{t=2}^{k} \alpha^{t-2}(\widetilde{y}_t - y_t) + \sum_{t=1}^{k} \alpha^{t-1}(\widetilde{w}_t - w_t) \\
&\leq y - \widetilde{y} + (1-\alpha)\sum_{t=2}^{k} \alpha^{t-2}(\widetilde{y}_t - y_t) + (n-1)M.
\end{aligned}$$

By Proposition 1, we have $\widetilde{y}_t \leq \max\{\widetilde{y}, (l+1)M\}$ and $y_t \geq -M$, $\forall t \in \mathbb{Z}_+$. Hence:

$$\begin{aligned}
\sum_{t=1}^{k} \alpha^{t-1}(\widetilde{u}_t - u_t) &\leq y - \widetilde{y} + (1-\alpha)\sum_{t=2}^{k} \alpha^{t-2}[\max\{\widetilde{y}, (l+1)M\} + M] + (n-1)M \\
&\leq y - \widetilde{y} + (k-1)\max\{\widetilde{y} + M, (l+2)M\} + (n-1)M. \qquad (4.6)
\end{aligned}$$

For ease of notation, let $A := \max\{\widetilde{y}h, (l+1)Mh, (l+1)Mb\}$ and $B := \max\{\widetilde{y} + M, (l+2)M\}$. We then obtain from (4.4)-(4.6) the following inequalities.

$$\begin{aligned}
&\delta_\alpha(\widetilde{\pi}, \widetilde{y}) - \delta_\alpha(\pi, y) \\
&\leq \sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N}}\sum_{n=1}^{\infty}\sum_{k=n}^{\infty}[(k-1)A + c(y-\widetilde{y}) + c(k-1)B + (n-1)cM] \\
&\qquad\qquad \mathbb{P}\{\tilde{K}_{ij} = k\}\mathbb{P}\{\eta_{ij} = n\}\pi_i\widetilde{\pi}_j \\
&\leq (A + cB)\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N}}\mathbb{E}[\tilde{K}_{ij} - 1]\pi_i\widetilde{\pi}_j + cM\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N}}\mathbb{E}[\eta_{ij} - 1]\pi_i\widetilde{\pi}_j + c(y - \widetilde{y}).
\end{aligned} \qquad (4.7)$$

We make the following two observations regarding the inequality in (4.7):

(1) Under Assumption 1, Borkar [37] states that there exists a finite $\Gamma > 0$ and $\mu \in (0,1)$ such that $\mathbb{E}[\eta_{ij} - 1] \leq \frac{\Gamma}{\mu}$, $\forall i, j \in \mathcal{N}$. If $i = j$, $\mathbb{E}[\eta_{ij}] = 1$. If $i \neq j$, $|i - j| \geq 1$. Consequently, $\mathbb{E}[\eta_{ij} - 1] \leq \frac{\Gamma}{\mu}|i - j|$, $\forall i, j \in \mathcal{N}$.

(2) If $i = j$, $\tilde{K}_{ij} = \min_{k\in\mathbb{Z}_+}\{k : y_k + u_k = \widetilde{y}_k + \widetilde{u}_k | \pi^1 = \widetilde{\pi}^1 = \hat{\pi}, y_1, \widetilde{y}_1\}$. Hence, if $i = j$ and $y_1 = \widetilde{y}_1$, $\mathbb{E}[\tilde{K}_{ij}] = 1$. If $i = j$ but $y_1 \neq \widetilde{y}_1$, the inventory positions become equal to each other after orders are observed in both systems. Without loss of generality, suppose that $y_1 \leq \widetilde{y}_1$. Since $S^\pi_\alpha \geq 0$, $\forall \pi \in \Pi$, by Proposition 1, we place an order no later than the period up to which a total of $\widetilde{y}_1 + 1$ units of demand is observed. Hence, $\tilde{K}_{ij} \leq \min\{n : \sum_{t=1}^n w_t \geq \widetilde{y}_1 + 1 |\ \pi^1 = \hat{\pi}\}$. For a sample path starting with belief $\pi \in \Pi$, and for a finite $\xi \in \mathbb{Z}_+$, let $\tau_{\pi,\xi} := \min\{n : \sum_{t=1}^n w_t \geq \xi |\ \pi^1 = \pi\}$ be the first period when the cumulative demand is no less than ξ. As we assume $\exists i \in \mathcal{N}$ such that $r_i(k) > 0$ for some $k > 0$, and by Assumption 1, we have $\mathbb{P}\{w_t \geq 1 | \pi^t\} > 0$, $\forall t \in \mathbb{Z}_+$. Thus $\mathbb{P}\left\{\sum_{t=1}^\xi w_t \geq \xi \middle| \pi^1 = \pi\right\} > 0$, $\forall \pi \in \Pi$. We define

$$\rho_\xi := \max_{\pi\in\Pi}\left\{\mathbb{P}\left[\sum_{t=1}^{\xi} w_t < \xi \middle| \pi^1 = \pi\right]\right\} < 1.$$

Notice that

$$\begin{aligned}
\mathbb{E}[\tau_{\pi,\xi}] &= \sum_{n=0}^{\infty} \mathbb{P}\left\{\tau_{\pi,\xi} > n\right\} = \sum_{n=0}^{\xi-1} \mathbb{P}\left\{\tau_{\pi,\xi} > n\right\} + \sum_{n=\xi}^{\infty} \mathbb{P}\left\{\tau_{\pi,\xi} > n\right\} \\
&\leq \xi + \sum_{n=\xi}^{\infty} \mathbb{P}\left\{\tau_{\pi,\xi} > n\right\} \\
&= \xi + \sum_{n=\xi}^{\infty} \mathbb{P}\left\{\sum_{t=1}^{n} w_t < \xi \;\middle|\; \pi^1 = \pi\right\} \\
&= \xi + \sum_{k=1}^{\infty} \sum_{m=k\xi}^{(k+1)\xi-1} \mathbb{P}\left\{\sum_{t=1}^{m} w_t < \xi \;\middle|\; \pi^1 = \pi\right\} \\
&\leq \xi + \sum_{k=1}^{\infty} \sum_{m=k\xi}^{(k+1)\xi-1} \mathbb{P}\left\{\sum_{t=1}^{k\xi} w_t < \xi \;\middle|\; \pi^1 = \pi\right\} \\
&= \xi + \sum_{k=1}^{\infty} \xi\, \mathbb{P}\left\{\sum_{t=1}^{k\xi} w_t < \xi \;\middle|\; \pi^1 = \pi\right\} \\
&\leq \xi + \xi \sum_{k=1}^{\infty} \prod_{m=1}^{k} \mathbb{P}\left\{\sum_{t=(m-1)\xi+1}^{m\xi} w_t < \xi \;\middle|\; \pi^1 = \pi\right\} \\
&\leq \xi + \xi \sum_{k=1}^{\infty} \rho_\xi^k \\
&= \frac{\xi}{1-\rho_\xi} < \infty.
\end{aligned}$$

Thus if $i = j$ but $y_1 \neq \widetilde{y}_1$, because $\tilde{K}_{ij} \leq \tau_{\pi,\widetilde{y}_1+1}$, we obtain $\mathbb{E}[\tilde{K}_{ij}] < \infty$. If $i \neq j$, because $\tilde{K}_{ij} \leq \eta_{ij} + \tau_{\pi,\widetilde{y}_1+1}$ and $\mathbb{E}[\eta_{ij}] < \infty$, we again obtain $\mathbb{E}[\tilde{K}_{ij}] < \infty$. Hence there exists a finite $C \in \mathbb{R}_+$ such that $\mathbb{E}[\tilde{K}_{ij} - 1] \leq C(|i-j| + |y - \widetilde{y}|)$.

Now recall the inequality in (4.7):

$$
\begin{aligned}
&\delta_\alpha(\widetilde{\pi},\widetilde{y}) - \delta_\alpha(\pi, y) \\
&\leq (A + cB) \sum_{i\in\mathcal{N}} \sum_{j\in\mathcal{N}} \mathbb{E}[\tilde{K}_{ij} - 1]\pi_i\widetilde{\pi}_j + cM \sum_{i\in\mathcal{N}} \sum_{j\in\mathcal{N}} \mathbb{E}[\eta_{ij} - 1]\pi_i\widetilde{\pi}_j + c(y - \widetilde{y}) \\
&\leq (A + cB) \sum_{i\in\mathcal{N}} \sum_{j\in\mathcal{N}} C(|i - j| + |y - \widetilde{y}|)\pi_i\widetilde{\pi}_j + cM\frac{\Gamma}{\mu} \sum_{i\in\mathcal{N}} \sum_{j\in\mathcal{N}} |i - j|\pi_i\widetilde{\pi}_j + c|y - \widetilde{y}| \\
&= \left(AC + cBC + cM\frac{\Gamma}{\mu}\right) \sum_{i\in\mathcal{N}} \sum_{j\in\mathcal{N}} |i - j|\pi_i\widetilde{\pi}_j + (AC + cBC + c)|y - \widetilde{y}| \\
&= \left(AC + cBC + cM\frac{\Gamma}{\mu}\right) \mathbb{E}[|d_1 - \tilde{d}_1|] + (AC + cBC + c)|y - \widetilde{y}|. \qquad (4.8)
\end{aligned}
$$

By an appropriate choice of the joint mass function of $(d_1, \tilde{d}_1)$, we can obtain

$$\delta_\alpha(\widetilde{\pi},\widetilde{y}) - \delta_\alpha(\pi, y) \leq \left(AC + cBC + cM\frac{\Gamma}{\mu}\right)[\Delta(\pi,\widetilde{\pi}) + \varepsilon] + (AC + cBC + c)|y - \widetilde{y}|$$

for some $\varepsilon > 0$. Thus there exists a finite $D \in \mathbb{R}_+$ such that $\delta_\alpha(\widetilde{\pi},\widetilde{y}) - \delta_\alpha(\pi, y) \leq D[\Delta(\pi,\widetilde{\pi}) + |y - \widetilde{y}|]$. As we assume that $0 \leq \delta_\alpha(\widetilde{\pi},\widetilde{y}) - \delta_\alpha(\pi, y)$, we have $|\delta_\alpha(\widetilde{\pi},\widetilde{y}) - \delta_\alpha(\pi, y)| \leq D[\Delta(\pi,\widetilde{\pi}) + |y - \widetilde{y}|]$. Thus $\delta_\alpha(\cdot,\cdot)$ is locally Lipschitz continuous for $\alpha \in (0, 1)$.

Because the inequality in (4.8) holds for any $\pi, \widetilde{\pi} \in \Pi$ and for any finite $y, \tilde{y} \in \mathbb{Z}$, and $\delta_\alpha(\bar{\pi}, 0) = v_\alpha(\bar{\pi}, 0) - v_\alpha(\bar{\pi}, 0) = 0$, the following inequality hold.

$$
\begin{aligned}
|\delta_\alpha(\pi, y)| &= |\delta_\alpha(\pi, y) - \delta_\alpha(\bar{\pi}, 0)| \\
&\leq \left(AC + cBC + cM\frac{\Gamma}{\mu}\right) \mathbb{E}[|d_1 - \tilde{d}_1|] + (AC + cBC + c)|y|.
\end{aligned}
$$

Since $|d_1 - \tilde{d}_1| \leq N$, there exists a finite $E \in \mathbb{R}_+$ such that $|\delta_\alpha(\pi, y)| \leq E$. Thus $\delta_\alpha(\cdot,\cdot)$ is uniformly bounded with respect to $\alpha \in (0, 1)$. Since $\delta_\alpha(\cdot,\cdot)$ is also locally Lipschitz continuous for $\alpha \in (0, 1)$, by the Arzela-Ascoli Theorem, there exists a subsequence $\alpha_t \to 1$ (which can be the same as in Lemma 1) and a locally Lipschitz continuous function $\delta^*(\pi, y)$ such that $\delta_{\alpha_t}(\pi, y) \to \delta^*(\pi, y)$, for all $\pi \in \Pi$ and for any finite $y \in \mathbb{Z}$. ■

We are now ready to state the main result of this thesis that builds upon Lemmas 1 and 2:

Theorem 1. *Under Assumption 1,* (λ^*, δ^*) *satisfies the average-cost optimality equation*

$$\delta(\pi, y) + \lambda = \min_{u \geq 0} \left\{ cu + g(\pi, y+u) + \sum_{w=0}^{M} \delta(T(\pi, w), y+u-w)\hat{r}_\pi(w) \right\}. \quad (4.9)$$

Furthermore, there exists an optimal stationary inventory replenishment policy that can be described as a belief-dependent base-stock policy with base-stock levels $S^\pi, \forall \pi \in \Pi$.

Proof. Proof. We take the limit on both sides of the equation in (4.2) as $\alpha_t \to 1$:

$$\lim_{\alpha_t \to 1} \{\delta_{\alpha_t}(\pi, y) + (1-\alpha_t) v_{\alpha_t}(\bar{\pi}, 0)\}$$
$$= \lim_{\alpha_t \to 1} \left\{ \min_{u \geq 0} \left\{ cu + (\alpha_t)^l g(\pi, y+u) + \alpha_t \sum_{w=0}^{M} \delta_{\alpha_t}(T(\pi, w), y+u-w)\hat{r}_\pi(w) \right\} \right\}.$$

By Lemma 1, $\lim_{\alpha_t \to 1}(1-\alpha_t) v_{\alpha_t}(\bar{\pi}, 0) = \lambda^*$. By Lemma 2, $\lim_{\alpha_t \to 1} \delta_{\alpha_t}(\pi, y) = \delta^*(\pi, y)$. By Proposition 1, $y + u - w \in [y - M, \max\{y, (l+1)M\}]$. Thus, again by Lemma 2, $\lim_{\alpha_t \to 1} \delta_{\alpha_t}(T(\pi, w), y+u-w) = \delta^*(T(\pi, w), y+u-w)$. Hence:

$$\delta^*(\pi, y) + \lambda^* = \min_{u \geq 0} \left\{ cu + g(\pi, y+u) + \sum_{w=0}^{M} \delta^*(T(\pi, w), y+u-w)\hat{r}_\pi(w) \right\}. \quad (4.10)$$

Theorem 1 in [35] states that if $\frac{1}{n}\mathbb{E}[\delta(\pi^n, y_n) | \pi^1 = \pi, y_1 = y] \to 0$ for all $\pi \in \Pi$ and for all finite $y \in \mathbb{Z}$, there exists an optimal stationary policy. We know from Lemma 2 that there exists a finite E such that $\delta(\pi, y) \leq E$ for all $\pi \in \Pi$ and for all finite $y \in \mathbb{Z}$. Hence:

$$\lim_{n \to \infty} \left\{ \frac{1}{n} \mathbb{E}[\delta(\pi^n, y_n) | \pi^1 = \pi, y_1 = y] \right\} \leq \lim_{n \to \infty} \left\{ \frac{1}{n} E \right\} = 0.$$

We thus verify that there exists an optimal stationary policy. By definition of δ_α and Proposition 1, $\delta_{\alpha_t}(\pi, y)$ is discrete-convex in y, $\forall \pi \in \Pi$. Because the limit of a sequence of discrete-convex functions is discrete-convex, $\delta^*(\pi, y)$ is also discrete-convex in y, $\forall \pi \in \Pi$. Thus the optimal stationary policy is a belief-dependent base-stock policy. ■

In the literature, several authors have identified the optimal policy structure for average-cost inventory systems with Markov-modulated demand when the state of the underlying Markov chain is *perfectly* observed; see [28] and [29]. To our knowledge, however, we are the first to characterize the optimal policy structure for average-cost inventory systems with non-stationary demand and *partial* observation.

Chapter 5

Discretized Approximation

Solving the optimality equation in (4.9) for each state $(\pi, y) \in \Pi \times \mathbb{Z}$ and finding the optimal base-stock level for each belief $\pi \in \Pi$ is a computational challenge since Π is an *uncountable* space and $\mathbb{Z}$ is a countably *infinite* set. We know from Proposition 1 that the optimal base-stock levels are bounded between 0 and $(l+1)M$ in the discounted-cost problem. Following the same proof steps as in Proposition 1, we are able to extend these bounds to the average-cost problem. Therefore, in the average-cost problem, the inventory positions can be restricted to take values between $-M$ and $(l+1)M$. Notice that if the initial inventory position is above $(l+1)M$ or below $-M$, it will eventually fall into this range after a finite number of periods. The contribution of the cost caused by the excess or insufficient inventory in those initial periods to the average cost can thus be disregarded in our infinite-horizon planning. Hence, without loss of generality, the optimality equations in (4.9) can be restricted to the state space $\Pi \times \mathbb{Z}_M^l$ where $\mathbb{Z}_M^l := \{y \in \mathbb{Z} : -M \leq y \leq (l+1)M\}$.

We next discretize the uncountable space Π, on which the beliefs are defined, based on the regular grid approach developed by Lovejoy [20]: Let Q_n be a regular grid for a given $n \in \mathbb{Z}_+$ such that the convex hull of Q_n is Π. Specifically, Q_n is

defined by

$$Q_n := \left\{ [q_1, ..., q_N] \in \mathbb{Q}^N \;\middle|\; q_i = \frac{k_i}{n}, \sum_{i=1}^{N} k_i = n, k_i \in \mathbb{Z}_+ \cup \{0\} \right\},$$

where $\mathbb{Q}$ denotes the set of rational numbers. The number of grid points in Q_n is $\kappa_n = |Q_n| = \frac{(N-1+n)!}{(N-1)!n!}$. We thus denote the elements of Q_n by $\{q^1, \ldots, q^{\kappa_n}\}$.

By Carathéodory's Fundamental Theorem, any point in Π can be written as a convex combination of at most N elements of Q_n. We utilize a linear program (LP) to determine the convex combination multipliers. Let $\gamma_i(\pi)$, $i = 1, \ldots, \kappa_n$, be the decision variables of the following LP:

$$\begin{aligned}
\min \; & \sum_{i=1}^{\kappa_n} \gamma_i(\pi) ||\pi - q^i|| \\
\text{s.t.} \; & \sum_{i=1}^{\kappa_n} \gamma_i(\pi) q^i = \pi, \\
& \sum_{i=1}^{\kappa_n} \gamma_i(\pi) = 1, \\
& \gamma_i(\pi) \geq 0, \qquad \forall i = 1, \ldots, \kappa_n,
\end{aligned}$$

where $||\cdot||$ denotes the Euclidean distance. Solution of the above LP yields the convex representation scheme $\underline{\gamma}_n := (\gamma_1^*(\cdot), \ldots, \gamma_{\kappa_n}^*(\cdot))$.

Following [21], let ϵ_n denote the fineness of the discretization scheme $(Q_n, \underline{\gamma}_n)$ that is defined by

$$\epsilon_n := \max_{\pi \in \Pi} \max_{q^i \in Q_n : \gamma_i(\pi) > 0} ||\pi - q^i||.$$

Because Q_n is a regular grid and any belief can only be represented by the closest grid points to that belief according to our construction of $\underline{\gamma}_n$, it can be shown that $\epsilon_n = \frac{\sqrt{N}}{n\sqrt{N+1}}$. Note that $\epsilon_n \to 0$ as $n \to \infty$. For any $n \in \mathbb{Z}_+$, we can compute the optimal average cost λ_{ϵ_n} associated 6with an ϵ_n-discretization scheme $(Q_n, \underline{\gamma}_n)$ by

solving the following optimality equations:

$$\delta(q,y)+\lambda=\min_{u\geq 0}\left\{cu+g(q,y+u)+\sum_{i=1}^{\kappa_n}\sum_{w=0}^{M}\gamma_i(T(q,w))\delta(q^i,y+u-w)\hat{r}_q(w)\right\},$$
$$\forall q\in Q_n,\ \forall y\in\mathbb{Z}_M^l.$$

Following Theorems 1 and 3 in [21], we verify that λ_{ϵ_n} increasingly converges to the optimal average cost λ^* as n grows large. We will use the lower bound λ_{ϵ_n} obtained from this discretization scheme in our performance evaluation of a myopic base-stock policy in Chapter 6.

Chapter 6

Numerical Results

For our MDP formulation in Chapter 3, we conduct numerical experiments to investigate the value of implementing the Bayesian updating mechanism (see Chapter 6.1) and the performance of a myopic belief-dependent base-stock policy as a heuristic replenishment policy (see Chapter 6.2). We consider instances with three demand states, i.e., $\mathcal{N}$=$\{1,\ 2,\ 3\}$, such that the demand distributions are $Binomial(20,\ p)$, $Binomial(20,\ 0.5)$, and $Binomial(20,\ 1-p)$ for the demand states 1, 2, and 3, respectively. We then generate instances in which $c = 1$, $b = 20$, $h \in \{2, 5, 10\}$, $l \in \{0, 1, 2\}$, $p \in \{0.1, 0.2, 0.3, 0.4\}$, and the transition matrix P is

$$P_1 = \begin{bmatrix} 0.5 & 0.25 & 0.25 \\ 0.25 & 0.5 & 0.25 \\ 0.25 & 0.25 & 0.5 \end{bmatrix},\ P_2 = \begin{bmatrix} 0.7 & 0.15 & 0.15 \\ 0.15 & 0.7 & 0.15 \\ 0.15 & 0.15 & 0.7 \end{bmatrix}, P_3 = \begin{bmatrix} 0.9 & 0.05 & 0.05 \\ 0.05 & 0.9 & 0.05 \\ 0.05 & 0.05 & 0.9 \end{bmatrix}.$$

Note that Assumption 1 holds in each of our 108 instances. For each of our instances, we calculate the average costs λ_{ϵ_n} associated with our discretization scheme for $n \in \{1, 2, 4, 8, 16, 32\}$ and their percentage differences from the average cost λ_{ϵ_1}, i.e., $100 \times \frac{\lambda_{\epsilon_n} - \lambda_{\epsilon_1}}{\lambda_{\epsilon_1}}$. Note that λ_{ϵ_1} is the worst lower bound that can be obtained from our discretization scheme. Also, λ_{ϵ_1} is the minimum average cost that could be achieved if the demand state were *perfectly* observed. Figures 6.1-6.3 exhibit the percentage gaps $100 \times \frac{\lambda_{\epsilon_n} - \lambda_{\epsilon_1}}{\lambda_{\epsilon_1}}$ for our instances with $l = 0,\ 1,$

and 2, respectively: The average cost $\lambda_{\epsilon_{32}}$ is a sufficiently tight lower bound on the optimal average cost for our MDP in Section 3, in each our instances. Consequently, and since computational burden increases rapidly with n for our discretization scheme, in this chapter we base our optimality gap calculations on this lower bound. Our simulation runs consist of 30 replications of 10000 periods each in all numerical experiments.

6.1 The Value of Bayesian Updating

In order to investigate the value of implementing the Bayesian updating mechanism in our MDP formulation, first, we consider a much simpler MDP with a stationary demand distribution that is obtained by compounding the demand distributions based on the stationary distribution of the underlying Markov chain. For such an MDP, a myopic base-stock policy with a single stationary base-stock level is optimal (see [22]) and the optimal base-stock level can be easily found using the newsvendor formula applied on the convolution of demand over $(l+1)$ periods (see [15]). For each of our 108 instances, we calculate the optimal base-stock level for this MDP and simulate the inventory system under this base-stock level. Our simulation results indicate that the average cost under this base-stock level is on average 16.3% greater than the lower bound $\lambda_{\epsilon_{32}}$ on our test bed, highlighting the importance of taking into account the non-stationarity of demand distribution and employing the Bayesian updating mechanism.

Another simplification of our MDP with Bayesian updating (and thus uncountable state space) is to formulate an MDP with perfectly observed demand states (and thus countable state space), the optimal policy of which is used to determine the action to be taken in the estimated demand state in every period. As our estimate of the demand state in this method, following Chapter 9 in [17], we choose the state with the highest posterior probability based on the entire demand history. Our simulation results show that the average cost under this simplification is on average 8.48% greater than the lower bound $\lambda_{\epsilon_{32}}$ on our test bed, indicating the significance of Bayesian updating in our problem.

Figure 6.1: $100 \times \frac{\lambda_{\epsilon_n} - \lambda_{\epsilon_1}}{\lambda_{\epsilon_1}}$ vs. n when $c = 1$, $b = 20$, $l = 0$, $P \in \{P_1, P_2, P_3\}$, $p \in \{0.1, 0.2, 0.3, 0.4\}$, $h \in \{2, 5, 10\}$.

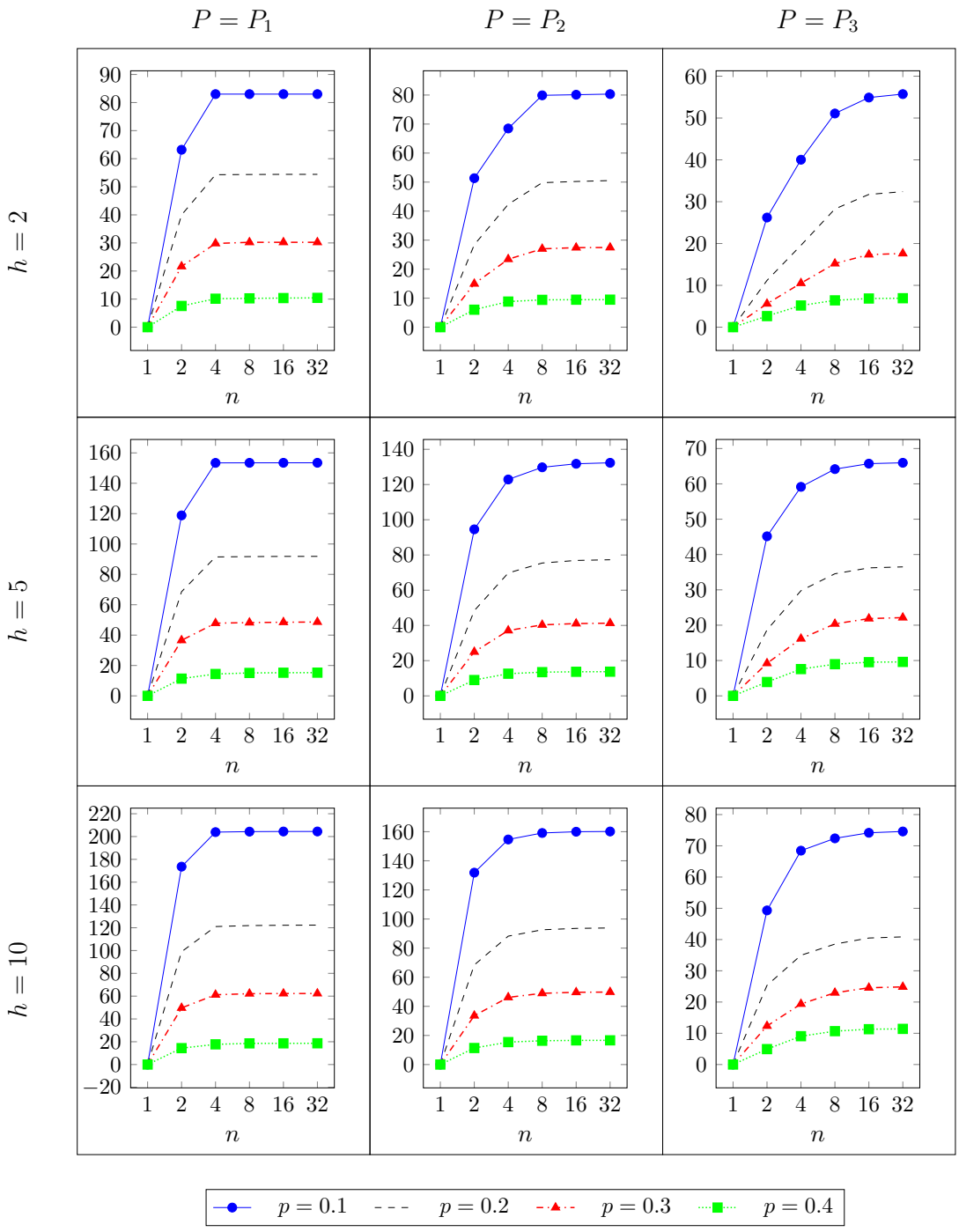


Figure 6.2: $100 \times \frac{\lambda_{\epsilon_n} - \lambda_{\epsilon_1}}{\lambda_{\epsilon_1}}$ vs. n when $c = 1$, $b = 20$, $l = 1$, $P \in \{P_1, P_2, P_3\}$, $p \in \{0.1, 0.2, 0.3, 0.4\}$, $h \in \{2, 5, 10\}$.

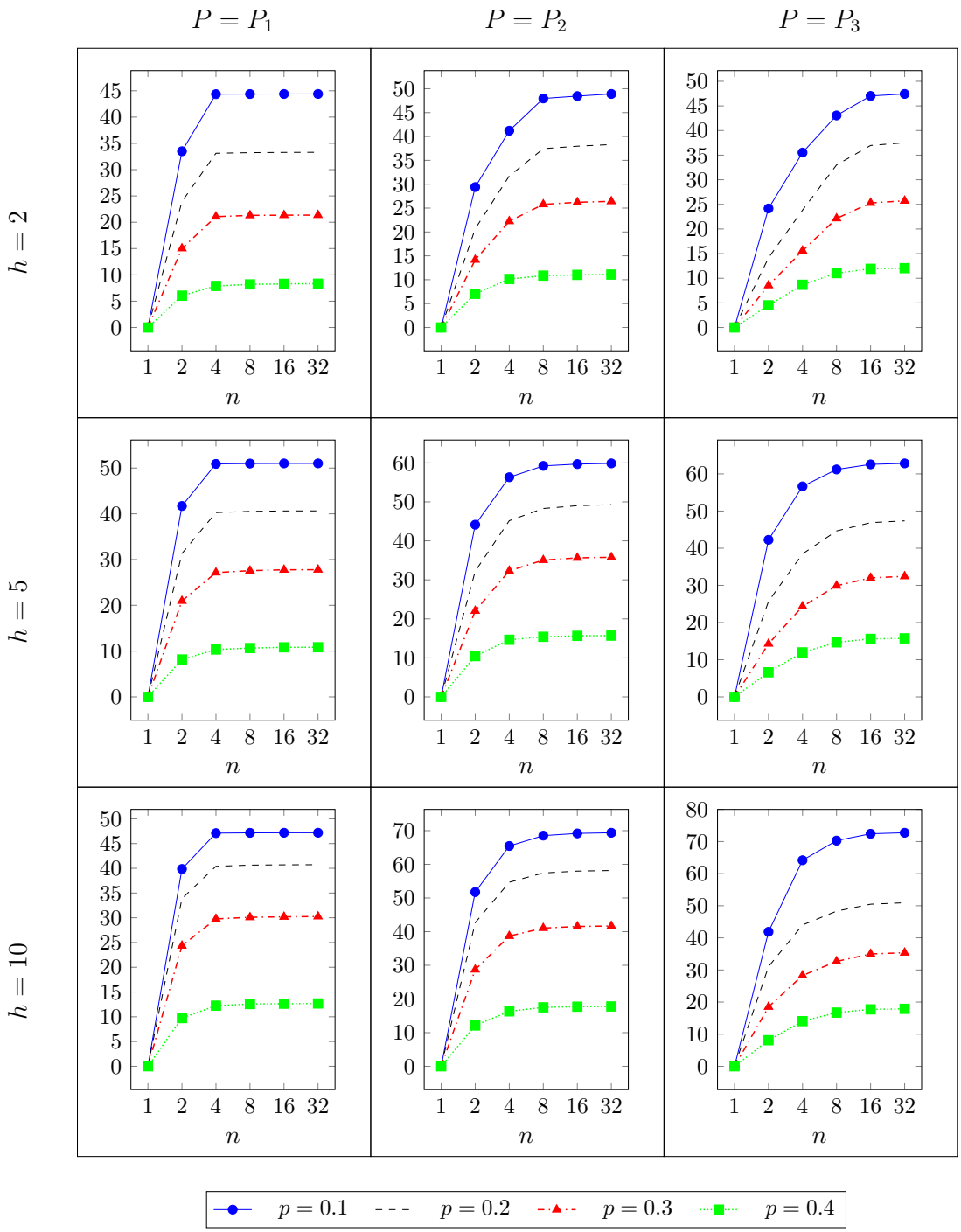
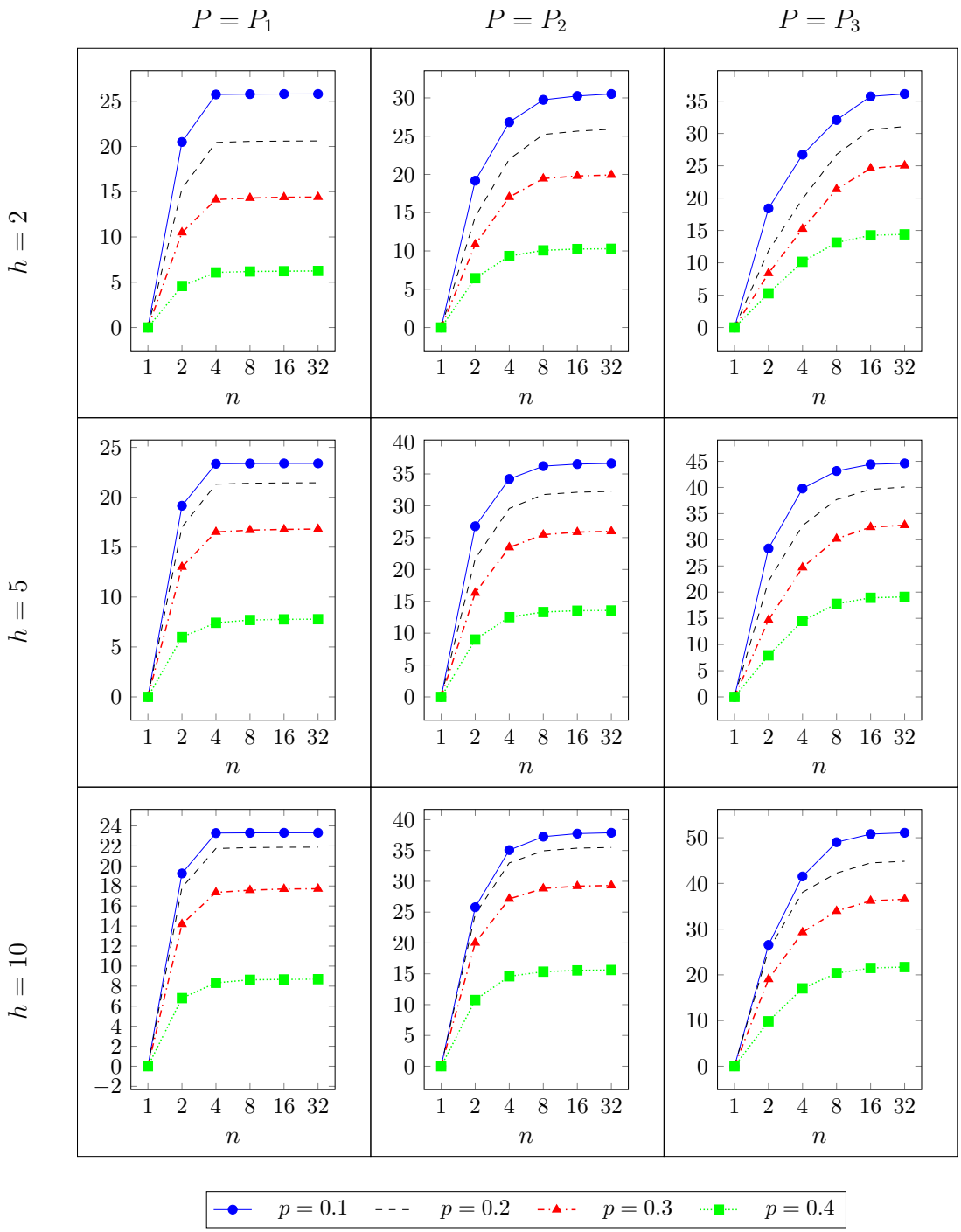


Figure 6.3: $100 \times \frac{\lambda_{\epsilon_n} - \lambda_{\epsilon_1}}{\lambda_{\epsilon_1}}$ vs. n when $c = 1$, $b = 20$, $l = 2$, $P \in \{P_1, P_2, P_3\}$, $p \in \{0.1, 0.2, 0.3, 0.4\}$, $h \in \{2, 5, 10\}$.



6.2 Performance Evaluation of the Myopic Base-Stock Policy

We now adapt the myopic base-stock policy introduced by Veinott [22] to our inventory model as a heuristic replenishment policy. In this heuristic, the order quantity in period t is determined according to a myopic belief-dependent base-stock level that is calculated as follows:

$$\tilde{S}^{\pi^t} = \underset{k\in\{0,\ldots,(l+1)M\}}{\arg\min} \left(\mathbb{P}\left\{ \sum_{n=0}^{l} w_{t+n} \leq k \middle| \pi^t \right\} \geq \frac{b}{h+b} \right).$$

For each of our instances, we simulate the inventory system under this myopic belief-dependent base-stock policy, calculating the average cost denoted by $\tilde{\lambda}$ in each replication. Figures 6.4-6.6 exhibit the 95% confidence intervals for the percentage difference from our lower bound, i.e., $100 \times \frac{\tilde{\lambda}-\lambda_{\epsilon_{32}}}{\lambda_{\epsilon_{32}}}$, for our instances with $l = 0$, 1, and 2, respectively.

We observe from Figures 6.4-6.6 that the confidence intervals contain zero in 92 of the 108 instances: The myopic base-stock policy is optimal at a confidence level of 95% in those instances. We also observe that the largest optimality gaps (no more than 2.33%) tend to occur when $p = 0.1$ and $h = 10$: The myopic base-stock policy can be shown to be optimal if $y_t \leq \tilde{S}^{\pi^t}$ with probability one (see [25]). It performs worse as the likelihood of excess inventory at the beginning of any period, i.e., $\mathbb{P}\{y_t \geq \tilde{S}^{\pi^t}\}$, increases. For the instances with $p = 0.1$, in a single period, the lowest possible expected demand is $20 \times p = 2$ while the highest possible expected demand is $20 \times (1-p) = 18$. For these instances with highly fluctuating demand, the base-stock levels are likely to vary more significantly over time, leading to a larger $\mathbb{P}\{y_t \geq \tilde{S}^{\pi^t}\}$. Hence, and since the holding cost is high, a worse performance results.

Figure 6.4: 95% confidence intervals for $100 \times \frac{\tilde{\lambda}-\lambda_{\epsilon_{32}}}{\lambda_{\epsilon_{32}}}$ vs. p when $c = 1$, $b = 20$, $l = 0$, $P \in \{P_1, P_2, P_3\}$, $h \in \{2, 5, 10\}$.

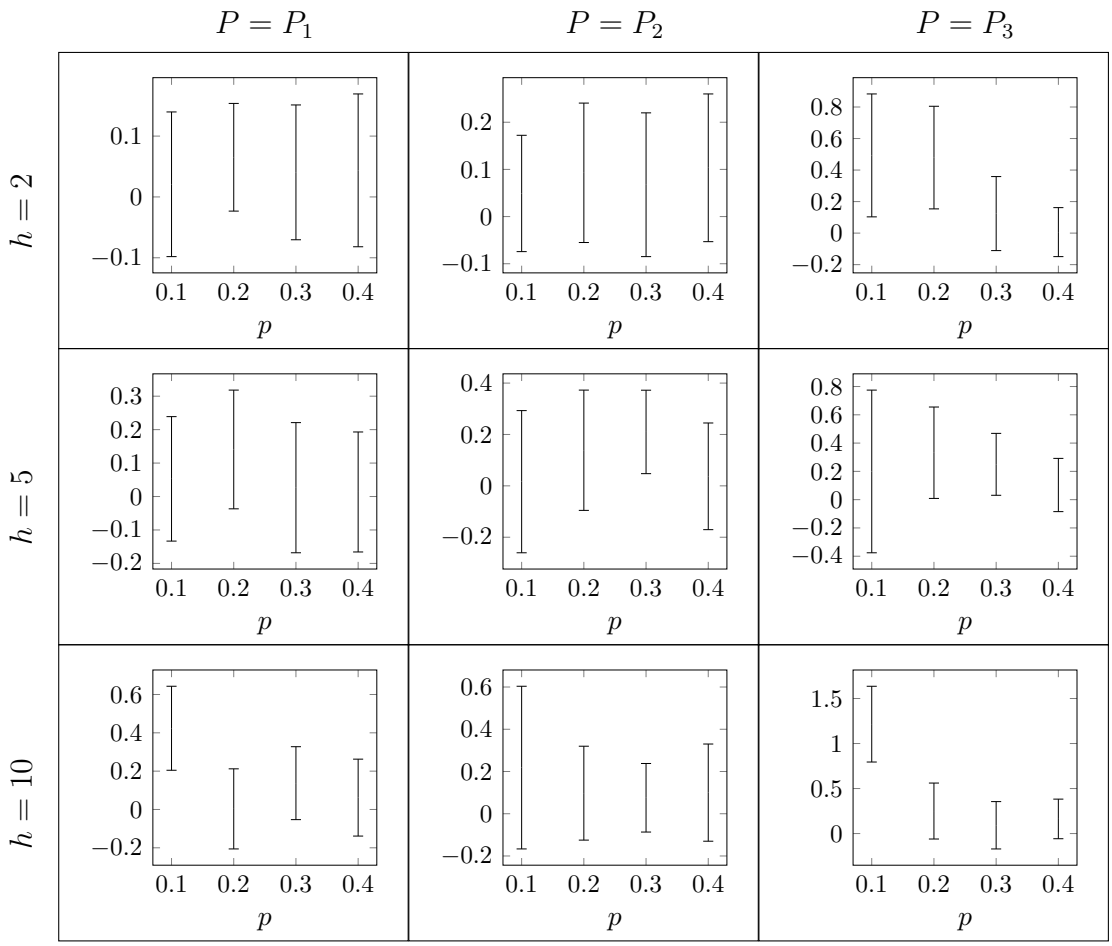


Figure 6.5: 95% confidence intervals for $100 \times \frac{\tilde{\lambda}-\lambda_{\epsilon_{32}}}{\lambda_{\epsilon_{32}}}$ vs. p when $c = 1$, $b = 20$, $l = 1$, $P \in \{P_1, P_2, P_3\}$, $h \in \{2, 5, 10\}$.

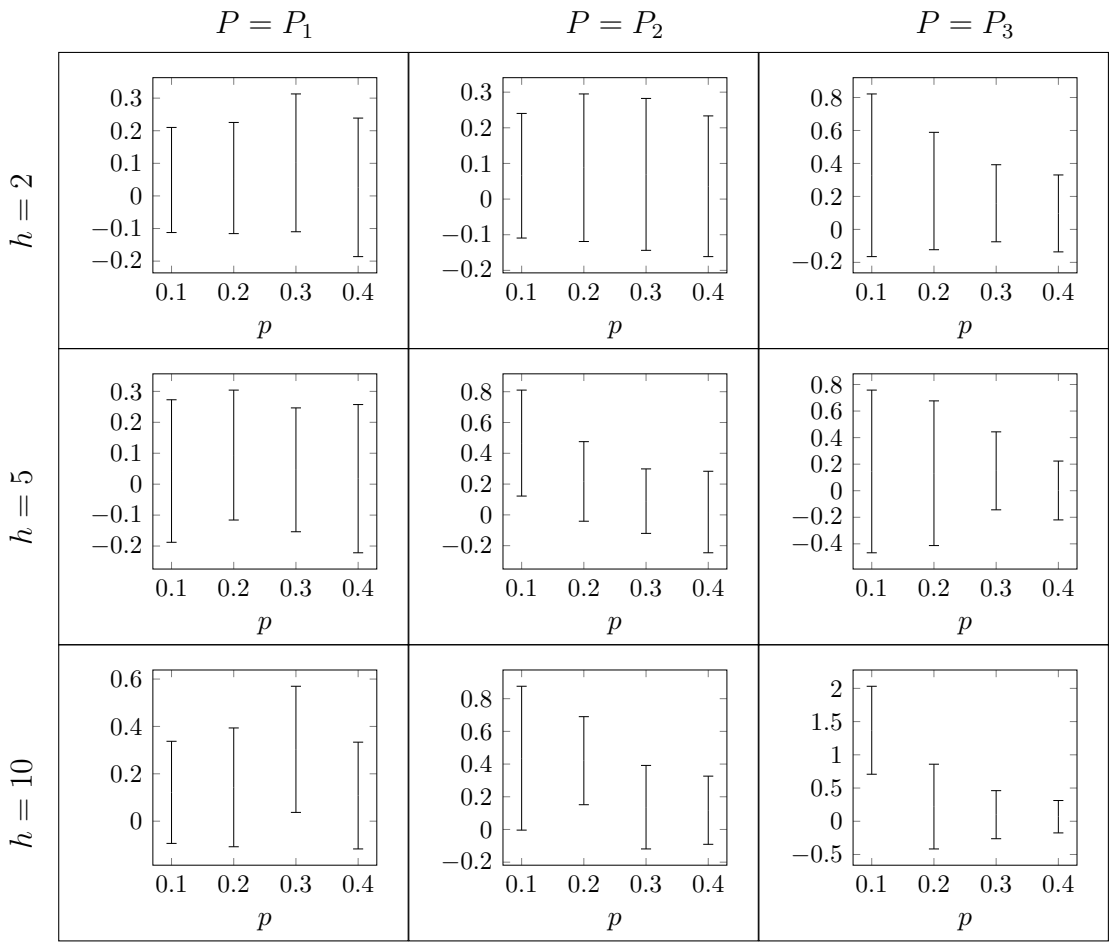
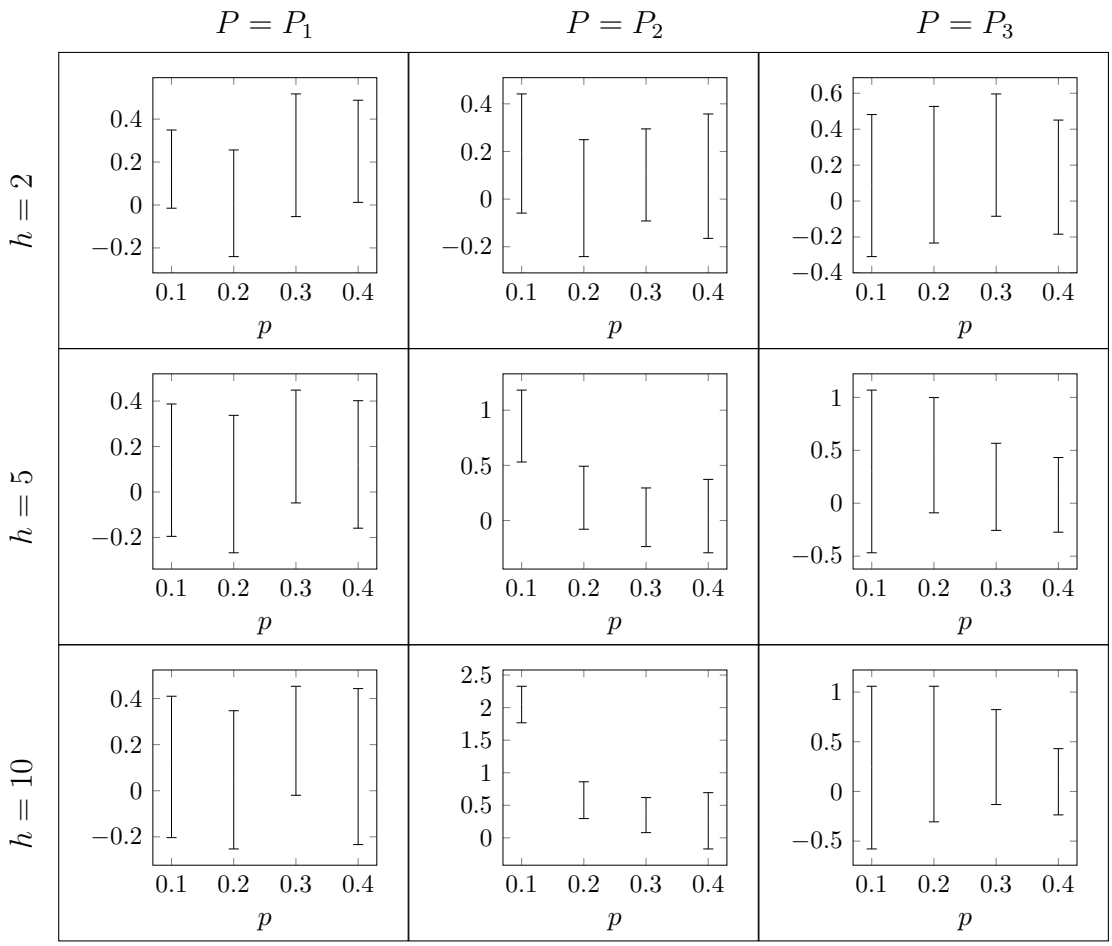


Figure 6.6: 95% confidence intervals for $100 \times \frac{\tilde{\lambda}-\lambda_{\epsilon_{32}}}{\lambda_{\epsilon_{32}}}$ vs. p when $c = 1$, $b = 20$, $l = 2$, $P \in \{P_1, P_2, P_3\}$, $h \in \{2, 5, 10\}$.



Chapter 7

Conclusions

We have studied the inventory replenishment problem when the demand distribution undergoes Markovian transitions over time. The state of the underlying Markov chain can be only partially observed based on past demand data. After formulating this problem as an MDP with Bayesian updating, we have established the optimality of a belief-dependent base-stock policy in the discounted-cost case. Using the vanishing discount method when the underlying Markov chain is ergodic, we have extended the optimality of the belief-dependent base-stock policy to the average-cost case. Our numerical experiments have revealed the outstanding cost performance of the myopic belief-dependent base-stock policy, which is easy to implement in practice.

Future extensions of this thesis could consider inventory models with fixed replenishment order costs. In the literature dealing with fixed ordering costs, Iglehart [13] and Zheng [45] have established the optimality of an (s, S) policy for average-cost inventory models with stationary demand, and Beyer and Sethi [28] have shown the optimality of a state-dependent (s, S) policy for average-cost inventory models with perfectly observed Markov-modulated demand. Leveraging our structural analysis, the optimality of (s, S) policies may be extended to average-cost models with partially observed Markov-modulated demand. Our

research may also guide future research aimed at characterizing the optimal policy structure in more complex average-cost inventory models, such as multi-item and/or multi-echelon inventory systems with partial demand information. Lastly, future research could study the inventory replenishment problem under more limited information about demand. Examples include inventory models with unknown demand distributions and unknown transition matrices for the underlying Markov chain, and inventory models with unknown numbers of demand states. The Baum-Welch and Viterbi algorithms may be usefully employed in estimation of such unknown parameters, enabling optimal policy characterizations. See Chapter 9 in [46] for detailed descriptions of these algorithms.

Bibliography

[1] K. H. Shang, "Single-stage approximations for optimal policies in serial inventory systems with nonstationary demand," *Manufacturing Service Oper. Management*, vol. 14, no. 3, pp. 414–422, 2012.

[2] S. Kesavan and T. Kushwaha, "Differences in retail inventory investment behavior during macroeconomic shocks: Role of service level," *Production Oper. Management*, vol. 23, no. 12, pp. 2118–2136, 2014.

[3] J. Hu, C. Zhang, and C. Zhu, "(s, S) inventory systems with correlated demands," *INFORMS J. Comput.*, vol. 28, no. 4, pp. 603–611, 2016.

[4] D. Beyer, F. Cheng, S. P. Sethi, and M. Taksar, *Markovian demand inventory models*. Springer, 2010.

[5] K. Arifoğlu and S. Özekici, "Inventory management with random supply and imperfect information: A hidden Markov model," *Int. J. Prod. Econ.*, vol. 134, no. 1, pp. 123–137, 2011.

[6] R. Levi, G. Perakis, and J. Uichanco, "The data-driven newsvendor problem: new bounds and insights," *Oper. Res.*, vol. 63, no. 6, pp. 1294–1306, 2015.

[7] X. Ding, M. L. Puterman, and A. Bisi, "The censored newsvendor and the optimal acquisition of information," *Oper. Res.*, vol. 50, no. 3, pp. 517–527, 2002.

[8] D. P. Bertsekas, *Dynamic Programming and Optimal Control, Vols. II.* Athena Scientific, 2012.

[9] B. Sandıkçı, "Reduction of a POMDP to an MDP," *Wiley Encyclopedia of Oper. Res. and Management Sci.*, 2010.

[10] H. Scarf, "Bayes solutions of the statistical inventory problem," *Ann. Math. Statist.*, vol. 30, no. 2, pp. 490–508, 1959.

[11] H. E. Scarf, "Some remarks on Bayes solutions to the inventory problem," *Naval Res. Logist.*, vol. 7, no. 4, pp. 591–596, 1960.

[12] S. Karlin, "Dynamic inventory policy with varying stochastic demands," *Management Sci.*, vol. 6, no. 3, pp. 231–258, 1960.

[13] D. L. Iglehart, "The dynamic inventory problem with unknown demand distribution," *Management Sci.*, vol. 10, no. 3, pp. 429–440, 1964.

[14] K. S. Azoury, "Bayes solution to dynamic inventory models under unknown demand distribution," *Management Sci.*, vol. 31, no. 9, pp. 1150–1160, 1985.

[15] J. T. Treharne and C. R. Sox, "Adaptive inventory control for nonstationary demand and partial information," *Management Sci.*, vol. 48, no. 5, pp. 607–624, 2002.

[16] K. Arifoğlu and S. Özekici, "Optimal policies for inventory systems with finite capacity and partially observed Markov-modulated demand and supply processes," *Eur. J. Oper. Res.*, vol. 204, no. 3, pp. 421–438, 2010.

[17] D. Barber, *Bayesian reasoning and machine learning*. Cambridge University Press, 2012.

[18] R. Zhou and E. A. Hansen, "An improved grid-based approximation algorithm for POMDPs," in *Proc. 17th Internat. Joint Conf. Artificial Intelligence*, pp. 707–716, 2001.

[19] N. Saldı, S. Yüksel, and T. Linder, "On the asymptotic optimality of finite approximations to Markov decision processes with Borel spaces," *Math. Oper. Res.*, vol. 42, no. 4, pp. 945–978, 2017.

[20] W. S. Lovejoy, "Computationally feasible bounds for partially observed Markov decision processes," *Oper. Res.*, vol. 39, no. 1, pp. 162–175, 1991.

[21] H. Yu and D. P. Bertsekas, "Discretized approximations for POMDP with average cost," pp. 619–627, 20th Conf. UAI, 7 2004.

[22] A. F. Veinott, "Optimal policy for a multi-product, dynamic, nonstationary inventory problem," *Management Sci.*, vol. 12, no. 3, pp. 206–222, 1965.

[23] W. S. Lovejoy, "Myopic policies for some inventory models with uncertain demand distributions," *Management Sci.*, vol. 36, no. 6, pp. 724–738, 1990.

[24] G. D. Johnson and H. Thompson, "Optimality of myopic inventory policies for certain dependent demand processes," *Management Sci.*, vol. 21, no. 11, pp. 1303–1307, 1975.

[25] W. S. Lovejoy, "Stopped myopic policies in some inventory models with generalized demand processes," *Management Sci.*, vol. 38, no. 5, pp. 688–707, 1992.

[26] J. Song and P. Zipkin, "Inventory control in a fluctuating demand environment," *Oper. Res.*, vol. 41, no. 2, pp. 351–370, 1993.

[27] S. P. Sethi and F. Cheng, "Optimality of (s,S) policies in inventory models with Markovian demand," *Oper. Res.*, vol. 45, no. 6, pp. 931–939, 1997.

[28] D. Beyer and S. P. Sethi, "Average cost optimality in inventory models with Markovian demands," *J. Optim. Theory Appl.*, vol. 92, no. 3, pp. 497–526, 1997.

[29] W. T. Huh, G. Janakiraman, and M. Nagarajan, "Average cost single-stage inventory models: An analysis using a vanishing discount approach," *Oper. Res.*, vol. 59, no. 1, pp. 143–155, 2011.

[30] F. Chen and J.-S. Song, "Optimal policies for multiechelon inventory problems with Markov-modulated demand," *Oper. Res.*, vol. 49, no. 2, pp. 226–234, 2001.

[31] A. Muharremoglu and J. N. Tsitsiklis, "A single-unit decomposition approach to multiechelon inventory systems," *Oper. Res.*, vol. 56, no. 5, pp. 1089–1103, 2008.

[32] L. Chen, J.-S. Song, and Y. Zhang, "Serial inventory systems with Markov-modulated demand: Derivative bounds, asymptotic analysis, and insights," *Oper. Res.*, vol. 65, no. 5, pp. 1231–1249, 2017.

[33] E. Bayraktar and M. Ludkovski, "Inventory management with partially observed nonstationary demand," *Ann. Oper. Res.*, vol. 176, no. 1, pp. 7–39, 2010.

[34] H. M. Taylor, "Markovian sequential replacement processes," *Ann. Math. Statist.*, pp. 1677–1694, 1965.

[35] S. M. Ross, "Arbitrary state Markovian decision processes," *Ann. Math. Statist.*, vol. 39, no. 6, pp. 2118–2122, 1968.

[36] L. K. Platzman, "Optimal infinite-horizon undiscounted control of finite probabilistic systems," *SIAM J. Control Optim.*, vol. 18, no. 4, pp. 362–380, 1980.

[37] V. S. Borkar, "Average cost dynamic programming equations for controlled Markov chains with partial observations," *SIAM J. Control Optim.*, vol. 39, no. 3, pp. 673–681, 2000.

[38] A. Arapostathis, V. S. Borkar, E. Fernández-Gaucherand, M. K. Ghosh, and S. I. Marcus, "Discrete-time controlled Markov processes with average cost criterion: a survey," *SIAM J. Control Optim.*, vol. 31, no. 2, pp. 282–344, 1993.

[39] E. A. Feinberg, P. O. Kasyanov, and N. V. Zadoianchuk, "Average cost Markov decision processes with weakly continuous transition probabilities," *Math. Oper. Res.*, vol. 37, no. 4, pp. 591–607, 2012.

[40] E. Fernández-Gaucherand, A. Arapostathis, and S. I. Marcus, "On the average cost optimality equation and the structure of optimal policies for partially observable Markov decision process," *Ann. Oper. Res.*, vol. 29, no. 1, pp. 439–470, 1991.

[41] G. E. Monahan, "State of the art—a survey of partially observable Markov decision processes: theory, models, and algorithms," *Management Sci.*, vol. 28, no. 1, pp. 1–16, 1982.

[42] D. P. Bertsekas, *Dynamic Programming and Optimal Control, Vols. I.* Athena Scientific, 2017.

[43] D. Rhenius, "Incomplete information in Markovian decision models," *Ann. Statist.*, pp. 1327–1334, 1974.

[44] M. L. Puterman, *Markov decision processes: Discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

[45] Y.-S. Zheng, "A simple proof for optimality of (s, S) policies in infinite-horizon inventory systems," *J. Appl. Probab.*, vol. 28, no. 4, pp. 802–810, 1991.

[46] D. Jurafsky and J. H. Martin, *Speech and language processing*, vol. 3. Pearson London, 2014.