VALUE OF INFORMATION AND SUPPLY UNCERTAINTY IN SUPPLY CHAINS

A Thesis
Presented to
The Academic Faculty

by

Tae Su Cheong

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Industrial and Systems Engineering

Georgia Institute of Technology
December 2011

VALUE OF INFORMATION AND SUPPLY UNCERTAINTY IN SUPPLY CHAINS

Approved by:

Professor Chelsea C. White III, Advisor
School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Alan L. Erera, Advisor
School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor John J. Bartholdi III
School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Mark E. Ferguson
College of Management
Georgia Institute of Technology

Professor Julie L. Swann
School of Industrial and Systems Engineering
Georgia Institute of Technology

Date Approved: 5 August 2011

To my lovely wife, Hwajung Oh, and

our two daughters, Emily and Claire,

for their love and patience

ACKNOWLEDGEMENTS

I would like to thank my advisors for their encouragement and helpfulness during my time at Georgia Tech. First and foremost, I would like to express my deepest gratitude to Dr. Chelsea C. White III for his sincere guidance, motivation, encouragements, and never ending patience with me throughout the entire course of my research. He has been and will be my best role model as a researcher, mentor and teacher for the rest of my life. I also would like to express my sincere appreciation to Dr. Alan Erera for his excellent insights and invaluable suggestions. Besides the academic guidance, I am especially grateful for his willingness to provide me professional advice for my success in career. Without them, this dissertation would not exist.

I would like to be grateful to Dr. John Bartholdi, Dr. Mark Ferguson and Dr. Julie Swann for serving as members of my dissertation committee, and their insightful comments and suggestions.

I would like to thank my fellow students – especially, all Korean students during my Ph.D. studies – and friends for helping make the time I spent at Georgia Tech some of the best years of my life. They are always there listening to me, sharing emotions, and offering help.

Lastly, I would specially like to thank my wife, Hwajung Oh, parents and parents-in-law for their unconditional love and constant encouragement. I deeply appreciate all the sacrifices they did to support me over many years of study. I owe special thanks to my two lovely princesses, Emily (Chaeyoon) and Claire (Yoonseo), as well.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

This dissertation focuses on topics related to the value of real-time information and/or to supply uncertainties due to uncertain lead-times and yields in supply chains. The first two of these topics address issues associated with freight transportation, while the remaining two topics are concerned with inventory replenishment.

We first assess the value of dynamic tour determination for the traveling salesman problem (TSP). Given a network with traffic dynamics that can be modeled as a Markov chain, we present a policy determination procedure that optimally builds a tour dynamically. We then explore the potential for expected total travel cost reduction due to dynamic tour determination, relative to two *a priori* tour determination procedures.

Second, we consider the situation where the decision to continue or abort transporting perishable freight from an origin to a destination can be made at intermediate locations, based on real-time freight status monitoring. We model the problem as a partially observed Markov decision process (POMDP) and develop an efficient procedure for determining an optimal policy. We determine structural characteristics of an optimal policy and upper and lower bounds on the optimal reward function.

Third, we analyze a periodic review inventory control problem with lost sales and random yields and present conditions that guarantee the existence of an optimal policy having a so-called staircase structure. We make use of this structure to accelerate both value iteration and policy evaluation.

Lastly, we examine a model of inventory replenishment where both lead time and supply qualities are uncertain. We model this problem as an MDP and show that the weighted sum of inventory in transit and inventory at the destination is a sufficient statistic, assuming that random shrinkage can occur from the origin to the supply system or destination, shrinkage is deterministic within the supply system and from the supply system to the destination,

and no shrinkage occurs once goods reach the destination.

CHAPTER I

INTRODUCTION

This dissertation focuses on topics related to the value of real-time information and/or to supply uncertainties due to uncertain lead-times and yields in supply chains. The first two of these topics address issues associated with freight transportation, while the remaining two topics are concerned with inventory replenishment.

Our interest in the value of information in supply chain control has grown as the application of sensor, communications, and control technologies has rapidly spread throughout the freight transportation, logistics, and supply chain industries, providing data in real time that enable the real time control of supply chains. The sources of potentially useful data in a supply chain include: inventory levels; production rates; vehicle, vessel, or trailer position, speed, direction, engine status, oil or air pressure; traffic congestion; weather; freight status and visibility.

With regard to supply uncertainty, there are many causes of lead-time variability: congestion on highways, at ports (both air and sea), and at intermodal sites; extreme weather; labor disputes; accidents; intentional disruptions; natural disasters; resource (e.g., fuel, labor, equipment) shortages; mistakes. This variability tends to increase as the mean of the lead-time increases, particularly when the supply chain requires multiple mode changes and international border crossings. Not surprisingly, globalization has been a significant contributor to supply chain lead-time mean and variability and the need for related supply chain control.

A significant portion of freight (e.g., types of food, particularly fresh produce, and vaccines) moved nationally and internationally is perishable (often due to poor temperature control during transit), and there are a variety of manufacturing processes (e.g., semiconductor manufacturing) that produce a significant and random percentage of unusable products (due, for example, to poor quality process control). Damage in-transit and pilferage can also contribute to a reduction in the amount of freight received, compared to the amount of freight ordered. In such situations, a challenge is to place orders that optimally take into account fact that the amount of useable goods received is only randomly related to the amount of goods ordered.

Assuming that network traffic congestion can be described by a Markov chain, we examine the value of traffic information for pickup and delivery in Chapter 2 by comparing the optimal expected value function for a traveling salesman problem (TSP) tour determined dynamically to a TSP tour determined a priori. In Chapter 3, we investigate the expected improvement of supply chain performance if actions to continue or abort the transport of goods from origin to destination can be taken, based on perishable freight status (e.g., temperature) monitored in transit.

Chapters 4 and 5 are more specifically related to supply uncertainty when the quality of all freight throughout the supply chain is completely observed and known to the decision maker. Chapter 4 examines the structure of an optimal replenishment policy under the assumption that replenishment of a perishable product is instantaneous. Chapter 5 considers a much more complex supply system model that allows lead-time variability and extends known results to the case where the product in the supply chain is perishable.

We now present more in-depth overviews of each of the remaining chapters in the dissertation. In Chapter 2, we study a variant of a classical TSP, called *dynamic traveling salesman problem* (DTSP), to identify an optimal policy in a stochastic TSP network where the network congestion dynamics are governed by a stationary Markov chain to determine the next stop, given the following information: set of nodes visited previously, the current node, and the current status of network congestion. More specifically, Chapter 2

- Investigates the value of choosing the next stop to visit in a multi-stop trip, based on current traffic conditions, in order to minimize expected total travel time of the tour;

- Develops a MDP decision model for the concomitant DTSP, and a solution approach based on a best-first search algorithm called AO*; and

- Provides two benchmark fixed tours in order to assess the value of constructing a tour dynamically, relative to the benchmark tours.

Our numerical results indicate that the proposed solution methodology is capable of solving realistically sized problems (e.g., for less-than-truckload pickup and delivery) fast enough for operational purposes while the standard dynamic programming approach can be computationally impractical or intractable for the same sized problems in terms of the number of states evaluated and CPU times.

Chapter 3 focuses on how real time in-transit monitoring of perishable freight for a single vehicle traveling from an origin to a destination can improve cold supply chain performance. Typically, such freight is either accepted or not accepted at the destination, based on temperature data collected during transit. However, if the condition of the freight is known in-transit, actions can be taken in-transit to improve overall supply chain performance. We investigate the expected improvement in supply chain performance if at intermediate points during the trip actions can be taken to either allow the trip to continue or to abort the trip. Allowing the trip to continue to the next intermediate point also involves deciding whether or not to observe the freight for a fee. Depending on circumstances, the action to abort can have one of many interpretations, including immediately disposing of the freight and returning directly back to the origin, disposing of the freight and expediting a fresh load to the destination, or selling the freight to a secondary market. Specifically, Chapter 3

- Examines the expected improvement of cold supply chain productivity if access is available to inspections of perishable freight for a single vehicle in transit from a source to a destination through intermediate locations;
- Models the problem as a partially-observed Markov decision process (POMDP);
- Develops an efficient procedure to solve the POMDP model;
- Presents structural properties for both the optimal expected cost function and an optimal policy; and
- Examines two special cases of the original model, the case reflecting current practice and the case where observations are free, and present additional structural properties

of the corresponding expected cost functions and optimal policies for each of these two special cases.

We note that the expected cost function is used to determine the expected improvement in supply chain performance if the freight can be monitored in transit while an optimal policy provides what actions should be taken, based on currently available data, in order to optimally increase expected supply chain productivity.

Chapter 4 studies the effects of the supply uncertainty on the optimal order quantity in supply chains. Specifically, we consider a discrete state, discrete decision epoch inventory replenishment control problem under random yield. We assume that there is no backlogging, the single period demand d is deterministic, and once an item is placed in inventory, it will not perish in order to investigate the impact of supply uncertainty in procurement. Chapter 4

- Develops a MDP decision model for an inventory control system with random yields and lost sale over discrete state and action spaces;
- Presents conditions that guarantee that an optimal replenishment policy δ^* is such that $\delta^*(z) = 0$ for $z < 0$, $\delta^*(z) - z > 0$, and $\delta^*(z) - z$ is monotonically non-decreasing for $z > 0$ where $z = d - x$ and x is the current inventory level; and
- Develops an algorithm for determining the optimal policy and the optimal expected total discounted cost over infinite horizon.

Chapter 5 further investigates the effects of the supply uncertainty during a shipment on the optimal order quantity when the lead-time between a source and a destination is uncertain and non-negligible. Specifically, Chapter 5

- Develops a MDP decision model for an multi-staged inventory control model with both lead-time and yield uncertainty over infinite horizon; and
- Presents conditions that guarantee that the weighted sum of inventory in-transit and inventory at the destination is a sufficient statistic for the MDP model, assuming that random shrinkage can occur from the origin to the supply system or destination,

shrinkage is deterministic within the supply system and from the supply system to the destination, and no shrinkage occurs once goods reach the destination.

Finally, we summarize our results and present topics for future research in Chapter 6.

CHAPTER II

DYNAMIC TRAVELING SALESMAN PROBLEM: VALUE OF REAL-TIME TRAFFIC INFORMATION

2.1 Introduction

The traveling salesman problem (TSP) has a long history of capturing the interests of researchers, in large part because of its usefulness in modeling a variety of important real-world problems (e.g., problems in logistics, genetics, manufacturing, telecommunications, and neuroscience) and because of a myriad of intellectual challenges. Applegate et al. [2] present the history, applications, and computational solution techniques for the TSP; we paraphrase their definition of the TSP as follows:

Given a set of cities along with the cost of travel between each pair of them, the TSP is to find the cheapest way of visiting all the cities and returning to the starting point,

where the order of the cities to be visited is called a *tour* (or circuit) and the cost of travel between each pair of cities is stationary and deterministic.

In reality for pick up and delivery (PUD) in an urban (typically congested) environment (the application that has motivated our research), travel times can change unpredictably and hence are reasonably modeled as random variables. Further, information technologies that sense traffic conditions in real time (perhaps an integration of infrastructure-based and vehicle-based intelligent transportation systems) can provide data that permit dynamic tour determination, the real-time construction of the tour as the trip progresses based on real-time traffic congestion data. We call the concomitant problem the dynamic TSP (DTSP).

In order to illustrate the potential value of real-time traffic information, consider a three location network, where location 1 is the origin and destination location. Assume the costs of traversing the arcs are as given in Table 1, where the cost of arc (n, n') represents the cost of traveling from location n to location n'.

Table 1: Cost Parameters

Arc	Cost
$(1,2)$	$\begin{cases} 1, & \text{if the arc is not congested} \\ 3, & \text{if the arc is congested} \end{cases}$
All other arcs	2

Clearly, if arc $(1,2)$ is not congested, the optimal tour is $(1,2,3,1)$, and if arc $(1,2)$ is congested, the optimal tour is $(1,3,2,1)$. If we let α be the probability that arc $(1,2)$ is not congested, then we note that the expected cost of always selecting tour $(1,2,3,1)$ is $5\alpha+7(1-\alpha)$, the expected cost of always selecting tour $(1,3,2,1)$ is 6, and the expected cost of selecting $(1,2,3,1)$ when arc $(1,2)$ is not congested and selecting $(1,3,2,1)$ when arc $(1,2)$ is congested is $5\alpha+6(1-\alpha)$. Thus, the value of selecting the congestion dependent policy, relative to selecting the best of the two congestion independent policies, is $\min\{6, 5\alpha+7(1-\alpha)\} - (5\alpha+6(1-\alpha)) = \min\{\alpha, 1-\alpha\}$. We remark that $\min\{\alpha, 1-\alpha\}/\min\{6, 5\alpha+7(1-\alpha)\}$, the ratio of gain from congestion information to the best cost of the two congestion independent policies, is in the interval $[0, 1/12]$, depending on α. This productivity gain could be significant for a (often low margin) PUD carrier and is quite similar to theoretical productivity gains estimated to result if vehicles were allowed to dynamically route (i.e., re-route in-route between origin and destination, based on real-time traffic information; see Kim et al. [33][34]). Both dynamic routing and touring together could produce significant productivity increases.

The objective of the research presented in this chapter is to better understand the value of real-time traffic data in developing tours for PUD vehicles in real-time (i.e., dynamic touring) and hence better understand the implications of incorporating this capability into the management of a PUD fleet. Our interest in determining the value of dynamic touring is due to

(i) research presented in Kim et al. [33] and Kim et al. [34], which indicated the possibility of significant productivity improvement from dynamic routing,

(ii) the fact that package express PUD drivers in Tokyo construct tours and routes dynamically using their experience and intuition, based on current traffic congestion information reported to them by their dispatchers and other company PUD drivers in the area (private communications).

Although we demonstrate that there may be material advantage in using current traffic data for dynamic touring, we recognize two challenges for eventual successful implementation. First, we note that PUD dispatchers in large urban areas often have wireless communications connectivity with their drivers and web access to real-time traffic conditions. Thus, we anticipate that the challenge of inserting and sustaining dynamic touring into a PUD fleet's operating procedures would be primarily behavioral and organizational, rather than financial.

The second challenge, which we take an initial step to address, is to determine how to turn current traffic data into dynamic touring decisions in light of the fact that the TSP has long represented a significant computational challenge, and stochastic optimization problems tend to be more computationally demanding than deterministic optimization problems. Due to the limitations of dynamic programming (DP) for solving realistically sized DTSPs, we transform the DTSP into an AND/OR graph and apply the best-first heuristic search algorithm AO* for optimal policy determination. This application has proved capable of solving many realistically sized problems (e.g., for LTL PUD) fast enough for operational purposes on a laptop with standard configuration. We note, however, that the DTSP represents a formidable computational challenge, and expanding the space of solvable, realistically sized problems would require further algorithmic development. Improved computational procedures and the development of good sub-optimal designs for the DTSP are topics for future research and development.

2.2 Related Literature

The deterministic TSP and its variants have been extensively studied Applegate et al. [2]. However, there has been limited research on solving the TSP problem with stochastic travel times. Leipala [37] proposed a method to estimate the expected length of an optimal tour when travel times between two nodes are assumed to be identically and independently distributed (i.i.d) random variables. Bellman and Roosta [6] presented a stochastic dynamic programming formulation for the TSP, and Percus and Martin [47] proposed a method called the *cavity method* to determine a tour of minimum expected length under the same assumptions on arcs whose travel times are random variables. Jula et al. [30] examined the stochastic TSP with time windows (STSPTW) having stochastic travel and service times. To seek a least-cost tour while meeting the service requirements for each customer, they proposed a method to estimate means and variances of arrival times at each node, and then presented an approximate algorithm to determine the least-cost STSPTW tour based on an estimate of arrival time at each node. Chang et al. [10] proposed a heuristic approach for the STSPTW having tight time windows. For most previous research, we observe that the travel time random variables on each arc are usually assumed to be i.i.d random variables, and it has been a common approach to approximate travel times in order to handle the complexity caused by the stochastic travel time assumption.

A comprehensive overview of the vehicle routing problem (VRP) and its stochastic variants can be found in Toth and Vigo [60]. Ichoua [28] proposed a VRP with time-dependent travel speed that satisfies the FIFO property. Laporte et al. [35] introduced a stochastic travel-time version of a VRP. Haghani and Jung [22] developed a genetic algorithm for the VRP having a continuous travel time function and proposed a strategy to adjust the vehicle route at certain times in the planning horizon, based on newly available traffic or demand information.

This chapter differs from existing research in that we focus on seeking an optimal *policy* (not an *a priori tour*) that is dependent on currently observed network status in a *stochastically evolving* network. To the best of our knowledge, there have been only a few studies regarding tours that dynamically change depending on network status. Taniguchi and Shimamoto [58] observed that the dynamic adjustment of a tour in transit, based on up-to-date travel time information, can decrease expected travel time and hence can reduce total costs. Recently, Bartin and Ozbay [5] performed a case study whose objective is to determine the subset of fixed routes in a highway network based on real-time traffic information in order to bring maximal benefits of traveler information systems, and modeled the problem with a nonlinear integer programming technique. Stochastic routing problems and their variants have been studied in Kim et al. [33] and Kim et al. [34]; however, this research does not address dynamic tour determination in a stochastic network. Related research is presented in Hall [23].

We use the AND/OR graph heuristic search algorithm AO* to identify an optimal policy for the DTSP. AO* is an extension of the OR graph heuristic search algorithm A*, which is a generalization of Dijkstra's algorithm. Heuristic search techniques for AND/OR graphs are presented in Pearl [46]. Mahanti and Bagchi [43] presented conditions for AO* to terminate with admissible solutions, and Bagchi and Mahanti [3] showed that if a heuristic function used for search guidance optimistically estimates the true cost-to-go function, the algorithm terminates with an optimal solution. Bander and White [4] investigated the application of AO* to a stochastic shortest path problem with random and time-dependent travel times.

2.3 Problem Statement and Preliminary Results

Let $\mathcal{N}$ be the set of locations to visit, and let $\sigma \in \mathcal{N}$ be the origin and destination location. Let (n, n') be the arc from location n to location n', and assume for every n and $n'(\neq n)$ in $\mathcal{N}$, there exists an arc (n, n'). Let $\mathcal{A} = \{(n, n') : n, n' \in \mathcal{N}, n \neq n'\}$, the set of all arcs in the network. Assume $\mathcal{A}_{prob} \subseteq \mathcal{A}$ is the set of all arcs monitored for traffic congestion. Thus, $s \equiv \{s(n, n') : (n, n') \in \mathcal{A}_{prob}\}$ represents the *state of network*, where $s(n, n')$ is the state of arc (n, n') (e.g., not congested, slightly congested, heavily congested, etc).

The (expected) cost of traversing arc (n, n') is $c(n, s, n')$, where $s \in \mathcal{S}$ represents the current state of the network of arcs when the vehicle departs from n to n' on arc (n, n') and where $\mathcal{S}$ is the set of all network states. We assume that the network state is completely observed by the decision maker and evolves according to the probability $q(s'|n, s, n')$, which is the probability that the network state will be s' when location n' is reached by traversing arc (n, n'), given that the network state was s upon departure from location n. Appendix B.1 presents a derivation of $q(s'|n, s, n')$ from probabilities more directly available from traffic data. Throughout, we assume that $q(s'|n, s, n')$ is time-invariant. It is straightforward, but notationally complicated, to extend all the results of this chapter to the time dependent case (an example of which is presented in Kim et al. [33]). Adding an additional variable can be significant computationally. If adding time is significant computationally for the DTSP, then clever use of the time-invariant model may serve for the development of good sub-optimal policies.

A policy π is a set of rules that selects the next location to visit, given (N, n, s), where $N \subseteq \mathcal{N}$ is the set of locations previously visited, $n \in \mathcal{N} \backslash N$ is the current location, and s is the current network state. The triple (N, n, s) serves as the *state of the DTSP*. The objectives of the DTSP are:

- To determine a policy (called an *optimal* policy) that minimizes the expected cost of visiting all of the locations in $\mathcal{N}$, starting and ending at σ, where the total cost of visiting all of the locations is the sum of the costs accrued by traversing all of the individual arcs in the tour.

- To determine the expected total cost accrued by an optimal policy.

We now present optimality equations and boundary conditions for the DTSP. Let $v(N, n, s)$ be the expected minimum cost of visiting all locations in $N' = \mathcal{N} \backslash (N \cup \{n\}) (\neq \phi)$ and then proceeding directly to σ, given the current location of the vehicle is n and the current network state is s. Then, the optimality equation for the concomitant dynamic program is

$$v(N, n, s) = \min_{n' \in N'} \left\{ c(n, s, n') + \sum_{s' \in \mathcal{S}} q(s'|n, s, n') v(N \cup \{n\}, n', s') \right\}$$

with boundary condition

$$v(N, n, s) = c(n, s, n') + \sum_{s' \in \mathcal{S}} q(s'|n, s, n') c(n', s', \sigma)$$

for $\{n'\} = \mathcal{N} \backslash (N \cup \{n\})$. Results in Puterman [50] guarantee that

- $v(\phi, \sigma, s)$ is the minimum expected total cost to be accrued by visiting all of the locations, assuming the network state is s when the vehicle departs from location σ (and that ϕ is the null set), and

- $\pi^*(N,n,s) \in \arg\min_{n' \in N'} \left\{ c(n,s,n') + \sum_{s' \in \mathcal{S}} q(s'|n,s,n') v(N \cup \{n\}, n', s') \right\}$ for all (N,n,s) is an optimal policy.

2.4 Computing π^* and $v(\pi, \sigma, s)$

Although our focus is on understanding the value of real-time traffic information for dynamic tour determination, it is important to at least initially investigate the potentially quite formidable computational challenges presented by the DTSP. DP is an approach for determining $v(\phi, \sigma, s)$ and hence, π^*. However, DP has not been the approach of choice for the deterministic TSP due to the large cardinality of the state space associated with the optimality equation. This state space cardinality enlarges significantly for the DTSP, relative to the deterministic TSP, by a multiple of the cardinality of the set of all network states $\mathcal{S}$. We remark that $|\mathcal{S}|$, the cardinality of $\mathcal{S}$, can be quite large. Assume each arc in $\mathcal{A}_{prob}$ can be in one of γ states. Then, $|\mathcal{S}| = \gamma^{|\mathcal{A}_{prob}|}$, and hence, the size of the state space for the DTSP can be as large as

$$\left(1 + \sum_{j=0}^{|\mathcal{N}|-2} \left[\binom{|\mathcal{N}|-1}{j} \times (|\mathcal{N}| - j - 1) \right] \right) \times |\mathcal{S}|.$$

We remark that, for most cases, it is unnecessary to identify an optimal policy for all states (N,n,s); i.e., there is no reason to know an optimal action for a state (N,n,s) if an optimally-executed tour will never visit it. The DP approach, however, evaluates all possible states (N,n,s) and determines an optimal policy for every state.

We now present an alternative approach for determining $v(\phi, \sigma, s)$ and π^* that has proven capable of solving problems of practical size fast enough for operational purposes. AO* is a best-first search algorithm for finding an optimal solution graph in an AND/OR graph [46]. The DTSP is an example of a finite horizon Markov decision process (MDP) [50]. It is straightforward to show that a finite horizon MDP can be represented as an AND/OR graph having the following special structure: each OR node corresponds to a state in the state space; each immediate successor of an OR node is an AND node; each immediate successor of an AND node is an OR node; each arc from an OR node to an AND node represents an action; each arc from an AND node to an OR node has a probability and a consequence associated with it. Figure 1 illustrates an AND/OR graph for the DTSP.

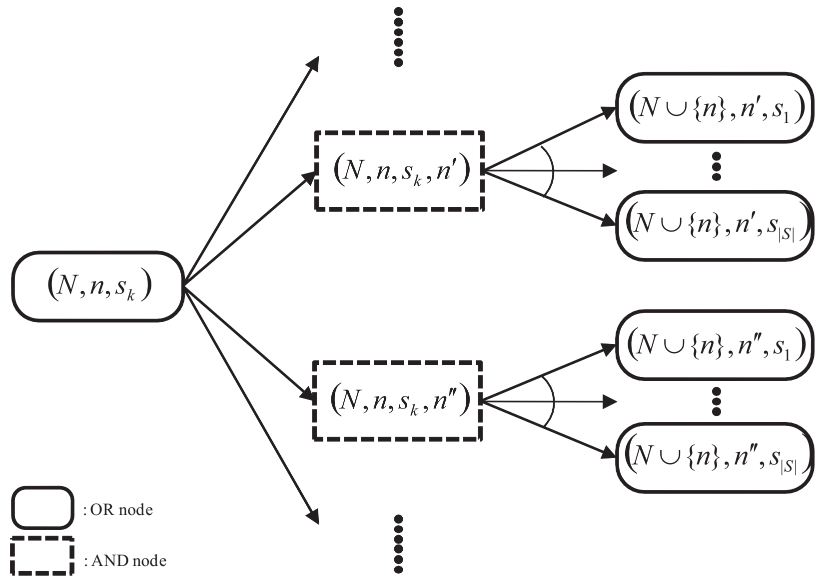

Figure 1: Part of a AND/OR graph for DTSP

AO* iteratively constructs what it needs from the AND/OR representation of the DTSP, starting from a single node (ϕ, σ, s), building partial solution graphs (equivalent to partial policies) until it finds a solution graph (equivalent to a policy) that satisfies the optimality criterion. This entire process is guided by a *heuristic function*. During the search process, the graph that AO* has thus far constructed (a subset of the AND/OR graph representation of the DTSP) is called the *explicit graph* G'. Let Γ be the goal (or terminal) node set (i.e., set of states (N, n, s) where $N \cup \{n\} = \mathcal{N}$). In the context of the AND/OR graph representation of the DTSP, a *solution graph* G: (i) contains the (start) OR node (ϕ, σ, s), (ii) every OR node not in Γ has exactly one AND node as its immediate successor, (iii) every AND node has all of its immediate OR node successors, and (iv) every directed path from the start OR node in the solution graph ends at a node in Γ. A *partial solution graph* G'_{psg} has the same definition as a solution graph except a directed path from the start OR node may end with a non-terminal (i.e., not in Γ) tip node of the explicit graph where a tip node is an OR node having no successors in the current explicit graph. We remark that AO* restricts its interest only to OR nodes that an optimal tour may visit with positive probability (i.e., the *reachable* states). Thus, AO* determines an optimal action for each reachable state

(N, n, s).

AO* expands the current explicit graph using two procedures, *forward expansion* and *back propagation*. The forward expansion procedure involves determining which tip node to expand and expanding it and the partial solution graph of current interest. Tip node expansion involves adding to the current explicit graph all AND nodes that are immediate successors of the selected tip node and all OR nodes that are immediate successors to these AND nodes. The heuristic function provides an estimate of the expected cost function for each newly added OR node; the estimate of the optimal expected cost function of the newly expanded tip node can be subsequently revised. The newly expanded tip node is then no longer a tip node, and the OR nodes that are newly added to the explicit graph become new tip nodes. The back propagation procedure may now choose a new partial solution graph to pursue and update, as needed, the estimates of the optimal expected cost function of the OR nodes throughout the new explicit graph. Based on these two procedures, AO* iteratively expands the best partial solution graph until an optimal solution graph is found.

Typically when AO* finds an optimal solution graph, the concomitant explicit graph is a very small subset of the AND/OR representation of the DTSP, whereas DP in essence must construct the entire AND/OR representation of the DTSP. Details are presented in Section 2.4.2.

Proposition 1 will state that if a heuristic function is a lower bound of the optimal expected cost function of the DTSP, then AO* is guaranteed to find an optimal policy for the DTSP. The determination of an easily computed heuristic function that is a tight lower bound of the optimal expected cost of the DTSP will be of considerable interest below.

2.4.1 Heuristic Function

In order to guide the search in an AND/OR graph so as to identify an optimal solution graph (and equivalently an optimal policy), AO* explicitly utilizes a so-called *heuristic* function. We focus on heuristic functions $h(N, n, s)$ having the following property for all (N, n, s).

$$h(N, n, s) \leq c(n, s, n') + \sum_{s' \in \mathcal{S}} q(s'|n, s, n') h(N \cup \{n\}, n', s') \tag{1}$$

for all $n' \in \mathcal{N}\backslash(N \cup \{n\})$, and, for $N \cup \{n\} \equiv \mathcal{N}$, $h(N, n, s) = c(n, s, \sigma)$. Then, it is straightforward to show that for all (N, n, s), $h(N, n, s) \leq v(N, n, s)$: i.e., $h(N, n, s)$ is a lower bound on $v(N, n, s)$. The next result follows from this observation, the proof of which is essentially identical to the proof of Theorem 4.4 in Chakrabarti and Ghose [9].

Proposition 1 *Suppose that a heuristic function h satisfies $h(N, n, s) \leq v(N, n, s)$ for all (N, n, s). Then, AO* finds an optimal solution graph.*

Proposition 1 states that AO* is guaranteed to terminate with an optimal solution graph if h is a lower bound on the solution of the optimality equation. We now seek such a function. Let $\mathcal{P}_{HP}(N, n, s)$ be a set of all (directed) Hamiltonian paths from n to σ that visit every node in $\mathcal{N}\backslash(N \cup \{n\})$ exactly once. Let $\mathcal{C}_{HP}(p)$ be the travel cost incurred by path $p \in \mathcal{P}_{HP}(N, n, s)$, where the cost of traversing arc (n, n') is $c(n, s, n')$ and the cost of traversing arc (n', n'') is $\min_{s \in \mathcal{S}} c(n', s, n'')$ when $n' \neq n$. Let $h_{HP}(N, n, s) = \min_{p \in \mathcal{P}_{HP}(N,n,s)} \{\mathcal{C}_{HP}(p)\}$ which clearly satisfies Equation (1).

Unfortunately, the Hamiltonian path problem is NP-Complete [16] and hence does not represent a computationally desirable approach for determining a lower bound on $v(N, n, s)$. We now turn our attention to a computationally efficient approach for determining a lower bound on $h_{HP}(N, n, s)$. Given a state (N, n, s), we construct a directed complete graph $T(N, n, s) = (\widetilde{N}, \widetilde{A})$ where $\widetilde{N} = (\mathcal{N}\backslash N) \cup \{\sigma, \kappa\}$ such that κ is a dummy node, and $\widetilde{A}$ is a set of arcs (n', n'') for all $n', n''(\neq n') \in \widetilde{N}$. We set the cost $c'(n', n'')$ for each arc $(n', n'') \in \widetilde{A}$ as follows:

- $c'(\kappa, n) = c'(\sigma, \kappa) = 0$, $c'(n, \kappa) = c'(\kappa, \sigma) = \infty$, and $c'(\kappa, n') = c'(n', \kappa) = \infty$ for all $n' \in \widetilde{N}\backslash\{n, \sigma\}$,
- $c'(n, n') = c(n, s, n')$, and $c'(n', n) = \min_{s \in \mathcal{S}}\{c(n', s, n)\}$ for all $n' \in \mathcal{N}\backslash(N \cup \{n\})$, and
- $c'(n', n'') = \min_{s \in \mathcal{S}}\{c(n', s, n'')\}$ for all $n', n''(\neq n') \in \widetilde{N}\backslash\{\kappa, n\}$.

Then, the travel cost of an optimal TSP tour in the graph $T(N, n, s)$ is equal to $h_{HP}(N, n, s)$. We now discuss a lower bound on an optimal TSP tour in the graph $T(N, n, s)$, and hence a lower bound on $h_{HP}(N, n, s)$. An optimal solution for either the corresponding assignment problem or the minimum spanning tree problem is a lower bound on the travel cost of an optimal TSP tour [11]. Our preliminary computational experiments suggest that a bound from the assignment problem is generally tighter than a bound from the minimum spanning tree problem, and, for an instance with less than 30 stops, the bound from the assignment problem is reasonably tight. Thus, we use a bound from the assignment problem as a heuristic function h. The detailed procedure is presented in Algorithm 1. It has been our experience that a well-chosen (lower bound) heuristic function h can improve performance significantly.

Data: State (N, n, s)
Result: $h(N, n, s)$, a lower bound for $v(N, n, s)$
Construct a $|\widetilde{N}| \times |\widetilde{N}|$ cost matrix $[c_{ij}]$ from $T(N, n, s) = (\widetilde{N}, \widetilde{A})$ where a (i, j)-th entry, c_{ij}, $i, j \in \widetilde{N}$, is determined by

$$c_{ij} = \begin{cases} c'(i,j) & \text{if } i, j (\neq i) \in \widetilde{N} \\ \infty & \text{if } i = j \end{cases};$$

Apply the Hungarian method [1] to the cost matrix $[c_{ij}]$, and let $[\tilde{c}_{ij}]$ be the resulting cost matrix from the method;
Set $h(N, n, s) = \sum_{i \in \widetilde{N}} \sum_{j \in \widetilde{N}} [c'(i, j) \times I(\tilde{c}_{ij} = 0)]$ where $I(\cdot)$ is an indicator function.

Algorithm 1: Heuristic function $h(N, n, s)$

2.4.2 Adaptations of the AO* Algorithm

Based on the proposed lower bound function above, we now adapt AO* to serve as a solution procedure for the MDP model of the DTSP. Throughout its operation, AO* maintains two sets of information for each node in the explicit graph G':

(i) $\omega_O(\cdot)$ and $\omega_A(\cdot)$: For an OR node (N, n, s), $\omega_O(N, n, s)$ is the current best estimate of $v(N, n, s)$. For an AND node (N, n, s, n'), $\omega_A(N, n, s, n')$ is the current best estimate of $v(N, n, s)$, assuming n' is the node to be visited next. The relationship between $\omega_O(N, n, s)$ and $\omega_A(N, n, s, n')$ is given below.

(ii) $l_O(\cdot)$ and $l_A(\cdot)$: For an OR node (N, n, s), $l_O(N, n, s)$ is *true* if the OR node (N, n, s) is either (a) in Γ or (b) in $\omega_O(N, n, s) = v(N, n, s)$, and is *false* otherwise. For an AND node (N, n, s, n'), $l_A(N, n, s, n')$ is *true* if all immediate successors of the

AND node (N, n, s, n') are labeled true, and *false* otherwise. As explained below, the OR node (N, n, s) is labeled *true* if an AND node (N, n, s, n') is labeled *true* and $\omega_O(N, n, s) = \omega_A(N, n, s, n')$.

For a partial solution graph, the estimates of nodes, $\omega_O(\cdot)$ or $\omega_A(\cdot)$, in the graph can be calculated by dynamic programming. When an OR node (N, n, s) is initially introduced in G', the value of $\omega_O(N, n, s)$ is determined by $h(N, n, s)$. In other cases, the current best estimate ω and the corresponding label l for each node in G' are determined as follows:

- OR node (N, n, s):

$$\begin{aligned} \omega_O(N, n, s) &= \min_{n' \in N'} \omega_A(N, n, s, n'), \\ l_O(N, n, s) &= l_A(N, n, s, \tilde{n}) \end{aligned} \tag{2}$$

  where $\tilde{n} = \arg\min_{n' \in N'} \omega_A(N, n, s, n')$.

- AND node (N, n, s, n') for $n' \in \mathcal{N} \backslash (N \cup \{n\})$:

$$\begin{aligned} \omega_A(N, n, s, n') &= c(n, s, n') + \sum_{s' \in \mathcal{S}} q(s'|n, s, n') \omega_O(N \cup \{n\}, n', s'), \\ l_A(N, n, s, n') &= \bigwedge_{s' \in \mathcal{S}} l_O(N \cup \{n\}, n', s') \end{aligned} \tag{3}$$

  where $\bigwedge_{s' \in \mathcal{S}} l(s')$ is *true* if all $l(s')$ for $s' \in \mathcal{S}$ are labeled true, and *false* otherwise.

These calculations are used for both forward expansion and back propagation.

Basically, for each step, AO* expands a (non-terminal) tip node labeled as *false* in a current partial solution graph G'_{psg}, and its successors that are connected through an AND node are added to G'_{psg} (and G'), where each successor AND node corresponds to a node to visit next and each AND node again adds its successor OR nodes to G'_{psg} (and G'). As mentioned above, if such an OR node is introduced in G' for the first time, its value of $\omega_O(\cdot)$ is given by the heuristic function h with $l_O(\cdot)$ labeled as *false*, and this value never decreases throughout the operation of the algorithm. Then, each successor AND node of the tip node determines its value of $\omega_A(\cdot)$ and $l_A(\cdot)$, according to Equation (3). We remark that there are several ways to select such a tip node to expand for each step. In our adaptation, a tip node that is farther from the start node in the graph G'_{psg} has higher priority to be selected first (as is the case for depth-first search) in order to increase the likelihood of reaching a terminal node earlier. Since there are multiple numbers of such the tip nodes due to the explicit graph construction, a tip node with the highest value of the estimate $\omega_O(\cdot)$ will be selected, based on the rationale that $\omega_O(\cdot)$ never decreases. If several tip nodes have the same $\omega_O(\cdot)$, one is arbitrarily chosen. We remark that such an arbitrary choice may lead to constructing different partial solution graphs and eventually may result in different optimal solution graphs if there does not exist a unique optimal solution graph. This forward expansion procedure continues until a newly expanded tip node is labeled to be *true* or its current best estimate ω is updated. Then, the back propagation procedure from the node is performed, and the revised estimate or label for the node is propagated back to all of its ancestors in G'_{psg} until an ancestor is reached whose estimate or label is not altered. We remark that when the forward expansion hits one of the terminal nodes, the value of $\omega_O(\cdot)$ for the node is equal to the true value $v(\cdot)$. Hence, its value of $l_O(\cdot)$ becomes *true*, which immediately triggers the back propagation.

As mentioned before, the back propagation procedure may alter the current partial solution graph. In the altered resulting G'_{psg}, each non-terminal OR node (N, n, s) has exactly one immediate AND node successor $(N, n, s, \tilde{n})$ where $\tilde{n} = \arg\min_{n' \in \mathcal{N} \setminus (N \cup \{n\})}$ $\omega_A(N, n, s, n')$, and, for the AND node $(N, n, s, \tilde{n})$, all of its immediate OR node successors are in the graph. The algorithm terminates when the start node (ϕ, σ, s) is labeled true (i.e., $\omega_O(\phi, \sigma, s) = v(\phi, \sigma, s)$), and the partial solution graph G'_{psg} at the point is indeed an optimal solution graph G. One of the interesting features in this model is that any two OR nodes (N, n, s_1) and (N, n, s_2) where $s_1 \neq s_2$ share the same immediate OR successors through AND nodes (N, n, s_1, n') and (N, n, s_2, n') respectively for all $n' \in \mathcal{N} \setminus (N \cup \{n\})$. Algorithm 2 describes the detailed adaptation of AO* to the DTSP.

2.4.3 Computational Evaluation

In this section, we briefly evaluate the performance of AO* on randomly generated networks, relative to DP. We developed a program that generates a random network over a square grid. Each node is randomly chosen from the intersections of the grid network, and an arc length

Data: State (ϕ, σ, s)
Result: an optimal solution graph G
Create an explicit graph G', and a partial solution graph G'_{psg} where both consist of an OR node (ϕ, σ, s) whose label $l_O(\phi, \sigma, s) = false$;
repeat
  Choose a non-terminal tip node (N, n, s) of the current partial solution graph G'_{psg};
  Expand the OR node (N, n, s) by generating all immediate AND nodes (N, n, s, n') for all $n' \in \mathcal{N} \backslash (N \cup \{n\})$ and $\sum_{s'} q(s'|n, s, n') > 0$, and generate subsequent OR nodes which are the immediate successors $(N \cup \{n\}, n', s')$ of each generated AND node (N, n, s, n') where $q(s'|n, s, n') > 0$. For each newly generated OR node, set $\omega_O(N \cup \{n\}, n', s') = h(N \cup \{n\}, n', s')$, and if $(N \cup \{n\}, n', s') \in \Gamma$, set $l_O(N \cup \{n\}, n', s')$ *true*; otherwise, *false*. Append all the newly generated AND and OR nodes in G';
  Evaluate all the immediate AND nodes of (N, n, s) by updating the corresponding $\omega_A(\cdot)$ and $l_A(\cdot)$;
  Evaluate the OR node (N, n, s) by updating $\omega_O(N, n, s)$ and $l_O(N, n, s)$;
  if *either* $\omega_O(N, n, s)$ *or* $l_O(N, n, s)$ *is changed* **then**
    Create a queue Q and put (N, n, s);
    repeat
      Remove a node from Q;
      if *the node is a OR node* **then**
        Let (N'', n'', s'') be the OR node;
        Revise $\omega_A(\cdot)$ and $l_A(\cdot)$ of all the AND ancestors that have the node (N'', n'', s'') as an immediate successor in G'_{psg}.;
        Put all the AND ancestors that either the corresponding $\omega_A(\cdot)$ or $l_A(\cdot)$ are revised to Q;
      else
        Let (N'', n'', s'', n''') be the AND node;
        Reevaluate $\omega_O(\cdot)$ and $l_O(\cdot)$ of all the OR ancestors that have the node (N'', n'', s'', n''') as an immediate successor in G'_{psg};
        Update the immediate AND successor node of all the OR ancestors that either $\omega_O(\cdot)$ or $l_O(\cdot)$ is revised;
        Put all the OR ancestors that either the corresponding $\omega_O(\cdot)$ or $l_O(\cdot)$ are revised to Q;
      end
    until Q *is empty* ;
    Reconstruct new partial solution graph G'_{psg};
  end
until $l_O(\phi, \sigma, s)$ *is labeled true* ;
Set G'_{psg} as G;

Algorithm 2: Description of AO* algorithm adapted to the DTSP

between two nodes is assigned by calculating the Euclidean distance between the nodes. To generate such networks, we determined several input parameters including the size of the grid I_1, the number of nodes (including an origin) I_2, and the number of probabilistic arcs I_3. For example, if $I_1 = 10$, then I_2 nodes are randomly selected from 100 intersections on a 10×10 grid. A complete graph based on the selected nodes is then constructed, and I_3 arcs that are elements of $\mathcal{A}_{prob}$ are also randomly selected from the constructed complete graph. The value of I_3 is determined such that at least 90% of the arcs have single deterministic travel times. We remark that in most urban areas in the U.S., only a small number of arcs are monitored for congestion (e.g., highways and major arterials), which lends support to this assumption. For arcs in $\mathcal{A}_{prob}$, the probability distributions over travel times have two or three travel times, with equal probability to each travel time for both idle and congested cases. For each parameter combination, we constructed 50 networks, applied AO*, and recorded the number of OR node expansions and CPU time required for each network. Both AO* and DP were implemented in Sun Java JDK 1.6.1 and were run on a LINUX machine with a Intel Xeon 2.66 GHz processor and 4GB of RAM. Table 2 presents performance-related statistics on the number of OR node expansions and average CPU times such as the ratios $\frac{\text{(Number of OR nodes expanded)}}{\text{(Size of state space)}}$ and $\frac{\text{(Average CPU time of AO*)}}{\text{(Average CPU time of DP)}}$ (columns (8) and (9) in Table 2). We note that the average CPU times for DP is determined from only 10 network instances out of 50 test cases for each parameter combination because DP evaluates all the states and thus the execution time is invariant over instances. The results, presented in Table 2, indicate that AO* significantly outperforms DP on the basis of the number of OR nodes expanded and CPU time on average. Specifically, AO* evaluates less than 10% of all possible states on average, and maximally less than 30% of states in the state space while DP is required to evaluate all the possible states. In addition, for most instances, AO* requires significantly less amount of execution time than DP (see column (9) in Table 2). However, we remark that AO* can require more computation time than DP (e.g., $(I_1, I_2, I_3) = (15, 10, 7)$) when AO* is required to expand a large percentage of OR nodes, due to the large amount of CPU time required to expand an OR node by AO*, relative to DP. We note that a single node expansion by AO* is more computationally extensive than a node expansion by DP. Usually, however, AO* will expand far fewer nodes than DP, thus resulting in reduced CPU times for AO*, relative to DP.

Table 2: Comparison of AO* to Dynamic Programming on Randomly Generated Instances (For columns (4) and (5), $(a/b/c/d/e)$ where a is minimum, b is 25-percentile, c is average, d is 75-percentile and e is maximum of 50 random instances; Average CPU times for DP is the average of 10 instances)

(1) Size of Grid	(2) Size of $\|\mathcal{N}\|$	(3) Size of $\|\mathcal{A}_{prob}\|$	(4) Number of OR Nodes Expanded by AO*	(5) CPU Time of AO* (sec)	(6) Size of States in DP	(7) Average CPU Time of DP (sec)	(8) $(4)_{avg}/(6)$ $((4)_{max}/(6))$ $\times 100$ (%)	(9) $(5)_{avg}/(7)$ $\times 100$ (%)
10×10	8	3	(64/151.25/327.41/448/1131)	(0/0/0/0/1)	6160	2.92	5.32% (18.36%)	≈ 0%
10×10	8	4	(128/358/685.4/862.5/2298)	(0/0/0/0/2)	12320	19.55	5.56% (18.65%)	≈ 0%
10×10	8	5	(256/778.25/1464.08/1946.25/5469)	(0/1/4.28/6/26)	24640	71.20	5.94% (22.19%)	6.01%
10×10	8	6	(512/2169/3540.98/4330.75/12987)	(5/20.75/70.68/78.25/312)	49280	203.20	7.19% (26.35%)	34.78%
15×15	10	5	(320/1530/3068.13/3925/10777)	(0/5.75/20.45/25.5/136)	131136	204.78	2.34% (8.22%)	9.98%
15×15	10	6	(1916/3316/6812.32/8703/29330)	(22/45.5/220.28/277/1355)	262272	622.50	2.59% (11.18%)	35.39%
15×15	10	7	(3815/11572.25/23604.84 /32319.75/80817)	(118/739.25/3433.84 /4824/15991)	524544	4589.84	4.51% (15.41%)	74.81%
15×15	10	8	(16017/34667.75/58616.64 /80310.25/117251)	(1068/7995/21369.71 /33791.5/47405)	1049088	> 1 day	5.59% (11.18%)	-
15×15	12	5	(679/2161/5781.57/7453.5/28356)	(2/16.5/83.96/100.5/698)	655424	4634.10	0.88% (4.33%)	1.81%
15×15	12	6	(1890/5321/14337.55/15455/67312)	(8/144.5/858.06/747.5/6953)	1310848	> 1 day	1.09% (5.14%)	-

Lastly, we perform a sensitivity analysis for investigating the impact on CPU times as either $|\mathcal{N}|$ or $|\mathcal{A}_{prob}|$ changes. Figure 2 presents such the changes in average CPU times of 50 instances which are randomly generated in the same manner as above for each parameter combination when (a) I_2 changes for $(I_1, I_3) = (10, 5)$ and (b) I_3 changes for $(I_1, I_2) = (10, 8)$. Let $t^{(a)}_{|\mathcal{N}|}$ and $t^{(b)}_{|\mathcal{A}_{prob}|}$ be such the average CPU times of 50 instances when $|\mathcal{N}|$ (I_2) and $|\mathcal{A}_{prob}|$ (I_3) are given for (a) and (b) respectively. Then, Figure 2 indicates $t^{(a)}_{|\mathcal{N}|+1} \simeq 2 \times t^{(a)}_{|\mathcal{N}|}$ while $t^{(b)}_{|\mathcal{A}_{prob}|+1} \simeq 6 \times t^{(b)}_{|\mathcal{A}_{prob}|}$, implying that CPU times are more sensitive to $|\mathcal{A}_{prob}|$ than to $|\mathcal{N}|$. We remark that each AND node has $2^{|\mathcal{A}_{prob}|}$ number of OR successors in G', and a label for the AND node, $l_A(\cdot)$, is determined according to Equation (3); on the other hand, $|\mathcal{N}|$ corresponds to the depth of a solution graph. We also remark that although the impact of either $|\mathcal{N}|$ or $|\mathcal{A}_{prob}|$ on CPU times is significant, as presented in numerical experiments above, AO* can identify an optimal policy for problem instances of a practical size within a reasonable time for operational purposes while DP for the same size problems is impractical operationally or intractable.

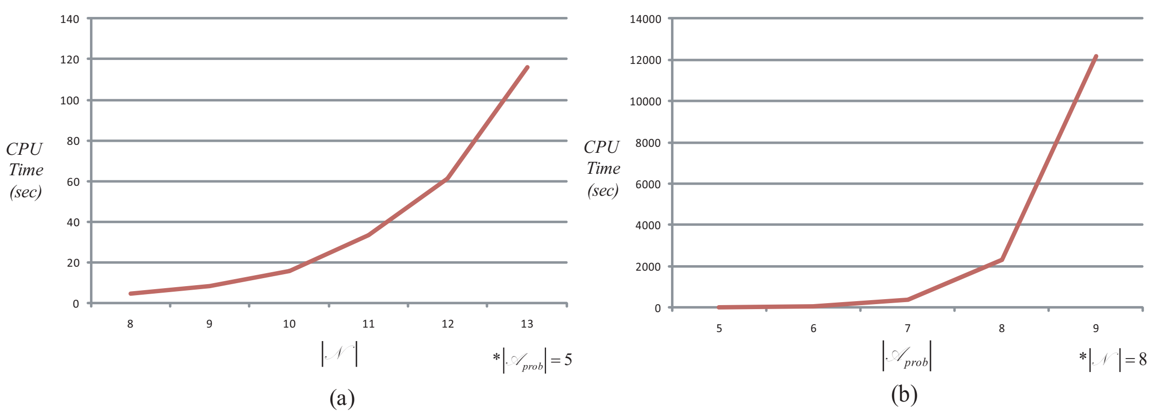

Figure 2: Changes of CPU times in (a) $|\mathcal{N}|$ and (b) $|\mathcal{A}_{prob}|$

2.5 Benchmarks

We now analyze the reduction of expected total cost due to dynamic tour determination, relative to the expected cost accrued by two fixed tours. The value of using real-time network information is the difference between the expected cost generated by the fixed tour and the expected cost generated by an optimal dynamic tour.

2.5.1 Fixed Tour 1

Consider the deterministic cost structure $\{c'(n,n')\}$, where $c'(n,n') \leq c(n,s,n')$ for all s. Determine an optimal fixed (network state invariant) tour, based on this cost structure, and call it $\{\sigma, n^1, \ldots, n^K, \sigma\}$ where $K = |\mathcal{N}| - 1$. If $c'(n,n') = d(n,n')/SL(n,n')$ where $d(n,n')$ is the distance from n to n', and $SL(n,n')$ is the speed limit on arc (n,n'), the resulting TSP represents the tour typically used in industry. This tour generates expected total cost $v'(\phi,\sigma,s)$, where

$$v'(\{\sigma, n^1, \ldots, n^{K-1}\}, n^K, s) = c(n^K, s, \sigma),$$

and for $k < K$,

$$\begin{aligned} v'(\{\sigma, n^1, \ldots, n^{k-1}\}, n^k, s) &= c(n^k, s, n^{k+1}) \\ &\quad + \sum_{s' \in \mathcal{S}} q(s'|n^k, s, n^{k+1}) v'(\{\sigma, n^1, \ldots, n^k\}, n^{k+1}, s'). \end{aligned}$$

Since the set of all policies for the DTSP contains the set of all network state invariant policies, the optimality of π^* guarantees that $v'(\phi,\sigma,s) - v(\phi,\sigma,s) \geq 0$ for all s, and this difference represents the value of real-time network information, using $\{\sigma, n^1, \ldots, n^K, \sigma\}$ as the benchmark.

2.5.2 Fixed Tour 2

An undesirable feature of the tour $\{\sigma, n^1, \ldots, n^K, \sigma\}$ is that it may not be optimal within the set of all network state invariant policies. We remark that the existence of an optimal network state invariant policy is assured by results in Smallwood and Sondik [54]. We now seek the fixed tour that is optimal within the class of all fixed tours. An optimal fixed tour is of interest in its own right since it would be easier to implement than a dynamic tour and would give clear guidance as to how rear-loaded trailers (as is typically the case for the U.S. less-than-truckload industry) should be loaded.

We recall that it was assumed that the decision maker for the DTSP knows exactly the current state of the network when deciding what stop to visit next (complete state observability). A generalization of the DTSP would assume the decision maker observes noise-corrupted observations of the network state (partial state observability). A modeling tool for such a generalization would be the partially observed Markov decision process (POMDP; see Smallwood and Sondik[54] and White [63]). There are two extreme cases of the POMDP, the completely *observed* MDP and the completely *unobserved* MDP. The DTSP is a completely observed MDP. The completely unobserved MDP assumes there are no state observations (or equivalently, there is no information content in the observations) and hence its optimal policy for the DTSP is a fixed tour. The optimality equation for the completely unobserved version of the DTSP is

$$v''(N,n,x) = \min_{n' \in N'} \Big\{ \sum_{s \in \mathcal{S}} x_s c(n,s,n') + v''(N \cup \{n\}, n', xQ(n,n')) \Big\}$$

with boundary condition

$$v''(N,n,x) = \sum_{s \in \mathcal{S}} x_s c(n,s,n') + \sum_{s \in \mathcal{S}} x_s \sum_{s' \in \mathcal{S}} q(s'|n,s,n') c(n',s',\sigma)$$

for $\{n'\} = \mathcal{N} \backslash (N \cup \{n\})$, where $x = \{x_s \geq 0, s \in \mathcal{S}\}$ is a probability mass vector (pmv) over $\mathcal{S}$ (i.e., $\sum_{s \in \mathcal{S}} x_s = 1$), x_s is the probability that s is the current network state, and the (s,s')-th entry of the matrix $Q(n,n')$ is $q(s'|n,s,n')$.

Assume that the state of the network evolves according to a stationary Markov chain P, having $P(s(t+1) = s'|s(t) = s)$ as its (s,s')-the entry. Appendix B.1 presents how $Q(n,n')$ can be constructed from P and probabilities of the form $P(\Delta(t+1)|\Delta(t), s(t))$, where $\Delta(t)$ is the distance remaining to n' at time t. Robust assumptions, i.e., $\mathcal{S}$ is a single ergodic class, that would invariably be satisfied in applications guarantee the existence of a unique pmv x^* that satisfies $x^* = x^* P$. Note also that $x^* Q(n,n') = x^*$ (see Appendix B.1). Thus, assuming x^* is the a priori, the optimality equation for the completely unobserved DTSP becomes

$$v''(N,n) = \min_{n' \in N'} \Big\{ \sum_{s \in \mathcal{S}} x_s^* c(n,s,n') + v''(N \cup \{n\}, n') \Big\}$$

with boundary condition

$$v''(N,n) = \sum_{s \in \mathcal{S}} x_s^* c(n,s,n') + \sum_{s \in \mathcal{S}} x_s^* c(n',s,\sigma)$$

for $\{n'\} = \mathcal{N} \backslash (N \cup \{n\})$ where for notational simplicity, we have dropped explicit dependence of $v''(N,n)$ on x^*. We now observe that the solutions of the completely unobserved DTSP, assuming a priori pmv x^*, is the deterministic TSP having (expected) arc cost structure $\{\sum_s x_s^* c(n, s, n')\}$.

We remark that

$$\sum_{s\in\mathcal{S}} x_s^* v'(\phi, \sigma, s) \geq v''(\phi, \sigma) \geq \sum_{s\in\mathcal{S}} x_s^* v(\phi, \sigma, s). \tag{4}$$

Note that the first inequality holds due to the fact that a tour corresponding to $v''(\phi, \sigma)$ is indeed an optimal fixed tour when the a priori distribution is x^*. The second inequality holds since the expected cost function for the completely unobserved MDP is always an upper bound on the expected cost function for the completely observed MDP [64]. Again, $v''(\phi, \sigma) - \sum_s x_s^* v(\phi, \sigma, s)$ represents the value of real-time information, using an optimal network state invariant policy (or fixed tour) as the benchmark policy.

2.5.3 Value of Dynamic Touring: Illustrative Examples

The objectives of the following two illustrative examples are, respectively, (i) to illustrate the use of the adapted AO* algorithm and (ii) to indicate the value of dynamic touring relative to the two aforementioned fixed tours.

1) Example 1

We now examine a simple four-node network where $\mathcal{N} = \{\sigma, 1, 2, 3\}$. Suppose that only one arc $(1, 2)$ is in $\mathcal{A}_{prob}$, and that this arc can be in one of two network states – *not congested* (0) or *congested* (1). Thus, $\mathcal{S} = \{(0), (1)\}$. The cost matrix of traversing arcs (n, n'), $[c(n, n')]$, and the stationary Markov chain P are:

$[c(n, n')] =$

$n' =$	σ	1	2	3
$n = \sigma$	∞	4	3	5
1	3	∞	$*$	4
2	3	6	∞	3
3	2	4	3	∞

, $P =$

$s' =$	(0)	(1)
$s =$(0)	0.89	0.11
(1)	0.15	0.85

for all t.

We also assume that the traversal time on arc $(1, 2)$ is

- either 3,4 or 5 with probabilities 0.9, 0.05, 0.05 respectively if the arc is *not congested* at the time the vehicle departs from node 1 to node 2, and
- either 4,5 or 6 with probabilities 0.05, 0.45, 0.5 respectively if the arc is *congested* at the time the vehicle departs from node 1 to node 2.

Therefore, if $(n, n') \neq (1,2)$, $c(n,s,n') = c(n,n')$ for all $s \in \mathcal{S}$, and if $(n,n') = (1,2)$, $c(1,(0),2) = 3.15$ and $c(1,(1),2) = 5.45$. Lastly, the a priori pmv x^* is $x^* = [0.5769,\ 0.4231]$.

We now determine the fixed tours 1 and 2 and an optimal policy for this DTSP.

Table 3: Cost matrix for determining fixed tours 1 (a) and 2 (b)

(a) $[c'(n,n')] =$

$n' =$	σ	1	2	3
$n = \sigma$	∞	4	3	5
1	3	∞	**3.15**	4
2	3	6	∞	3
3	2	4	3	∞

(b) $[\sum_s x^*_s c(n,s,n')] =$

$n' =$	σ	1	2	3
$n = \sigma$	∞	4	3	5
1	3	∞	**4.115**	4
2	3	6	∞	3
3	2	4	3	∞

.

Sub-optimal Policy 1 (Fixed Tour 1): To determine the fixed tour 1, we first construct the deterministic cost matrix $\{c'(n,n')\}$, which is presented in Table 3(a). Note that $c'(1,2) = \min\{c(1,(0),2), c(1,(1),2)\} = \min\{3.15, 5.45\}$. DP can be used to identify the minimum travel cost tour, which is $\{\sigma, 1, 2, 3, \sigma\}$.

Sub-optimal Policy 2 (Fixed Tour 2): Similarly, to determine the fixed tour 2, we first construct the deterministic cost matrix $\{\sum_s x^*_s c(n,s,n')\}$, which is presented in Table 3(b). Again, DP can be used to identify the minimum travel cost tour, which is $\{\sigma, 2, 3, 1, \sigma\}$.

Optimal Policy: Algorithm 2 terminates with an *optimal* solution graph G, presented in Figure 3. The optimal policy for this DTSP from the graph G' is as follows: When the vehicle starts a tour from the origin σ,

(i) if the arc $(1,2)$ is not congested, then travel toward node 1. When the vehicle departs from node 1, if the arc $(1,2)$ is not congested, then follow the path $\{1,2,3,\sigma\}$;

otherwise, follow the path $\{1, 3, 2, \sigma\}$.

(ii) if the arc $(1, 2)$ is congested, then follow the tour $\{\sigma, 2, 3, 1, \sigma\}$.

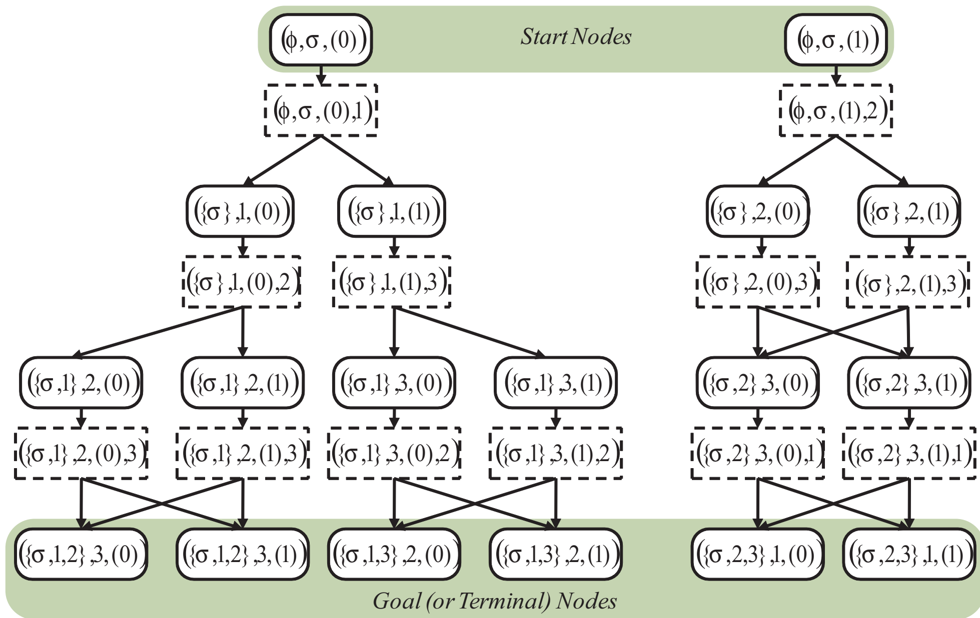

Figure 3: Optimal solution graph G for Example 1

Results and Discussion: Table 4 presents the expected total travel cost for each policy. We remark that these costs satisfy Equation (4). In order to quantify the value of dynamic touring, the ratio $\Delta_1 = \dfrac{\sum_s x_s^* v'(\phi, \sigma, s) - \sum_s x_s^* v(\phi, \sigma, s)}{\sum_s x_s^* v'(\phi, \sigma, s)} \times 100$ (%) and $\Delta_2 = \dfrac{v''(\phi, \sigma) - \sum_s x_s^* v(\phi, \sigma, s)}{v''(\phi, \sigma)} \times 100$ (%) are introduced as measures of the value of information. Table 5 indicates that the percent reduction in expected total travel time due to dynamic touring, relative to static tours 1 and 2, are 2.27% and 1.34%. respectively.

Table 4: Expected Total Travel Costs for Each Policy

Optimal Policy	$v(\phi, \sigma, (0))$	12.697989684
	$v(\phi, \sigma, (1))$	13
	$\sum_{s \in \mathcal{S}} x_s^* v(\phi, \sigma, s)$	12.82577025
Fixed Tour 1	$v'(\phi, \sigma, (0))$	12.831284472
	$v'(\phi, \sigma, (1))$	13.52097572
	$\sum_{s \in \mathcal{S}} x_s^* v'(\phi, \sigma, s)$	13.12309284
Fixed Tour 2	$v''(\phi, \sigma)$	13

Table 5: Value of Information: Percentage Decrease of Total Travel Cost

$\Delta_1 = \dfrac{\sum_s x_s^* v'(\phi, \sigma, s) - \sum_s x_s^* v(\phi, \sigma, s)}{\sum_s x_s^* v'(\phi, \sigma, s)} \times 100(\%)$	2.27
$\Delta_2 = \dfrac{v''(\phi, \sigma) - \sum_s x_s^* v(\phi, \sigma, s)}{v''(\phi, \sigma)} \times 100(\%)$	1.34

2) Example 2

In this example, we examine a twelve-node network, where $\mathcal{N} = \{\sigma, 1, 2, \ldots, 11\}$, $\mathcal{A}_{prob} = \{(\sigma, 3), (3, 6), (5, 9)\}$, and each arc can be in one of two network states, *not congested* (0) or *congested* (1). Thus, $\mathcal{S} = \{(000), (001), (010), (011), (100), (101), (110), (111)\}$ where, for state (abc), a is the network state of $(\sigma, 3)$, b is the state of $(3, 6)$, and c is the state of $(5, 9)$. The cost matrix of traversing arcs (n, n'), $[c(n, n')]$, and the stationary Markov chain P are given in Tables 6 and 7 respectively.

Table 6: Cost matrix used in Example 2

$[c(n, n')] =$

$n' =$	σ	1	2	3	4	5	6	7	8	9	10	11
$n = \sigma$	∞	9	19	$\ast$	10	8	8	9	12	4	5	20
1	8	∞	14	18	4	9	14	9	15	3	8	1
2	15	15	∞	16	11	16	3	2	8	15	6	10
3	6	20	20	∞	13	18	$\ast$	12	18	6	11	5
4	4	9	14	6	∞	5	13	12	18	4	9	7
5	10	3	13	4	18	∞	19	9	15	$\ast$	6	4
6	17	10	6	6	13	10	∞	4	7	17	12	20
7	11	16	9	13	12	17	4	∞	6	11	16	11
8	16	17	14	14	7	12	20	16	∞	16	11	15
9	4	5	19	2	4	9	17	6	12	∞	5	13
10	15	4	18	4	3	8	16	5	4	20	∞	19
11	18	14	13	17	6	3	6	15	20	15	18	∞

Travel times with corresponding probabilities for the arcs in $\mathcal{A}_{prob}$ are presented in Table 8. The a priori pmv is $x^* = [0.1838, 0.0459, 0.1963, 0.0557,\ 0.1588, 0.0889, 0.1474, 0.1232]$. We now determine the expected total costs for fixed tours 1 and 2 and for an optimal policy.

Sub-optimal Policy 1 (Fixed Tour 1): To determine fixed tour 1 and the expected total cost, we first construct the deterministic cost matrix $\{c'(n, n')\}$ which, for $(n, n') \in \mathcal{A}_{prob}$,

Table 7: State transition probability matrix P ((abc) where a is the network state of $(\sigma, 3)$, b is the network state of $(3, 6)$ and c is the network state of $(5, 9)$)

$P =$

$s' =$	(000)	(001)	(010)	(011)	(100)	(101)	(110)	(111)
$s =$(000)	0.85	0.05	0.05	0.0	0.05	0.0	0.0	0.0
(001)	0.0	0.8	0.1	0.0	0.1	0.0	0.0	0.0
(010)	0.1	0.0	0.8	0.05	0.0	0.0	0.05	0.0
(011)	0.0	0.0	0.05	0.7	0.0	0.0	0.1	0.15
(100)	0.05	0.0	0.05	0.0	0.8	0.05	0.05	0.0
(101)	0.0	0.0	0.0	0.05	0.05	0.8	0.0	0.1
(110)	0.0	0.0	0.1	0.0	0.05	0.0	0.8	0.05
(111)	0.0	0.0	0.0	0.02	0.05	0.08	0.05	0.8

Table 8: Travel times and corresponding probabilities when a network state is given on each probabilistic arc: in the third column, $a(b)$ where a is a travel time with probability b on the corresponding arc (n, n') if the arc is in a state $s_{(n,n')}$ of either *Not Congested* (0) or *Congested* (1) at the time the vehicle departs node n to node n'

$(n, n') \in \mathcal{A}_{prob}$	Network State $s_{(n,n')}$	Time (Probability)			$c(n, s, n')$
$(\sigma, 3)$	Not Congested	2 (0.9)	3 (0.1)		2.1
	Congested	9 (0.4)	10 (0.4)	11 (0.2)	9.8
$(3, 6)$	Not Congested	5 (0.8)	6 (0.2)		5.2
	Congested	13 (0.3)	14 (0.6)	15 (0.1)	13.8
$(5, 9)$	Not Congested	5 (0.8)	6 (0.1)	8 (0.1)	5.4
	Congested	15 (0.1)	16 (0.9)		15.9

$c'(n, n')$ is determined by $c'(n, n') = \min_{s \in \mathcal{S}}\{c(n, s, n')\}$, which is presented in Table 9. DP was used to identify this tour, which is $\{\sigma, 10, 1, 11, 5, 9, 3, 6, 2, 7, 8, 4, \sigma\}$.

Sub-optimal Policy 2 (Fixed Tour 2): Similarly, to determine fixed tour 2, we first construct the deterministic cost matrix $\{\sum_s x_s^* c(n, s, n')\}$, which is presented in Table 10. DP was used to identify a tour of minimum travel cost, which is $\{\sigma, 9, 10, 1, 11, 5, 3, 6, 2, 7, 8, 4, \sigma\}$.

Optimal Policy: The algorithm 2 terminates with an *optimal* solution graph, which specifies an optimal policy for the DTSP. Details of the policy are presented in Appendix B.2.

Results and Discussion: Figure 4 presents the expected total travel cost for each initial

Table 9: Cost matrix for fixed tour 1 in Example 2

$[c'(n,n')] =$

$n' =$	σ	1	2	3	4	5	6	7	8	9	10	11
$n = \sigma$	∞	9	19	**2.1**	10	8	8	9	12	4	5	20
1	8	∞	14	18	4	9	14	9	15	3	8	1
2	15	15	∞	16	11	16	3	2	8	15	6	10
3	6	20	20	∞	13	18	**5.2**	12	18	6	11	5
4	4	9	14	6	∞	5	13	12	18	4	9	7
5	10	3	13	4	18	∞	19	9	15	**5.4**	6	4
6	17	10	6	6	13	10	∞	4	7	17	12	20
7	11	16	9	13	12	17	4	∞	6	11	16	11
8	16	17	14	14	7	12	20	16	∞	16	11	15
9	4	5	19	2	4	9	17	6	12	∞	5	13
10	15	4	18	4	3	8	16	5	4	20	∞	19
11	18	14	13	17	6	3	6	15	20	15	18	∞

Table 10: Cost matrix for fixed tour 2 in Example 2

$[\sum_s x_s^* c(n,n')] =$

$n' =$	σ	1	2	3	4	5	6	7	8	9	10	11
$n = \sigma$	∞	9	19	**6.1**	10	8	8	9	12	4	5	20
1	8	∞	14	18	4	9	14	9	15	3	8	1
2	15	15	∞	16	11	16	3	2	8	15	6	10
3	6	20	20	∞	13	18	**9.7**	12	18	6	11	5
4	4	9	14	6	∞	5	13	12	18	4	9	7
5	10	3	13	4	18	∞	19	9	15	**8.7**	6	4
6	17	10	6	6	13	10	∞	4	7	17	12	20
7	11	16	9	13	12	17	4	∞	6	11	16	11
8	16	17	14	14	7	12	20	16	∞	16	11	15
9	4	5	19	2	4	9	17	6	12	∞	5	13
10	15	4	18	4	3	8	16	5	4	20	∞	19
11	18	14	13	17	6	3	6	15	20	15	18	∞

network state. The results suggest that the larger the cardinality of $\mathcal{A}_{prob}$, the greater the impact of optimal dynamic touring. We remark that the inequalities in (4) hold as presented in Figure 4 and Table 11. We again use the ratios Δ_1 and Δ_2 to measure the value of information. Table 11 indicates that the percentage reduction in expected total travel time due to dynamic touring, relative to static tours 1 and 2, is 11% and 7%, respectively.

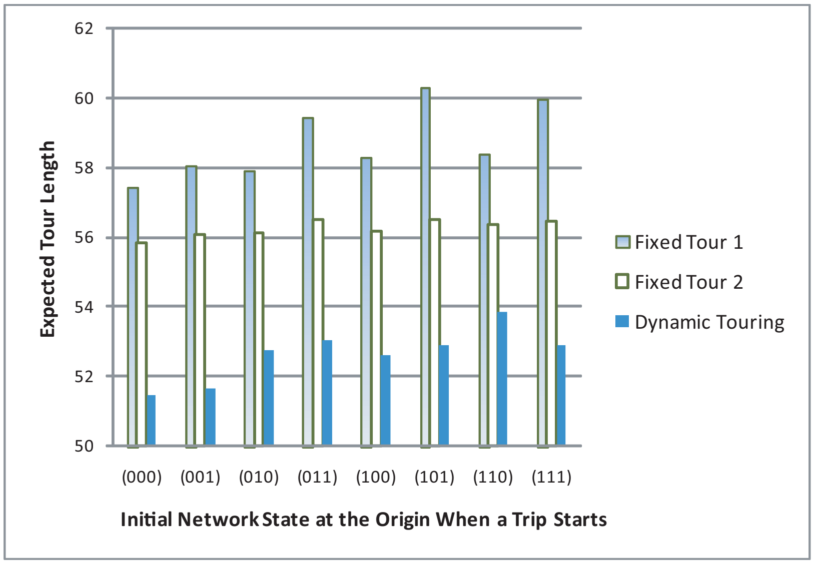

Figure 4: Expected total travel costs for each policy (Example 2)

Table 11: Expected Total Travel Costs for Each Policy and Value of Information (Percentage Decrease of Total Travel Cost)

Optimal Policy	$\sum_{s \in \mathcal{S}} x_s^* v(\phi, \sigma, s)$	52.654
Fixed Tour 1	$\sum_{s \in \mathcal{S}} x_s^* v'(\phi, \sigma, s)$	58.924
Fixed Tour 2	$v''(\phi, \sigma)$	56.324
$\Delta_1 = \dfrac{\sum_s x_s^* v'(\phi, \sigma, s) - \sum_s x_s^* v(\phi, \sigma, s)}{\sum_s x_s^* v'(\phi, \sigma, s)} \times 100(\%)$		10.641(%)
$\Delta_2 = \dfrac{v''(\phi, \sigma) - \sum_s x_s^* v(\phi, \sigma, s)}{v''(\phi, \sigma)} \times 100(\%)$		6.516(%)

2.6 Summary

Chapter 2 has examined the so-called *dynamic traveling salesman problem* in which the vehicle chooses the next stop to visit, based on current traffic conditions at each stop of a multi-stop trip. In the problem, travel time along each arc in the network has been modeled as a random variable, and we have assumed that network congestion dynamics can be described by a stationary Markov chain. In this chapter, we presented an efficient algorithm based on AO* for dynamically determining a tour that minimizes the expected total travel time and showed that this algorithm outperformed the standard dynamic programming approach, in terms of CPU times and states evaluated, in determining an optimal policy. We have also investigated the potential value of using traffic congestion information for dynamic tour determination, relative to two benchmark tours developed prior to departure, through numerical examples.

CHAPTER III

IN-TRANSIT PERISHABLE PRODUCT INSPECTION

3.1 Introduction

A significant portion of freight moved nationally and internationally is perishable. For example, the United States imports and exports annually about US$ 40 billion of perishable food, such as produce and fresh meat and fish [12, 14]. In 2007, the pharmaceutical market was estimated to be a US$ 712 billion market worldwide [17] and many important pharmaceuticals are perishable, e.g., vaccines. Such freight is typically transported and stored in temperature controlled trailers, containers, and warehouses, which are part of the cold supply chain. However, cold supply chains are only partially successful in insuring perishable freight arrives at its destination fresh. According to a study by the United Nations Environment Program (UNEP), *over half of the food produced globally is lost, wasted or discarded as a result of inefficiency in the human-managed food chain* [13]. The percentage of perishable food that perishes during storage and transit in the U.S. is approximately $10\% - 15\%$ of perishable freight tonnage and $25\% - 50\%$ of total economic value due to the degraded quality of goods [42]. A recent study by the University of Florida Food Distribution and Retailing Resource Center[1] showed that one-third of shipped food production is wasted annually (equivalent to a US$ 35 billion loss each year), and half of the loss is mainly due to temperature control problems in a shipment between the grower and the retailer [29]. Further, approximately 25% of all vaccine products worldwide degrade before reaching their destination [65, 66]. These percentages can be substantially higher in countries with a less well-developed cold supply chain infrastructure (e.g., China [7]). According to the United Kingdom's Medicines and Healthcare Products Regulatory Agency (MHRA), as much as 43% of major deficiencies in pharmaceutical cold chain shipping and distribution are related to poor control and monitoring of storage and transportation temperature [41].

[1] http://cfdr.ifas.ufl.edu/

There are a variety of ways of improving the productivity of cold supply chains; we mention two. One way is to improve the likelihood that a perishable is kept at a temperature within its temperature range during transport and storage. Another way is to base the real-time control of the cold supply chain on real-time data resulting from monitoring the state of the product during transport and storage.

With respect to this second approach, which is the focus of this chapter, there may be advantage in knowing when freight in-transit materially degrades. For example, consider perishable freight being transported from South America to Miami and then to its final destination, Chicago. Assume that the freight perishes in-route between South America and Miami. If the state of the freight is monitored in Miami and it is determined that the freight has perished, then transportation and perhaps disposal costs may be reduced significantly, relative to determining upon delivery in Chicago that the freight has perished. Clearly, though, deciding whether or not to monitor freight quality in Miami is dependent on a variety of information, including the cost of monitoring, the likelihood of the freight perishing en-route to Miami, and the cost of the transportation.

Our communications with individuals in the food supply chain and logistics industry include discussions with the vice president for logistics and the vice president for food safety and quality for a large food shipper, the manager of systems development and design for a well-known third party logistics provider of temperature sensitive freight transportation, and a senior vice president for a large food re-distributor. These communications indicate that it is common that food temperature is recorded but not transmitted in transit and that these data are used at the destination to determine whether or not the freight is accepted or rejected (see Harrington [25] for further detail). These communications also indicate that the primary barrier to requiring freight status data be transmitted in transit is the perception that the value of such data currently does not justify the investment in the enabling information technology infrastructure and the cost of the concomitant service, although it is commonly felt that this perception may change as these investment and service costs are reduced and the value of these data is better understood. In contrast, we have learned from contact with *Euroscan*[2], a supplier of systems solutions for mobile temperature monitoring, which provides solutions for both the case where no temperature (and other) data are transmitted in transit and the case where data are transmitted in transit [24], that the latter service is provided to a variety of customers, including a large fast food company. All of our industry contacts believe that regulation, e.g., the Hazard Analysis and Critical Control Points (HACCP), may eventually require in-transit temperature records for a significant amount of food transportation, and several of our industry contacts envision when some combination of reduced infrastructure and service costs and a more demanding regulatory environment will motivate growing use of temperature transmission in transit.

We, and the individuals in the private sector of the food industry with whom we have communicated, are unaware of any formal approach for determining the value of transmitting temperature data during transit relative to simply waiting until the food arrives at the destination before examining the in-transit or arrival temperature data. In this chapter, we take an initial step in developing such a formal approach with the intent of providing an approach to better inform (1) the private sector regarding the expected value of such data and (2) the regulator regarding the economic impact of potential regulatory environments on the private sector.

Food transportation typically involves multiple: distribution centers, products, carriers, vehicles, and customers. Perishable food is often transported in temperature controlled and monitored vehicles. Temperature measurements are often noise corrupted. Pickup and delivery vehicles often visit several customers per delivery cycle. We take an initial step in this chapter to develop an understanding of the role of real-time information in the management of this complex distribution system by considering transportation of perishable food from a single distribution center (the origin) to a single customer (the destination) with multiple intermediate locations on a prescribed route from an origin to a destination, assuming inspections, when they occur, perfectly describe the state of the freight. At each intermediate location, the decision maker (DM) can decide to go forward to the next location or to alter the plan to deliver the freight to the destination (the 'alternative'

[2] http://www.euroscan-group.com/

decision). If the decision is to go forward, we also decide whether or not the freight will be inspected (for a fee) upon arrival at the next location. We assume that an inspection is a perfect determinant of freight quality. Once the destination is reached, then a reward is received based on the quality of the freight. The cost and reward structure is composed of transportation, inspection, wholesale costs, and a destination delivery reward and may include, depending on the alternative decision, disposal cost, an expedite fee, and/or a reward structure for a secondary market sell.

The alternative decision is essentially a decision to abort (in some manner) transporting the current (presumably poor quality) load to the destination and restart, with a new load, the process of transporting goods from origin to destination. The alternative decision can model any one of several scenarios. These scenarios include:

- Dispose of the freight, return the vehicle directly to the origin, and begin again with a fresh load of freight. This scenario assumes that there is a single vehicle (e.g., an aircraft) tasked with providing freight to the customer and that freight may be delivered aperiodically.

- Dispose of the freight and immediately expedite a fresh load of freight to the destination. This scenario allows for periodic freight delivery and permits the cost of expedited freight to be dependent on when the decision to expedite is made during the delivery cycle.

- Sell the freight to a secondary market at the current location. This scenario allows for the case where, for example, fresh strawberries being shipped from the West Coast to the East Coast in the U.S. have value in St. Louis but are likely to be spoiled by the time they would reach their destination.

The problem objective is to find a rule (or policy) for selecting a decision at each intermediate location that maximizes the expected total discounted reward over the infinite planning horizon, where decisions are based on the results of all the inspections made and at what locations they were made since the freight last left the origin.

We remark that when freight quality is measured but not transmitted in transit is the special case where no monitoring is permitted or equivalently where the cost of data transmission is prohibitively high. This special case will serve as the basis for comparison with the more general problem in the analysis later in this chapter.

We also remark that data transmitted in real-time can have benefit ancillary to productivity improvement. For example, temperature monitoring equipment can be bundled with information technology that monitors, for example, whether or not the trailer door is open or closed, perhaps for very little additional cost. Then, real-time data can be used to insure food quality and perhaps, additionally help to reduce pilferage and/or avoid tampering. More generally, we believe that information technology for food protection and defense is much more likely to be embraced by the private sector and inserted in cold supply chains if it is bundled with information technology for productivity improvement.

For modeling simplicity, we will assume that product quality is perfectly observed. A topic for future research would assume that product quality is probabilistically related to measurement (e.g., the output of a temperature gauge and perhaps additionally the output of other sensors) and hence is partially observed. We remark that studies have shown significant temperature variation inside a trailer [51], which further supports the claim that food quality may not be perfectly observed in transit. We further remark that the level of understanding of the relationship between product quality and temperature history is highly variable across foods and this relationship is of considerable current interest to the food industry and the food regulatory community.

We model the problem considered in this chapter as an infinite horizon, expected total discounted reward, partially observed Markov decision process (POMDP). The sufficient statistic of the POMDP is the pair $(n, \vec{x})$, where n is the current location of the loaded vehicle traveling toward the destination and $\vec{x}$ is the probability mass vector (pmv) of the state of the freight, conditioned on all the inspections made and at what locations they were made since the freight last left the origin. Taking advantage of the special structure of the problem, we develop an efficient procedure for determining the optimal reward function that avoids the tractability issues typically associated with solution procedures for the POMDP.

We show that the optimal reward function and an optimal policy have several intuitively appealing and implementable characteristics but provide a counter example to an intuitive claim regarding the structure of an optimal policy. We determine the value of having access to inspections at the intermediate locations and an upper bound on this value. Such results could be useful to a company deciding whether or not to purchase access to such information, how to optimally extract value from inspection data, and how much value the company can expect to extract.

3.2 Literature Review

Several research studies have focused on inventory control for perishable products. Comprehensive literature surveys on this topic have been provided by Goyal and Giri [20] and Nahmias [44]. Both studies classified perishable inventory problems in terms of shelf-life (fixed or random) and reviewed various types of inventory control policy models.

Research has also focused on perishable product logistics. Lin and Chen [39] examined a dynamic logistics control model for perishable products in order to maximize total net profit of all supply chain parties. Specifically, their focus was on the dynamic allocation of perishable commodities to retailers when supply availability may be uncertain.

Regarding the transportation and delivery of perishable products, much previous research has dealt with the transportation problem as a variant of the vehicle routing problem (VRP). Such studies are concerned with the distribution of perishable products in order to prevent the reduction of quality or value of fresh products and focus on reducing travel time or distance. Tarantilis and Kiranoudis [59] examined the problem of fresh meat distribution in Athens, Greece, and formulated the problem as an open multi-depot VRP. The authors proposed a stochastic threshold-accepting meta-heuristic to solve the problem. Osvald and Stirn [45] considered the problem of distributing fresh vegetables using a VRP formulation with time windows and time-dependent travel time constraints and used a tabu-search based algorithm for solution determination. They showed that the proposed algorithm reduced the amount of perished products in a Slovenian food market by 47%. Hemmelmayr et al. [26] considered a blood delivery problem of the Austrian Red Cross and investigated the potential benefit by changing their delivery process into a vendor-managed inventory system. They provided an integer programming formulation that is a variant of the periodic VRP and used variable neighborhood search (VNS) algorithms as a solution approach. In their formulation, they did not consider blood product spoilage during delivery.

The use of RFID technology for real-time information collection in supply chains has attracted considerable interest. Lee and Ozer [36] have provided an overview of the impact of RFID on the improvement of inventory visibility. Kang and Gershwin [31] have examined the impact of RFID technology on theft prevention in an inventory system. Ustundag and Tanyas [61] evaluated supply chain performance enhancement from RFID technology in a three-echelon supply chain through a simulation model by quantifying the expected increase of benefits from its deployment. Karkkainen [32] presented a case study, conducted at Sainsbury's in the UK, that investigated the benefits of RFID deployment for short shelf-life products. The objectives were to reduce labor costs associated with stock counting and rotation monitoring and to reduce spoilage in the supply chain. Sahin et al. [52] evaluated an existing monitoring tool, called the 'Time Temperature Integrator (TTI)', to check perishable product freshness. The authors identified several potential benefits from TTI deployment but did not develop an operational strategy based on product freshness.

Regarding the use of real-time information for cold chain management, Ferguson and Ketzenberg [15] considered the value of information sharing between suppliers and retailers for perishable products. The authors examined the impact on retailer inventory replenishment of fixed lifetime perishable products that are discarded if they are unsold before their life time. LiKehoe and Drake [38] gathered product quality information using RFID tags that also measured other environmental indicators and used the information to make decisions for dynamic product (re)allocations in order to maximize product value to customers. This research and research found in Lin and Chen [39] also focused on perishable product allocations. To the best of our knowledge, no previous research has addressed the potential benefit derived from product inspection in cold chain freight transportation. Recently, Cai et al. [8] studied the impact of preserving product freshness on the order quantity and selling price, based on a newsvendor model. The authors also investigated an incentive structure for supply chain coordination between a producer and a retailer. That is different from the cost and reward structure of this chapter.

We model our problem as a partially observed Markov Decision Process (POMDP) (see Puterman [50] and White [63] for details). Related analysis regarding a production process control problem modeled as a POMDP can be found in White [62], where the structure of an optimal policy was characterized.

3.3 Problem Statement

A vehicle is to transport a perishable product from the origin (0) to the destination (N) through intermediate locations, $n = 1, \ldots, N-1$, along the prescribed path $0, 1, \ldots, N$. Let $s(n)$ be the state or quality of the product at location n, where $s(n) \in \{0, 1, \ldots, S\}$. We think of state s as being at least as preferred as state s' if and only if $s \leq s'$. Thus, state 0 can be thought of as "fresh", state S "spoiled", and the remaining states as varying degrees or percentages of fresh and spoiled.

We assume that the quality of the product changes (deteriorates) randomly during transport according to the stochastic matrix $P = \{p_{ij}\}$ where $p_{ij} = P(s(n+1) = j \mid s(n) = i)$ for all n. Let x_s be the probability that the product is in state s at the current location, conditioned on the outcome of all inspections made and at what locations these inspections were made since the freight last left the origin. Since an inspection is assumed to perfectly reveal the state of the freight, all that is needed to determine the probability mass vector (pmv) of the product state $\vec{x} = (x_0, x_1, \ldots, x_S)$ is the last location where the freight was inspected, the resultant observation, and P. More specifically, the pmv at location n, given that an observation took place most recently at location $k < n$, and that the observation was j (i.e., $s(k) = j$) is $\vec{e}_j P^{n-k}$, where $\vec{e}_j$ is the $(S+1)$-vector having 1 as its j-th entry and 0 otherwise. If an observation has not been made since departure from the origin, then the pmv at location n is $\vec{x}^* P^n$, where $\vec{x}^*$ is the (a priori) pmv of the freight when it leaves the origin.

One might assume that the freight leaves the origin fresh (i.e., $x_0^* = 1$). However, this may not be the case for a variety of reasons: e.g., lengthy time on the dock when moving the freight from the temperature controlled warehouse to the temperature controlled trailers. Presumably, though, x_0^* equals or is close to 1.

At the origin, the DM can decide to inspect ($\mathcal{I}$) or not inspect ($\mathcal{NI}$) the product at location 1. If the decision is $\mathcal{I}$, then the actual (perfectly observed) state (or quality) of the product at location 1 is revealed once the vehicle arrives at location 1. At any intermediate location $n \in \{1, \ldots, N-1\}$, the DM can decide $\mathcal{I}$, $\mathcal{NI}$, or can select the alternative decision $\mathcal{A}$. Both the $\mathcal{I}$ and $\mathcal{NI}$ decisions assume the vehicle travels from n to $n+1$. Again, if the decision is $\mathcal{I}$, the state of the product at $n+1$ is revealed once the vehicle arrives at $n+1$. Decision $\mathcal{A}$ represents the decision to abort transporting the current load to the destination and restart the process of transporting goods from origin to destination.

The cost and reward structure is as follows. Assume each inspection costs M and the wholesale purchase cost is W, where both M and W are assumed to be negative. If the product arrives at the destination in state s, then a reward of R_s is accrued where R_s may be positive or negative. The product can only be purchased wholesale at the origin. Assume that the cost of traveling from n to $n' \in \{n-1, n+1\}$, whether the vehicle is loaded or empty, is $c(n, n') < 0$.

Costs associated with action $\mathcal{A}$ might include: a freight disposal cost, the cost (or reward) of selling the current load at a secondary market (perhaps dependent on freight quality), an expedite cost, and the cost of returning the vehicle to the origin (location dependent). We assume all of these costs can be described by the function $l_n + \vec{x}\vec{l}_n^x$, where n is the current location and $\vec{x}$ is the pmv of the current load at location n, and hence, the scalar l_n reflects the location-dependent costs while the vector $\vec{l}_n^x$ reflects both location and quality dependent costs (or rewards).

We discount costs and rewards as follows: Assume that at location n, it is decided to continue to location $n+1$ (i.e., either $\mathcal{I}$ or $\mathcal{NI}$ is selected at location n). Then, all costs incurred after reaching location $n+1$ are discounted by discount factor $\beta \in [0, 1)$. Both selection of action $\mathcal{A}$ and reaching the destination cause the current process of transporting goods from origin to destination to terminate and the next load of product to begin transport from origin to destination, either immediately or at some time or event in the future, depending on the situation. All costs incurred after the next load of product begins transport are discounted by discount factor $\beta_n \in [0,1)$ when action $\mathcal{A}$ is selected at location $n \in \{1,2,\ldots,N-1\}$, and are discounted by discount factor $\beta_N \in [0,1)$ when the destination is reached.

A policy is a rule that maps the current location n and the current pmv of the state of the freight at location n into the available actions. Our criterion is expected total discounted reward over the infinite horizon. Our objective is to determine a policy that maximizes the criterion and the resultant value of the criterion. Throughout, we assume that P and β are location invariant for notational simplicity. It is straightforward but notationally complicated to extend all of the results of this chapter to the more realistic case where P and β are location dependent. Proofs of all results are presented in Appendix A.1.

We remark that the operational issue of deciding what to do, based on real-time inspection data, with a perishable product that is in-transit from an origin to a destination is inherently a finite horizon problem while the more strategic firm level issue of managing a perishable product logistics business is reasonably modeled as an infinite horizon problem. Both the operational and firm level issues are captured and linked by the infinite horizon expected total discounted cost criterion presented above.

3.4 Preliminary Results

Results in Sondik [55] guarantee that it is sufficient to base action selection on $(n, \vec{x})$, where n is the current location and $\vec{x}$ is the pmv of the state of the freight at location n. Let $\omega(n, \vec{x})$ be the optimal expected discounted reward over the infinite horizon, given $(n, \vec{x})$. We seek $\omega(n, \vec{x})$ for all $(n, \vec{x})$ and a policy that generates these criterion values.

We begin by constructing $\omega(N, \vec{x})$. An expected quality-dependent reward/cost of $\vec{x}\vec{R}$ is accrued once the destination N is reached. Other costs may be incurred additionally (e.g., the cost of relocating the vehicle from the destination back to the origin). Let $l_N + \vec{x}\vec{R}$ represent the total cost and reward accrued between the time the destination is reached and the time the next movement of goods from origin to destination begins. Then,

$$\omega(N, x) = l_N + \vec{x}\vec{R} + \beta_N \omega(0, \vec{x}^*), \tag{5}$$

where $\omega(0, \vec{x}^*)$ is the optimal expected discounted reward over the infinite horizon when the next movement of goods from 0 to N begins.

For $n = 1, \ldots, N-1$,

$$\omega(n, \vec{x}) = \max \begin{cases} c(n, n+1) + \beta\omega(n+1, \vec{x}P), \\ c(n, n+1) + M + \beta\vec{x}P\vec{\omega}(n+1), \\ l_n + \vec{x}\vec{l}_n^x + \beta_n\omega(0, \vec{x}^*), \end{cases} \tag{6}$$

where $\vec{\omega}(n+1) = \text{col}\{\omega(n+1, \vec{e}_0), \omega(n+1, \vec{e}_1), \ldots, \omega(n+1, \vec{e}_S)\}$ and where the three terms on the RHS of Equation (6) are associated with actions $\mathcal{NI}$, $\mathcal{I}$, and $\mathcal{A}$ respectively. Similarly,

$$\omega(0, \vec{x}) = W + \max \begin{cases} c(0, 1) + \beta\omega(1, \vec{x}P), \\ c(0, 1) + M + \beta\vec{x}P\vec{\omega}(1). \end{cases} \tag{7}$$

Equations (5), (6) and (7) constitute the optimality equations and boundary conditions for the problem. Results in Chapter 6 of Puterman [50] guarantee that an optimal policy can be constructed from the actions that cause the maxima in the optimality equations to be achieved.

3.4.1 Determination of $\omega(0, \vec{x}^*)$

We observe that the key to determining $\omega(n, \vec{x})$ for all $(n, \vec{x})$ is determining $\omega(0, \vec{x}^*)$. Once $\omega(0, \vec{x}^*)$ is known, determining $\omega(n, \vec{x})$ for all $(n, \vec{x})$ becomes a finite horizon dynamic program.

We now construct a process for finding $\omega(0, \vec{x}^*)$. Let z be the current approximation of $\omega(0, \vec{x}^*)$ and define

$$\omega'(N, \vec{x}, z) = l_N + \vec{x}\vec{R} + \beta_N z,$$

$$\omega'(n, \vec{x}, z) = \max \begin{cases} c(n, n+1) + \beta\omega'(n+1, \vec{x}P, z), \\ c(n, n+1) + M + \beta\vec{x}P\vec{\omega}'(n+1, z), \\ l_n + \vec{x}\vec{l}_n^x + \beta_n z, \end{cases} \quad \text{for } n = 1, \ldots, N-1,$$

$$\omega'(0, \vec{x}, z) = W + \max \begin{cases} c(0, 1) + \beta\omega'(1, \vec{x}P, z), \\ c(0, 1) + M + \beta\vec{x}P\vec{\omega}'(1, z), \end{cases}$$

where the s-th element of the $(S+1)$-vector $\vec{\omega}'(n, z)$ is $\omega'(n, \vec{e}_s, z)$.

We note that if $z = \omega(0, \vec{x}^*)$, then $\omega'(n, \vec{x}, z) = \omega(n, \vec{x})$ for all $(n, \vec{x})$. Let $H(z) = \omega'(0, \vec{x}^*, z)$. Our objective then becomes to find a fixed point for the operator H. It is straightforward to show that H is a contraction mapping. Hence, there exists a unique z^* such that $z^* = H(z^*)$, and this fixed point is $\omega(0, x^*)$. Further, for any bounded z_0, $\lim_{n\to\infty} |\omega(0, x^*) - z_n| = 0$, where $z_{n+1} = H(z_n)$. This limit forms the basis for determining $\omega(0, \vec{x}^*)$ and hence $\omega(n, \vec{x})$ for all $(n, \vec{x})$. We now focus on finding $H(z)$, given z.

3.5 Structural Results

3.5.1 Optimal Reward Function Determination

We now present an approach for determining $\omega(n, \vec{x})$ for all $(n, \vec{x})$. We begin by determining $H(z)$, given z. For given z, results in Smallwood and Sondik [54] guarantee that $\omega'(n, \vec{x}, z)$ is piecewise linear and convex in $\vec{x}$. Thus, for each n, there is a finite set of $(S+1)$-vectors $V'(n, z)$ such that

$$\omega'(n, \vec{x}, z) = \max \Big\{ \vec{x}\vec{m} : \vec{m} \in V'(n, z) \Big\}.$$

We now determine the vectors in $V'(n, z)$. For $n = 1, \ldots, N$, let the $(S+1)$-vector $\vec{l}(n, z) = (l_0(n, z), \ldots, l_S(n, z)) = \vec{l}_n^x + (l_n + \beta_n z)\vec{e}$, where $\vec{e}$ is the $(S+1)$-vector having 1 in all entries and $\vec{l}_N^x = \vec{R}$. Let the vector $\vec{m}(\rho, n, z)$ for all n and ρ such that $N \geq \rho \geq n \geq 0$, be defined as follows:

$$\vec{m}(N, N, z) = \vec{l}(N, z),$$

$$\vec{m}(\rho, n, z) = c(n, n+1)\vec{e} + \beta P \vec{m}(\rho, n+1, z) \text{ for } N \geq \rho > n \geq 0, \text{ and}$$

$$\vec{m}(n, n, z) = (c(n, n+1) + M)\vec{e} + \beta P \vec{\eta}(n+1, z) \text{ for } N > n \geq 0,$$

where the s-th element of the $(S+1)$-vector $\vec{\eta}(n+1, z)$ is $\max\{l_s(n+1, z), \vec{m}_s(n+1, n+1, z), \ldots, \vec{m}_s(N, n+1, z)\}$, and $\vec{m}_s(\rho, n, z)$ is the s-th element of the vector $\vec{m}(\rho, n, z)$ (i.e., $\vec{m}_s(\rho, n, z) = \vec{e}_s \vec{m}(\rho, n, z)$).

Let A1 (Assumption 1) be defined as follows:

A1 : *For* $n = 1, 2, \ldots, N-2$, $\vec{l}(n, z) \geq c(n, n+1)\vec{e} + \beta P \vec{l}(n+1, z)$.

A1 indicates that for all intermediate locations, the earlier the decision $\mathcal{A}$ is made, the less corresponding loss is incurred. This is clearly the case for transportation cost back to the origin or the cost of expediting freight when the freight is to be delivered periodically (i.e., the less time-urgent the expedite transport, the less expensive the expedite decision).

A straightforward induction argument proves the following result.

Proposition 2 *Assume A1. Then,*

$$\omega'(N, \vec{x}, z) = \vec{x}\vec{l}(N, z),$$

$$\omega'(n, \vec{x}, z) = \max\{\vec{x}\vec{l}(n, z), \vec{x}\vec{m}(n, n, z), \vec{x}\vec{m}(n+1, n, z), \ldots, \vec{x}\vec{m}(N, n, z)\}$$

$$\text{for } n = 1, \ldots, N-1, \text{ and}$$

$$\omega'(0, \vec{x}, z) = W + \max\{\vec{x}\vec{m}(0, 0, z), \vec{x}\vec{m}(1, 0, z), \ldots, \vec{x}\vec{m}(N, 0, z)\}.$$

Thus, $V'(N, z) = \{\vec{l}(N, z)\}$, $V'(n, z) = \{\vec{l}(n, z), \vec{m}(n, n, z), \vec{m}(n+1, n, z), \ldots, \vec{m}(N, n, z)\}$ for $n = 1, \ldots, N-1$, and $V'(0, z) = \{W\vec{e} + \vec{m}(0, 0, z), W\vec{e} + \vec{m}(1, 0, z), \ldots, W\vec{e} + \vec{m}(N, 0, z)\}$. Let $|V|$ be the cardinality of the set V. Note that $|V'(N, z)| = 1$, $|V'(N-k, z)| = k+2$, $k = 1, \ldots, N-1$, and $|V'(0, z)| = N+1$. The extremely modest increase in the cardinality of the $V'(n, z)$ sets as n gets smaller is in stark contrast to the potential increase in the cardinality of concomitant sets for the general finite horizon POMDP and to the fact that this cardinality can be (countably) infinite for the infinite horizon POMDP [55].

We summarize the determination of $\omega(n, \vec{x})$ as Algorithm 3:

0. Select z_0; set $k = 0$;
1. Determine $\{V'(n, z_k), n = N, \ldots, 0\}$;
2. Set $z_{k+1} = \max\{\vec{x}^*\vec{m} : \vec{m} \in V'(0, z_k)\}$;
3. If $|z_{k+1} - z_k| < \epsilon$ for a sufficiently small ϵ, then stop, use z_k as an approximation of z^*, and use $\max\{\vec{x}\vec{m} : \vec{m} \in V'(n, z_k)\}$ as an approximation of $\omega(n, \vec{x})$. Otherwise, set $k = k+1$ and go to Step 1;

Algorithm 3: Procedure for determining $\omega(n, \vec{x})$

3.5.2 Optimal Reward Function Structure

We now show that the model reflects the statement: the fresher the product, the greater the expected profit. We use stochastic dominance to compare two pmvs for freshness: $\vec{x}'$ is at least as fresh as $\vec{x}$ ($\vec{x} \prec \vec{x}'$) if and only if $\sum_{s \geq k} x_s \geq \sum_{s \geq k} x'_s$ for all k. Let $\vec{p}_s = (p_{s0}, \ldots, p_{sS})$ be the s-th row of P. We now present Assumption 2 (A2):

A2 : *P is stochastic and $\vec{p}_{s+1} \prec \vec{p}_s$ for all s.*

We remark that A2 has the following interpretation: the fresher the current product, the more likely the product will be fresher one decision epoch in the future.

We now present conditions that guarantee $\omega'(n, \vec{x}, z)$ is monotonically non-decreasing in $\vec{x}$ with respect to stochastic dominance. Let a vector $\vec{v}$ be NIV (a non-increasing vector) if and only if $\vec{v}_s \geq \vec{v}_{s+1}$ for all s.

Proposition 3 *Assume A1, A2, and $\vec{l}_n^x$ are NIV for all $n = 1, 2, \ldots, N$. If $\vec{x} \prec \vec{x}'$, then $\omega'(n, \vec{x}, z) \leq \omega'(n, \vec{x}', z)$ for all $n = 0, 1, \ldots, N$.*

We remark that the assumptions that $\vec{l}_n^x$ are NIV for $n = 1, \ldots, N$ are consistent with the assumption that the greater the state of the freight, the lower its quality, which, in turn, induces higher cost or lower reward.

3.5.3 Optimal Policy Structure

We now investigate the structure of an optimal policy. It is intuitive that if, at any intermediate location, the product is spoiled (i.e., if $\vec{x} = \vec{e}_S$), then the trip should be aborted (i.e., select action $\mathcal{A}$). This characteristic holds under robust conditions, as we show in Proposition 4. It is also intuitive that if it is known that the product is fresh (i.e., if $\vec{x} = \vec{e}_0$), then it is unlikely to be necessary to inspect (select action $\mathcal{NI}$). For situations in between, there may be a reason to inspect the product (select action $\mathcal{I}$). Thus, as the product's state changes from spoiled to fresh, it is reasonable to expect that an optimal action would change from $\mathcal{A}$ to $\mathcal{I}$ to $\mathcal{NI}$ (i.e., the fresher the product, the less the need to inspect). More formally, and letting π^* be an optimal policy, we interpret these statements as follows:

(i) If $\pi^*(n, \vec{x}) \neq \mathcal{A}$, and $\vec{x} \prec \vec{x}'$, then $\pi^*(n, \vec{x}') \neq \mathcal{A}$.

(ii) If $\pi^*(n, \vec{x}) \notin \{\mathcal{A}, \mathcal{I}\}$ and $\vec{x} \prec \vec{x}'$, then $\pi^*(n, \vec{x}') \notin \{\mathcal{A}, \mathcal{I}\}$.

We will show (i) and (ii) hold when $S = 1$ in Proposition 5, and we will show (i) holds for all S in Proposition 6. However, Appendix C.2 presents a counter-example to the claim that (ii) holds for all S. We begin by defining an optimal policy based on the optimality equations.

Define the optimal policy $\pi^*(n, \vec{x})$ as follows:

(i) if $\omega'(n, \vec{x}, z^*) = \vec{x}\vec{l}(n, z^*)$, then let $\pi^*(n, \vec{x}) = \mathcal{A}$,

(ii) if $\omega'(n, \vec{x}, z^*) = \vec{x}\vec{m}(n, n, z^*)$, then let $\pi^*(n, \vec{x}) = \mathcal{I}$, and

(iii) if $\omega'(n, \vec{x}, z^*) = \vec{x}\vec{m}(\rho, n, z^*)$ for any $\rho \in \{n+1, \ldots, N\}$, then let $\pi^*(n, \vec{x}) = \mathcal{NI}$.

The policy π^* is optimal by results in Smallwood and Sondik [54] and by the construction of the $\vec{m}(\rho, n, \vec{x})$ and essentially reflects the statement that if an action causes the maximum in the optimality equations to be attained, then it is an optimal action.

We now show that under weak conditions, it is always optimal to select action $\mathcal{A}$ when the product has spoiled in route to the destination.

Proposition 4 *Assume A1, $p_{SS} = 1$, and $c(N-1, N) + \beta l_S(N, z^*) \leq l_S(N-1, z^*)$. Then, $\pi^*(n, \vec{e}_S)$ can be selected to equal $\mathcal{A}$ for all $n = 1, \ldots, N-1$.*

We remark that $p_{SS} = 1$ is consistent with the statement: once the freight has spoiled, its quality will never improve. Additionally, the assumption that $c(N-1, N) + \beta l_S(N, z^*) \leq l_S(N-1, z^*)$ is consistent with assumption A1.

We now investigate the $S = 1$ case. Let $\sigma(\rho, n, z) = \vec{m}_0(\rho, n, z) - \vec{m}_1(\rho, n, z)$, the slope of $\vec{x}\vec{m}(\rho, n, z)$ versus x_0. Proposition 3 guarantees that this slope is always non-negative.

Proposition 5 *Assume:*

(i) $S = 1$, $p_{01} = \alpha$, $p_{11} = 1$, and $\gamma = \beta(1-\alpha)$,

(ii) A1, and $c(N-1, N) + \beta l_1(N, z) \leq l_1(N-1, z)$,

(iii) for all $n = 1, \ldots, N-1$, $l_0(n, z) < \omega'(n, \vec{e}_0, z)$.

Then, for all $n = 1, \ldots, N-1$, $\sigma(\rho, n, z) \leq \sigma(\rho+1, n, z)$ for $\rho = n, \ldots, N-1$.

We interpret Proposition 5, with the help of Proposition 4 restricted to the case where $S = 1$, as follows. Proposition 4 guarantees that there is a $x_0' \in [0, 1]$ where if $x_0 \leq x_0'$, the optimal action is $\mathcal{A}$. Note that the facet of $\omega'(n, \vec{x}, z)$ with the least non-negative slope has slope $\sigma(n, n, z)$ and is associated with action $\mathcal{I}$. This suggests that there is a $x_0'' \in [x_0', 1]$ such that if $x_0 \in [x_0', x_0'']$, then choose action $\mathcal{I}$, and for all $x_0 \in [x_0'', 1]$, choose action $\mathcal{NI}$. Figure 5 is a graphical depiction of this discussion.

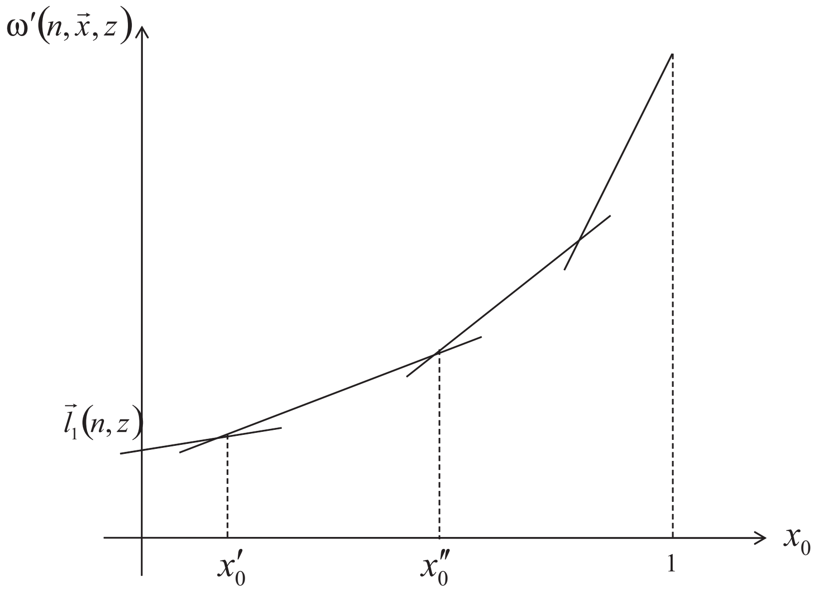

Figure 5: Graph of $\omega'(n, \vec{x}, z)$ versus x_0 for the $S = 1$ case.

Observe that assumption (iii) in Proposition 5 implies that it is optimal to continue to travel toward the destination if, at location n, the probability that the product is fresh is 1 (i.e., $x_0 = 1$). If this assumption does not hold, then it would always be optimal to select action $\mathcal{A}$, one would never make a profit, and hence there would be no reason to enter into or remain in the business.

Proposition 5 implies for the $S = 1$ case that if $\pi^*(n, \vec{x}) = \mathcal{NI}$ and $x_0 \leq x_0'$, then $\pi^*(n, \vec{x}') \neq \mathcal{I}$. However, this statement does not hold in general when $S > 1$. Given that the current decision epoch and pmv is $(n, \vec{x})$, the expected discounted value of knowing the state of the freight at epoch $n+1$ is $\beta[\vec{x}P\overline{\omega}(n+1) - \omega(n+1, \vec{x}P)]$. If $|M| \geq \beta[\vec{x}P\overline{\omega}(n+1) - \omega(n+1, \vec{x}P)]$, then the cost of obtaining freight state information exceeds its expected value, and hence it would be better to select $\mathcal{NI}$ than $\mathcal{I}$. The convexity of $\omega(n, \vec{x})$ in $\vec{x}$ for all n assures there are $\vec{x}$ and $\vec{x}'$ such that, for $\vec{x} \prec \vec{x}'$ and P satisfying A2,

$$\vec{x}'P\overline{\omega}(n+1) - \vec{x}P\overline{\omega}(n+1) \geq \omega(n+1, \vec{x}'P) - \omega(n+1, \vec{x}P).$$

Thus, for any M such that

$$\beta[\vec{x}'P\overline{\omega}(n+1) - \omega(n+1, \vec{x}'P)] \geq |M| \geq \beta[\vec{x}P\overline{\omega}(n+1) - \omega(n+1, \vec{x}P)],$$

$\pi^*(n, \vec{x}) = \mathcal{NI}$ and $\pi^*(n, \vec{x}') = \mathcal{I}$ (assuming $\pi^*(n, \vec{x}) \neq \mathcal{A}$ and $\pi^*(n, \vec{x}') \neq \mathcal{A}$). Appendix C.2 presents an example where $\vec{x}$ and $\vec{x}'$ are such that $\vec{x} \prec \vec{x}'$, $\pi^*(n, \vec{x}) = \mathcal{NI}$ and $\pi^*(n, \vec{x}') = \mathcal{I}$ when $S > 1$.

We now give conditions that guarantee for all S that if it is suboptimal to select $\mathcal{A}$ in state $(n, \vec{x})$, and if $\vec{x}'$ is at least as fresh as $\vec{x}$, then it is suboptimal to choose $\mathcal{A}$ when in state $(n, \vec{x}')$.

Lemma 1 *Let X be a (multi-dimensional) convex set and $\{\omega_a(\vec{x})\}_{a \in A}$ be an arbitrary family of (real-valued) convex functions over X where A is a (convex) index set where each element corresponds to an action. In addition, let $\omega^*(\vec{x}) = \sup_{a \in A}\{\omega_a(\vec{x})\}$ for all $\vec{x} \in X$, and $X_a^* = \{\vec{x} \in X : \omega_a(\vec{x}) = \omega^*(\vec{x})\}$, an optimal region in X which is associated with $\omega_a(\vec{x})$, for each $a \in A$. Then, if $\omega_a(\vec{x})$ is an affine function for any $a \in A$, X_a^* is a convex set.*

Thus, there exists a convex set $\mathcal{X}_{\mathcal{A}} \subseteq \mathcal{X}$ containing $\vec{e}_S$ such that $\pi^*(\vec{x}) = \mathcal{A}$ for all $\vec{x} \in \mathcal{X}_{\mathcal{A}}$ due to Lemma 1 and Propositions 4 and 6.

Proposition 6 *Assume A1, A2, and $\beta P\vec{l}^{x}_{n+1} - \vec{l}^{x}_{n}$ is NIV for all $n \leq N-1$. If $\pi^*(n, \vec{x}) \neq \mathcal{A}$ and $\vec{x} \prec \vec{x}'$, then $\pi^*(n, \vec{x}') \neq \mathcal{A}$.*

We now examine the conditions on $\{\vec{l}^{x}_{n}\}$ found in the assumptions of Proposition 6. These conditions hold trivially when $\vec{l}^{x}_{n} = \vec{0}$. Observe that A1 holds when the following two conditions hold:

(i) $l_n + \beta_n z^* \geq c(n, n+1) + \beta(l_{n+1} + \beta_{n+1} z^*)$

(ii) $\vec{l}^{x}_{n} \geq \beta P \vec{l}^{x}_{n+1}$.

If condition (i) is intended to capture discounted transportation cost, as it would for our first scenario (where $\vec{l}^{x}_{n} = \vec{0}$ for all n), then (i) simply represents the triangle inequality when $\beta = 1$.

We remark that $\beta P\vec{l}^x_{n+1}$ can be interpreted as the expected secondary market reward structure one decision epoch in the future, taking into account expected freight deterioration. Thus, condition (ii) indicates that this reward structure is expected to be uniformly reduced going forward. We remark that if $\vec{l}^x_n$ is NIV and location invariant (all secondary markets share the same reward structure) and if $p_{ij} = 0$ for $j < i$ (freight quality never improves), then condition (ii) holds. The condition that $\beta P\vec{l}^x_{n+1} - \vec{l}^x_n$ is NIV indicates that we can expect the state of secondary reward structure deterioration to be greater as freight quality degrades. We remark that when secondary markets are highly dynamic and/or non-uniform, then these assumptions on $\vec{l}^x_n$ may not hold, and hence the result of Proposition 6 may not be true for the $\vec{l}^x_n \neq 0$ case.

3.5.4 Sensitivity Analysis

We now examine how the cost of inspection (M), the structure of the rewards and costs received when the freight is delivered to the destination ($\vec{R}$), and the description of freight deterioration (P) affect the optimal total reward function $\omega(\cdot)$. Let $\omega'(n, x, z; M)$, $\omega'(n, x, z; \vec{R})$, and $\omega'(n, x, z; P)$ be defined as before, explicitly recognizing the dependence of ω' on the parameters M, $\vec{R}$, and P, respectively.

We begin by considering the cost of inspection. It can be shown that for any n, $\vec{x}$ and z, (i) $\omega'(n, \vec{x}, z; M)$ is monotonically non-decreasing in M, and (ii) there is a set $V_M(n, \vec{x}, z)$ (countable but not necessarily finite) where $\omega'(n, \vec{x}, z; M) = \max\{aM + b : (a, b) \in V_M(n, \vec{x}, z)\}$ and noting that an element in $V_M(n, \vec{x}, z)$ is a pair of scalars. Hence, $\omega'(n, \vec{x}, z; M)$ is both monotonically non-decreasing and convex in M. Thus, an increase in the cost of an inspection has more of a negative impact on the optimal reward function when this cost is close to zero than otherwise. This characteristic is illustrated in Figure 6, where the slope of the graph increases as M increases. We also remark that inspection frequency will decrease, and the time between inspections will increase as the inspection cost increases.

We now examine the impact of changes of the reward vector $\vec{R}$ on $\omega'(n, \vec{x}, z; \vec{R})$. Following an argument identical to the above case for the cost of inspection, we can show that for

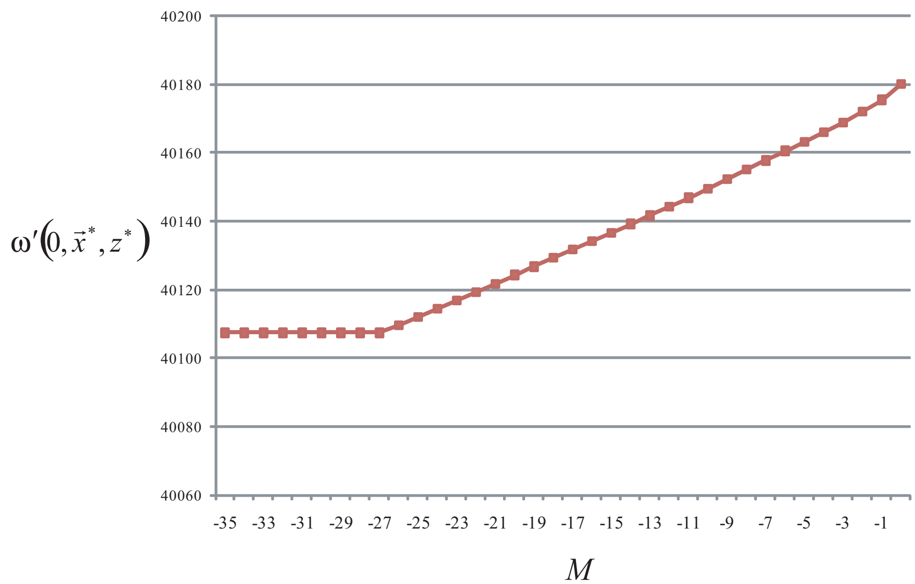

Figure 6: Graph of $\omega'(0, \vec{x}^*, z^*; M)$ versus M.

any n, $\vec{x}$ and z, $\omega'(n, \vec{x}, z; \vec{R})$ is monotonically non-decreasing in $\vec{R}$, and there is a countable but not necessarily finite set $V_R(n, \vec{x}, z)$ where $\omega'(n, \vec{x}, z; \vec{R}) = \max\{\vec{a}\vec{R} + b : (\vec{a}, b) \in V_R(n, \vec{x}, z)\}$ where in this case $\vec{a}$ is an $(S+1)$-vector and b is a scalar. Thus, $\omega'(n, \vec{x}, z; \vec{R})$ is also convex in $\vec{R}$, and hence the incremental increase of expected profit per increase in unit reward gets larger as the reward gets larger.

We next investigate the impact of variations in the transition matrix P on $\omega'(n, \vec{x}, z; P)$. Let $P \prec P'$ if and only if $\vec{p}_s \prec \vec{p}'_s$ for all $s = 0, 1, \ldots, S$. We can interpret this ordering on P as follows: A transportation mode with P' is more reliable than one with P in terms of freight quality degradation. Then, we can show an intuitive result stating that more reliable transportation mode yields higher expected reward.

The following result summarizes all the discussions above:

Proposition 7 *Given any n, $\vec{x}$ and z,*

(a) $\omega'(n, \vec{x}, z; M)$ is monotonically non-decreasing and convex in M.

(b) $\omega'(n, \vec{x}, z; \vec{R})$ is monotonically non-decreasing and convex in $\vec{R}$.

(c) if $P \prec P'$, then $\omega'(n, \vec{x}, z; P) \leq \omega'(n, \vec{x}, z; P')$.

3.6 Benchmarks

It follows from Proposition 7(a) that, for any $M \leq 0$, $\omega'(n, \vec{x}, z; -\infty) \leq \omega'(n, \vec{x}, z; M) \leq \omega'(n, \vec{x}, z; 0)$. In addition to serving as bounds, both $\omega'(n, \vec{x}, z; -\infty)$ and $\omega'(n, \vec{x}, z; 0)$ are special cases useful in understanding the value of being able to access freight status in transit. If observing freight status in transit and communicating this information in real time is cost prohibitive, then no observations will be made and transmitted. Since it is currently common business practice to make and transmit no observations of freight status in transit, $\omega'(n, \vec{x}, z^*; -\infty)$ represents the current practice benchmark, and $\omega'(n, \vec{x}, z^*; M) - \omega'(n, \vec{x}, z^*; -\infty)$ represents the value of having access to freight status in transit.

If observations are free, then (as we will show) it is optimal to observe freight status at every intermediate location. Hence, $\omega'(n, \vec{x}, z^*; 0) - \omega'(n, \vec{x}, z^*; -\infty)$ represents the maximum value that can be achieved of having access to freight status in transit. If this value does not justify acquiring the capability of monitoring freight status in transit, then it is unlikely that a business case can be made for in-transit freight status monitoring.

3.6.1 Lower Bounds: Current Practice in Industry

Let $\omega_L(n, \vec{x})$ be the optimal expected total discounted reward assuming that there is no access to product inspections in-transit from the origin to the destination. We remark that $\omega_L(n, \vec{x}) = \omega'(n, \vec{x}, z^*; -\infty)$ since having no access to product inspections in-transit is equivalent from a decision making perspective to having access to prohibitively expensive product inspections in-transit. As indicated by Proposition 7(a), $\omega_L(n, \vec{x}) \leq \omega(n, \vec{x})$ for all n and $\vec{x}$. The difference $\omega(n, \vec{x}) - \omega_L(n, \vec{x})$ is the value of having access to inspection at the intermediate locations and taking optimal advantage of this access. We now investigate $\omega_L(n, \vec{x})$.

If product inspection is not available at the intermediate locations, then the optimality equations simplify to

$$\omega_L(0,\vec{x}) = W + c(0,1) + \beta\omega_L(1,\vec{x}P),$$

$$\omega_L(n,\vec{x}) = \max\begin{cases} c(n,n+1) + \beta\omega_L(n+1,\vec{x}P) \\ l_n + \vec{x}\vec{l}_n^x + \beta_n\omega_L(0,\vec{x}^*), \end{cases} \quad \text{for } 1 \le n < N, \text{ and}$$

$$\omega_L(N,\vec{x}) = l_N + \vec{x}\vec{R} + \beta_N\omega_L(0,\vec{x}^*).$$

Define

$$\omega_{LL}(0,\vec{x}) = W + c(0,1) + \beta\omega_{LL}(1,\vec{x}P),$$

$$\omega_{LL}(n,\vec{x}) = c(n,n+1) + \beta\omega_{LL}(n+1,\vec{x}P) \text{ for } 1 \le n < N, \text{ and}$$

$$\omega_{LL}(N,\vec{x}) = l_N + \vec{x}\vec{R} + \beta_N\omega_{LL}(0,\vec{x}^*).$$

Clearly, $\omega_{LL}(n,\vec{x}) \le \omega_L(n,\vec{x})$ for all n and $\vec{x}$. Note that $\omega_{LL}(\cdot)$ is the value function for the policy "always travel directly from the origin to the destination," which is the policy typically used in industry currently. Letting $\vec{x} = \vec{x}^*$ in the equations above, it is straightforward to show that for $1 \le n < N$,

$$\omega_{LL}(n,\vec{x}^*P^n) = \sum_{k=n}^{N-1} \beta^{k-n}c(k,k+1) + \beta^{N-n}\left[l_N + \vec{x}^*P^N\vec{R} + \beta_N\omega_{LL}(0,\vec{x}^*)\right],$$

where $\omega_{LL}(0,\vec{x}^*) = \left\{W + \sum_{n=0}^{N-1}\beta^n c(n,n+1) + \beta^N\left[l_N + \vec{x}^*P^N\vec{R}\right]\right\}/(1-\beta^N\beta_N)$.

We now investigate the following question: assuming there is no opportunity to inspect the product in transit, when is the "always travel from origin to destination" policy optimal? An answer is when the following inequality holds for all $1 \le n < N$:

$$c(n,n+1) + \beta\omega_{LL}(n+1,\vec{x}^*P^{n+1}) \ge l_n + \vec{x}^*P^n\vec{l}_n^x + \beta_n\omega_{LL}(0,\vec{x}^*). \tag{8}$$

We now give conditions that imply this inequality holds for all $1 \le n < N$, assuming the inequality holds when $n = 1$.

Proposition 8 *If* $c(n,n+1) + \beta\omega_{LL}(n+1,\vec{x}^*P^{n+1}) \ge l_n + \vec{x}^*P^n\vec{l}_n^x + \beta_n\omega_{LL}(0,\vec{x}^*)$, *then* $c(n+1,n+2) + \beta\omega_{LL}(n+2,\vec{x}^*P^{n+2}) \ge l_{n+1} + \vec{x}^*P^{n+1}\vec{l}_{n+1}^x + \beta_{n+1}\omega_{LL}(0,\vec{x}^*)$.

Thus, if it is optimal to travel from location 1 to location 2 (rather than select action $\mathcal{A}$), then it is optimal to travel from the origin to the destination, which suggests a bound on the initial product quality $\vec{x}^*$. Appendix C.3 presents the conditions that the inequality (8) holds when $n = 1$, and hence $\omega_L(n, \vec{x}) = \omega_{LL}(n, \vec{x})$ due to Proposition 8.

We remark that $\omega(n, \vec{x})$ represents an upper bound on the optimal reward function for the case where product quality inspections are noise-corrupted. Hence, $\omega(n, \vec{x}) - \omega_L(n, \vec{x})$ represents an upper bound on the value of having access to inspections at the intermediate locations if inspections are partial observations of product quality.

3.6.2 Upper Bounds

We now seek upper bounds on $\omega(n, \vec{x})$. Let $\omega_U(n, \vec{x})$ be the expected optimal total discounted value function when there is no cost of inspection (i.e., $M = 0$). By Proposition 7(a), $\omega_U(n, \vec{x}) = \omega'(n, \vec{x}, z^*; 0)$ and hence $\omega(n, \vec{x}) \leq \omega_U(n, \vec{x})$ for all n and $\vec{x}$. We remark that $\omega_U(n, \vec{x}) - \omega_L(n, \vec{x})$ represents an upper bound on the value of being able to inspect the product in transit.

We also note that since $\omega(n, \vec{x})$ is convex in $\vec{x}$, $\omega(n, \vec{x}) \leq \vec{x}\vec{\omega}(n)$. Thus, when $M = 0$, action $\mathcal{NI}$ is always suboptimal and can be eliminated as an action. Hence,

$$\omega_U(0, \vec{x}) = W + c(0, 1) + \beta \vec{x} P \vec{\omega}_U(1),$$

$$\omega_U(n, \vec{x}) = \max \begin{cases} c(n, n+1) + \beta \vec{x} P \vec{\omega}_U(n+1) \\ l_n + \vec{x}\vec{l}_n^x + \beta_n \omega_U(0, \vec{x}^*) = \vec{x}\vec{l}(n, \omega_U(0, \vec{x}^*)), \end{cases} \quad \text{for } 1 \leq n < N, \text{ and}$$

$$\omega_U(N, \vec{x}) = l_N + \vec{x}\vec{R} + \beta_N \omega_U(0, \vec{x}^*) = \vec{x}\vec{l}(N, \omega_U(0, \vec{x}^*))$$

where $\vec{\omega}_U(n)$ is an $(S+1)$-vector whose j-th element is $\omega_U(n, \vec{e}_j)$.

Since the underlying product quality is assumed to be perfectly observable, the optimality equations above become as follows: for $1 \leq n < N$,

$$\omega_U(n, \vec{e}_s) = \max \begin{cases} c(n, n+1) + \beta \sum_{s'} p_{ss'} \omega_U(n+1, \vec{e}_{s'}) \\ \vec{e}_s \vec{l}(n, \omega_U(0, \vec{x}^*)) \end{cases}$$

and $\omega_U(N, \vec{e}_s) = \vec{e}_s \vec{l}(N, \omega_U(0, \vec{x}^*))$.

We now show that there is an optimal policy for the upper bound case with an intuitively appealing structure.

Proposition 9 *Let $M = 0$. Then,*

(i) for each $1 \le n < N$*, there is an* $\overline{s}(n) \in \{0, 1, \dots, S\}$ *for which*

(a) if $s \ge \overline{s}(n)$*, then choose action* $\mathcal{A}$*,*

(b) if $s < \overline{s}(n)$*, then choose action* $\mathcal{I}$*.*

Additionally, assume P *is upper-triangular (i.e.,* $p_{ij} = 0$ *if* $j < i$*). Then,*

(ii) $\overline{s}(n) \le \overline{s}(n+1)$*,* $n = 1, 2, \dots, N-2$*.*

Thus, there exists an optimal control-limit rule for each intermediate location, and the control-limit is non-increasing as the intermediate location gets closer to the destination (i.e., the farther we are from the destination, the fresher the product must be in order to be eventually profitable). These structural characteristics are potentially useful in constructing good, easily implementable suboptimal policies when the cost of inspection is small. In Appendix C.4, we present a closed-form formula of $\omega_U(n, \vec{e}_s)$ when such a policy having the structural properties in Proposition 9 is assumed.

We remark that since determining the upper bound $\omega_U(n, \vec{x})$ involves solving for $\omega(n, \vec{x})$ for the special case where $M = 0$, the solution procedure for determining $\omega(n, \vec{x})$ presented earlier can be used to determine $\omega_U(n, \vec{x})$. Thus, $\lim_{n \to \infty} |\omega_U(0, \vec{x}^*) - z_n| = 0$, where $z_{n+1} = H_U(z_n)$ for any bounded z_0, $H_U(z) = \omega_U'(0, \vec{x}^*, z)$,

$$\omega_U'(0, \vec{x}, z) = W + c(0,1) + \beta \vec{x} P \vec{\omega}_U(1, z),$$

$$\omega_U'(n, \vec{e}_s, z) = \max \begin{cases} c(n, n+1) + \beta \sum_{s'} p_{ss'} \omega_U'(n+1, \vec{e}_{s'}, z) \\ \vec{e}_s \vec{l}(n, z) \end{cases} \quad \text{for } 1 \le n < N, \text{ and}$$

$$\omega_U'(N, \vec{e}_s, z) = l_N + R_s + \beta_N z.$$

3.7 Illustrative Example

Consider a distribution system where fresh strawberries in California are loaded on a refrigerated truck for delivery to a destination on the East Coast of the U.S. Suppose that transit takes 4 days, and the refrigerated truck is equipped with a GPS-enabled communication module and temperature monitoring system. We assume that the decision maker can poll for real-time temperature data captured by the monitoring system from the vehicle via GPS/Satellite communication link, and the minimum inter-transmission interval is 1 hour. In this scenario, we assume that a new shipment is initiated every week. Typically, the procurement cost c_w of strawberries from suppliers is \$0.98/lb, their retail price c_p is \$2.29/lb, and the disposal cost c_d is \$0.2/lb. We also assume that transportation costs (including driver wages, fuel, and other costs) is approximately \$2/mile, and the driver drives 11 hours per day at an average speed of 50 miles/hour.

Suppose that the strawberries can be in one of eight states ($S = 7$). We can think of state 0 as the freshest state of the strawberries, freshness is reduced as the state increases, and state 7 is the least fresh (perished) state. The strawberries deteriorate in a Markovian fashion over each of the minimum inter-transmission intervals with the corresponding state transition probability matrix P as follows:

$$P = \begin{bmatrix} 0.97 & 0.03 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.96 & 0.04 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.94 & 0.06 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.92 & 0.08 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.88 & 0.12 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.86 & 0.12 & 0.02 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.8 & 0.2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

When strawberries are loaded onto the refrigerated truck at the origin, it is assumed that the strawberries are in state 0 with probability 0.95 and in state 1 with probability 0.05.

Finally, we consider a scenario for action $\mathcal{A}$ that an expedite delivery is arranged while the current freight is disposed of, and that the expedite decision incurs a fixed cost of d_S per pound at any intermediate location. Thus, we consider sets of freight sizes Q (lb) and (fixed) costs d_S (\$/lb) for action $\mathcal{A}$: $Q \in \{30000\text{lb}, 40000\text{lb}, 50000\text{lb}\}$ and $d_S \in \{1.5c_d, 2.5c_d, 3c_d, 3.5c_d, 4c_d, 6c_d, 10c_d\}$ where c_d is the disposal cost per pound.

We also consider three different levels of reward per pound (denoted c_r (\$/lb)), where $c_r \in \{1.4, 1.5, 1.6\}$ when the freight state at the destination is in state 0. We remark that the penalty $6c_d$ is larger than the procurement cost per pound, c_w, and $10c_d$ is larger than the reward per pound, c_r. We further assume a state-dependent reward structure as follows: Let R_0 be the total reward when delivered goods are in state 0. Then, if the delivered freight at the destination is in state s ($s = 1, 2, \ldots, 7$), the corresponding reward R_s is determined by $\rho_s R_0$ where $(s, \rho_s) \in \{(1, 0.99), (2, 0.97), (3, 0.95), (4, 0.9), (5, 0.8), (6, 0.3), (7, 0)\}$. We note that when the strawberries are in state 6, the per-pound reward becomes $0.3c_r$, which is less than the per-pound procurement cost c_w, implying that a net loss is incurred. We also vary values of the inspection cost M(\$) over $\{0, -0.5, -1, -2, -5, -10, -20, -50, -70, -100\}$ in order to determine how the cost of inspection affects the expected total discounted profit over the infinite horizon. In practice, the overall inspection cost is determined as follows. For example, Euroscan[3] charges at least (i) a data transmission cost ($\approx$ \$0.1/message), (ii) a data hosting and web access fee (\$6/month), and (iii) an one-time configuration fee (\$22 for terminal activation fee, and \$50 for one-time setup fee per terminal), in addition to a fee related to the monitoring and communication modules. Table 12 presents the summary of parameters used in this example.

For each freight size and per-pound reward, we present the expected total discounted reward $\omega(0, \vec{x}^*)$ if the pmv of the initial freight state at the origin is $\vec{x}^*$ in Figure 7, depending on the values of costs for action $\mathcal{A}$, d_S, and inspection cost M. We observe that, for given reward vector R, $\omega(0, \vec{x}^*)$ becomes more sensitive to the change of the inspection cost M as the penalty for delivery failure increases. This supports the intuitive idea that as the cost d_S increases, the decision maker will want to inspect more frequently in order to make an appropriate action as early as possible if the quality of the freight deteriorates significantly. Numerical experiments show that $\omega(0, \vec{x}^*)$ gradually decreases as d_S increases and eventually, $\omega(0, \vec{x}^*)$ converges for reasonable inspection cost levels. This suggests that there may exist a threshold of d_S such that $\omega(0, \vec{x}^*)$ no longer decreases when the inspection cost is not sufficiently large. This result seems initially counter-intuitive. However, it may

[3]http://www.euroscanweb.com

Table 12: Value of parameters in the illustrative example

Parameters	Description	Value(s)
M	Inspection cost (\$)	$\{0, -0.5, -1, -2, -5, -10, -20, -50, -70, -100\}$
Q	Freight size (lb)	$\{30000, 40000, 50000\}$
c_p	Retail price (\$/lb)	2.29
c_w	Procurement cost (\$/lb); $W = -c_w Q$	0.98
c_r	Reward if the freight state is in 0 (\$/lb); $R_0 = c_r Q$	$\{1.4, 1.5, 1.6\}$
ρ_s	State-dependent reward rate; $R_s = \rho_s R_0, s = 0, 1, \ldots, S$	$\{1, 0.99, 0.97, 0.95, 0.9, 0.8, 0.3, 0\}$
c_d	Disposal cost (\$/lb)	0.2
d_S	Fixed cost for action $\mathcal{A}$ per pound (\$/lb); $l_n = -d_S Q$	$\{1.5c_d, 2.5c_d, 3c_d, 3.5c_d, 4c_d, 6c_d, 10c_d\}$
$\vec{x}^*$	pmv of the initial product quality at the origin	$\begin{bmatrix}0.95 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0\end{bmatrix}$
β	Discount factor	0.995

imply that the larger penalty impedes selecting a decision $\mathcal{A}$, and hence, the vehicle is less likely to abort delivery since the freight may still generate a small amount of revenue. In addition, as the reward vector $\vec{R}$ increases, $\omega(0, \vec{x}^*)$ becomes more sensitive to a change in the inspection cost M since the net profit upon arrival at the destination increases. Thus, the decision maker becomes more interested in the freight status in transit as $\vec{R}$ increases.

We also evaluate how the value of inspection in transit depends on the changes of rewards or penalties. An upper bound on the value of information (VI), $\Delta_{VI} = \Big\{(\omega_U(0, \vec{x}^*) - \omega_L(0, \vec{x}^*))/\omega_L(0, \vec{x}^*)\Big\} \times 100$ (%), is introduced to measure the percentage increase of the expected total discounted reward when in-transit inspection is available, and we use $\overline{\Delta}_{VI} = \Big\{(\omega_U(0, \vec{x}^*) - \omega(0, \vec{x}^*; -100))/\omega(0, \vec{x}^*; -100)\Big\} \times 100$ (%) as its approximation. Figure 8 presents the value of this bound as per-pound rewards, penalties and freight sizes vary. This analysis provides several intuitively appealing results: the percentage gains in profit from inspection increases as per-pound penalty increases, and a smaller per-pound reward induces a larger information gap (see Appendix C.5 for further discussion). In addition, the benefits from inspection decline, and the increasing rate of the bound, as a function of

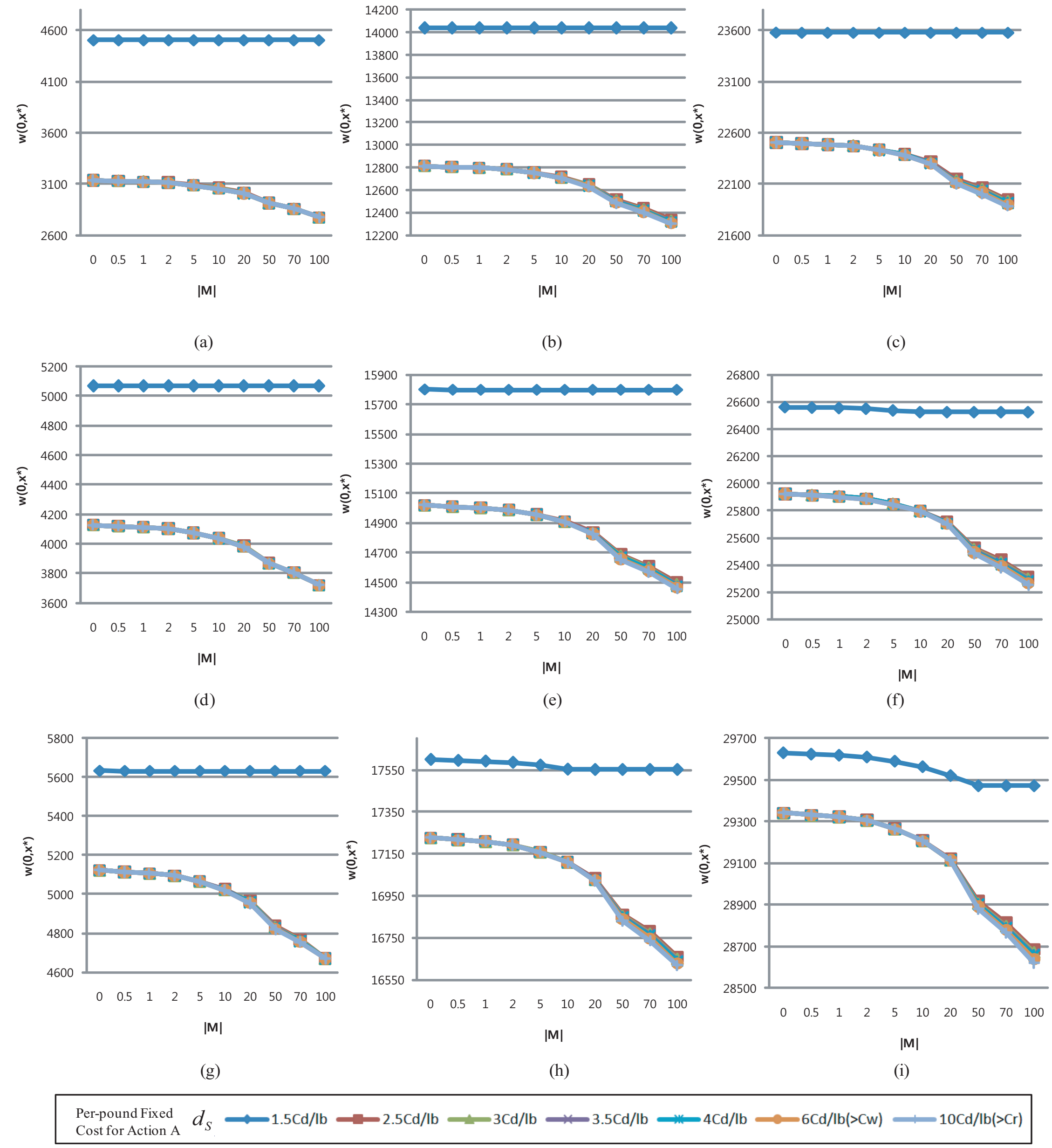

Figure 7: Changes of $\omega(0, \vec{x}^*)$ according to the changes of freight size Q, inspection cost M, per-pound reward c_r, and per-pound cost for action $\mathcal{A}$ d_S: The parameter values used in this figure are (i) freight size $Q = 30000lb$ for (a)(b)(c), $Q = 40000lb$ for (d)(e)(f), and $Q = 50000lb$ for (g)(h)(i); (ii) per-pound reward $c_r = \$1.4/lb$ for (a)(d)(g), $c_r = \$1.5/lb$ for (b)(e)(h), and $c_r = \$1.6/lb$ for (c)(f)(i).

per-pound penalty, decreases as the freight size increases.

We now investigate the impact of variations in the transition matrix P on the optimal expected total discounted reward over the infinite horizon. For the numerical studies, we consider the following three transition matrices P_l, P_m, P_h where $P_l \prec P_m \prec P_h$ such that

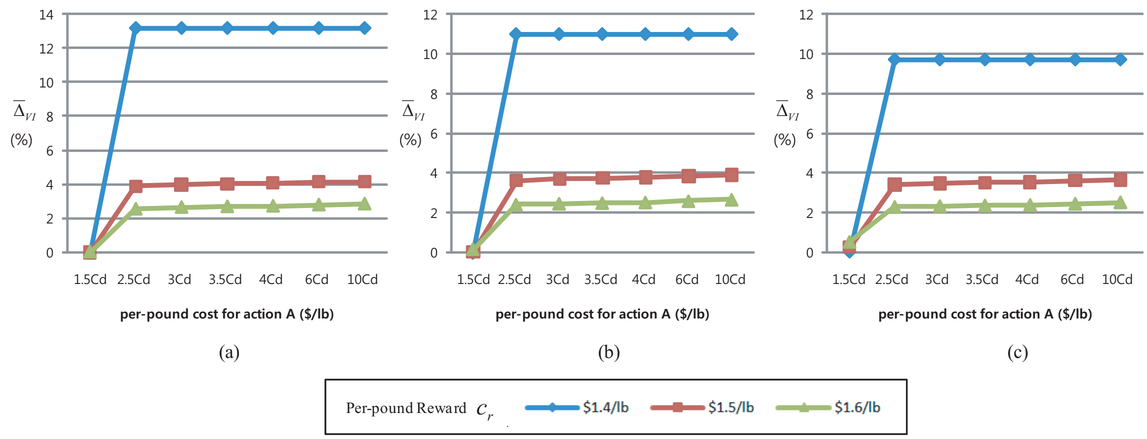

Figure 8: Changes of $\overline{\Delta}_{VI}$ according to the changes of freight size Q, per-pound reward c_r, and per-pound cost for action $\mathcal{A}$ d_S: The parameter values used in this figure are as follows: freight size $Q = 30000lb$ for (a), $Q = 40000lb$ for (b), and $Q = 50000lb$ for (c).

P_l and P_h are P after replacing $\vec{p}_4$ and $\vec{p}_5$ of P with

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0.88 & 0.12 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.85 & 0.13 & 0.02 \end{bmatrix},$$

and

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0.89 & 0.11 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.87 & 0.11 & 0.02 \end{bmatrix}$$

respectively, where $P_m = P$. Figure 9 indicates that the information gap decreases as shipment process reliability increases, which implies that the information from freight status inspection in transit is more valuable when the freight is delivered with a less reliable transportation mode and the benefits from inspection decline as reliability increases.

3.8 Summary

In this chapter, we have determined the value of monitoring perishable freight in-transit for a single vehicle traveling from an origin to a destination. We developed a computationally practical approach for determining the optimal expected cost function and an optimal policy, based on an infinite horizon partially observed Markov decision process model. Structural properties of the optimal expected cost function and optimal policy have been determined. These results can lend insight when deciding whether to acquire the capacity to monitor

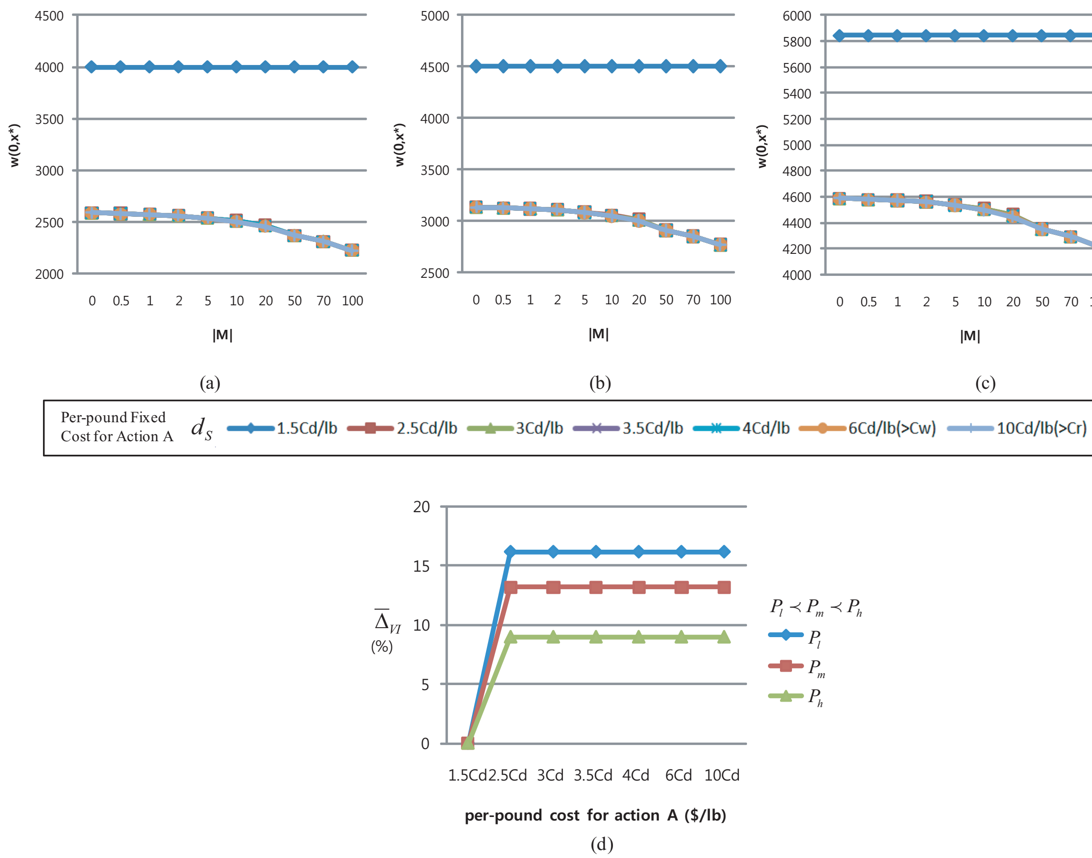

Figure 9: Changes of $\omega(0, \vec{x}^*)$ (a)(b)(c) and $\overline{\Delta}_{VI}$ (d) according to the changes of the freshness-keeping effort P, inspection cost M, and per-pound cost for action $\mathcal{A}$ d_S: The parameter values used in this figure are as follows: (i) state transition matrix P_l for (a), P_m for (b), and P_h for (c); (ii) freight size $Q = 40000lb$; (iii) per-pound reward $c_r = \$1.5/lb$.

freight status in transit and what actions to take, based on the data from the in-transit monitoring, that optimally increase expected supply chain productivity.

Acknowledgement

This research in this chapter is based upon work supported by the U.S. Department of Homeland Security under Grant Award Number 2007-ST-061-000003-02 through a grant awarded by the National Center for Food Protection and Defense at the University of Minnesota. The views and conclusions contained in this chapter should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security or the National Center for Food Protection and Defense.

CHAPTER IV

INVENTORY REPLENISHMENT CONTROL UNDER SUPPLY UNCERTAINTY

4.1 Introduction

Models of sequential decision making under uncertainty have been widely applied to a variety of inventory control problems. Many of these models are such that optimal order policies exist with structures that are easy to understand, implement, and/or enable efficient computation. For example, when demand is random and replenishment can be selected precisely, then there is a broad class of single product inventory control models having optimal (s, S) policies [53, 48]. Structured policy results have been investigated when replenishment decisions result in random replenishment (e.g., lot sizing with random yields) and the state space is continuous. Primary focus has been on stochastically proportional yield models where if Q is the replenishment decision, then QU replenishment results, where U is a random variable independent of Q [27, 67, 19].

In this chapter, we consider a discrete state and action infinite horizon, expected total discounted cost Markov decision process model of a single product, periodic review inventory problem with lost-sales, deterministic demand and random yield. We model random yield by the conditional probability $P(\alpha|a)$, where a is the number of items ordered and $\alpha \leq a$ is the number of items that are acceptable to add to the inventory. As our interest in this problem has been motivated by perishable inventory replenishment, we think of $(a - \alpha)$ as the amount of goods ordered that perish (or spoil) before inventory replenishment occurs. If d is demand and x is the current inventory level, then $z = d - x$ represents the amount of items required to satisfy demand prior to placement of an order. If $P(a|a) = 1$, i.e., if there is no spoilage, then an optimal inventory decision at each decision epoch is to order z. More generally, when yield is described by the conditional probability $P(\alpha|a)$, we present conditions that guarantee the existence of an optimal policy δ^* such that $\delta^*(z) = 0$ for

$z \leq 0$, and for $z \geq 0$, $\delta^*(z) \geq z$ and $(\delta^*(z) - z)$ is monotonically non-decreasing. Thus, $\delta^*(z) - z$ can be described as a staircase function in z and hence has a simple parametric characterization. We also show that this policy structure can reduce the computational demands of a single iteration of value iteration and of both the policy improvement and policy evaluation steps in policy iteration.

4.2 Problem Formulation

Let $x(t) \in \mathcal{S} = \{0, 1, 2, \ldots\}$ be the number of items of a single product in inventory at the beginning of period t. Based on this inventory level, the decision maker orders $a(t) \in \mathcal{A} = \{0, 1, 2, \ldots\}$ items, of which $\alpha(t) \leq a(t)$ items are immediately added to the inventory with probability $P(\alpha(t)|a(t))$, where $\sum_{\alpha(t)=0}^{a(t)} P(\alpha(t)|a(t)) = 1$. We can think of $a(t) - \alpha(t)$ as the number of ordered items that perish before inventory replenishment occurs. We assume that once an item is placed in inventory, it will not perish. We assume that there is no backlogging and that single period demand d is deterministic and stationary. Thus,

$$x(t+1) = \max\Big\{0, x(t) + \alpha(t) - d\Big\}$$

with probability $P(\alpha(t)|a(t))$.

Let $c > 0$ be the wholesale price, $h > 0$ the single period holding cost, and $p > 0$ the retail price for each item. For notational simplicity, we drop explicit dependence on t. Then, the single period expected total cost is

$$ca + h \sum_{\alpha} (x + \alpha) P(\alpha|a) - p \sum_{\alpha} \min\{x + \alpha, d\} P(\alpha|a) \tag{9}$$

where we assume that the retailer pays the wholesaler on the basis of the number of items ordered, perished items are discarded before any holding costs are accrued, and perished goods have no retail value (see Appendix D.1).

A *policy* maps the set of current inventory levels into the set of replenishment actions $\mathcal{A}$. The problem *objective* is to determine a policy, an *optimal* policy, that minimizes the expected total discounted cost over the infinite horizon. Further details can be found in Chapter 6 of Puterman [50]. Proofs of all results are presented in Appendix A.2

4.3 Preliminary Results

The optimality equation, given discount factor $\beta \in [0, 1)$, is

$$\bar{v}(x) = \min_{a \geq 0} \Bigg\{ ca + h \sum_{\alpha} (x + \alpha) P(\alpha|a) - p \sum_{\alpha} \min\{x + \alpha, d\} P(\alpha|a) \\ + \beta \sum_{\alpha} P(\alpha|a) \bar{v}(\max\{0, x + \alpha - d\}) \Bigg\}.$$

Let $z = d - x$, the amount of items required to satisfy demand prior to placement of an order, which we will use as our state. Note that $z(t+1) = d - x(t+1) = d - \max\{0, x(t) + \alpha(t) - d\} = d - \max\{0, \alpha(t) - z(t)\}$. We observe that $z(t) \in \mathcal{Z} = \{\ldots, -1, 0, 1, \ldots, d\}$. In addition, we note that $\min\{y, d\} = y - \max\{0, y - d\}$. Let $Y_a \in \{0, 1, \ldots, a\}$ be the number of acceptable items, given that a number of items are ordered. Denote $\mathcal{Y}(a) = \mathbf{E}[Y_a] = \sum_{\alpha} \alpha P(\alpha|a)$, and define $v(z) = \bar{v}(x) + (p - h)x$. Thus, $v(z) = \bar{v}(x) - (p - h)z + d(p - h)$. Straightforward algebraic manipulation implies that the optimality equation can be expressed as $v = Hv$, where for $\bar{p} = p - \beta(p - h)$,

$$[Hv](z) = \min_{a \geq 0} h(z, a, v),$$

$$h(z, a, v) = f(z, a) + \beta \sum_{\xi} P(\xi|z, a) v(\xi),$$

$$P(\xi|z, a) = \begin{cases} \sum_{\alpha=0}^{z} P(\alpha|a) & \text{if } \xi = d \\ P(z + k|a) & \text{if } \xi = d - k, k > 0 \end{cases},$$

$$f(z, a) = ca - (p - h)\mathcal{Y}(a) + \bar{p} \sum_{\alpha} P(\alpha|a) \max\{0, \alpha - z\}.$$

We note that $\sum_{\xi} P(\xi|z, a) v(\xi) = \sum_{\alpha} P(\alpha|a) v(d - \max\{0, \alpha - z\})$. Results in Puterman [50] imply that there exists a unique solution to $v = Hv$; let v^* represent this solution. Then, $\bar{v}(x) = v^*(z) - (p - h)x$ represents the minimum expected total infinite horizon discounted cost. Furthermore, let v_0 be any bounded, real-valued function on $\mathcal{Z}$, and define the sequence $\{v_n\}$ by $v_{n+1} = Hv_n$. Then, $\lim_{n \to \infty} \| v_n - v^* \| = 0$, where $\| \cdot \|$ is the supremum norm. Further, an action selection rule that causes the minimization in Hv^* to be attained, as a function of z, is a stationary policy $\delta^* : \mathcal{Z} \to \mathcal{A}$ that is an optimal policy. We remark that finding a policy δ^*, which is the problem objective, requires the determination of v^*. Thus, we seek both v^* and an δ^*.

4.4 Structured Results

We now present conditions that gurantee that there exists an optimal policy δ^* such that

(a) $\delta^*(z) = 0$ if $z \leq 0$,

(b) $\delta^*(z) \geq z \geq 0$, and

(c) $\delta^*(z) - z$ is isotone (monotonically non-decreasing) for $z \geq 0$,

following the proof of a structural property satisfied by v^*.

We remark that if $P(a|a) = 1$, and hence there is no spoilage with probability 1, an optimal policy would be $\delta(z) = z$. Thus, $\delta^*(z) - z \geq 0$ represents the additional items ordered to compensate for spoilage. The isotonicity of $\delta^*(z) - z$ implies that the more items ordered, the more items we would expect would perish. (This assumes, as we do in this paper, that all orders are treated identically, irrespective of order size. If large orders receive special treatment, relative to small orders, then $\delta^*(z) - z$ may not be isotone in z.) We note that if $\delta^*(z) - z$ is isotone, then $\delta^*(z)$ is isotone, but not conversely, since the isotonicity of $\delta^*(z) - z$ is equivalent to $\delta^*(z+1) \geq \delta^*(z) + 1$.

We begin by proving that v^* is antitone, or equivalently, the more inventory in stock, the greater the minimum expected total infinite horizon discounted cost.

Proposition 10 *If v is antitone, then Hv is antitone and therefore, v^* is antitone.*

We now consider the $z \leq 0$ case. Define A3, Assumption 3, as follows:

A3 : $ca + (\overline{p} - (p - h))\mathcal{Y}(a)$ *is non-negative for all* $a \geq 0$.

Proposition 11 *Assume $z \leq 0$, v is antitone, and A3. Then, for all $a \geq 0$, $h(z, a, v) \geq h(z, 0, v) = \overline{p}|z| + \beta v(d - |z|)$, and hence $\delta^*(z) = 0$.*

We remark that A3 may not hold but that $h(z, a, v^*) - h(z, 0, v^*) \geq 0$ for all $z \leq 0$ and $a \geq 0$ due to the antitonicity of v^* and that fact that for $z \leq 0$,

$$h(z, a, v^*) - h(z, 0, v^*) = ca + (\overline{p} - (p - h))\mathcal{Y}(a) + \beta \sum_{\alpha} P(\alpha|a)\Big[v^*(d - |z| - \alpha) - v^*(d - |z|)\Big].$$

Determination of an easy-to-compute, more easily satisfied A3 that takes into account the non-negativity of $v^*(d-|z|-\alpha) - v^*(d-|z|)$ is a topic of continuing interest.

We now present conditions that guarantee that there exists an optimal policy δ^* such that $\delta^*(z) \geq z$ when $z \geq 0$. Let

A4 : $ca - (p-h)\mathcal{Y}(a) \geq cz - (p-h)\mathcal{Y}(z)$ *for all* $a \leq z$.

Proposition 12 *Assume* $z \geq 0$ *and A4. Then, any element in argmin*$\{h(z,a,v)\}$ *is bounded below by* z.

We remark that A4 is a reasonably robust assumption, and that A3 and A4 can hold simultaneously under reasonable assumption (e.g., $c = p/2$, $h = p/20$, $\beta = 0.9$, and $\mathcal{Y}(a)/a = 0.6$ for all a, a ratio not uncommon in semiconductor wafer manufacturing and fresh produce).

We now give conditions that guarantee the existence of an optimal policy δ^* such that $(\delta^*(z) - z)$ is isotone for $z \geq 0$. Let $q(k|a) = \sum_{\alpha \geq k} P(\alpha|a)$,

A5 : $ca - (p-h)\mathcal{Y}(a)$ *is antitone in* a.

A6 : $\mathcal{Y}(a) - \mathcal{Y}(a+1)$ *is antitone in* a.

A7 : $q(k|a) - q(k|a+1)$ *is isotone in* a *for all* k.

Proposition 13 *Assume A5,A6, and A7. Then, there exists an optimal policy* δ^* *such that* $\delta^*(z+1) \geq \delta^*(z) + 1$ *for all* $z \geq 0$.

As observed earlier, A5 implies A4. Lemma 4.7.2 in Puterman [50] and A7 imply that $\mathcal{Y}(a) - \mathcal{Y}(a+1)$ is isotone in a. Noting A6, it follows that $\mathcal{Y}(a) - \mathcal{Y}(a+1)$ is a constant and hence the concomitant random variable has the proportional-yield-in-expectation property [21]. We remark that uniformly and binomially distributed discrete random variables possess this property and also satisfy A5.

4.5 Parametric Characterization of δ^*

We have presented conditions that guarantee the existence of an optimal policy δ^* such that

(i) $\delta^*(z) = 0$ if $z \leq 0$,

(ii) $\delta^*(z) \geq z$ if $z \geq 0$, and

(iii) $\delta^*(z) - z$ is isotone if $z \geq 0$.

Since both state and action spaces are discrete, the graph of $\delta^*(z) - z$ versus z will have the shape of a staircase, which has the following simple parametric characterization.

Definition 1 A policy $\delta : \mathcal{Z} \to \mathcal{A}$ is a *staircase* (SC) policy if and only if:

(i) there is a non-negative integer L,

(ii) there are states z^l, $l = 0, \ldots, L$ such that $0 \leq z^0, z^L = d$, and $z^l < z^{l+1}$ for $l = 0, \ldots, L-1$, and

(iii) there are integers γ^l, $l = 1, \ldots, L$ such that $0 < \gamma^l < \gamma^{l+1}$ for $l = 1, \ldots, L-1$

such that

(i) $\delta(z) = 0$ for $z \leq 0$,

(ii) $\delta(z) - z = 0$ for $0 < z \leq z^0$,

(iii) $\delta(z) - z = \gamma^l$ for $z^{l-1} < z \leq z^l$, $l = 1, \ldots, L-1$, and

(iv) $\delta(z) - z = \gamma^L$ for $z^{L-1} < z \leq z^L = d$.

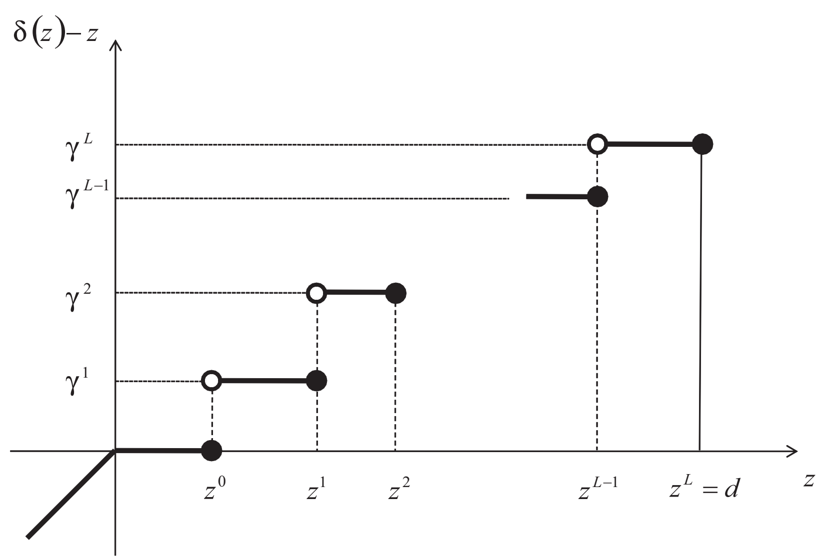

Figure 10: Graph of an SC-policy

Thus, we have determined conditions guaranteeing that the intersection of the set of all SC-policies and the set of all optimal policies is non-null. Hence, we can restrict our search for an optimal policy to a search for an optimal L, $\{z^l, l = 0, \ldots, L\}$, and $\{\gamma^l, l = 1, \ldots, L\}$.

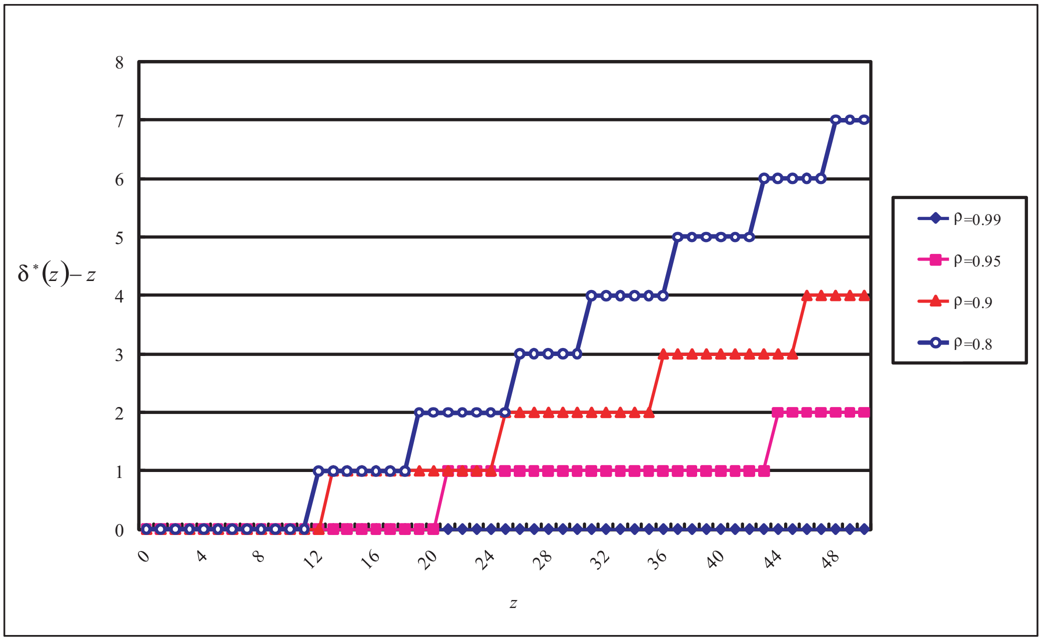

Figure 11: Graph of SC-policies for binomially distributed $P(\alpha|a)$ with success probability ρ when $d = 50$

4.6 Computational Implications

Value iteration and policy iteration are two valuable computational procedures for the Markov decision process (see Sections 6.3 and 6.4 in Puterman [50]). We now show that the structure of an optimal policy described in the last section can reduce the computational demands of key steps in both value iteration and policy iteration.

4.6.1 Single-Step Value Iteration and Policy Improvement

Let the operator H_δ be defined as

$$[H_\delta v](z) = h(z, \delta(z), v).$$

For given function v, a value iteration step involves determining $v' = Hv$, and policy improvement involves determining a δ such that $H_\delta v = Hv$ which by Proposition 13, equivalently involves determining a concomitant L, $\{\gamma^l : l = 1, 2, \ldots, L\}$, and $\{z^l : l = 0, 1, \ldots, L\}$.

We note that γ^L and $v'(d)$ satisfy

$$v'(d) = \min_{a \geq d} h(d, a, v) = h(d, d + \gamma^L, v).$$

A recursive procedure to find L, $\{\gamma^l : l = 1, \ldots, L\}$, and $\{z^l : l = 0, 1, \ldots, L-1\}$ is as follows:

Data: *Antitone* v

Result: L, $\{\gamma^l : l = 1, \ldots, L\}$, and $\{z^l : l = 0, 1, \ldots, L\}$

Set $j \leftarrow 0$;

Find $\widetilde{\gamma}^j$ such that

$$v'(d) = \min_{a \geq d} h(d, a, v) = h(d, d + \widetilde{\gamma}^j, v);$$

Set $z \leftarrow d$ and $\widetilde{z}^j \leftarrow d$;

while $z > 0$ **do**

  $z \leftarrow z - 1$;

  Find $\delta(z)$ such that

$$v'(z) = \min_{z \leq a \leq z + \widetilde{\gamma}^j} h(z, a, v) = h(z, \delta(z), v);$$

  if $\delta(z) < z + \widetilde{\gamma}^j$ **then**

    Set $j \leftarrow j + 1$, $\widetilde{z}^j \leftarrow z$ and $\widetilde{\gamma}^j \leftarrow \delta(z) - z$;

  end

end

Set $L \leftarrow j$, $z^l \leftarrow \widetilde{z}^{L-l}$ for $l = 0, 1, \ldots, L$ and $\gamma^l \leftarrow \widetilde{\gamma}^{L-l}$ for $l = 1, \ldots, L$;

Algorithm 4: Finding L, $\{\gamma^l\}$ and $\{z^l\}$

We remark that the structure of an optimal policy leads to an increasingly reduced action space cardinality in determining the z^l, γ^l and v'.

4.6.2 Policy Evaluation

Assume that L, $\{\gamma^l : l = 1, \ldots, L\}$ and $\{z^l : l = 0, 1, \ldots, L\}$ are given, where $0 < \gamma^l < \gamma^{l+1}$ and $0 \leq z^0, z^l < z^{l+1}, z^L = d$. Define $\delta(\cdot)$ as follows:

$$\begin{aligned} \delta(z) &= z && \text{for } \; 0 \leq z \leq z^0 \\ \delta(z) &= z + \gamma^l && \text{for } \; z^{l-1} < z \leq z^l, \; l = 1, 2, \ldots, L. \end{aligned}$$

We seek the (unique) fixed point of H_δ, v_δ (i.e., $v_\delta = H_\delta v_\delta$). We consider this policy evaluation procedure under two possible cases – (i) $d \geq \gamma^L$ and (ii) $d < \gamma^L$.

<u>*Case when $d - \gamma^L \geq 0$*</u>

First, assume $d - \gamma^L \geq 0$. Let V be the $(\gamma^L + 1)$-vector $V = \text{col}\{v_\delta(d), v_\delta(d-1), \ldots, v_\delta(d - \gamma^L)\}$. Note that

(i) if $0 \leq z \leq z^0$, then

$$v_\delta(z) = f(z, z) + \beta v_\delta(d), \tag{10}$$

(ii) if $z^{l-1} < z \leq z^l$, then

$$v_\delta(z) = f(z, z+\gamma^l) + \beta \Big[\sum_{\alpha=0}^{z} P(\alpha | z+\gamma^l) v_\delta(d) + \sum_{\alpha=z+1}^{z+\gamma^l} P(\alpha | z+\gamma^l) v_\delta(d-\alpha+z) \Big]. \tag{11}$$

Thus, if V is known, then $v_\delta(z)$ is known for all $0 \leq z \leq d$. To calculate V, we can construct a $(\gamma^L + 1)$-vector $\mathcal{A}$ and a $(\gamma^L + 1) \times (\gamma^L + 1)$ stochastic matrix $\mathcal{P}$ such that $V = \mathcal{A} + \beta \mathcal{P} V$ and hence $V = (I - \beta \mathcal{P})^{-1} \mathcal{A}$. Details of its construction are presented in Appendix D.2.

Recalling that for $z \leq 0$, $v_\delta(z) = -\overline{p} z + \beta v_\delta(d + z)$, it is straightforward to show that if $-md \leq z \leq -(m-1)d$, then

$$v_\delta(z) = \overline{p} \left(\sum_{k=0}^{m-1} \beta^k |kd + z| \right) + \beta^m v_\delta(md + z). \tag{12}$$

Thus, given $v_\delta(z)$, $0 \leq z \leq d$, we can determine $v_\delta(z)$, for all $z \leq 0$.

<u>*Case when $d - \gamma^L < 0$*</u>

Second, assume $d - \gamma^L < 0$, and let $M > 0$ be such that $(M+1)d - \gamma^L \geq 0$ and $Md - \gamma^L < 0$. Then, V can be described as $V = \text{col}\{V_0, V_1, \ldots, V_M\}$, where

(i) $V_0 = \text{col}\{v_\delta(z), 0 \le z \le d\}$, a $(d+1)$-vector,

(ii) for $1 \le m < M$, $V_m = \text{col}\{v_\delta(z), -md \le z < -(m-1)d\}$, a d-vector, and

(iii) $V_M = \text{col}\{v_\delta(z), d - \gamma^L \le z < -(M-1)d\}$, a $(\gamma^L - Md)$-vector.

Let $\mathcal{A}_0$ and $\mathcal{P}_{0m}$, $0 \le m \le M$ be such that $V_0 = \mathcal{A}_0 + \beta \sum_{m=0}^{M} \mathcal{P}_{0m} V_m$ (see Equation (11)), where $\mathcal{A}_0$ is a $(d+1)$-vector, $\mathcal{P}_{00}$ is $(d+1) \times (d+1)$ matrix, $\mathcal{P}_{0m}$ is a $(d+1) \times d$ matrix for $1 \le m < M$, and $\mathcal{P}_{0M}$ is $(d+1) \times (\gamma^L - Md)$ matrix. Note that all elements in the matrix $\left[\begin{array}{c|c|c|c} \mathcal{P}_{00} & \mathcal{P}_{01} & \cdots & \mathcal{P}_{0M} \end{array}\right]$ are non-negative and each row sums to one.

Equation (12) indicates that for each $1 \le m \le M$, there is a vector $\mathcal{A}'_m$ and a matrix I'_m such that $V'_m = \mathcal{A}'_m + \beta^m I'_m V_0$. Note the $\mathcal{A}'_m$ is a d-vector for $1 \le m < M$, $\mathcal{A}'_M$ is a $(\gamma^L - Md)$-vector, I'_m is a $d \times (d+1)$ matrix for $1 \le m < M$, and I'_M is a $(\gamma^L - Md) \times (d+1)$ matrix. Further note that $I'_1 = \left[\begin{array}{c|c} 0 & I \end{array}\right]$ (a matrix with all zeros in the first column, where I is the identity matrix), $I'_m = I$ for $1 < m < M$, and $I'_M = \left[\begin{array}{c|c} I & 0 \end{array}\right]$. Then,

$$V_0 = \mathcal{A}_0 + \beta\Big[\mathcal{P}_{00} V_0 + \sum_{m=1}^{M} \mathcal{P}_{0m} \left(\mathcal{A}'_m + \beta^m I'_m V_0\right)\Big] = \Lambda + \beta \mathcal{P} V_0,$$

where $\Lambda = \mathcal{A}_0 + \beta \sum_{n=1}^{M} \mathcal{P}_{0m} \mathcal{A}'_m$ and $\mathcal{P} = \mathcal{P}_{00} + \sum_{m=1}^{M} \beta^m \mathcal{P}_{0m} I'_m$, which, by its construction, is substochastic. Thus, there exists a unique solution to $V_0 = (I - \beta\mathcal{P})^{-1}\Lambda$, which can be used to determine $v_\delta(z)$ for all z.

In summary, the overall procedure for identifying an optimal SC policy is presented in Algorithm 5.

Data: *Antitone* v^0

Set $j \leftarrow 0$;

repeat

(*Policy Improvement*) Given v^j, perform Algorithm 4, and obtain $\delta^j \equiv (L, \{\gamma^l\}, \{z^l\})$;

(*Policy Evaluation*) Given δ^j, evaluate $v_{\delta^j}(z)$ for all z based on the discussion in Section 4.6.2, and let $v^{j+1} \leftarrow v_{\delta^j}$;

$j \leftarrow j + 1$;

until $\max_z |v^j(z) - v^{j-1}(z)| < \epsilon$ *for a sufficiently small* ϵ ;

Algorithm 5: Procedure for identifying an optimal SC policy

4.6.3 Numerical Analysis

In this section, we briefly compare the performance of the proposed approach with dynamic programming. We use the following parameter values for the computational experiments: $c = 9$, $p = 20$, $h = 3$ and $\beta = 0.99$. We vary the size of demand d and compare a SC policy-based approach with a standard dynamic programming approach in terms of CPU times. Both approaches were implemented in Sun Java JDK 1.6.1, and their implementations were run on a LINUX machine with a Intel Xeon 2.66 GHz processor and 4GB of RAM.

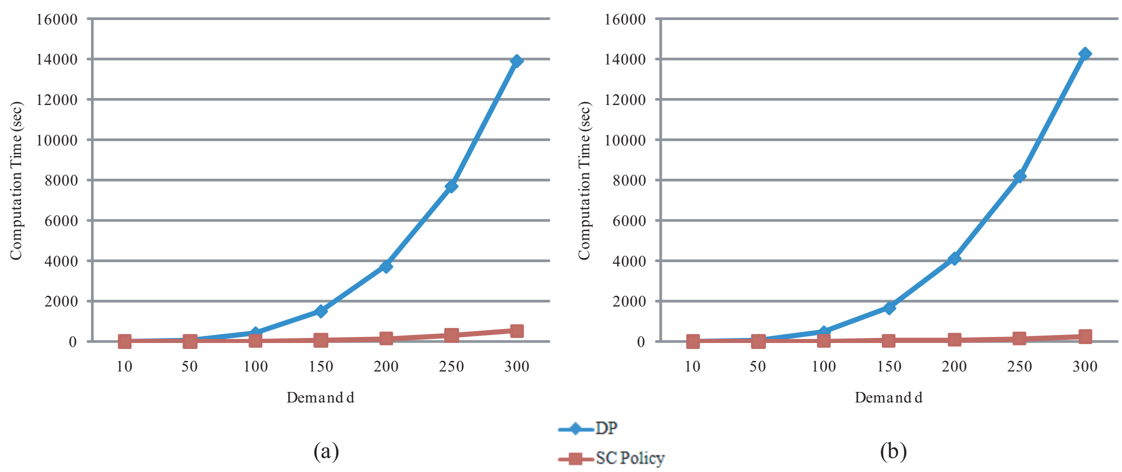

Figure 12: Comparisons of CPU times between dynamic programming and SC policy as demand d varies, assuming random yield is binomially distributed with success probability ρ: (a) $\rho = 0.8$ and (b) $\rho = 0.9$

Figure 12 shows that the solution approach developed in this section significantly reduces CPU times, relative to dynamic programming (DP). Specifically, the SC policy approach requires less than 5% of the CPU time required by DP to determine an optimal policy. Since DP evaluates every possible state, its CPU times grow exponentially as d increases (known as the *curse of dimensionality*). The solution procedure proposed in this section requires significantly less information (i.e., L, $\{\gamma^l\}$, and $\{z^l\}$) to identify an optimal policy since we limit the search for an optimal policy having a staircase structure. We remark that computational times seem insensitive to changes in the parameter ρ, as indicated in Figure 12.

We now investigate how CPU times for the SC policy approach are sensitive to changes in d. Figure 13 presents such an example of CPU time changes in d. The figure indicates that, unlike DP approaches, the increasing rate of CPU times gradually decreases as d increases, and hence, for larger d, CPU times seem increasing almost linearly in d. Therefore, the SC policy-based solution approach outperforms DP approaches significantly, and its computational times increase moderately unlike DP approaches, which is desirable in terms of computational tractability.

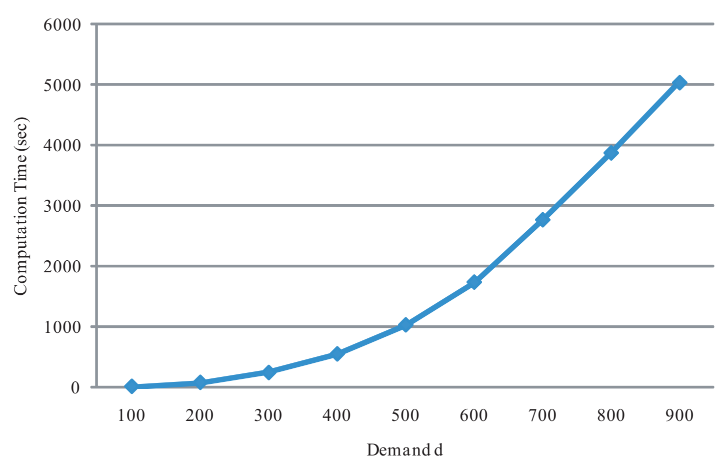

Figure 13: Changes of CPU time for the SC policy approach in d, assuming random yield is binomially distributed with success probability $\rho = 0.9$

4.7 Summary

Chapter 4 has investigated a discrete state, discrete decision epoch inventory replenishment control problem under supply uncertainty. We have assumed that there is no backlogging, the single period demand d is deterministic and stationary, and once an item is placed in inventory, it will not perish. By letting $z = d - x$ as a state where x is the inventory level at the beginning of each period in a MDP model with infinite-horizon total expected discounted costs, we have presented conditions that guarantee that an optimal replenishment policy δ^* is such that $\delta^*(z) = 0$ for $z \leq 0$, $\delta^*(z) \geq z \geq 0$, and $\delta^*(z) - z$ is monotonically non-decreasing for $z \geq 0$. Such a "staircase" structure has a simple parametric description, which can help to significantly accelerate value iteration and policy iteration.

CHAPTER V

INVENTORY CONTROL WITH SUPPLY CHAIN YIELD INFORMATION

5.1 Introduction

Firms often face the challenge of receiving a random amount of an order placed. As examples,

(a) In semiconductor manufacturing, the percentage of defective wafers produced can be significant and as high as 30% to 50% [18],

(b) The percentage of perishable food that perishes during storage and transit in the U.S. is typically 10% to 15% by weight, and the reduction of the total economic value of perishable food due to the reduction of food quality caused by storage and transit is typically 25% to 50% [42].

In order to mitigate this type of risk, firms typically order or manufacture more than is needed. The challenge is to order just the right amount to insure that an adequate amount of the desired product is on hand at the right time. This challenge is exacerbated if the length of time needed to deliver an order (lead time) is also uncertain.

In this chapter, we consider a model of inventory replenishment and delivery where both lead time and the amount of the order delivered are uncertain. The firm periodically places orders to a supplier (the *origin*, 0) for delivery to a plant or retail outlet (the *destination*, N). There are at most a finite number of in-transit states $n \in \{1, 2, \ldots, N-1\}$, which we call the *supply system*. We assume the stochastic lead time model proposed by Song and Zipkin [56] for the case where the number of in-transit states is finite and where the supply system is governed by an exogenous Markov chain. Further, we assume that the product is subject to shrinkage as it travels from origin to destination. Our objective is to place orders so as to minimize expected total discounted cost over the infinite horizon, where the usual fixed order, wholesale, and inventory holding costs and backlogging penalties apply.

This chapter is organized as follows: Section 5.2 introduces an infinite-horizon Markov decision process model for the multiple-staged inventory control problem under yield and lead time uncertainties. In Section 5.3, we present conditions where there exists a sufficient statistic for the problem introduced in the previous section by assuming that random shrinkage can occur from the origin to the supply system or destination, shrinkage is deterministic within the supply system and from the supply system to the destination, and no shrinkage occurs once goods reach the destination. We summarize our results in Section 5.4.

5.2 Problem Statement and Preliminaries

Consider the following periodic review inventory and supply system problem. An order placed at decision epoch t, $a(t)$, is transported from the origin (0) to the destination (N) through the supply system as follows. Goods at location $k \in \{0, 1, \ldots, N-1\}$ at time t will move to location $M(k|i(t), e(t)) \in \{1, 2, \ldots, N\}$ during period $(t, t+1)$, where:

(a) The $\{e(t)\}$ are independent, identically distributed random variables,

(b) $\{i(t), t = 0, 1, \ldots\}$ is an exogenous Markov process such that $i(t+1) = Q(i(t), e(t))$.

The process $\{i(t), t = 0, 1, \ldots\}$ serves as a surrogate for anything that can affect lead time and/or freight shrinkage (e.g., weather, labor strikes, congestion).

With regard to the function M, and consistent with assumptions made in Song and Zipkin [56], we assume that for each pair (i, e):

(a) $k < k'$ implies $M(k|i, e) \leq M(k'|i, e)$. Hence, "cross-over" is not permitted,

(b) $M(N|i, e) = N$. Thus, once products reach the destination, they never move back into the supply system.

We now describe our model of shrinkage as the product proceeds from origin to destination. We remark that no shrinkage occurs in the model considered by Song and Zipkin [56]. Let $g_k^{n-1}(x|i, e)$ be the amount of goods that moves to in-transit state n from in-transit state k during period $(t, t+1)$, assuming the amount of goods at in-transit state k

at epoch t is x, $i(t) = i$, and $e(t) = e$. It is reasonable to assume that for all k, n, i, and e, $0 \leq g_k^{n-1}(x|i, e) \leq x$. We further assume that no shrinkage occurs once the destination is reached, and hence $g_N^N(x|i, e) = x$ for all (i, e).

We now present the dynamic equations for the system. Let $\mathcal{M}(n|i, e) = \{k|M(k|i, e) = n\}$, and let $x_n(t)$ represent the number of items at in-transit state $n \geq 1$ at epoch t. Then, for $1 \leq n < N$,

$$x_n(t+1) = \sum_{k \in \mathcal{M}(n|i,e)} g_k^{n-1}(x_k(t)|i, e),$$

for $i(t) = i$ and $e(t) = e$, where $x_0(t) = a(t)$. At the destination,

$$x_N(t+1) = x_N(t) - D(t) + \sum_{k \in \mathcal{M}(N|i,e)} g_k^{N-1}(x_k(t)|i, e),$$

for $i(t) = i$, and $e(t) = e$, and where $\{D(t)\}$ are stationary, independent, and identically distributed random variables that represent demand.

Let $x(t) = \{x_n(t), 1 \leq n \leq N\}$, and assume the function f is such that $x(t+1) = f\big(x(t), i(t), a(t), e(t), D(t)\big)$.

Let $c > 0$ be the wholesale cost per item, $h > 0$ be the holding cost per item per period for an item not lost to shrinkage, and $\sigma > 0$ be the backlogging penalty per item. Let $K > 0$ be the fixed order cost, and $\delta(a) = 1$ (0) if $a > 0$ $(a = 0)$. Then, the cost accrued during period $(t, t+1)$ is:

$$\begin{aligned} &K\delta(a) + ca + h \sum_{1 \leq n < N} x_n(t+1) + h(x_N(t+1))^+ - \sigma(x_N(t+1))^- \\ &\qquad = K\delta(a) + ca + h \sum_{1 \leq n \leq N} x_n(t+1) - (h+\sigma)(x_N(t+1))^-, \end{aligned}$$

where $x^+ = \max\{0, x\}$ and $x^- = \min\{0, x\}$. Let $c(x, i, a)$ be the expected cost accrued during period $(t, t+1)$, given $x = x(t)$, $i = i(t)$, and $a = a(t)$.

We assume the decision maker selects $a(t)$, knowing the realizations of the random variable $x(t)$ and $i(t)$, but before knowing the realization of the random variables $D(t)$, $e(t)$, and hence $g_k^{n-1}(x_k(t)|i(t), e(t))$ for all n and k. Thus, a policy determines $a(t)$, given $x(t)$ and $i(t)$. The criterion is the expected total discounted cost over the infinite horizon. The problem objective is (i) to determine a policy, an *optimal* policy, that minimizes the criterion and (ii) to determine the value of the criterion generated by an optimal policy.

The optimality equation is

$$v(x,i) = [Hv](x,i) = \min_a \Big\{ c(x,i,a) + \beta \mathbb{E}\Big[v(f(x,i,a,e,D), Q(i,e)) \Big] \Big\},$$

where the expectation operator $\mathbb{E}$ is with respect to the random variables (e, D), and $\beta \in [0, 1)$ is the discount factor. Results in Puterman [50] guarantee the existence of a unique solution to the optimality equation, this solution is the value of the criterion generated by an optimal policy, and a policy that causes the minimum to be obtained in the optimality equation is an optimal policy. Further, for bounded v_0, the sequence $\{v_n\}$ is such that $\lim_{n\to\infty} \parallel v_n - v^* \parallel = 0$, where $\parallel \cdot \parallel$ is the supremum norm, v^* is the unique fixed point of the operator H, and $v_{n+1} = Hv_n$. Thus, we seek both v^* and a policy that attains the minimum in Hv^*. Proofs of all results are presented in Appendix A.3.

5.3 Yield Model

One of the key results in Song and Zipkin [56] is that the pipeline inventory position, $\sum_{n=1}^{N} x_n$, is a sufficient statistic for this problem. That is, it is sufficient to know only the sum of the inventory in-transit plus the inventory at the destination (a scalar), rather than the entire vector x, in order to select an optimal inventory replenishment level. We now extend this result by showing that a natural extension of the pipeline inventory position is a sufficient statistic when shrinkage is assumed deterministic for goods moving within the supply system and moving from the supply system to the destination.

More precisely, we assume throughout this section that:

(a) $g_k^{n-1}(x_k|i,e)$ is independent of (i,e) for all $k \geq 1$ and all n (i.e., $g_k^{n-1}(x_k|i,e) = g_k^{n-1}(x_k)$),

(b) $g_k^{n-1}(x_k) = g_m^{n-1}\left(g_k^{m-1}(x_k)\right)$ for all $k \geq 1$, m and $n (> m)$,

and remark that under these assumptions $g_0^{n-1}(a|i,e)$, for all $n \geq 1$, can remain dependent on (i,e). We further assume that $g_0^{n-1}(a|i,e) = g_m^{n-1}(g_0^{m-1}(a|i,e))$. The extension of the pipeline inventory position that will be of interest is

$$\mathcal{I}(x) = \sum_{n=1}^{N-1} g_n^{N-1}(x_n) + x_N,$$

and we call it *effective pipeline inventory position.*

We begin by showing that $\mathcal{I}$ has an important property under the following assumptions:

A8 :

(i) $g_k^{n-1}(x'+x'') = g_k^{n-1}(x') + g_k^{n-1}(x'')$ for all $1 \leq k < n \leq N$.

(ii) $|\mathcal{M}(n|i,e)| \leq 1$ for all $n < N$ and all (i,e).

We remark that the deterministic version of the multiplicative random yield model can be an example of A8 (i). A8 (ii) indicates that each in-transit location is the destination over a single period of goods from at most one other in-transit location, and synchronous production lines in which all the movements of jobs are coordinated serves as such an example.

Lemma 2 *Suppose that A8 (i) or (ii) holds. Then,* $\mathcal{I}(f(x,i,a,e,D)) = g_0^{N-1}(a|i,e) + \mathcal{I}(x) - D$ *for all* (i,e).

We now present several preliminary results and definitions. Let $\mathcal{M}^l(n|i,e_l) = \{k : M^l(k|i,e_l) = n\}$, where $e_l = \{e(0),\ldots,e(l-1)\}$, $M^l(n(0)|i(0),e_l) = n(l)$, $n(k+1) = M(n(k)|i(k),e(k))$, $i(k+1) = Q(i(k),e(k))$, and where $i(0)$ and $n(0)$ are given. Let $M^l(n|i)$ be the l-step movement random variable where $M^1 = M$. Then, $L_N(i) = \min\{l : M^l(0|i) = N, l \geq 1\}$, which is a random variable representing the lead time for an order placed, given that i is the supply system status when the order is placed. Let

$$
\begin{aligned}
v'(x,i) &= \mathbb{E}_{L_N}\Big[\sum_{j=0}^{L_N-2}\beta^j h \mathbb{E}_{e_{j+1},D_{j+1}}\Big\{\sum_{n=1}^{N}\sum_{k\in\mathcal{M}^{j+1}(n|i,e_{j+1})} g_k^{n-1}(x_k) + x_N - D_{j+1}\Big\}\Big] \\
v''(x,i) &= -\mathbb{E}_{L_N}\Big[\sum_{j=0}^{L_N-2}\beta^j (h+\sigma)\mathbb{E}_{e,D_{j+1}}\Big\{\Big(\sum_{l=1}^{j+1}\sum_{k\in\mathcal{M}^{l}(N|i,e_l)} g_k^{N}(x_k) + x_N - D_{j+1}\Big)^-\Big\}\Big]
\end{aligned}
\tag{13}
$$

$$
\begin{aligned}
c^\circ(\mathcal{I}(x),i,a) = K\delta(a) + ca + h\mathbb{E}_e\Big[\sum_{k=1}^{N-1}\sum_{0\in\mathcal{M}(k|i,e)} g_0^{k-1}(a|i,e)\Big] \\
+ \mathbb{E}_{L_N}\Big[\sum_{j=1}^{L_N-2}\beta^j h\mathbb{E}_{e,e_j}\Big\{\sum_{n=1}^{N-1}\sum_{k\in\mathcal{M}^j(n|i,e_j)} g_k^{n-1}\Big(\sum_{0\in\mathcal{M}(k|i,e)} g_0^{k-1}(a|i,e)\Big)\Big\}\Big] \\
+ \mathbb{E}_{L_N}\Big[\beta^{L_N-1}h\mathbb{E}_{e,D_{L_N}}\Big\{g_0^N(a|i,e) + \mathcal{I}(x) - D_{L_N}\Big\}\Big] \\
- \mathbb{E}_{L_N}\Big[\beta^{L_N-1}(h+\sigma)\mathbb{E}_{e,D_{L_N}}\Big\{\Big(g_0^N(a|i,e) + \mathcal{I}(x) - D_{L_N}\Big)^-\Big\}\Big],
\end{aligned}
$$

where $L_N = L_N(i)$. In addition, the explicit expression for $c(x,i,e)$ is

$$c(x,i,e) = K\delta(a) + ca + h\mathbb{E}_e\Big[\sum_{n=1}^{N-1} \sum_{0\in\mathcal{M}(n|i,e)} g_0^{n-1}(a|i,e)\Big]$$
$$+ h\mathbb{E}_{e,D}\Big[\sum_{n=1}^{N} \sum_{k\in\mathcal{M}(n|i,e)} g_k^{n-1}(x_k) + x_N - D\Big]$$
$$- (h+\sigma)\mathbb{E}_{e,D}\Big[\Big(\sum_{k\in\mathcal{M}(N|i,e)} g_k^{N}(x_k) + x_N - D\Big)^{-}\Big].$$

We now present a preliminary result that is useful in the proof of Proposition 14, the main result of this chapter.

Lemma 3 *It follows that:*

(a)

$$h\mathbb{E}_{e,D}\Big[\sum_{n=1}^{N} \sum_{k\in\mathcal{M}(n|i,e)} g_k^{n-1}(x_k) + x_N - D\Big] + \beta\mathbb{E}_{e,D}\Big[v'(f(x,i,a,e,D),Q(i,e))\Big]$$
$$= \mathbb{E}_{L_N}\Big[\sum_{j=1}^{L_N-2} \beta^j h\mathbb{E}_{e,e_j}\Big\{\sum_{n=1}^{N-1} \sum_{k\in\mathcal{M}^j(n|i,e_j)} g_k^{n-1}\Big(\sum_{0\in\mathcal{M}(k|i,e)} g_0^{k-1}(a|i,e)\Big)\Big\}\Big]$$
$$+ \mathbb{E}_{L_N}\Big[\beta^{L_N-1} h\mathbb{E}_{e,D_{L_N}}\Big\{g_0^N(a|i,e) + \mathcal{I}(x) - D_{L_N}\Big\}\Big] + v'(x,i).$$

(b)

$$-(h+\sigma)\mathbb{E}_{e,D}\Big[\Big(\sum_{k\in\mathcal{M}(N|i,e)} g_k^N(x_k) + x_N - D\Big)^{-}\Big] + \beta\mathbb{E}_{e,D}\Big[v''(f(x,i,a,e,D),Q(i,e))\Big]$$
$$= -\mathbb{E}_{L_N}\Big[\beta^{L_N-1}(h+\sigma)\mathbb{E}_{e,D_{L_N}}\Big\{\Big(g_0^N(a|i,e) + \mathcal{I}(x) - D_{L_N}\Big)^{-}\Big\}\Big] + v''(x,i).$$

(c)

$$c(x,i,a) + \beta\mathbb{E}_{e,D}\Big[v'(f(x,i,a,e,D),Q(i,e))\Big] + \beta\mathbb{E}_{e,D}\Big[v''(f(x,i,a,e,D),Q(i,e))\Big]$$
$$= c^{\circ}(\mathcal{I}(x),i,a) + v'(x,i) + v''(x,i).$$

Define the operator H° as

$$[H^{\circ}v](x,i) = \min_a \Big\{c^{\circ}(\mathcal{I}(x),i,a) + \beta\mathbb{E}\Big[v(f(x,i,a,e,D),Q(i,e))\Big]\Big\}.$$

Let v°^*} be the fixed point of H°, $\{v_n\}$ be the sequence $v_{n+1} = Hv_n$ for $v_0 = v' + v''$, and $\{v_n^\circ\}$ be the sequence $v_{n+1}^\circ = H^\circ v_n^\circ$, for $v_0^\circ = 0$. We now show finding v°^*} is equivalent to finding v^*. We now present the main result of the chapter.

Proposition 14 *Suppose that A8 holds. Then,* $v_n = v_n^\circ + v' + v''$ *for all* n*. Hence,* $v^* = v^{\circ^*} + v' + v''$*.*

We remark that v_0° trivially depends on x only through $\mathcal{I}(x)$. Assume $v_n^\circ(x,i) = v_n^\circ(\mathcal{I}(x),i)$. Then, from Lemma 3,

$$v_n^\circ\left(f(x,i,a,e,D), Q(i,e)\right) = v_n^\circ\left(g_0^N(a|i,e) + \mathcal{I}(x) - D, Q(i,e)\right)$$

for all (i,e). Since $c^\circ(\mathcal{I}(x),i,a)$ depends on $\vec{x}$ only through $\mathcal{I}(x)$, $[H^\circ v_n^\circ](x,i)$ depends on x only through $\mathcal{I}(x)$. Thus, v_{n+1}° and the decision rule that causes the minimum in $[H^\circ v_n^\circ](x,i)$ to be attained both depends on x only through $\mathcal{I}(x)$. An induction argument then can conclude that v°^*} depends on x only through $\mathcal{I}(x)$, and that there exists an optimal policy that depends on x only through $\mathcal{I}(x)$. Thus, $\mathcal{I}(x)$ is a sufficient statistic. Of further interest is the fact that the solution of the problem associated with the operator H° is guaranteed to have an optimal policy that has an (s,S) structure. Thus, there is an s and S $(s < S)$ such that it is optimal to order up to S once $\mathcal{I}(x)$ falls below s and not to order otherwise.

5.4 Summary

Chapter 5 has considered a periodic review inventory and supply system with an origin, a destination, and a supply system composed of a finite number of intermediate locations between the origin and the destination. We have assumed that the model of the supply system is the model proposed by Song and Zipkin [56] for the case where there are a finite number of intermediate locations. Further, we have assumed that random shrinkage can occur from the origin to the supply system or destination, shrinkage is deterministic within the supply system and from the supply system to the destination, and no shrinkage occurs once goods reach the destination. We have shown that under these conditions, the effective pipeline inventory position is a sufficient statistic, extending the results of Song and Zipkin [56].

CHAPTER VI

CONCLUSION AND FUTURE RESEARCH

This dissertation has focused on two related issues of importance in supply chain control, the value of information and supply uncertainty. With regard to supply uncertainty, we have investigated issues surrounding uncertain quantity and/or quality with instantaneous delivery and with uncertain lead-times.

Chapter 2 investigated the value of dynamically determining a tour for the TSP, based on current network conditions and assuming network arc costs are random variables. A numerical study has shown the potential for significant reduction in expected total travel costs due to dynamic tour determination, relative to two benchmarks. The first of the benchmarks reflected the current practice in industry of constructing a fixed tour prior to departure from the starting point based on known distances and speed limits; the second benchmark represented the optimal tour within the class of all fixed tours determined prior to departure from the starting point, based on historic travel time data.

Due to the limitations of DP for solving realistically sized DTSPs, we transformed the DTSP into an AND/OR graph and applied the best-first heuristic search algorithm AO* for optimal policy determination. This application proved capable of solving realistically sized problems (e.g., for LTL PUD) fast enough for operational purposes on a laptop with standard configuration. We note, however, that the DTSP represents a formidable computational challenge, and expanding the space of solvable, realistically sized problems would require further algorithmic development. Possible next steps in this regard include improving the current AO* implementation and examining other algorithms for optimal and sub-optimal (e.g, applications of approximate DP [49]) policy determination.

In Chapter 3, we have taken an initial step in analyzing the impact of being able to monitor perishable freight quality in transit by examining a model of a single vehicle transporting a perishable from a single origin to a single destination through several intermediate predetermined locations. Taking advantage of the special structure of the problem, we developed an efficient procedure for determining the optimal reward function that avoids the tractability issues typically associated with solution procedures for the general POMDP. We showed that the optimal reward function and an optimal policy have several intuitively appealing and implementable structural characteristics but provide a counter example to an intuitive claim regarding the structure of an optimal policy. We remark that structured policy results for POMDPs are a relatively unexplored area of research. We determined the value of having access to inspections at the intermediate locations and an upper bound on this value. These results can provide guidance when deciding whether to acquire the capacity to monitor freight status in transit, how to optimally extract value from inspection data, and how much value we can expect to extract. Several extensions of these results for more detailed models are topics for future research.

Chapter 4 presented conditions that guarantee the existence of an optimal policy having a staircase structure for a discrete state and action, single product periodic review inventory problem with random yield. We have used this structure to reduce the computational demands of key steps in both value iteration and policy evaluation. The numerical impact of this computational reduction appears encouraging. Future research will focus on problems involving multiple products, which quickly become intractable even if each product has a small state space, in order to understand the computational advantage of restricting attention to staircase policies for determining good suboptimal designs.

Chapter 5 discussed an inventory control model which includes the two aspects of supply-side uncertainties – uncertain lead-time and yield – under the assumption that the supply system is evolved in a Markovian fashion, and such the evolution of the system affects the two aspects. For the corresponding system with yield model presented in Section 5.3, we showed that the effective pipeline inventory position, which is a scalar, is indeed a sufficient statistic for the model in order to determine an optimal inventory replenishment level. Future research includes the development of good sub-optimal algorithm designs for the original problem based on the results from the deterministic yield model.

APPENDIX A

PROOFS

A.1 Proofs in Chapter 3

Before presenting proofs in detail, we begin by introducing several useful properties of a vector that is NIV:

Lemma 4 *(i) If $\vec{v}$ is NIV, then $\vec{v} + c\vec{e}$ is NIV for any constant c.*

(ii) If $\vec{v}$ and $\vec{v}'$ are both NIV, then $\vec{v} + \vec{v}'$ is NIV.

(iii) If $\vec{v}$ is NIV and $\vec{x} \prec \vec{x}'$, then $\vec{x}\vec{v} \leq \vec{x}'\vec{v}$.

(iv) If $\vec{v}$ is NIV, $\beta \geq 0$, and A2 holds, then $\beta P\vec{v}$ is NIV.

(v) If $\vec{v}^k$ is NIV for all $k = 1, 2, \ldots, K < \infty$, and $\vec{v}'$ is such that $\vec{v}'_s = \max_k\{\vec{v}^k_s\}$, then $\vec{v}'$ is NIV.

Proof: Proofs of (iii) and (iv) follow from Lemma 4.7.2 in Puterman [50]. Regarding (v), assume there is an $\tilde{s}$ such that $v'_{\tilde{s}} < v'_{\tilde{s}+1}$. Let k be such that $v'_{\tilde{s}+1} = v^k_{\tilde{s}+1}$. Then, $v^k_{\tilde{s}} \leq v'_{\tilde{s}} < v'_{\tilde{s}+1} = v^k_{\tilde{s}+1}$, which is a contradiction. ■

A.1.1 Proof of Proposition 3

By assumption, $\vec{m}(N, N, z)$ is NIV, and hence by Lemma 4(iii) and Proposition 2, $\vec{x} \prec \vec{x}'$ implies $\omega'(N, \vec{x}, z) \leq \omega'(N, \vec{x}', z)$. Assume $\vec{m}(\rho, n+1, z)$ is NIV for all $\rho \geq n+1$ and hence $\omega'(n+1, \vec{x}, z) \leq \omega'(n+1, \vec{x}', z)$. Due to Lemma 4(i) and (iv), $\vec{m}(\rho, n, z)$ is NIV. Moreover, $\vec{l}(n, z)$ is NIV due to assumption that $\vec{l}^x_n$ is NIV and Lemma 4(i). Therefore, Lemma 4(iii) again implies that $\vec{x}\vec{m}(\rho, n, z) \leq \vec{x}'\vec{m}(\rho, n, z)$ for $\rho \geq n$, and $\vec{x}\vec{l}(n, z) \leq \vec{x}'\vec{l}(n, z)$. The maximum of monotonically nondecreasing functions is monotonically nondecreasing. Hence, the results hold by induction.

A.1.2 Proof of Proposition 4

It follows from the definitions and assumption of $c(N-1,N) + \beta R_S \le \vec{e}_S \vec{l}(N-1,z) = l_S(N-1,z)$ that $\omega'(N-1,\vec{e}_S,z) = l_S(N-1,z)$. Assume $\omega'(n+1,\vec{e}_S,z) = l_S(n+1,z) \ge \vec{m}_S(\rho,n+1,z)$, $\rho = n+1,\ldots,N$. Then,

$$\omega'(n,\vec{e}_S,z) = \max\left\{\begin{array}{l} \vec{m}_S(\rho,n,z), \rho = n,\ldots,N \\ l_S(n,z) \end{array}\right\},$$

where for $\rho > n$,

$$\vec{m}_S(\rho,n,z) = c(n,n+1) + \beta\vec{m}_S(\rho,n+1,z) \le c(n,n+1) + \beta l_S(n+1,z) \le l_S(n,z)$$

due to the induction hypothesis, Assumption A1, and $p_{SS} = 1$.

For $\rho = n$,

$$\begin{aligned}
\vec{m}_S(n,n,z) &= c(n,n+1) + M + \beta\omega'(n+1,\vec{e}_S,z) \\
&= c(n,n+1) + M + \beta\max\left\{\begin{array}{l} l_S(n+1,z), \\ \vec{m}_S(\rho,n+1,z), \forall\rho \ge n+1 \end{array}\right\} \\
&= c(n,n+1) + M + \beta l_S(n+1,z) \le c(n,n+1) + \beta l_S(n+1,z) \le l_S(n,z).
\end{aligned}$$

Therefore, $\omega'(n,\vec{e}_S,z) = l_S(n,z)$, and proof follows by induction.

A.1.3 Proof of Proposition 5

We now show

(i) $\omega'(n,\vec{e}_0,z) - \omega'(n,\vec{e}_1,z) \le \gamma[\omega'(n+1,\vec{e}_0,z) - \omega'(n+1,\vec{e}_1,z)]$,

(ii) $\sigma(n,n,z) \le \gamma\sigma(n+1,n+1,z)$,

(iii) $\sigma(\rho,n,z) = \gamma\sigma(\rho,n+1,z)$ for $\rho > n$,

and then prove the result by induction. Result (ii) follows from (i) and the definition of $\sigma(n,n,z)$. It is straightforward to show (iii). We now show (i). It follows that

$$\begin{aligned}
\omega'(n,\vec{e}_0,z) &= \max\left\{\begin{array}{l} c(n,n+1) + \beta\omega'(n+1,[1-\alpha,\alpha],z) \\ c(n,n+1) + M + \beta[(1-\alpha)\omega'(n+1,\vec{e}_0,z) + \alpha\omega'(n+1,\vec{e}_1,z)] \end{array}\right. \\
&\le c(n,n+1) + \beta[(1-\alpha)\omega'(n+1,\vec{e}_0,z) + \alpha\omega'(n+1,\vec{e}_1,z)]
\end{aligned}$$

where the equality is due to assumption (iii) that $\omega'(n, \vec{e}_0, z) > l_0(n, z)$ $(\geq l_1(n, z))$ and the inequality follows from the convexity of $\omega'(n+1, \cdot, z)$ and the assumption $M \leq 0$. Thus,

$$\begin{aligned}
\omega'(n, \vec{e}_0, z) &\leq c(n, n+1) + \beta\omega'(n+1, \vec{e}_1, z) + \gamma[\omega'(n+1, \vec{e}_0, z) - \omega'(n+1, \vec{e}_1, z)] \\
&= c(n, n+1) + \beta l_1(n+1, z) + \gamma[\omega'(n+1, \vec{e}_0, z) - \omega'(n+1, \vec{e}_1, z)] \\
&\leq l_1(n, z) + \gamma[\omega'(n+1, \vec{e}_0, z) - \omega'(n+1, \vec{e}_1, z)] \\
&= \omega'(n, \vec{e}_1, z) + \gamma[\omega'(n+1, \vec{e}_0, z) - \omega'(n+1, \vec{e}_1, z)]
\end{aligned}$$

where the first and second equalities are due to Proposition 4, and the second inequality follows from Assumption A1.

We now prove the Proposition. It is straightforward to show that $\sigma(N-1, N-1, z) = \sigma(N, N-1, z)$; hence, the result holds for $N-1$. Assume $\sigma(\rho, n+1, z) \leq \sigma(\rho+1, n+1, z)$, $\rho \geq n+1$. Then, for $\rho \geq n$,

$$\sigma(\rho, n, z) = \gamma\sigma(\rho, n+1, z) \leq \gamma\sigma(\rho+1, n+1, z) = \sigma(\rho+1, n, z),$$

where the equalities are due to (iii), and the inequality is due to (ii) for the $\rho = n$ case and by assumption for $\rho > n$.

A.1.4 Proof of Lemma 1

Let $\vec{x}_1, \vec{x}_2 \in X_a^*$. It is sufficient to show that $\lambda\vec{x}_1 + (1-\lambda)\vec{x}_2 \in X_a^*$ where $\lambda \in [0, 1]$. Observe that $\omega^*(\vec{x})$ is convex and hence,

$$\begin{aligned}
\omega^*(\lambda\vec{x}_1 + (1-\lambda)\vec{x}_2) &\leq \lambda\omega^*(\vec{x}_1) + (1-\lambda)\omega^*(\vec{x}_2) \\
&= \lambda\omega_a(\vec{x}_1) + (1-\lambda)\omega_a(\vec{x}_2) \\
&= \omega_a(\lambda\vec{x}_1 + (1-\lambda)\vec{x}_2)
\end{aligned} \tag{14}$$

where the first equality is due to the definition of $\vec{x}_1, \vec{x}_2$, and the second equality is due to the assumption that $\omega_a(\vec{x})$ is an affine function. On the other hand, by definition of $\omega^*(\vec{x})$, we have

$$\omega^*(\lambda\vec{x}_1 + (1-\lambda)\vec{x}_2) \geq \omega_a(\lambda\vec{x}_1 + (1-\lambda)\vec{x}_2). \tag{15}$$

By Equations (14) and (15), $\omega^*(\lambda\vec{x}_1 + (1-\lambda)\vec{x}_2) = \omega_a(\lambda\vec{x}_1 + (1-\lambda)\vec{x}_2)$, and hence, $\lambda\vec{x}_1 + (1-\lambda)\vec{x}_2 \in X_a^*$.

A.1.5 Proof of Proposition 6

We first show that the assumptions of Proposition 6 lead to the result that $\vec{m}(\rho, n, z) - \vec{l}_n^x$ is NIV for all $\rho \geq n$ and $n \leq N-1$. It is straightforward to show that $\vec{m}(\rho, N-1, z) - \vec{l}_{N-1}^x$ is NIV for all $\rho \geq N-1$. Assume for all $\rho \geq n+1$, $\vec{m}(\rho, n+1, z) - \vec{l}_{n+1}^x$ is NIV. There are two cases:

Case 1 $\rho > n$: By Lemma 4(iv), $\beta P(\vec{m}(\rho, n+1, z) - \vec{l}_{n+1}^x)$ is NIV, and by Lemma 4(ii), it is sufficient that

$$[\vec{m}(\rho, n, z) - \vec{l}_n^x] - \beta P[\vec{m}(\rho, n+1, z) - \vec{l}_{n+1}^x] \tag{16}$$

is NIV in order for $[\vec{m}(\rho, n, z) - \vec{l}_n^x]$ to be NIV. We note that Equation (16) equals $c(n, n+1)\vec{e} + \beta P\vec{l}_{n+1}^x - \vec{l}_n^x$, which is NIV by assumption.

Case 2 $\rho = n$: Let $\vec{\eta}(n+1, z) = (\eta_0, \eta_1, \ldots, \eta_S)$ be such that $\eta_s = \max\{\vec{m}_s(\rho, n+1, z), \rho \geq n+1, l_s(n+1, z)\}$. Then, $(\vec{\eta}(n+1, z) - \vec{l}_{n+1}^x)_s = \max\{(\vec{m}(\rho, n+1, z) - \vec{l}_{n+1}^x)_s, \rho \geq n+1, l_{n+1} + \beta_{n+1} z\}$, which is NIV due to Lemma 4(v). Again, it is sufficient that

$$[\vec{m}(n, n, z) - \vec{l}_n^x] - \beta P[\vec{\eta}(n+1, z) - \vec{l}_{n+1}^x] \tag{17}$$

is NIV in order for $(\vec{m}(n, n, z) - \vec{l}_n^x)$ to be NIV. We note that Equation (17) equals $(c(n, n+1) + M)\vec{e} + \beta P\vec{l}_{n+1}^x - \vec{l}_n^x$, which is NIV by assumption.

Thus, the result holds by induction.

Based on the result above, we next prove the result in this proposition. Since $\pi^*(n, \vec{x}) \in \{\mathcal{NI}, \mathcal{I}\}$, there is a $\rho \geq n$ such that $\vec{x}\vec{l}(n, z) \leq \vec{x}\vec{m}(\rho, n, z)$ or equivalently, $l_n + \beta_n z \leq \vec{x}(\vec{m}(\rho, n, z) - \vec{l}_n^x)$. By Lemma 4(iii), $\vec{x}'\vec{l}(n, z) \leq \vec{x}'\vec{m}(\rho, n, z) \leq \max\{\vec{x}'\vec{m}(\rho, n, z), \forall \rho \geq n, \vec{x}'\vec{l}(n, z)\}$, and hence $\pi^*(n, \vec{x}')$ can be chosen so that $\pi^*(n, \vec{x}') \neq \mathcal{A}$.

Remark 1 *In this proof, we show that the condition that $\beta P\vec{l}_{n+1}^x - \vec{l}_n^x$ is NIV implies the condition that $(\vec{m}(\rho, n, z) - \vec{l}_n^x)$ is NIV. The implication of the latter condition (i.e., $(\vec{m}(\rho, n, z) - \vec{l}_n^x)$ is NIV for all $\rho \geq n$) is that as quality of a freight decreases at a location n, an incentive to move toward the destination (actions $\mathcal{NI}$ or $\mathcal{I}$) relative to abort the current trip (action $\mathcal{A}$) decreases. In general, rewards at the destination must be much higher than ones at a secondary market; otherwise, the shipping business to the destination may not be valid, and no further discussion is required.*

A.1.6 Proof of Proposition 7

(a) We remark that for each $n \in \{0, 1, \ldots, N\}$, there are only a finite number of points in $\mathcal{X} = \{(x_0, \ldots, x_S) : \sum_{i=0}^{S} x_i = 1, x_i \geq 0, i = 0, \ldots, S\}$ that can be reached. That is, it is sufficient to only consider the finite subset $\mathcal{S}$ of the state space where $\mathcal{S} \equiv \{0, 1, \ldots, N\} \times (\mathcal{X}_0 \cup \mathcal{X}_1 \cup \cdots \cup \mathcal{X}_N)$, $\mathcal{X}_0 = \{\vec{x}^*\}$, and for $n \geq 1$, $\mathcal{X}_n = \{\vec{x}^* P^n, \vec{e}_s P^{n-1}, \ldots, \vec{e}_s, s = 0, 1, \ldots, S\}$. We now prove the result by induction. The result is clearly true for any $(N, \vec{x}, z)$ such that $\vec{x} \in \mathcal{X}_N$. Assume that the result holds for any $(n+1, \vec{x}, z)$ such that $\vec{x} \in \mathcal{X}_{n+1}$. Thus, there is a finite set $V_M(n+1, \vec{x}, z)$ such that

$$\omega'(n+1, \vec{x}, z; M) = \max\{aM + b : (a, b) \in V_M(n+1, \vec{x}, z)\}$$

where $a \geq 0$ for all $(a, b) \in V_M(n+1, \vec{x}, z)$ for all $\vec{x} \in \mathcal{X}_{n+1}$. Note that

$$\begin{aligned} & c(n, n+1) + \beta\omega'(n+1, \vec{x}P, z; M) \\ = \;& c(n, n+1) + \beta \max\{aM + b : (a, b) \in V_M(n+1, \vec{x}P, z)\} \\ = \;& \max\{\beta a M + c(n, n+1) + \beta b : (a, b) \in V_M(n+1, \vec{x}P, z)\}. \end{aligned}$$

In addition,

$$\begin{aligned} & c(n, n+1) + M + \beta \vec{x} P \vec{\omega}'(n+1, z; M) \\ = \;& c(n, n+1) + M + \beta \sum_{s=0}^{S} \vec{e}_s \vec{x} P \max\{a_s M + b_s : (a_s, b_s) \in V_M(n+1, \vec{e}_s, z)\} \\ = \;& \max\{\left(1 + \beta \sum_{s=0}^{S} \vec{e}_s \vec{x} P a_s\right) M + c(n, n+1) \\ & \qquad + \beta \sum_{s=0}^{S} \vec{e}_s \vec{x} P b_s : (a_s, b_s) \in V_M(n+1, \vec{e}_s, z)\}. \end{aligned}$$

Further,

$$\begin{aligned} & l_n + \vec{x}\vec{l}_n^{\vec{x}} + \beta_n \max\{aM + b : (a, b) \in V_M(0, \vec{x}^*, z)\} \\ = \;& \max\{\beta_n a M + l_n + \vec{x}\vec{l}_n^{\vec{x}} + \beta_n b : (a, b) \in V_M(0, \vec{x}^*, z)\}. \end{aligned}$$

We observe that if a is nonnegative, then so are βa, $(1 + \beta \sum_{s=0}^{S} \vec{e}_s \vec{x} P a_s)$, and $\beta_n a$. Thus, the expressions $c(n, n+1) + \beta\omega'(n+1, \vec{x}P, z; M)$, $c(n, n+1) + M + \beta \vec{x} P \vec{\omega}'(n+

$1, z; M)$ and $l_n + \vec{x}\vec{l}_n^x + \beta z$ are piecewise-linear, convex and isotone in M, which is also true for their maximum, $\omega'(n, \vec{x}, z; M)$. The result then follows by induction.

(b) Proof for this case follows (a) analogously.

(c) Let $\vec{m}(\rho, n, z; P)$ be $\vec{m}(\rho, n, z)$ with explicit dependence on P. When $n = N - 1$, according to C.1, $\vec{m}(N-1, N-1, z; P) \leq \vec{m}(N, N-1, z; P)$, implying that action $\mathcal{NI}$ dominates action $\mathcal{I}$. Thus, we are only interested in whether $\vec{m}(N, N-1, z; P) \leq \vec{m}(N, N-1, z; P')$ holds so that $\omega'(N-1, x, z; P) \leq \omega'(N-1, x, z; P')$. Note that $\vec{l}(n, z)$ is independent of P. Since $\vec{p}_s \prec \vec{p}'_s$, $\vec{m}(N, N, z; P) = \vec{m}(N, N, z; P')$, and $\vec{m}(N, N, z)$ is NIV, $\vec{p}_s\vec{m}(N, N, z; P) \leq \vec{p}_s\vec{m}(N, N, z; P')$ holds for all s, and hence, $\vec{m}(N, N-1, z; P) \leq \vec{m}(N, N-1, z; P')$ holds. Therefore, $\omega'(N-1, x, z; P) \leq \omega'(N-1, x, z; P')$ for all x. Assume that $\vec{m}(\rho, n+1, z; P) \leq \vec{m}(\rho, n+1, z; P')$ for all $\rho \geq n+1$ so that $\omega'(n+1, x, z; P) \leq \omega'(n+1, x, z; P')$ holds. Since $\vec{p}_s \prec \vec{p}'_s$ for all s, $P\vec{m}(\rho, n+1, z; P) \leq P'\vec{m}(\rho, n+1, z; P)$, and the induction hypothesis leads to $P'\vec{m}(\rho, n+1, z; P) \leq P'\vec{m}(\rho, n+1, z; P')$. Thus, for $\rho \geq n+1$, $\vec{m}(\rho, n, z; P) \leq \vec{m}(\rho, n, z; P')$ by its definition. Likewise, due to the induction hypothesis, $\vec{\eta}(n+1, z; P) \leq \vec{\eta}(n+1, z; P')$ where $\vec{\eta}(n, z; P)$ is $\vec{\eta}(n, z)$ defined as before, explicitly recognizing the dependence on P, and hence, it is straightforward to show that $\vec{m}(n, n, z; P) \leq \vec{m}(n, n, z; P')$ also holds. Therefore, $\omega'(n, x, z; P) \leq \omega'(n, x, z; P')$ holds, and this completes the proof by induction.

A.1.7 Proof of Proposition 8

Note that

$$\begin{aligned} c(n, n+1) + \beta\omega_{LL}(n+1, \vec{x}^* P^{n+1}) &= c(n, n+1) + \beta(c(n+1, n+2) + \beta\omega_{LL}(n+2, \vec{x}^* P^{n+2})) \\ &\geq \vec{x}^* P^n \vec{l}(n, \omega_{LL}(0, \vec{x}^*)) \\ &\geq \vec{x}^* P^n \left[c(n, n+1)\vec{e} + \beta P\vec{l}(n+1, \omega_{LL}(0, \vec{x}^*)) \right] \\ &= c(n, n+1) + \beta\vec{x}^* P^{n+1}\vec{l}(n+1, \omega_{LL}(0, \vec{x}^*)), \end{aligned}$$

which implies the result.

A.1.8 Proof of Proposition 9

(i) It follows from Proposition 4 ($\pi^*(n, \vec{e}_S) = \mathcal{A}$) and Proposition 6 (both $\pi^*(n, \vec{x}) = \mathcal{I}$ and $\vec{x} \prec \vec{x}'$ imply $\pi^*(n, \vec{x}') \neq \mathcal{A}$, where we note $\vec{e}_{s+1} \prec \vec{e}_s$).

(ii) Let $z_U = \omega_U(0, \vec{x}^*)$. By the definition of $\bar{s}(n)$, $\omega_U(n, \vec{e}_k) = \vec{e}_k \vec{l}(n, z_U) = l_k(n, z_U)$ if and only if $k \geq \bar{s}(n)$. Note that

$$\omega_U(n, \vec{e}_i) = \max\left\{\vec{e}_i \vec{m}(n, n, z_U), \vec{e}_i \vec{l}(n, z_U)\right\} = \max\left\{\begin{array}{l} c(n, n+1) + \beta \sum_j p_{ij} \omega_U(n+1, \vec{e}_j), \\ \vec{e}_i \vec{l}(n, z_U) = l_i(n, z_U) \end{array}\right\}.$$

Let $i = \bar{s}(n+1)$. Then,

$$\sum_j p_{ij} \omega_U(n+1, \vec{e}_j) = \sum_{j \geq i} p_{ij} l_j(n+1, z_U).$$

Thus,

$$\omega_U(n, \vec{e}_i) = \max\left\{\begin{array}{l} c(n, n+1) + \beta \sum_{j \geq i} p_{ij} l_j(n+1, z_U), \\ l_i(n, z_U) \end{array}\right\} = l_i(n, z_U)$$

due to A1, and hence, $\bar{s}(n+1) \geq \bar{s}(n)$.

A.2 Proofs in Chapter 4

A.2.1 Proof of Proposition 10

We note that $d - \max\{0, \alpha - z\}$ is isotone in z, and hence $v(d - \max\{0, \alpha - z\})$ is antitone in z if v is antitone. Note also that $f(z, a)$ is antitone in z for fixed a. The weighted sum of antitone functions is antitone. Thus, $h(z, a, v)$ is antitone in z for fixed a if v is antitone. The minimum of antitone functions is antitone. Thus, if v is antitone, Hv is antitone. Since the limit of antitone functions is antitone, v^* is antitone.

A.2.2 Proof of Proposition 11

The result follows directly from the fact that if $z \leq 0$, then for all $a \geq 0$,

$$h(z, a, v) = ca + (\bar{p} - (p - h))\mathcal{Y}(a) + \bar{p}|z| + \beta \sum_\alpha P(\alpha|a) v(d - |z| - \alpha).$$

A.2.3 Proof of Proposition 12

If $0 \le a \le z$, then $h(z,a,v) = ca-(p-h)\mathcal{Y}(a)+\beta v(d) \ge cz-(p-h)\mathcal{Y}(a)+\beta v(d) = h(z,z,v)$.

A.2.4 Proof of Proposition 13

Noting that A5 implies A4, Proposition 12 guarantees that for $z \ge 0$, there is an optimal policy δ^* such that $\delta^*(z) \ge z$. Hence, there is a $\gamma \ge 0$ such that $\delta^*(z) = z+\gamma$. It is therefore sufficient for

$$h(z, z+\gamma, v) - h(z, z+\gamma+1, v) \le h(z+1, (z+1)+\gamma, v) - h(z+1, (z+1)+\gamma+1, v) \quad (18)$$

given v is antitone. Note that this inequality is associated with subadditivity, given the structure of an optimal policy described in Proposition 12 (see Puterman [50]). We note that the equation (18) is equivalent to

$$\begin{aligned}
&-(p-h)[\mathcal{Y}(z+\gamma) - \mathcal{Y}(z+\gamma+1)] \\
&\quad + \bar{p}\sum_{\alpha}[P(\alpha|z+\gamma) - P(\alpha|z+\gamma+1)]\max\{0, \alpha - z\} \\
&\quad\quad + \beta\sum_{\xi}[P(\xi|z, z+\gamma) - P(\xi|z, z+\gamma+1)]v(\xi) \\
\le &-(p-h)[\mathcal{Y}(z+1+\gamma) - \mathcal{Y}(z+1+\gamma+1)] + \bar{p}\sum_{\alpha}[P(\alpha|z+1+\gamma) \\
&\quad - P(\alpha|z+1+\gamma+1)]\max\{0, \alpha - z - 1\} \\
&\quad\quad + \beta\sum_{\xi}[P(\xi|z+1, z+1+\gamma) - P(\xi|z+1, z+1+\gamma+1)]v(\xi).
\end{aligned}$$

This inequality holds if

(a) $\mathcal{Y}(z+\gamma+1) - \mathcal{Y}(z+\gamma) \le \mathcal{Y}(z+\gamma+2) - \mathcal{Y}(z+\gamma+1)$

(b) $\sum_{\alpha}[P(\alpha|z+\gamma) - P(\alpha|z+\gamma+1)]\max\{0, \alpha - z\} \le \sum_{\alpha}[P(\alpha|z+1+\gamma) - P(\alpha|z+2+\gamma)]\max\{0, \alpha - z - 1\}$

(c) $\sum_{\xi}[P(\xi|z, z+\gamma) - P(\xi|z, z+\gamma+1)]v(\xi) \le \sum_{\xi}[P(\xi|z+1, z+1+\gamma) - P(\xi|z+1, z+1+\gamma+1)]v(\xi)$.

Note that (a) holds by assumption. Lemma 4.7.2 in Puterman [50] guarantees that (b) holds and that (c) holds if $Q(k|z, z+\gamma+1) - Q(k|z, z+\gamma)$ is isotone in z where $Q(k|z, a) =$

$\sum_{\xi \geq k} P(\xi|z,a) = \sum_{\alpha=0}^{z+d-k} P(\alpha|a)$. It is straightforward to show that $Q(k|z,a) = 1 - \sum_{\alpha \geq k'} P(\alpha|a)$ for $k' = z+d-k+1$. Thus, the isotonicity of $q(k|a) - q(k|a+1)$ guarantees the isotonicity of $Q(k|z,a+1) - Q(k|z,a)$.

A.3 Proofs in Chapter 5

A.3.1 Proof of Lemma 2

Let m be such that $0 \in \mathcal{M}(m|i,e)$, and assume $m < N$. Then,

$$\begin{aligned}
\mathcal{I}(f(x,i,a,e,D)) &= \sum_{n=1}^{N} g_n^{N-1}(f_n(x,i,a,e,D)) + f_N(x,i,a,e,D) \\
&= g_m^{N-1}\Big(g_0^{m-1}(a|i,e) + \sum_{\substack{k\in\mathcal{M}(m|i,e)\\ k\neq 0}} g_k^{m-1}(x_k)\Big) \\
&\qquad + \sum_{n>m}^{N-1} g_n^{N-1}\Big(\sum_{k\in\mathcal{M}(n|i,e)} g_k^{n-1}(x_k)\Big) + x_N - D + \sum_{k\in\mathcal{M}(N|i,e)} g_k^{N-1}(x_k) \\
&= g_0^{N-1}(a|i,e) + \sum_{n=1}^{N-1} g_n^{N-1}(x_n) + x_N - D.
\end{aligned}$$

where the last equality is due to the assumption. Proof of the $m = N$ case is straightforward.

A.3.2 Proof of Lemma 3

(a)

$$\begin{aligned}
&\beta\mathbb{E}_{e,D}\Big[v'(f(x,i,a,e,D),Q(i,e))\Big] \\
&= \mathbb{E}_{L_N}\Bigg[\sum_{j=0}^{L_N-2}\beta^{j+1}h\mathbb{E}_D\Bigg\{\mathbb{E}_{e,e_{j+1},D_{j+1}}\Bigg[\sum_{n=1}^{N-1}\sum_{k\in\mathcal{M}^{j+1}(n|i,e_{j+1})}g_k^{n-1}\Big(\sum_{0\in\mathcal{M}(k|i,e)}g_0^{k-1}(a|i,e)\Big) \\
&\quad + \sum_{k\in\mathcal{M}^{j+1}(N|i,e_{j+1})}g_k^{N-1}\Big(\sum_{0\in\mathcal{M}(k|i,e)}g_0^{k-1}(a|i,e)\Big) + \sum_{n=1}^{N-1}\sum_{k\in\mathcal{M}^{j+1}(n|i,e_{j+1})}g_k^{n-1}\Big(\sum_{u\in\mathcal{M}(k|i,e)}g_u^{k-1}(x_u)\Big) \\
&\quad + \sum_{k\in\mathcal{M}^{j+1}(N|i,e_{j+1})}g_k^{N-1}\Big(\sum_{u\in\mathcal{M}(k|i,e)}g_u^{k-1}(x_u)\Big) + g_N^N(x_N) - D - D_{j+1}\Bigg]\Bigg\}\Bigg] \\
&= \mathbb{E}_{L_N}\Bigg[\sum_{j=0}^{L_N-2}\beta^{j+1}h\mathbb{E}_{e,e_{j+1},e_{j+2},D_{j+2}}\Bigg\{\sum_{n=1}^{N-1}\sum_{k\in\mathcal{M}^{j+1}(n|i,e_{j+1})}g_k^{n-1}\Big(\sum_{0\in\mathcal{M}(k|i,e)}g_0^{k-1}(a|i,e)\Big) \\
&\quad + \sum_{k\in\mathcal{M}^{j+1}(N|i,e_{j+1})}g_k^{N-1}\left(\sum_{0\in\mathcal{M}(k|i,e)}g_0^{k-1}(a|i,e)\right) + \sum_{n=1}^{N-1}\sum_{k\in\mathcal{M}^{j+2}(n|i,e_{j+2})}g_k^{n-1}(x_k) \\
&\quad + \sum_{k\in\mathcal{M}^{j+2}(N|i,e_{j+2})}g_k^{N-1}(x_k) + g_N^N(x_N) - D_{j+2}\Bigg\}\Bigg] \\
&= \mathbb{E}_{L_N}\Bigg[\sum_{j=1}^{L_N-1}\beta^{j}h\mathbb{E}_{e,e_j}\Bigg\{\sum_{n=1}^{N-1}\sum_{k\in\mathcal{M}^{j}(n|i,e_j)}g_k^{n-1}\Big(\sum_{0\in\mathcal{M}(k|i,e)}g_0^{k-1}(a|i,e)\Big) \\
&\qquad\qquad + \sum_{k\in\mathcal{M}^{j}(N|i,e_j)}g_k^{N-1}\Big(\sum_{0\in\mathcal{M}(k|i,e)}g_0^{k-1}(a|i,e)\Big)\Bigg\}\Bigg] \\
&\quad + \mathbb{E}_{L_N}\Bigg[\sum_{j=1}^{L_N-1}\beta^{j}h\mathbb{E}_{e_{j+1},D_{j+1}}\Bigg\{\sum_{n=1}^{N-1}\sum_{k\in\mathcal{M}^{j+1}(n|i,e_{j+1})}g_k^{n-1}(x_k) \\
&\qquad\qquad + \sum_{k\in\mathcal{M}^{j+1}(N|i,e_{j+1})}g_k^{N-1}(x_k) + g_N^N(x_N) - D_{j+1}\Bigg\}\Bigg]
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}_{L_N}\Bigg[\sum_{j=1}^{L_N-2}\beta^j h\mathbb{E}_{e_j}\Bigg\{\sum_{n=1}^{N-1}\sum_{k\in\mathcal{M}^j(n|i,e_j)} g_k^{n-1}\Big(\sum_{0\in\mathcal{M}(k|i,e)} g_0^{k-1}(a|i,e)\Big) \\
&\qquad\qquad + \underbrace{\sum_{k\in\mathcal{M}^j(N|i,e_j)} g_k^{N-1}\Big(\sum_{0\in\mathcal{M}(k|i,e)} g_0^{k-1}(a|i,e)\Big)}_{=0}\Bigg\}\Bigg] \\
&\quad + \mathbb{E}_{L_N}\Bigg[\beta^{L_N-1}h\mathbb{E}_{e,e_{L_N-1}}\Bigg\{\underbrace{\sum_{n=1}^{N-1}\sum_{k\in\mathcal{M}^{L_N-1}(n|i,e_{L_N-1})} g_k^{n-1}\Big(\sum_{0\in\mathcal{M}(k|i,e)} g_0^{k-1}(a|i,e)\Big)}_{=0} \\
&\qquad\qquad + \underbrace{\sum_{k\in\mathcal{M}^{L_N-1}(N|i,e_{L_N-1})} g_k^{N-1}\Big(\sum_{0\in\mathcal{M}(k|i,e)} g_0^{k-1}(a|i,e)\Big)}_{=g_0^{N-1}(a|i,e)}\Bigg\}\Bigg] \\
&\quad + \underbrace{\mathbb{E}_{L_N}\Bigg[\sum_{j=0}^{L_N-2}\beta^j h\mathbb{E}_{e_{j+1},D_{j+1}}\Bigg\{\sum_{n=1}^{N}\sum_{k\in\mathcal{M}^{j+1}(n|i,e_{j+1})} g_k^{n-1}(x_k) + g_N^N(x_N) - D_{j+1}\Bigg\}\Bigg]}_{=v'(x,i)} \\
&\quad + \mathbb{E}_{L_N}\Bigg[\beta^{L_N-1}h\mathbb{E}_{e_{L_N},D_{L_N}}\Bigg\{\underbrace{\sum_{n=1}^{N-1}\sum_{k\in\mathcal{M}^{L_N}(n|i,e_{L_N})} g_k^{n-1}(x_k)}_{=0} \\
&\qquad\qquad + \underbrace{\sum_{k\in\mathcal{M}^{L_N}(N|i,e_{L_N})} g_k^{N-1}(x_k) + g_N^N(x_N)}_{=\mathcal{I}(x)} - D_{L_N}\Bigg\}\Bigg] \\
&\quad - h\mathbb{E}_{e,D}\Bigg[\sum_{n=1}^{N}\sum_{k\in\mathcal{M}(n|i,e)} g_k^{n-1}(x_k) + g_N^N(x_N) - D\Bigg] \\
&= \mathbb{E}_{L_N}\Bigg[\sum_{j=1}^{L_N-2}\beta^j h\mathbb{E}_{e,e_j}\Bigg\{\sum_{n=1}^{N-1}\sum_{k\in\mathcal{M}^j(n|i,e_j)} g_k^{n-1}\Big(\sum_{0\in\mathcal{M}(k|i,e)} g_0^{k-1}(a|i,e)\Big)\Bigg\}\Bigg] + v'(x|i) \\
&\quad + \mathbb{E}_{L_N}\Bigg[\beta^{L_N-1}h\mathbb{E}_{e,D_{L_N}}\Bigg\{g_0^N(a|i,e) + \mathcal{I}(\vec{x}) - D_{L_N}\Bigg\}\Bigg] \\
&\quad - h\mathbb{E}_{e,D}\Bigg[\sum_{n=1}^{N}\sum_{k\in\mathcal{M}(n|i,e)} g_k^{n-1}(x_k) + x_N - D\Bigg].
\end{aligned}
$$

(b)

$$
\begin{aligned}
&\beta\mathbb{E}_{e,D}\Big[v''(f(x,i,a,e,D),Q(i,e))\Big]\\
&= -\mathbb{E}_{L_N}\Big\{\sum_{j=0}^{L_N-2}\beta^{j+1}\mathbb{E}_D\Big[\mathbb{E}_{e,D_{j+1}}\Big\{(h+\sigma)\Big[\sum_{l=1}^{j+1}\sum_{k\in\mathcal{M}^l(N|i,e_l)} g_k^N\Big(\sum_{0\in\mathcal{M}(k|i,e)} g_0^{k-1}(a|i,e)\Big)\\
&\quad+\sum_{l=1}^{j+1}\sum_{k\in\mathcal{M}^l(N|i,e_l)} g_k^N\Big(\sum_{u\in\mathcal{M}(k|i,e)} g_u^{k-1}(x_u)\Big)\\
&\quad+g_N^N\Big(\sum_{k\in\mathcal{M}(N|i,e)} g_k^{N-1}(x_k)+x_N-D\Big)-D_{j+1}\Big]^-\Big\}\Big]\Big\}\\
&= -\mathbb{E}_{L_N}\Big\{\sum_{j=0}^{L_N-2}\beta^{j+1}\mathbb{E}_{e,D_{j+2}}\Big[(h+\sigma)\Big(\sum_{l=1}^{j+1}\sum_{k\in\mathcal{M}^l(N|i,e_l)} g_k^N\Big(\sum_{0\in\mathcal{M}(k|i,e)} g_0^{k-1}(a|i,e)\Big)\\
&\quad+\sum_{l=1}^{j+1}\sum_{k\in\mathcal{M}^{l+1}(N|i,e_{l+1})} g_k^N(x_k)+\sum_{k\in\mathcal{M}(N|i,e)} g_k^N(x_k)+g_N^N(x_N)-D_{j+2}\Big)^-\Big]\Big\}\\
&= -\mathbb{E}_{L_N}\Big\{\sum_{j=0}^{L_N-2}\beta^{j+1}\mathbb{E}_{e,D_{j+2}}\Big[(h+\sigma)\Big(\sum_{l=1}^{j+1}\sum_{k\in\mathcal{M}^l(N|i,e_l)} g_k^N\Big(\sum_{0\in\mathcal{M}(k|i,e)} g_0^{k-1}(a|i,e)\Big)\\
&\quad+\sum_{l=1}^{j+2}\sum_{k\in\mathcal{M}^l(N|i,e_l)} g_k^N(x_k)+g_N^N(x_N)-D_{j+2}\Big)^-\Big]\Big\}\\
&= -\mathbb{E}_{L_N}\Big\{\sum_{j=1}^{L_N-1}\beta^{j}\mathbb{E}_{e,D_{j+1}}\Big[(h+\sigma)\Big(\sum_{l=1}^{j}\sum_{k\in\mathcal{M}^l(N|i,e_l)} g_k^N\Big(\sum_{0\in\mathcal{M}(k|i,e)} g_0^{k-1}(a|i,e)\Big)\\
&\quad+\sum_{l=1}^{j+1}\sum_{k\in\mathcal{M}^l(N|i,e_l)} g_k^N(x_k)+g_N^N(x_N)-D_{j+1}\Big)^-\Big]\Big\}\\
&= \underbrace{-\mathbb{E}_{L_N}\Big\{\textstyle\sum_{j=0}^{L_N-2}\beta^j\mathbb{E}_{e,D_{j+1}}\Big[(h+\sigma)\Big(\overbrace{\sum_{l=1}^{j}\sum_{k\in\mathcal{M}^l(N|i,e_l)} g_k^N\Big(\sum_{0\in\mathcal{M}(k|i,e)} g_0^{k-1}(a|i,e)\Big)}^{=0}
+\textstyle\sum_{l=1}^{j+1}\sum_{k\in\mathcal{M}^l(N|i,e_l)} g_k^N(x_k)+g_N^N(x_N)-D_{j+1}\Big)^-\Big]\Big\}}_{=v''(x,i)}\\
&\quad+(h+\sigma)\mathbb{E}_{e,D}\Big[\Big(\sum_{k\in\mathcal{M}(N|i,e)} g_k^N(x_k)+x_N-D\Big)^-\Big]\\
&\quad-\mathbb{E}_{L_N}\Big[\beta^{L_N-1}(h+\sigma)\mathbb{E}_{e,D_{L_N}}\Big\{\Big(\underbrace{\sum_{l=1}^{L_N-1}\sum_{k\in\mathcal{M}^l(N|i,e_l)} g_k^N\Big(\sum_{0\in\mathcal{M}(k|i,e)} g_0^{k-1}(a|i,e)\Big)}_{=g_0^N(a|i,e)}\\
&\quad+\underbrace{\sum_{l=1}^{L_N}\sum_{k\in\mathcal{M}^l(N|i,e_l)} g_k^N(x_k)}_{=\mathcal{I}(x)}-D_{L_N}\Big)^-\Big\}\Big]
\end{aligned}
$$

(c)

$$
\begin{aligned}
&c(x,i,a)+\beta\mathbb{E}_{e,D}\Big[v'(f(x,i,a,e,D),Q(i,e))\Big]+\beta\mathbb{E}_{e,D}\Big[v''(f(x,i,a,e,D),Q(i,e))\Big]\\
&=K\delta(a)+ca+h\mathbb{E}_e\Big[\sum_{n=1}^{N-1}\sum_{0\in\mathcal{M}(n|i,e)}g_1^{n-1}(\overline{g}(a|i,e))\Big]\\
&\quad+\Big\{h\mathbb{E}_{e,D}\Big[\sum_{n=1}^{N}\sum_{k\in\mathcal{M}(n|i,e)}g_k^{n-1}(x_k)+x_N-D\Big]+\beta\mathbb{E}_{e,D}\Big[v'(f(x,i,a,e,D),Q(i,e))\Big]\Big\}\\
&\quad+\Big\{-(h+\sigma)\mathbb{E}_{e,D}\Big[\Big(\sum_{k\in\mathcal{M}(N|i,e)}g_k^{N}(x_k)+x_N-D\Big)^{-}\Big]+\beta\mathbb{E}_{e,D}\Big[v''(f(x,i,a,e,D),Q(i,e))\Big]\Big\}\\
&=K\delta(a)+ca+h\mathbb{E}_e\Big[\sum_{n=1}^{N-1}\sum_{0\in\mathcal{M}(n|i,e)}g_0^{n-1}(a|i,e)\Big]\\
&\quad+\Big\{\mathbb{E}_{L_N}\Big[\sum_{j=1}^{L_N-2}\beta^j h\mathbb{E}_{e,e_j}\Big[\sum_{n=1}^{N-1}\sum_{k\in\mathcal{M}^j(n|i,e_j)}g_k^{n-1}\Big(\sum_{0\in\mathcal{M}(k|i,e)}g_0^{k-1}(a|i,e)\Big)\Big]\Big]\\
&\qquad+\mathbb{E}_{L_N}\Big[\beta^{L_N-1}h\mathbb{E}_{e,D_{L_N}}\Big[g_0^{N}(a|i,e)+\mathcal{I}(x)-D_{L_N}\Big]\Big]+v'(x,i)\Big\}\\
&\quad+\Big\{-\mathbb{E}_{L_N}\Big[\beta^{L_N-1}(h+\sigma)\mathbb{E}_{e,D_{L_N}}\Big[\Big(g_0^{N}(a|i,e)+\mathcal{I}(x)-D_{L_N}\Big)^{-}\Big]\Big]+v''(x,i)\Big\}\\
&=c^\circ(\mathcal{I}(x),i,a)+v'(x,i)+v''(x,i)
\end{aligned}
$$

where the second equality is due to (a) and (b).

A.3.3 Proof of Proposition 14

We prove this proposition by induction. Let $v_{k+1}=Hv_k$ with $v_0=v'+v''$, and $v^\circ_{k+1}=H^\circ v^\circ_k$ with $v^\circ_0=0$.

$$
\begin{aligned}
v_1(x,i)&=[Hv_0](x,i)\\
&=\min_a\Big\{c(x,i,a)+\beta\mathbb{E}\left[v_0(f(x,i,a,e,D),Q(i,e))\right]\Big\}\\
&=\min_a\Big\{c(x,i,a)+\beta\mathbb{E}\left[v'(f(x,i,a,e,D),Q(i,e))\right]+\beta\mathbb{E}\left[v''(f(x,i,a,e,D),Q(i,e))\right]\Big\}\\
&=\min_a\Big\{c^\circ(\mathcal{I}(x),i,a)+v'(x,i)+v''(x,i)\Big\}\\
&=v^\circ_1(x,i)+v'(x,i)+v''(x,i)
\end{aligned}
$$

where the third equality is due to Lemma 3(c).

We now assume that $v_{k-1}(x,i) = v^\circ_{k-1}(x,i) + v'(x,i) + v''(x,i)$. Then,

$$
\begin{aligned}
v_k(x,i) &= [Hv_{k-1}](x,i) \\
&= \min_a \Big\{ c(x,i,a) + \beta \mathbb{E}\left[v_{k-1}(f(x,i,a,e,D), Q(i,e))\right] \Big\} \\
&= \min_a \Big\{ c(x,i,a) + \beta \mathbb{E}\left[v^\circ_{k-1}(f(x,i,a,e,D), Q(i,e))\right] \\
&\qquad\qquad + \beta \mathbb{E}\left[v'(f(x,i,a,e,D), Q(i,e))\right] + \beta \mathbb{E}\left[v''(f(x,i,a,e,D), Q(i,e))\right] \Big\} \\
&= \min_a \Big\{ c^\circ(\mathcal{I}(x),i,a) + \beta \mathbb{E}\left[v^\circ_{k-1}(f(x,i,a,e,D), Q(i,e))\right] + v'(x,i) + v''(x,i) \Big\} \\
&= v^\circ_k(x,i) + v'(x,i) + v''(x,i)
\end{aligned}
$$

where the third equality is due to the induction hypothesis, and the fourth equality is again due to Lemma 3(c). This completess the proof by induction.

APPENDIX B

APPENDICES IN CHAPTER 2

B.1 Determination of $q(s'|n, s, n')$

We now determine $q(s'|n, s, n')$, assuming the conditional probabilities $P(s(t+1)|s(t))$ and $P(\Delta(t+1)|\Delta(t), s(t))$ are given (presumably derived from historic traffic data), where $\Delta(t) \in \{0, 1, \ldots, d(n, n')\}$ is the distance remaining to n' at time t, the trip from n to n' began at $t = 0$, and $d(n, n')$ is the total distance from n to n'. We remark that Section III in Kim et al. [33] presents the details of data collection (Section III-A) and the construction of the Markov chain $[P(s(t+1)|s(t))]$ or $[q(s'|n, s, n')]$ (Section III-B). Yeon et al. [68] also presents both data collection and the development of the discrete-time Markov chain $[P(s(t+1)|s(t))]$ whose state representation is same as ours (see Section 4 in Yeon et al. [68]). Let τ be the (random) length of time required to travel from n to n'.

Then,

$$\begin{aligned} q(s'|n, s, n') &= P(s(\tau) = s'|s(0) = s) \\ &= \textstyle\sum_{\xi \geq 1} P(s(\tau) = s', \tau = \xi|s(0) = s) \\ &= \textstyle\sum_{\xi \geq 1} P(s(\tau) = s'|\tau = \xi, s(0) = s)P(\tau = \xi|s(0) = s) \\ &= \textstyle\sum_{\xi \geq 1} P(s(\xi) = s'|s(0) = s)P(\tau = \xi|s(0) = s). \end{aligned}$$

We use Kolmogorov equations to determine $P(s(\xi) = s'|s(0) = s)$, given $P(s(t+1)|s(t))$. We now seek $P(\tau = \xi|s(0))$. Note that $\tau = \xi$ if and only if $\Delta(\xi) = 0$ and $\Delta(\xi - 1) \neq 0$. Then,

$$\begin{aligned} &P(\tau = \xi|s(0)) \\ &= \textstyle\sum_{\Delta(\xi-1)>0} P(\Delta(\xi) = 0, \Delta(\xi - 1)|s(0)) \\ &= \textstyle\sum_{\Delta(\xi-1)>0} \sum_{s(\xi-1)} P(\Delta(\xi) = 0, s(\xi - 1), \Delta(\xi - 1)|s(0)) \\ &= \textstyle\sum_{\Delta(\xi-1)>0} \sum_{s(\xi-1)} P(\Delta(\xi) = 0|s(\xi - 1), \Delta(\xi - 1))P(\Delta(\xi - 1)|s(0))P(s(\xi - 1)|s(0)). \end{aligned}$$

Therefore, it is sufficient to know $P(\Delta(t)|s(0))$ in order to determine $P(\tau = \xi|s(0))$ and hence $q(s'|n,s,n')$. We note that

$$\begin{aligned} P(\Delta(t+1)|s(0)) &= \sum_{\Delta(t)} \sum_{s(t)} P(\Delta(t+1), s(t), \Delta(t)|s(0)) \\ &= \sum_{\Delta(t)} \sum_{s(t)} P(\Delta(t+1)|s(t), \Delta(t)) P(s(t)|s(0)) P(\Delta(t)|s(0)). \end{aligned}$$

As an initial condition, $P(\Delta(0) = d(n,n')|s(0)) = 1$ for all $s(0)$. Hence,

$$P(\Delta(1)|s(0)) = P(\Delta(1)|\Delta(0) = d(n,n'), s(0)).$$

B.2 Optimal Policy for Example 2

State $N/n/s$	Action n'	Expected Cost $v(N,n,s)$	State $N/n/s$	Action n'	Expected Cost $v(N,n,s)$
/0/000	3	51.48	0.3/6/101	2	44.00
/0/100	10	52.73	0.3/6/100	2	44.00
/0/010	10	52.75	0.3/11/111	5	46.00
/0/001	3	51.64	0.3/11/000	5	46.00
/0/110	10	53.89	0.3/11/110	5	46.00
/0/101	10	52.93	0.3/11/001	5	46.00
/0/011	10	52.94	0.3/11/010	5	46.00
/0/111	10	52.93	0.3/11/011	5	46.00
0/3/100	6	49.20	0.3/11/101	5	46.00
0/3/101	6	49.20	0.3/11/100	5	46.00
0/3/111	11	51.00	0.10/8/100	4	44.00
0/3/010	11	51.00	0.10/8/101	4	44.00
0/3/011	11	51.00	0.10/8/111	4	44.00
0/3/000	6	49.20	0.10/8/000	4	44.00
0/3/001	6	49.20	0.10/8/110	4	44.00
0/3/110	11	51.00	0.10/8/010	4	44.00
0/10/110	8	48.00	0.10/8/011	4	44.00
0/10/001	3	47.62	0.10/8/001	4	44.00
0/10/111	8	48.00	0.10/3/100	6	42.20
0/10/101	8	48.00	0.10/3/010	11	47.00
0/10/010	8	48.00	0.10/3/011	9	47.00
0/10/011	8	48.00	0.10/3/101	6	42.20
0/10/100	3	47.66	0.10/3/110	9	47.00
0/10/000	3	47.08	0.10/3/111	9	47.00
0.3/6/010	2	44.00	0.10/3/000	6	42.20
0.3/6/011	2	44.00	0.10/3/001	6	42.20
0.3/6/001	2	44.00	0.3.6/2/111	7	38.00
0.3/6/111	2	44.00	0.3.6/2/011	7	38.00
0.3/6/000	2	44.00	0.3.6/2/010	7	38.00
0.3/6/110	2	44.00	0.3.6/2/110	7	38.00

0.3.6/2/000	7	38.00	0.2.3.6/7/000	8	36.00
0.3.6/2/001	7	38.00	0.2.3.6/7/001	8	36.00
0.3.6/2/101	7	38.00	0.2.3.6/7/111	8	36.00
0.3.6/2/100	7	38.00	0.2.3.6/7/110	8	36.00
0.3.11/5/110	1	43.00	0.2.3.6/7/010	8	36.00
0.3.11/5/111	1	43.00	0.2.3.6/7/101	8	36.00
0.3.11/5/011	1	43.00	0.2.3.6/7/100	8	36.00
0.3.11/5/010	1	43.00	0.3.5.11/1/111	9	40.00
0.3.11/5/101	1	43.00	0.3.5.11/1/101	9	40.00
0.3.11/5/001	1	43.00	0.3.5.11/1/100	9	40.00
0.3.11/5/100	1	43.00	0.3.5.11/1/011	9	40.00
0.3.11/5/000	1	43.00	0.3.5.11/1/010	9	40.00
0.8.10/4/110	9	37.00	0.3.5.11/1/001	9	40.00
0.8.10/4/100	9	37.00	0.3.5.11/1/110	9	40.00
0.8.10/4/101	9	37.00	0.3.5.11/1/000	9	40.00
0.8.10/4/111	9	37.00	0.4.8.10/9/100	1	33.00
0.8.10/4/010	9	37.00	0.4.8.10/9/010	1	33.00
0.8.10/4/011	9	37.00	0.4.8.10/9/011	1	33.00
0.8.10/4/000	9	37.00	0.4.8.10/9/101	1	33.00
0.8.10/4/001	9	37.00	0.4.8.10/9/000	1	33.00
0.3.10/6/010	2	37.00	0.4.8.10/9/001	1	33.00
0.3.10/6/011	2	37.00	0.4.8.10/9/111	1	33.00
0.3.10/6/101	2	37.00	0.4.8.10/9/110	1	33.00
0.3.10/6/100	2	37.00	0.3.6.10/2/100	7	31.00
0.3.10/6/001	2	37.00	0.3.6.10/2/101	7	31.00
0.3.10/6/000	2	37.00	0.3.6.10/2/010	7	31.00
0.3.10/6/111	2	37.00	0.3.6.10/2/011	7	31.00
0.3.10/6/110	2	37.00	0.3.6.10/2/110	7	31.00
0.3.10/11/100	6	42.00	0.3.6.10/2/001	7	31.00
0.3.10/11/101	6	42.00	0.3.6.10/2/111	7	31.00
0.3.10/11/000	6	42.00	0.3.6.10/2/000	7	31.00
0.3.10/11/001	6	42.00	0.3.10.11/6/100	2	36.00
0.3.10/11/010	6	42.00	0.3.10.11/6/011	2	36.00
0.3.10/11/111	6	42.00	0.3.10.11/6/111	2	36.00
0.3.10/11/011	6	42.00	0.3.10.11/6/110	2	36.00
0.3.10/11/110	6	42.00	0.3.10.11/6/010	2	36.00
0.3.10/9/101	11	41.00	0.3.10.11/6/101	2	36.00
0.3.10/9/111	11	41.00	0.3.10.11/6/000	2	36.00
0.3.10/9/100	11	41.00	0.3.10.11/6/001	2	36.00
0.3.10/9/110	11	41.00	0.3.9.10/11/101	6	38.00
0.3.10/9/000	11	41.00	0.3.9.10/11/100	6	38.00
0.3.10/9/011	11	41.00	0.3.9.10/11/111	6	38.00
0.3.10/9/010	11	41.00	0.3.9.10/11/110	6	38.00
0.3.10/9/001	11	41.00	0.3.9.10/11/011	6	38.00
0.2.3.6/7/011	8	36.00	0.3.9.10/11/010	6	38.00

0.3.9.10/11/000	6	38.00	0.3.9.10.11/6/100	2	32.00
0.3.9.10/11/001	6	38.00	0.3.9.10.11/6/110	2	32.00
0.2.3.6.7/8/000	10	30.00	0.3.9.10.11/6/001	2	32.00
0.2.3.6.7/8/001	11	30.00	0.3.9.10.11/6/000	2	32.00
0.2.3.6.7/8/110	10	30.00	0.3.9.10.11/6/111	2	32.00
0.2.3.6.7/8/111	10	30.00	0.2.3.6.7.8/10/001	4	19.00
0.2.3.6.7/8/101	4	30.00	0.2.3.6.7.8/10/110	4	19.00
0.2.3.6.7/8/100	4	30.00	0.2.3.6.7.8/10/000	4	19.00
0.2.3.6.7/8/011	4	30.00	0.2.3.6.7.8/10/111	4	19.00
0.2.3.6.7/8/010	10	30.00	0.2.3.6.7.8/10/101	4	19.00
0.1.3.5.11/9/111	7	37.00	0.2.3.6.7.8/10/100	4	19.00
0.1.3.5.11/9/000	7	37.00	0.2.3.6.7.8/10/010	4	19.00
0.1.3.5.11/9/101	7	37.00	0.2.3.6.7.8/10/011	4	19.00
0.1.3.5.11/9/110	7	37.00	0.2.3.6.7.8/11/100	5	21.00
0.1.3.5.11/9/001	7	37.00	0.2.3.6.7.8/11/101	5	21.00
0.1.3.5.11/9/011	7	37.00	0.2.3.6.7.8/11/000	5	21.00
0.1.3.5.11/9/010	7	37.00	0.2.3.6.7.8/11/011	5	21.00
0.1.3.5.11/9/100	7	37.00	0.2.3.6.7.8/11/001	5	21.00
0.4.8.9.10/1/100	11	28.00	0.2.3.6.7.8/11/111	5	21.00
0.4.8.9.10/1/101	11	28.00	0.2.3.6.7.8/11/010	5	21.00
0.4.8.9.10/1/111	11	28.00	0.2.3.6.7.8/11/110	5	21.00
0.4.8.9.10/1/110	11	28.00	0.2.3.6.7.8/4/111	9	23.00
0.4.8.9.10/1/000	11	28.00	0.2.3.6.7.8/4/110	9	23.00
0.4.8.9.10/1/001	11	28.00	0.2.3.6.7.8/4/101	9	23.00
0.4.8.9.10/1/010	11	28.00	0.2.3.6.7.8/4/011	9	23.00
0.4.8.9.10/1/011	11	28.00	0.2.3.6.7.8/4/000	9	23.00
0.2.3.6.10/7/000	8	29.00	0.2.3.6.7.8/4/001	9	23.00
0.2.3.6.10/7/001	8	29.00	0.2.3.6.7.8/4/010	9	23.00
0.2.3.6.10/7/010	8	29.00	0.2.3.6.7.8/4/100	9	23.00
0.2.3.6.10/7/011	8	29.00	0.1.3.5.9.11/7/100	6	31.00
0.2.3.6.10/7/100	8	29.00	0.1.3.5.9.11/7/101	6	31.00
0.2.3.6.10/7/101	8	29.00	0.1.3.5.9.11/7/110	6	31.00
0.2.3.6.10/7/110	8	29.00	0.1.3.5.9.11/7/001	6	31.00
0.2.3.6.10/7/111	8	29.00	0.1.3.5.9.11/7/111	6	31.00
0.3.6.10.11/2/010	7	30.00	0.1.3.5.9.11/7/000	6	31.00
0.3.6.10.11/2/100	7	30.00	0.1.3.5.9.11/7/011	6	31.00
0.3.6.10.11/2/011	7	30.00	0.1.3.5.9.11/7/010	6	31.00
0.3.6.10.11/2/101	7	30.00	0.1.4.8.9.10/11/100	5	27.00
0.3.6.10.11/2/110	7	30.00	0.1.4.8.9.10/11/101	5	27.00
0.3.6.10.11/2/001	7	30.00	0.1.4.8.9.10/11/111	5	27.00
0.3.6.10.11/2/111	7	30.00	0.1.4.8.9.10/11/110	5	27.00
0.3.6.10.11/2/000	7	30.00	0.1.4.8.9.10/11/001	5	27.00
0.3.9.10.11/6/011	2	32.00	0.1.4.8.9.10/11/000	5	27.00
0.3.9.10.11/6/101	2	32.00	0.1.4.8.9.10/11/010	5	27.00
0.3.9.10.11/6/010	2	32.00	0.1.4.8.9.10/11/011	5	27.00

0.2.3.6.7.10/8/000	4	23.00	0.2.3.4.6.7.8/9/111	11	19.00
0.2.3.6.7.10/8/111	4	23.00	0.2.3.4.6.7.8/9/100	11	19.00
0.2.3.6.7.10/8/101	4	23.00	0.2.3.4.6.7.8/9/101	11	19.00
0.2.3.6.7.10/8/001	4	23.00	0.1.3.5.7.9.11/6/011	2	27.00
0.2.3.6.7.10/8/100	4	23.00	0.1.3.5.7.9.11/6/110	2	27.00
0.2.3.6.7.10/8/110	4	23.00	0.1.3.5.7.9.11/6/010	2	27.00
0.2.3.6.7.10/8/010	4	23.00	0.1.3.5.7.9.11/6/111	2	27.00
0.2.3.6.7.10/8/011	4	23.00	0.1.3.5.7.9.11/6/100	2	27.00
0.2.3.6.10.11/7/111	8	28.00	0.1.3.5.7.9.11/6/101	2	27.00
0.2.3.6.10.11/7/101	8	28.00	0.1.3.5.7.9.11/6/001	2	27.00
0.2.3.6.10.11/7/110	8	28.00	0.1.3.5.7.9.11/6/000	2	27.00
0.2.3.6.10.11/7/100	8	28.00	0.1.4.8.9.10.11/5/110	2	24.00
0.2.3.6.10.11/7/000	8	28.00	0.1.4.8.9.10.11/5/111	2	24.00
0.2.3.6.10.11/7/011	8	28.00	0.1.4.8.9.10.11/5/100	2	24.00
0.2.3.6.10.11/7/010	8	28.00	0.1.4.8.9.10.11/5/010	2	24.00
0.2.3.6.10.11/7/001	8	28.00	0.1.4.8.9.10.11/5/011	2	24.00
0.3.6.9.10.11/2/010	7	26.00	0.1.4.8.9.10.11/5/101	2	24.00
0.3.6.9.10.11/2/011	7	26.00	0.1.4.8.9.10.11/5/000	2	24.00
0.3.6.9.10.11/2/110	7	26.00	0.1.4.8.9.10.11/5/001	2	24.00
0.3.6.9.10.11/2/001	7	26.00	0.2.3.6.7.10.11/8/111	4	22.00
0.3.6.9.10.11/2/111	7	26.00	0.2.3.6.7.10.11/8/101	4	22.00
0.3.6.9.10.11/2/101	7	26.00	0.2.3.6.7.10.11/8/100	4	22.00
0.3.6.9.10.11/2/100	7	26.00	0.2.3.6.7.10.11/8/001	4	22.00
0.3.6.9.10.11/2/000	7	26.00	0.2.3.6.7.10.11/8/000	4	22.00
0.2.3.6.7.8.10/4/101	9	16.00	0.2.3.6.7.10.11/8/011	4	22.00
0.2.3.6.7.8.10/4/100	9	16.00	0.2.3.6.7.10.11/8/110	4	22.00
0.2.3.6.7.8.10/4/001	9	16.00	0.2.3.6.7.10.11/8/010	4	22.00
0.2.3.6.7.8.10/4/000	9	16.00	0.2.3.6.9.10.11/7/111	8	24.00
0.2.3.6.7.8.10/4/011	9	16.00	0.2.3.6.9.10.11/7/110	8	24.00
0.2.3.6.7.8.10/4/010	9	16.00	0.2.3.6.9.10.11/7/010	8	24.00
0.2.3.6.7.8.10/4/111	9	16.00	0.2.3.6.9.10.11/7/101	8	24.00
0.2.3.6.7.8.10/4/110	9	16.00	0.2.3.6.9.10.11/7/100	8	24.00
0.2.3.6.7.8.11/5/100	1	18.00	0.2.3.6.9.10.11/7/011	8	24.00
0.2.3.6.7.8.11/5/101	1	18.00	0.2.3.6.9.10.11/7/001	8	24.00
0.2.3.6.7.8.11/5/111	1	18.00	0.2.3.6.9.10.11/7/000	8	24.00
0.2.3.6.7.8.11/5/110	1	18.00	0.2.3.4.6.7.8.10/9/001	11	12.00
0.2.3.6.7.8.11/5/001	1	18.00	0.2.3.4.6.7.8.10/9/000	11	12.00
0.2.3.6.7.8.11/5/010	1	18.00	0.2.3.4.6.7.8.10/9/011	11	12.00
0.2.3.6.7.8.11/5/000	1	18.00	0.2.3.4.6.7.8.10/9/111	11	12.00
0.2.3.6.7.8.11/5/011	1	18.00	0.2.3.4.6.7.8.10/9/110	11	12.00
0.2.3.4.6.7.8/9/110	11	19.00	0.2.3.4.6.7.8.10/9/100	11	12.00
0.2.3.4.6.7.8/9/000	11	19.00	0.2.3.4.6.7.8.10/9/101	11	12.00
0.2.3.4.6.7.8/9/001	11	19.00	0.2.3.4.6.7.8.10/9/010	11	12.00
0.2.3.4.6.7.8/9/011	11	19.00	0.2.3.5.6.7.8.11/1/010	9	15.00
0.2.3.4.6.7.8/9/010	11	19.00	0.2.3.5.6.7.8.11/1/111	9	15.00

0.2.3.5.6.7.8.11/1/110	9	15.00	0.2.3.6.7.9.10.11/8/010	4	18.00
0.2.3.5.6.7.8.11/1/100	9	15.00	0.2.3.4.6.7.8.9.10/11/100	5	9.00
0.2.3.5.6.7.8.11/1/011	9	15.00	0.2.3.4.6.7.8.9.10/11/101	5	9.00
0.2.3.5.6.7.8.11/1/000	9	15.00	0.2.3.4.6.7.8.9.10/11/000	5	9.00
0.2.3.5.6.7.8.11/1/001	9	15.00	0.2.3.4.6.7.8.9.10/11/001	5	9.00
0.2.3.5.6.7.8.11/1/101	9	15.00	0.2.3.4.6.7.8.9.10/11/010	5	9.00
0.2.3.4.6.7.8.9/11/100	5	16.00	0.2.3.4.6.7.8.9.10/11/011	5	9.00
0.2.3.4.6.7.8.9/11/101	5	16.00	0.2.3.4.6.7.8.9.10/11/111	5	9.00
0.2.3.4.6.7.8.9/11/011	5	16.00	0.2.3.4.6.7.8.9.10/11/110	5	9.00
0.2.3.4.6.7.8.9/11/010	5	16.00	0.1.2.3.5.6.7.8.11/9/001	10	12.00
0.2.3.4.6.7.8.9/11/110	5	16.00	0.1.2.3.5.6.7.8.11/9/100	10	12.00
0.2.3.4.6.7.8.9/11/111	5	16.00	0.1.2.3.5.6.7.8.11/9/101	10	12.00
0.2.3.4.6.7.8.9/11/001	5	16.00	0.1.2.3.5.6.7.8.11/9/000	10	12.00
0.2.3.4.6.7.8.9/11/000	5	16.00	0.1.2.3.5.6.7.8.11/9/010	10	12.00
0.1.3.5.6.7.9.11/2/011	10	21.00	0.1.2.3.5.6.7.8.11/9/011	10	12.00
0.1.3.5.6.7.9.11/2/100	10	21.00	0.1.2.3.5.6.7.8.11/9/110	10	12.00
0.1.3.5.6.7.9.11/2/101	10	21.00	0.1.2.3.5.6.7.8.11/9/111	10	12.00
0.1.3.5.6.7.9.11/2/010	10	21.00	0.2.3.4.6.7.8.9.11/5/001	10	13.00
0.1.3.5.6.7.9.11/2/000	10	21.00	0.2.3.4.6.7.8.9.11/5/000	10	13.00
0.1.3.5.6.7.9.11/2/111	10	21.00	0.2.3.4.6.7.8.9.11/5/010	10	13.00
0.1.3.5.6.7.9.11/2/110	10	21.00	0.2.3.4.6.7.8.9.11/5/011	10	13.00
0.1.3.5.6.7.9.11/2/001	10	21.00	0.2.3.4.6.7.8.9.11/5/101	10	13.00
0.1.4.5.8.9.10.11/2/010	7	15.00	0.2.3.4.6.7.8.9.11/5/100	10	13.00
0.1.4.5.8.9.10.11/2/110	7	15.00	0.2.3.4.6.7.8.9.11/5/111	10	13.00
0.1.4.5.8.9.10.11/2/111	7	15.00	0.2.3.4.6.7.8.9.11/5/110	10	13.00
0.1.4.5.8.9.10.11/2/011	7	15.00	0.1.2.3.5.6.7.9.11/10/110	8	15.00
0.1.4.5.8.9.10.11/2/100	7	15.00	0.1.2.3.5.6.7.9.11/10/111	8	15.00
0.1.4.5.8.9.10.11/2/000	7	15.00	0.1.2.3.5.6.7.9.11/10/000	8	15.00
0.1.4.5.8.9.10.11/2/001	7	15.00	0.1.2.3.5.6.7.9.11/10/001	8	15.00
0.1.4.5.8.9.10.11/2/101	7	15.00	0.1.2.3.5.6.7.9.11/10/100	8	15.00
0.2.3.6.7.8.10.11/4/000	5	15.00	0.1.2.3.5.6.7.9.11/10/011	8	15.00
0.2.3.6.7.8.10.11/4/001	5	15.00	0.1.2.3.5.6.7.9.11/10/010	8	15.00
0.2.3.6.7.8.10.11/4/011	5	15.00	0.1.2.3.5.6.7.9.11/10/101	8	15.00
0.2.3.6.7.8.10.11/4/010	5	15.00	0.1.2.4.5.8.9.10.11/7/100	6	13.00
0.2.3.6.7.8.10.11/4/110	5	15.00	0.1.2.4.5.8.9.10.11/7/101	6	13.00
0.2.3.6.7.8.10.11/4/111	5	15.00	0.1.2.4.5.8.9.10.11/7/000	6	13.00
0.2.3.6.7.8.10.11/4/100	5	15.00	0.1.2.4.5.8.9.10.11/7/110	6	13.00
0.2.3.6.7.8.10.11/4/101	5	15.00	0.1.2.4.5.8.9.10.11/7/001	6	13.00
0.2.3.6.7.9.10.11/8/001	4	18.00	0.1.2.4.5.8.9.10.11/7/111	6	13.00
0.2.3.6.7.9.10.11/8/000	4	18.00	0.1.2.4.5.8.9.10.11/7/011	6	13.00
0.2.3.6.7.9.10.11/8/101	4	18.00	0.1.2.4.5.8.9.10.11/7/010	6	13.00
0.2.3.6.7.9.10.11/8/100	4	18.00	0.2.3.4.6.7.8.10.11/5/010	1	10.00
0.2.3.6.7.9.10.11/8/111	4	18.00	0.2.3.4.6.7.8.10.11/5/101	1	10.00
0.2.3.6.7.9.10.11/8/110	4	18.00	0.2.3.4.6.7.8.10.11/5/100	1	10.00
0.2.3.6.7.9.10.11/8/011	4	18.00	0.2.3.4.6.7.8.10.11/5/011	1	10.00

0.2.3.4.6.7.8.10.11/5/001	1	10.00	0.1.2.4.5.7.8.9.10.11/6/110	3	9.00
0.2.3.4.6.7.8.10.11/5/000	1	10.00	0.1.2.4.5.7.8.9.10.11/6/111	3	9.00
0.2.3.4.6.7.8.10.11/5/111	1	10.00	0.1.2.4.5.7.8.9.10.11/6/100	3	9.00
0.2.3.4.6.7.8.10.11/5/110	1	10.00	0.1.2.4.5.7.8.9.10.11/6/101	3	9.00
0.2.3.6.7.8.9.10.11/4/110	5	11.00	0.1.2.4.5.7.8.9.10.11/6/000	3	9.00
0.2.3.6.7.8.9.10.11/4/111	5	11.00	0.1.2.4.5.7.8.9.10.11/6/001	3	9.00
0.2.3.6.7.8.9.10.11/4/101	5	11.00	0.2.3.4.5.6.7.8.10.11/1/100	9	7.00
0.2.3.6.7.8.9.10.11/4/011	5	11.00	0.2.3.4.5.6.7.8.10.11/1/101	9	7.00
0.2.3.6.7.8.9.10.11/4/010	5	11.00	0.2.3.4.5.6.7.8.10.11/1/001	9	7.00
0.2.3.6.7.8.9.10.11/4/100	5	11.00	0.2.3.4.5.6.7.8.10.11/1/000	9	7.00
0.2.3.6.7.8.9.10.11/4/001	5	11.00	0.2.3.4.5.6.7.8.10.11/1/111	9	7.00
0.2.3.6.7.8.9.10.11/4/000	5	11.00	0.2.3.4.5.6.7.8.10.11/1/011	9	7.00
0.2.3.4.6.7.8.9.10.11/5/111	1	6.00	0.2.3.4.5.6.7.8.10.11/1/110	9	7.00
0.2.3.4.6.7.8.9.10.11/5/101	1	6.00	0.2.3.4.5.6.7.8.10.11/1/010	9	7.00
0.2.3.4.6.7.8.9.10.11/5/100	1	6.00	0.2.3.4.5.6.7.8.9.10.11/1/010	0	3.00
0.2.3.4.6.7.8.9.10.11/5/110	1	6.00	0.2.3.4.5.6.7.8.9.10.11/1/011	0	3.00
0.2.3.4.6.7.8.9.10.11/5/001	1	6.00	0.2.3.4.5.6.7.8.9.10.11/1/000	0	3.00
0.2.3.4.6.7.8.9.10.11/5/010	1	6.00	0.2.3.4.5.6.7.8.9.10.11/1/001	0	3.00
0.2.3.4.6.7.8.9.10.11/5/011	1	6.00	0.2.3.4.5.6.7.8.9.10.11/1/101	0	3.00
0.2.3.4.6.7.8.9.10.11/5/000	1	6.00	0.2.3.4.5.6.7.8.9.10.11/1/100	0	3.00
0.1.2.3.5.6.7.8.9.11/10/011	4	7.00	0.2.3.4.5.6.7.8.9.10.11/1/110	0	3.00
0.1.2.3.5.6.7.8.9.11/10/100	4	7.00	0.2.3.4.5.6.7.8.9.10.11/1/111	0	3.00
0.1.2.3.5.6.7.8.9.11/10/101	4	7.00	0.1.2.3.5.6.7.8.9.10.11/4/010	0	4.00
0.1.2.3.5.6.7.8.9.11/10/010	4	7.00	0.1.2.3.5.6.7.8.9.10.11/4/011	0	4.00
0.1.2.3.5.6.7.8.9.11/10/111	4	7.00	0.1.2.3.5.6.7.8.9.10.11/4/000	0	4.00
0.1.2.3.5.6.7.8.9.11/10/110	4	7.00	0.1.2.3.5.6.7.8.9.10.11/4/001	0	4.00
0.1.2.3.5.6.7.8.9.11/10/001	4	7.00	0.1.2.3.5.6.7.8.9.10.11/4/110	0	4.00
0.1.2.3.5.6.7.8.9.11/10/000	4	7.00	0.1.2.3.5.6.7.8.9.10.11/4/100	0	4.00
0.2.3.4.5.6.7.8.9.11/10/111	1	7.00	0.1.2.3.5.6.7.8.9.10.11/4/111	0	4.00
0.2.3.4.5.6.7.8.9.11/10/000	1	7.00	0.1.2.3.5.6.7.8.9.10.11/4/101	0	4.00
0.2.3.4.5.6.7.8.9.11/10/101	1	7.00	0.1.2.4.5.6.7.8.9.10.11/3/111	0	6.00
0.2.3.4.5.6.7.8.9.11/10/001	1	7.00	0.1.2.4.5.6.7.8.9.10.11/3/011	0	6.00
0.2.3.4.5.6.7.8.9.11/10/100	1	7.00	0.1.2.4.5.6.7.8.9.10.11/3/110	0	6.00
0.2.3.4.5.6.7.8.9.11/10/110	1	7.00	0.1.2.4.5.6.7.8.9.10.11/3/010	0	6.00
0.2.3.4.5.6.7.8.9.11/10/010	1	7.00	0.1.2.4.5.6.7.8.9.10.11/3/101	0	6.00
0.2.3.4.5.6.7.8.9.11/10/011	1	7.00	0.1.2.4.5.6.7.8.9.10.11/3/100	0	6.00
0.1.2.3.5.6.7.9.10.11/8/001	4	11.00	0.1.2.4.5.6.7.8.9.10.11/3/000	0	6.00
0.1.2.3.5.6.7.9.10.11/8/000	4	11.00	0.1.2.4.5.6.7.8.9.10.11/3/001	0	6.00
0.1.2.3.5.6.7.9.10.11/8/110	4	11.00	0.1.2.3.4.5.6.7.8.10.11/9/101	0	4.00
0.1.2.3.5.6.7.9.10.11/8/111	4	11.00	0.1.2.3.4.5.6.7.8.10.11/9/010	0	4.00
0.1.2.3.5.6.7.9.10.11/8/100	4	11.00	0.1.2.3.4.5.6.7.8.10.11/9/011	0	4.00
0.1.2.3.5.6.7.9.10.11/8/011	4	11.00	0.1.2.3.4.5.6.7.8.10.11/9/001	0	4.00
0.1.2.3.5.6.7.9.10.11/8/010	4	11.00	0.1.2.3.4.5.6.7.8.10.11/9/100	0	4.00
0.1.2.3.5.6.7.9.10.11/8/101	4	11.00	0.1.2.3.4.5.6.7.8.10.11/9/000	0	4.00
0.1.2.4.5.7.8.9.10.11/6/010	3	9.00	0.1.2.3.4.5.6.7.8.10.11/9/110	0	4.00
0.1.2.4.5.7.8.9.10.11/6/011	3	9.00	0.1.2.3.4.5.6.7.8.10.11/9/111	0	4.00

APPENDIX C

APPENDICES IN CHAPTER 3

C.1 Facet Reduction

Returning to the general $S \geq 1$ case and to Proposition 2, we note that $\omega(n, \vec{x}, z)$ is described by $N - n + 1$ faces. We remark that a face $\vec{x}\vec{m}(z)$ is unnecessary to describe $\omega(\cdot)$ if there is another face $\vec{x}\vec{m}'(z)$ such that $\vec{x}\vec{m}(z) \leq \vec{x}\vec{m}'(z)$ for all $\vec{x} \in \mathcal{X}$. Thus, if $\vec{m}(z) \leq \vec{m}'(z)$, then $\vec{x}\vec{m}(z)$ is an unnecessary face and hence can be eliminated.

The proofs of the following result follow from straightforward application of the definition of $\vec{m}(\rho, n, z)$ and Proposition 4.

Lemma 5 *For* $0 \leq n + 1 \leq \rho \leq N - 1$, $\vec{m}_s(\rho, n, z) \leq \vec{m}_s(\rho - 1, n, z)$ *for* $s = 0, 1, \ldots, S$ *is equivalent to*

$$c(\rho, \rho + 1) + \beta \vec{e}_s P^{\rho+1-n} \vec{\eta}(\rho + 1, z) \leq \vec{e}_s P^{\rho-n} \vec{\eta}(\rho, z) + M \frac{1 - \beta}{\beta}.$$

Additionally, if $p_{SS} = 1$, $\vec{m}_S(\rho, n, z) \leq \vec{m}_S(\rho - 1, n, z)$ *is equivalent to*

$$c(\rho, \rho + 1) + \beta l_S(\rho + 1, z) \leq l_S(\rho, z) + M \frac{1 - \beta}{\beta}.$$

When $\rho = N$, $\vec{e}_s P^{N-n} \vec{m}(N, N, z) = \vec{e}_s P^{N-n} \vec{\eta}(N, z) > \vec{e}_s P^{N-n} \vec{\eta}(N, z) + \frac{M}{\beta}$ for $M < 0$ and all $s = 0, 1, \ldots, S$, which implies that $\vec{m}(N - 1, n, z) < \vec{m}(N, n, z)$ for all $n = 0, 1, \ldots, N-1$. For $S = 1$, $\vec{m}_1(N, N, z) > l_1(N, z) + \frac{M}{\beta}$, which also implies that $\vec{m}_1(N, n, z) > \vec{m}_1(N - 1, n, z)$, and Lemma 5 and Proposition 5 imply that $\vec{m}(N - 1, n, z) \leq \vec{m}(N, n, z)$ for all $n \leq N - 1$. Hence, for $M < 0$ and $S \geq 1$, all the faces $\vec{x}\vec{m}(N - 1, n, z)$ for all n are unnecessary. We observe, however, for sufficiently small $|M|(1 - \beta)/\beta$, $\vec{m}_S(\rho, n, z) \leq \vec{m}_S(\rho - 1, n, z)$ and hence, it is unlikely to hold that $\vec{m}(\rho, n, z) > \vec{m}(\rho - 1, n, z)$. Thus, with the exception of the $\rho = N - 1$ case, it is unlikely that faces can be eliminated because $\vec{m}_S(\rho, n, z) \leq \vec{m}_S(\rho - 1, n, z)$. We further remark that procedures are presented in Lin et al. [40] to identify and eliminate unnecessary faces for general POMDPs.

C.2 Counterexample that $\pi^*(n,\vec{x}') = \mathcal{NI}$ if $\pi^*(n,\vec{x}) = \mathcal{NI}$ and $\vec{x} \prec \vec{x}'$

We present an example where $\pi^*(n,\vec{x}) = \mathcal{NI}$, $\vec{x} \prec \vec{x}'$, and $\pi^*(n,\vec{x}') = \mathcal{I}$. Thus, there are situations where increased product freshness implies a greater need to inspect. Consider the following parameter values:

$$\beta = 0.99, \quad N = 10, \quad S = 4, \quad M = -8, \quad W = -300,$$

$$P = \begin{bmatrix} 0.8 & 0.1 & 0.1 & 0.0 & 0.0 \\ 0.0 & 0.8 & 0.2 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.7 & 0.3 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.7 & 0.3 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix}, \quad c(n,n+1) = \begin{cases} -20, & \text{if } n = 1 \\ -15, & \text{if } n = 3 \\ -40, & \text{if } n = 6 \\ -10, & \text{otherwise} \end{cases},$$

$$\vec{x}^* = (0.4, 0.3, 0.3, 0.0, 0.0), \quad \vec{R} = (1600, 1500, 1400, 1300, 0)^T,$$

$$\vec{l}_n^x = (30, 20, 15, 10, 0)^T, \quad l_n = \sum_{k=0}^{n-1} \beta^{n-1-k} c(k,k+1), \text{ and } \beta_n = \beta^n.$$

Suppose that

$$\vec{x}_1 = \begin{bmatrix} 0.00 & 0.00 & 0.00 & 0.00 & 1.00 \end{bmatrix}$$
$$\vec{x}_2 = \begin{bmatrix} 0.00 & 0.01 & 0.69 & 0.25 & 0.05 \end{bmatrix}$$
$$\vec{x}_3 = \begin{bmatrix} 0.00 & 0.25 & 0.45 & 0.25 & 0.05 \end{bmatrix}$$
$$\vec{x}_4 = \begin{bmatrix} 0.20 & 0.15 & 0.55 & 0.05 & 0.05 \end{bmatrix}$$

and note $\vec{x}_1 \prec \vec{x}_2 \prec \vec{x}_3 \prec \vec{x}_4$. Then, for $n = 6$,

$$\pi^*(6,\vec{x}) = \begin{cases} \mathcal{A} & \text{if } \vec{x} = \vec{x}_1 \\ \mathcal{NI} & \text{if } \vec{x} = \vec{x}_2 \\ \mathcal{I} & \text{if } \vec{x} = \vec{x}_3 \\ \mathcal{NI} & \text{if } \vec{x} = \vec{x}_4 \end{cases},$$

We note that A1 holds for all n.

This example may indicate a possible scenario that a freight which is less fresher than a certain quality threshold may not be worth to have a costly inspection although it is still expected to generate more revenue at the destination than at the secondary market near by the current location. However, fresher freight may be more valuable and thus warrant the cost of an inspection.

C.3 Conditions on $\omega_L(n,\vec{x}) = \omega_{LL}(n,\vec{x})$

We present an inequality that the reward vector $\vec{R}$ must satisfy in order that the inequality (8) holds when $n = 1$, and hence $\omega_{LL}(n,\vec{x}) = \omega_L(n,\vec{x})$ for all n and $\vec{x}$. It is straightforward to show that $\omega_{LL}(1,\vec{x}^*P) \geq l_1 + \vec{x}^*P\vec{l}_1^x + \beta_1\omega_{LL}(0,\vec{x}^*)$ is equivalent to

$$\left[\frac{\beta^{N-1}(1-\beta\beta_1)}{(1-\beta^N\beta_N)}\right]\vec{x}^*P^N\vec{R} \geq -\sum_{k=1}^{N-1}\beta^{k-1}c(k,k+1) - \beta^{N-1}[l_N + \vec{x}^*P^N\vec{l}_N^x + \beta_N\zeta_L]$$
$$+ l_1 + \vec{x}^*P\vec{l}_1^x + \beta_1\zeta_L,$$

where $\omega_{LL}(0,\vec{x}^*) = \zeta_L + \vec{\xi}_L\vec{R}$ such that $\zeta_L = \Big\{W + \sum_{n=0}^{N-1}\beta^n c(n,n+1) + \beta^N[l_N + \vec{x}^*P^N\vec{l}_N^x]\Big\}/(1-\beta^N\beta_N)$ and $\vec{\xi}_L = \beta^N\vec{x}^*P^N/(1-\beta^N\beta_N)$.

We remark that a reward structure that guarantees the "always travel from origin to destination" policy is optimal when there is no opportunity to inspect the product in-transit does not guarantee profitability. We now seek conditions that guarantee $\omega(n,\vec{x}) \geq 0$ for all n and $\vec{x}$, and note that $\omega(n,\vec{x}) \geq \omega_L(n,\vec{x}) \geq \vec{x}\vec{l}(n,\omega_L(0,\vec{x}^*))$. If, given $z = \omega_L(0,\vec{x}^*)$, $\vec{e}_S\vec{l}(n,z) \geq \vec{e}_S\vec{l}(n+1,z)$ for $n = 1,2,\ldots,N-1$ (i.e., for totally spoiled freight, the corresponding cost incurred to select an action $\mathcal{A}$ at a location is non-decreasing as the location at which action $\mathcal{A}$ is made gets closer to the destination), then $\vec{x}\vec{l}(n,z) \geq \vec{e}_S\vec{l}(n,z) \geq \vec{e}_S\vec{l}(N,z)$. Our next result follows from this argument and the fact that the inequality $\vec{e}_S\vec{l}(N,\omega_L(0,\vec{x}^*)) \geq 0$ is equivalent to $\vec{\xi}_L\vec{R} \geq -\zeta_L - (l_N + \vec{e}_S\vec{l}_N^x)/\beta_N$.

Proposition 15 *Let $z = \zeta_L + \vec{\xi}_L\vec{R} \geq 0$. If $l_S(n,z) \geq l_S(n+1,z)$ for $n = 1,2,\ldots,N-1$, and $\vec{\xi}_L\vec{R} \geq -\zeta_L - (l_N + \vec{e}_S\vec{l}_N^x)/\beta_N$, then $\omega(n,\vec{x}) \geq 0$ for all n and $\vec{x}$.*

C.4 Closed-form Formula for the Upper Bound Case When P is Upper-triangular

We now present a closed-form formula for the upper bound of $\omega(n, \vec{x})$ when P is upper-triangular. Thus, we restrict our attention to policies having the structural properties found in Proposition 9. Let $\{\overline{s}(n) : n = 1, \ldots, N-1\}$ be given, where $\overline{s}(n) \leq \overline{s}(n+1)$ for all n and where the policy under consideration, $\overline{\pi}$, is such that $\overline{\pi}(n, \vec{e}_s) = \mathcal{A}(\mathcal{I})$ if $s \geq \overline{s}(n)$ ($s < \overline{s}(n)$). Let $\vec{\omega}_U$ be the resulting reward function vector.

Let E_U, F_U and G_U be such that $\vec{\omega}_U(n, z_U) = E_U(n) + F_U(n)\vec{R} + z_U G_U(n)$ where $z_U = \omega_U(0, x^*)$ under the policy $\overline{\pi}$ above. Note that $E_U(N) = l_N\vec{e}$, $F_U(N) = I$, and $G_U(N) = \beta_N\vec{e}$. Define the diagonal $(S+1) \times (S+1)$ matrix $I(n)$ as having 1 (0) as its (s, s)-th entry if $s < \overline{s}(n)$ ($s \geq \overline{s}(n)$), and let $I'(n) = I - I(n)$. It is then straightforward to show that

$$E_U(n) = c(n, n+1)\vec{e} + \beta P I'(n+1)(\vec{l}^{x}_{n+1} + l_{n+1}\vec{e}) + \beta P I(n+1) E_U(n+1),$$

$$F_U(n) = \beta P I(n+1) F_U(n+1), \text{ and}$$

$$G_U(n) = \beta P \left[I(n+1) G_U(n+1) + \beta_{n+1} I'(n+1)\vec{e} \right]$$

where for $n = 0$, $W\vec{e}$ is added to $E_U(n)$.

Then, $z_U = \vec{x}^*(E_U(0) + F_U(0)\vec{R} + z_U G_U(0))$, and hence $z_U = \vec{x}^*(E_U(0) + F_U(0)\vec{R})/(1 - \vec{x}^* G_U(0))$, assuming $1 - \vec{x}^* G_U(0) \neq 0$. Therefore, a closed-form solution for z_U is given as $z_U = \zeta_U + \vec{\xi}_U \vec{R}$, where $\zeta_U = \vec{x}^* E_U(0)/(1 - \vec{x}^* G_U(0))$ and $\vec{\xi}_U = \vec{x}^* F_U(0)/(1 - \vec{x}^* G_u(0))$.

C.5 Bounds on the Productivity Measure by Inspection

As discussed in Section 3.5.4, with respect to the value of inspection information as a function of M, consider Figure 2, which is a graph of $\omega(0, \vec{x}^*; M)$ as a function of M. It is straightforward to show that $\omega(0, \vec{x}^*; M)$ is piecewise linear, convex and isotone in M. We recall that the ratio $(\omega_U(0, \vec{x}^*) - \omega_L(0, \vec{x}^*))/\omega_L(0, \vec{x}^*)$ represents an upper bound on the fraction of productivity gained by optimal freight inspection. Under the case when $S = 1$ and the suboptimal policies stated below are optimal for each lower and upper bound, we now show that this ratio, as a function of the reward vector $\vec{R}$, is antitone in general.

<u>**S=1 (2 states) Case**</u> We assume that the following suboptimal policies are optimal for lower and upper bounds:

(i) Lower bound: always travel directly from the origin to the destination

(ii) Upper bound: when in the fresh state ($s = 0$), proceed toward the destination and when in the spoiled state ($s = 1$), abort the trip (i.e., $\bar{s}(n) = 1$ for $n = 1, \ldots, N-1$).

According to Section 3.6.1 and C.4, $\omega_L(0, \vec{x}^*) = \vec{\xi}_L \vec{R} + \zeta_L$ and $\omega_U(0, \vec{x}^*) = \vec{\xi}_U \vec{R} + \zeta_U$. Moreover, since $\vec{R} = [R_0, 0]^T$, $\omega_L(0, \vec{x}^*) = (\vec{\xi}_L)_0 R_0 + \zeta_L$ and $\omega_U(0, x^*) = (\vec{\xi}_U)_0 R_0 + \zeta_L$ where $(\vec{\xi}_i)_0$ is the first entry of a vector $\vec{\xi}_i$, $i = L, U$. Let $E_L(0) = W + \sum_{n=0}^{N-1} \beta^n c(n, n+1) + \beta^N \left[\vec{x}^* P^N \vec{l}_N^x + l_N\right]$ so that $\zeta_L = E_L(0)/(1 - \beta^N \beta_N)$.

Lemma 6 *Assume (i) $S = 1$, (ii) $\vec{R}$ guarantees that the aforementioned suboptimal policies are optimal for lower and upper bounds respectively, and (iii) $\omega_L(0, \vec{x}^*) > 0$. Then, the ratio $(\omega_U(0, \vec{x}^*) - \omega_L(0, \vec{x}^*))/\omega_L(0, \vec{x}^*)$ is antitone in $\vec{R}$ if and only if $E_L(0) \le \vec{x}^* E_U(0)$.*

Proof: Define $\Delta_1 = \zeta_U - \zeta_L$ and $\Delta_2 = (\vec{\xi}_U)_0 - (\vec{\xi}_L)_0$, and note

$$\frac{\omega_U(0, \vec{x}^*) - \omega_L(0, \vec{x}^*)}{\omega_L(0, \vec{x}^*)} = \frac{\Delta_1 + \Delta_2 R_0}{(\vec{\xi}_L)_0 R_0 + \zeta_L},$$

which has as its derivative, with respect to R_0, the following:

$$\frac{\left[((\vec{\xi}_L)_0 R_0 + \zeta_L)\Delta_2 - (\Delta_1 + \Delta_2 R_0)\,(\vec{\xi}_L)_0\right]}{((\vec{\xi}_L)_0 R_0 + \zeta_L)^2} = \frac{\zeta_L (\vec{\xi}_U)_0 - \zeta_U (\vec{\xi}_L)_0}{((\vec{\xi}_L)_0 R_0 + \zeta_L)^2}.$$

We now show that the above numerator is negative, i.e., $\zeta_L (\vec{\xi}_U)_0 \le \zeta_U (\vec{\xi}_L)_0$, or equivalently $\frac{E_L(0)}{x_0^* \gamma^N} \le \frac{\vec{x}^* E_U(0)}{(x^* F_U(0))_0}$. It is straightforward to show that $(\vec{x}^* F_U(0))_0 = x_0^* \gamma^N$. Hence, the derivative of $(\omega_U(0, \vec{x}^*) - \omega_L(0, \vec{x}^*))/\omega_L(0, \vec{x}^*)$ with respect to $\vec{R}$, is negative if and only if $E_L(0) \le \vec{x}^* E_U(0)$. ∎

We now present circumstances for which $E_L(0) \le \vec{x}^* E_U(0)$.

Lemma 7 *Assume that $l_n + \vec{e}_1 \vec{l}_n^x \ge l_{n+1} + \vec{e}_1 \vec{l}_{n+1}^x$ for all $n = 1, \ldots, N-1$. If (i) $x^* = [1, 0]$ and $\alpha = 0$, or (ii) $\beta = 1$, then $E_L(0) \le \vec{x}^* E_U(0)$.*

Proof: First, it is straightforward to see that

$$(E_U(n))_0 = \begin{cases} l_N + \vec{e}_0 \vec{l}_N^x & \text{if } n = N \\ c(N-1,N) + \beta l_N + \beta \left[(1-\alpha)\vec{e}_0 \vec{l}_N^x + \alpha \vec{e}_1 \vec{l}_N^x\right] & \text{if } n = N-1 \\ c(n,n+1) + \alpha\beta(l_{n+1} + \vec{e}_1 \vec{l}_{n+1}^x) + \beta(1-\alpha)\left(E_U(n+1)\right)_0 & \text{if } 1 \le n < N-1 \end{cases}$$

and

$$\vec{x}^* E_U(0) = W + c(0,1) + \beta(1 - x_0^*(1-\alpha))(l_1 + \vec{e}_1 \vec{l}_1^x) + \beta x_0^*(1-\alpha)\left(E_U(1)\right)_0$$

since $x^* = [x_0^*, 1-x_0^*]$, $P = \begin{bmatrix} 1-\alpha & \alpha \\ 0 & 1 \end{bmatrix}$ and $I(n) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ for $n = 1, \ldots, N-1$. Observe that, for $1 \le n < N-1$,

$$(E_U(n))_0 = \sum_{k=n}^{N-1} \gamma^{k-n} c(k,k+1) + \alpha\beta \sum_{k=n}^{N-1} \gamma^{k-n}(l_{k+1} + \vec{e}_1 \vec{l}_{k+1}^x) + (1-\alpha)\beta\gamma^{N-1-n}(l_N + \vec{e}_0 \vec{l}_N^x)$$

and, for $n = 0$,

$$E_L(0) = W + \sum_{k=0}^{N-1} \beta^k c(k,k+1) + \beta^N \left[l_N + x_0^*(1-\alpha)^N \vec{e}_0 \vec{l}_N^x + (1 - x_0^*(1-\alpha)^N)\vec{e}_1 \vec{l}_N^x\right],$$

$$\begin{aligned} \vec{x}^* E_U(0) = W &+ \left\{c(0,1) + x_0^* \sum_{k=1}^{N-1} \gamma^k c(k,k+1)\right\} + \Big\{\beta(1 - x_0^*(1-\alpha))(l_1 + \vec{e}_1 \vec{l}_1^x) \\ &+ x_0^*\Big[\alpha\beta \sum_{k=1}^{N-1} \gamma^k (l_{k+1} + \vec{e}_1 \vec{l}_{k+1}^x) + (1-\alpha)\beta\gamma^{N-1}(l_N + \vec{e}_0 \vec{l}_N^x)\Big]\Big\}. \end{aligned}$$

Thus,

(i) If $\vec{x}^* = [1,0]$ and $\alpha = 0$, then

$$\begin{aligned} E_L(0) &= W + \sum_{k=0}^{N-1} \beta^k c(k,k+1) + \beta^N \left[l_N + \vec{e}_0 \vec{l}_N^x\right] \\ \vec{x}^* E_U(0) &= W + \left\{c(0,1) + \sum_{k=1}^{N-1} \beta^k c(k,k+1)\right\} + \beta^N \left[l_N + \vec{e}_0 \vec{l}_N^x\right] \\ &= W + \sum_{k=0}^{N-1} \beta^k c(k,k+1) + \beta^N \left[l_N + \vec{e}_0 \vec{l}_N^x\right] \end{aligned}$$

and hence, $E_L(0) = \vec{x}^* E_U(0)$.

(ii) If $\beta = 1$, for $n = 0$,

$$E_L(0) = W + \sum_{k=0}^{N-1} c(k,k+1) + [l_N + x_0^*(1-\alpha)^N \vec{e}_0 \vec{l}_N^x + (1 - x_0^*(1-\alpha)^N)\vec{e}_1 \vec{l}_N^x], \text{ and}$$

$$\vec{x}^* E_U(0) = W + \left\{ c(0,1) + x_0^* \sum_{k=1}^{N-1} (1-\alpha)^k c(k,k+1) \right\} + \left\{ (1 - x_0^*(1-\alpha))(l_1 + \vec{e}_1 \vec{l}_1^x) \right.$$
$$\left. + x_0^* \Big[(1-\alpha)^N (l_N + \vec{e}_0 \vec{l}_N^x) + \alpha \sum_{k=1}^{N-1} (1-\alpha)^k (l_{k+1} + \vec{e}_1 \vec{l}_{k+1}^x) \Big] \right\}$$

Thus,

$$\begin{aligned}
\vec{x}^* E_U(0) - E_L(0) &= -\sum_{k=1}^{N-1} ((1 - x_0^*(1-\alpha)^k) c(k,k+1)) + (1 - x_0^*(1-\alpha))(l_1 + \vec{e}_1 \vec{l}_1^x) \\
&\quad + x_0^* \alpha \sum_{k=1}^{N-2} (1-\alpha)^k (\vec{l}_{k+1} + \vec{e}_1 \vec{l}_{k+1}^x) - (1 - x_0^*(1-\alpha)^{N-1})(l_N + \vec{e}_1 \vec{l}_N^x) \\
&\geq -\sum_{k=1}^{N-1} ((1 - x_0^*(1-\alpha)^k) c(k,k+1)) + (1 - x_0^*(1-\alpha))(l_N + \vec{e}_1 \vec{l}_N^x) \\
&\quad + x_0^* \alpha \sum_{k=1}^{N-2} (1-\alpha)^k (l_N + \vec{e}_1 \vec{l}_N^x) - (1 - x_0^*(1-\alpha)^{N-1})(l_N + \vec{e}_1 \vec{l}_N^x) \\
&= -\sum_{k=1}^{N-1} ((1 - x_0^*(1-\alpha)^k) c(k,k+1)) + (1 - x_0^*(1-\alpha)^{N-1})(l_N + \vec{e}_1 \vec{l}_N^x) \\
&\quad - (1 - x_0^*(1-\alpha)^{N-1})(l_N + \vec{e}_1 \vec{l}_N^x) \\
&= -\sum_{k=1}^{N-1} ((1 - x_0^*(1-\alpha)^k) c(k,k+1)) \geq 0
\end{aligned}$$

where the first and second inequalities are due to the assumption, and the third inequality is due to the fact that $c(n, n+1) < 0$. ∎

Thus, for β sufficiently close to 1, $E_L(0) \leq \vec{x}^* E_U(0)$ under reasonable conditions.

APPENDIX D

APPENDICES IN CHAPTER 4

D.1 Alternative Representation of The Single-period Cost Function

The single-period cost function (Equation (9)) in Section 4.2 can model overage and underage costs accrued at the end of each period. Observe that

- $\min\{x+\alpha, d\} = d - (d-x-\alpha)^+$ where $(a)^+ = \max\{a, 0\}$, and
- $(x+\alpha) = \max\{x+\alpha, d\} - \max\{d-x-\alpha, 0\} = \max\{x+\alpha-d, 0\} + d - \max\{d-x-\alpha, 0\} = d + (x+\alpha-d)^+ - (d-x-\alpha)^+$.

Therefore,

$$
\begin{aligned}
& ca + h\sum_{\alpha}(x+\alpha)P(\alpha|a) - p\sum_{\alpha}\min\{x+\alpha, d\}P(\alpha|a) \\
= \quad & ca + h\sum_{\alpha}\left[d + (x+\alpha-d)^+ - (d-x-\alpha)^+\right]P(\alpha|a) - p\sum_{\alpha}\left[d - (d-x-\alpha)^+\right]P(\alpha|a) \\
= \quad & \underbrace{-(p-h)d}_{(a)} + \left\{\underbrace{ca}_{(b)} + \underbrace{h\sum_{\alpha}(x+\alpha-d)^+P(\alpha|a)}_{(c)} + \underbrace{(p-h)\sum_{\alpha}(d-x-\alpha)^+P(\alpha|a)}_{(d)}\right\}.
\end{aligned}
$$

Note that:

- The term (a) is constant, consistent for each period throughout infinite horizon, and independent of a.
- The term (b) is the ordering cost, assuming that the retailer pays the wholesaler on the basis of the number of items ordered.
- The term (c) is the overage cost with unit holding cost h at the end of a period.
- The term (d) is the underage (shortage) cost with unit shortage penalty $s = (p-h)$ at the end of a period, and this cost term indeed corresponds to penalties for the lost sales.

D.2 Construction of $V = \mathcal{A} + \beta \mathcal{P} V$

Assuming $d \geq \gamma^L$, for each $\zeta \in \{0, 1, \ldots, \gamma^L\}$, there exists $l(\zeta) \in \{0, 1, 2, \ldots, L\}$ such that $z^{l(\zeta)-1} < d - \zeta \leq z^{l(\zeta)}$. Then,

$$\begin{aligned} v_\delta(d-\zeta) = f(d-\zeta, d-\zeta+\gamma^{l(\zeta)}) + \beta \sum_{\alpha=0}^{d-\zeta} P(\alpha | d-\zeta+\gamma^{l(\zeta)}) v_\delta(d) \\ + \beta P(d-\zeta+1 | d-\zeta+\gamma^{l(\zeta)}) v_\delta(d-1) + \cdots \\ + \beta P(d-\zeta+\gamma^{l(\zeta)} | d-\zeta+\gamma^{l(\zeta)}) v_\delta(d-\gamma^{l(\zeta)}), \end{aligned}$$

which suggests the following matrix formulation:

$$\underbrace{\begin{bmatrix} v_\delta(d) \\ \vdots \\ v_\delta(d-\gamma^L) \end{bmatrix}}_{=V} = \underbrace{\begin{bmatrix} f(d, d+\gamma^L) \\ \vdots \\ f(d-\gamma^L, d-\gamma^L+\gamma^{l(\gamma^L)}) \end{bmatrix}}_{=\mathcal{A}}$$

$$+ \beta \underbrace{\begin{bmatrix} \sum_{\alpha=0}^{d} P(\alpha | d+\gamma^L) & \cdots & P(d+\gamma^L | d+\gamma^L) \\ \vdots & \cdots & \vdots \\ \sum_{\alpha=0}^{d-\gamma^L} P(\alpha | d-\gamma^L+\gamma^{l(\gamma^L)}) & \cdots & . \end{bmatrix}}_{=\mathcal{P}} \underbrace{\begin{bmatrix} v_\delta(d) \\ \vdots \\ v_\delta(d-\gamma^L) \end{bmatrix}}_{=V}.$$

REFERENCES

[1] Ahuja, R., Magnanti, T., and Orlin, J., *Network Flows.* Prentice Hall, 1993.

[2] Applegate, D., Bixby, R., Chvatal, V., and Cook, W., *The Traveling Salesman Problem: A Computational Study.* Princeton University Press, 2006.

[3] Bagchi, A. and Mahanti, A., "Three approaches to heuristic search in networks," *Journal of the ACM*, vol. 32, pp. 1–27, 1985.

[4] Bander, J. and White, C., "A heuristic search approach for a nonstationary stochastic shortest path problem with terminal cost," *Transportation Science*, vol. 36, no. 2, pp. 218–230, 2002.

[5] Bartin, B. and Ozbay, K., "Determining the optimal configuration of highway routes for real-time traffic information: a case study," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 1, pp. 225–231, 2010.

[6] Bellman, R. and Roosta, M., "A stochastic travelling salesman problem," *Stochastic Analysis and Applications*, vol. 1, no. 2, pp. 159–161, 1983.

[7] Bolton, J. and Liu, W., "Creating an effective china cold supply chain – current status, challenges and implementation considerations." Accenture Report, 2006.

[8] Cai, X., Chen, J., Xiao, Y., and Xu, X., "Optimization and coordination of fresh product supply chains with freshness-keeping effort," *Production and Operations Management*, vol. 19, no. 3, pp. 261–278, 2010.

[9] Chakrabarti, P., Ghose, S., and DeSarkar, S., "Admissibility of ao* when heuristics overestimate," *Artificial Intelligence*, vol. 34, pp. 97–113, 1988.

[10] Chang, T.-S., Wan, Y.-W., and Ooi, W., "A stochastic dynamic traveling salesman problem with hard time windows," *European Journal of Operational Research*, vol. 198, no. 3, pp. 748–759, 2009.

[11] Christofides, N., "Bounds for the travelling-salesman problem," *Operations Research*, vol. 20, no. 5, pp. 1044–1056, 1972.

[12] Coyle, W., Hall, W., and Ballenger, N., "Transportation technology and the rising share of u.s. perishable food trade," in *Changing Structure of Global Food Consumption and Trade*, Economic Research Service - United States Department of Agriculture, 2001. WRS-01-1.

[13] ENS, "The environmental food crisis: A crisis of waste." Environment News Service, http://www.ens-newswire.com/ens/feb2009/2009-02-17-01.asp, 2009.

[14] FATUS, "Foreign agricultural trade of the united states (fatus)." http://www.ers.usda.gov/Data/FATUS/.

[15] Ferguson, M. and Ketzenberg, M., "Information sharing to improve retail product freshness of perishables," *Production and Operations Management*, vol. 15, no. 1, pp. 57–73, 2006.

[16] Garey, M. and Johnson, D., *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W.H. Freeman, 1979.

[17] GARPC, "The pharmaceutical industry in germany: Statistics." German Association of Research-based Pharmaceutical Companies, 2008.

[18] Gavirneni, S., "Sypply chain management at a chip tester manufacturer," in *The Practice of Supply Chain Management*, (Norwell, MA), Kluwer Academic Publishing, 2003.

[19] Gerchak, Y., Vickson, R., and Parlar, M., "Periodic review production models with variable yield and uncertain demand," *IIE Transactions*, vol. 20, pp. 144–150, 1998.

[20] Goyal, S. and Giri, B., "Recent trends in modeling of deteriorating inventory," *European Journal of Operational Research*, vol. 134, pp. 1–16, 2001.

[21] Grosfeld-Nir, A. and Gerchak, Y., "Multiple lot-sizing in production to order with random yields: Review of recent advances," *Annals of Operations Research*, vol. 126, pp. 43–69, 2004.

[22] Haghani, A. and Jung, S., "A dynamic vehicle routing problem with time-dependent travel times," *Computers and Operations Research*, vol. 32, no. 11, pp. 2959–2986, 2005.

[23] Hall, R., "The fastest path through a network with random time-dependent travel times," *Transportation Science*, vol. 20, no. 3, pp. 182–188, 1986.

[24] Hamrani, K. and Famos, T., "Genossenschaft migros neuchatel/fribourg." http://www.ameriscangroup.com/Case_Study_Migros.pdf, 2006.

[25] Harrington, L., "Temperature controlled logistics: Provide or perish." http://www.inboundlogistics.com/articles/features/0807_feature02.shtml, 2007.

[26] Hemmelmayr, V., Doerner, K., Hartl, R., and Savelsbergh, M., "Delivery strategies for blood products supplies," *OR Spectrum*, vol. 31, no. 4, pp. 707–725, 2009.

[27] Henig, M. and Gerchak, Y., "The structure of periodic review policies in the presence of random yield," *Operations Research*, vol. 38, no. 4, pp. 634–643, 1990.

[28] Ichoua, S., Gendreau, M., and Potvin, J.-Y., "Vehicle dispatching with time-dependent travel times," *European Journal of Operational Research*, vol. 144, pp. 379–396, 2003.

[29] Intelleflex, "Temperature management and the perishable cold chain: Improving quality, shelf-life and revenues with pallet-level monitoring." http://info.intelleflex.com/rs/intelleflex/images/Intelleflex-TemperatureManagement.pdf, 2010.

[30] Jula, H., Dessouky, M., and Ioannou, P., "Truck route planning in nonstationary stochastic networks with time windows at customer locations," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 1, pp. 51–62, 2006.

[31] Kang, Y. and Gershwin, S., "Information inaccuracy in inventory systems: stock loss and stockout," *IIE Transactions*, vol. 37, pp. 843–859, 2005.

[32] Karkkainen, M., "Increasing efficiency in the supply chain for short shelf life goods using rfid tagging," *International journal of retail and distribution management*, vol. 31, no. 10, pp. 529–536, 2003.

[33] Kim, S., Lewis, M., and White, C., "Optimal vehicle routing with real-time traffic information," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 2, pp. 178–188, 2005a.

[34] Kim, S., Lewis, M., and White, C., "State space reduction for non-stationary stochastic shortest path problems with real-time traffic information," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 3, pp. 273–284, 2005b.

[35] Laporte, G., Louveaux, F., and Mercure, H., "The vehicle routing problem with stochastic travel times," *Transportation Science*, vol. 26, no. 3, pp. 161–170, 1992.

[36] Lee, H. and Ö. Özer, "Unlocking the value of rfid," *Production and Operations Management*, vol. 16, no. 1, pp. 40–64, 2007.

[37] Leipala, T., "On the solutions of stochastic traveling salesman problems," *European Journal of Operational Research*, vol. 2, no. 4, pp. 291–297, 1978.

[38] Li, D., Kehoe, D., and Drake, P., "Dynamic planning with a wireless product identification technology in food supply chains," *International Journal of Advanced Manufacturing Technology*, vol. 30, pp. 938–944, 2006.

[39] Lin, C. and Chen, H., "Dynamic allocation of uncertain supply for the perishable commodity supply chain," *International Journal of Production Research*, vol. 41, no. 13, pp. 3119–3138, 2003.

[40] Lin, Z., Bean, J., and White, C., "A hybrid genetic/optimization algorithm for finite-horizon, partially observed markov decision processes," *INFORMS Journal on Computing*, vol. 16, no. 1, pp. 27–38, 2004.

[41] Lucas, T., Bishara, R., and Seevers, R., "A stability program for the distribution of drug products," *Pharmaceutical Technology*, pp. 68–73, 2004.

[42] Lundqvist, J., Fraiture, C., and Molden, D., "Saving water: From field to fork – curbing losses and wastage in the food chain," in *SIWI Policy Brief*, SIWI, 2008.

[43] Mahanti, A. and Bagchi, A., "And/or graphs heuristic search methods," *Journal of the ACM*, vol. 32, no. 1, pp. 28–51, 1985.

[44] Nahmias, S., "Perishable inventory theory: A review," *Operations Research*, vol. 30, no. 4, pp. 680–708, 1982.

[45] Osvald, A. and Stirn, L., "A vehicle routing algorithm for the distribution of fresh vegetables and similar perishable food," *Journal of Food Engineering*, vol. 85, pp. 285–295, 2008.

[46] Pearl, J., *Heuristics: Intelligent Search Strategies for Computer Problem Solving*. Addison Wesley, 1984.

[47] Percus, A. and Martin, O., "The stochastic traveling salesman problem: Finite size scaling and the cavity prediction," *Journal of Statistical Physics*, vol. 94, no. 5/6, pp. 739–758, 1999.

[48] Porteus, E., "On the optimality of generalized (s,s) policies," *Management Science*, vol. 17, pp. 411–426, 1971.

[49] Powell, W., *Approximate Dynamic Programming: Solving the Curses of Dimensionality*. New Jersey: John Wiley & Sons, Inc., 2007.

[50] Puterman, M., *Markov Decision Process: Discrete Stochastic Dynamic Programming*. New York: John Wiley & Sons, Inc., 1994.

[51] Reed, C., "Cold chains are hot! mastering the challenges of temperature-sensitive distribution in supply chains." ChainLink Research (http://www.clresearch.com), 2005.

[52] Sahin, E., Babai, M., and Dallery, Y., "Ensuring supply chain safety through time temperature integrators," *The International Journal of Logistics Management*, vol. 18, no. 1, pp. 102–124, 2007.

[53] Scarf, H., "The optimality of (s, s) policies in the dynamic inventory problem," in *Mathematical Methods in The Social Sciences*, (Stanford, CA), Stanford University Press, 1959.

[54] Smallwood, R. and Sondik, E., "The optimal control of partially observable markov processes over a finite horizon," *Operations Research*, vol. 21, pp. 1071–1088, 1973.

[55] Sondik, E., "The optimal control of partially observable markov processes over the infinite horizon: Discounted costs," *Operations Research*, vol. 26, pp. 282–304, 1978.

[56] Song, J. and Zipkin, P., "Inventory control with information about supply conditions," *Management Science*, vol. 42, no. 10, pp. 1409–1419, 1996.

[57] Tajbakhsh, M., Zolfaghari, S., and Lee, C.-G., "Supply uncertainty and diversification: A review," in *Trends in Supply Chain Design and Management: Technologies and Methodologies*, Springer, 2007.

[58] Taniguchi, E. and Shimamoto, H., "Intelligent transportation system based dynamic vehicle routing and scheduling with variable travel times," *Transportation Research Part C*, vol. 12, pp. 235–250, 2004.

[59] Tarantilis, C. and Kiranoudis, C., "Distribution of fresh meat," *Journal of Food Engineering*, vol. 51, no. 1, pp. 85–91, 2002.

[60] Toth, P. and Vigo, D., *The Vehicle Routing Problem*. SIAM Publishing: SIAM Monographs on Discrete Mathematics and Applications, 2002.

[61] Ustundag, A. and Tanyas, M., "The impacts of radio frequency identification (rfid) technology on supply chain costs," *Transportation Research Part E*, vol. 45, pp. 29–38, 2009.

[62] White, C., "A markov quality control process subject to partial observation," *Management Science*, vol. 23, no. 8, pp. 843–852, 1977.

[63] White, C., "A survey of solution techniques for the partially observed markov decision process," *Annals of Operations Research*, vol. 32, pp. 215–230, 1991.

[64] White, C. and Harrington, D., "Application of jensen's inequality for adaptive suboptimal design," *Journal of Optimization Theory and Applications*, vol. 32, no. 1, pp. 89–99, 1980.

[65] WHO, "Ensuring the quality of vaccines at country level." World Health Organization, http://www.who.int/vaccines-documents/DocsPDF02/www693.pdf, 2003.

[66] WHO, "Temperature sensitivity of vaccines." World Health Organization, http://www.who.int/vaccines-documents/DocsPDF06/847.pdf, 2006.

[67] Yano, C. and Lee, H., "Lot sizing with random yields: A review," *Operations Research*, vol. 43, no. 2, pp. 311–334, 1995.

[68] Yeon, J., Elefteriadou, L., and Lawphongpanich, S., "Travel time estimation on a freeway using discrete time markov chains," *Transportation Research Part B*, vol. 42, pp. 325–338, 2008.